Imperial College London
Business School

Empirical Essays on Global Pharmaceutical Innovation

Eliana Barrenho

Submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy of Imperial College London

Abstract

The economic impact of the pharmaceutical industry is incontestable. The total drug bill across OECD countries has continuously risen over the last decades (OECD, 2011), and, consequently, healthcare payers are increasingly implementing stricter policies that promote access to cheaper treatments. These policies, however, may hinder incentives to innovate, especially in diseases of substantial public health importance for which there is either under-investment or where innovation is difficult.

Therefore, it is fundamental that health systems design policies that strike the right balance between promoting the development of affordable drugs, and allowing sufficient rents to innovators in order to incentivise R&D investment.

The debate on this balance has never been more pertinent, with a slowdown in the number of drugs in the pipeline for potential market launch. Around 90% of drug candidates do not successfully complete the mid-stage of drug discovery (Paul et al., 2010; Mestre-Ferrandiz et al., 2012), contributing to the escalation of R&D costs with potentially significant social welfare consequences.

The aim of this thesis is to contribute to this debate by exploring the nature of the R&D process, and assessing the factors associated with decreased productivity across disease areas and its equity implications.

We survey the literature on the determinants of pharmaceutical innovation and critically appraise the evidence on factors that influence innovation of new therapies. We identify gaps and contribute conceptually to the understanding of the determinants of innovation. We depart from the existing literature in three significant ways. The major contribution is the analysis of failure for all stages of the R&D process, using a unique global panel dataset built by merging data on the lifecycle of industry innovation processes with global health data (Chapter 4). Secondly, we have used methodological approaches that model the dynamic nature of R&D decisions to forecast drug availability in the coming decades (Chapter 5). Thirdly, we are the first to assess the global impact of innovation on equity of access to new therapies (Chapter 6).

Acknowledging the implications of data limitations (Chapter 3) we produce insight that contributes to understanding the determinants of failure in innovation and its implications for future availability of new therapies. Results suggest those

determinants differ across the R&D stages. Furthermore, market competition may intensify the level of failure if too many young drugs are competing in the market, whilst collaboration between firms has an unclear effect on innovation. Moreover, the distribution of the R&D activity and disease burden have not changed significantly over the last two decades with a concentration of innovation in more commercially attractive disease areas associated with high mortality in the richest countries. Finally, that those equity concerns are likely to persist unless new policy interventions are designed to address inequalities in R&D and access to new therapies.

Contents

List of Figures

List of Tables

Declaration of Originality

An earlier version of Chapter 4 has been presented in several conferences and seminars. I would like to thank to the attendants at internal seminars at Imperial College London, both in the Business School and Faculty of Medicine in 2012 and 2013, Bocconi CERGAS in February 2014, $13th$ Portuguese Health Economics Conference 2013 in Braga, International Health Economics Association conference 2013 in Sydney, $2nd$ London Industrial Organisation Day, Spanish Economics Association Symposium 2013 in Santander, Royal Economic Society Job Market 2014, and London Health Economics Group meeting 2014 for helpful comments. In particular, I would like to thank to Dr Simone Ghislandi, Dr Jorge Mestre-Ferrandiz, Dr Paolo Li Donni, Professor Deborah Ashby, Professor Rossana Tarricone, Professor Giovanni Fattore, and Dr Alessia Melegaro for the very helpful comments and interesting discussions.

I am also grateful to Simone Ghislandi and Eugenio Zuchelli for the excellent comments and suggestions for the micro-simulation model developed in Chapter 5.

Also, an earlier version of Chapter 6 has been presented at the Health Economics Study Group 2012 in Oxford, European Health Economics Conference 2012 in Zurich, and at an internal seminar at Imperial College Business School in 2013. I would like to thank to the attendants of those forums and Miqdad Asaria, Professor Tommaso Valletti, Professor Rifat Atun, and Professor Chris Chapman for the helpful suggestions and comments.

Finally, I am grateful for the excellent comments on the revision of the final draft of this thesis to Professor Marisa Miraldo, Professor Pete C. Smith, Søren, Peter and Paulo. All errors remain my own.

I herewith certify that this thesis constitutes my own work and that all material, which is not my own work, has been properly acknowledged.

Eliana Barrenho

Copyright

Eliana Barrenho

Acknowledgments

I gratefully acknowledge the *Fundação para a Ciência e a Tecnologia*, in Portugal, for financial support to pursue this project (under the Doctoral scholarship referenced POPH-QREN-SFRH-BD-69131-2010, funded by *Ministério para a Educação e Ciência, Governo da República Portuguesa*, and *Fundo Social Europeu*). I would also like to thank Imperial College Business School for the financial support I was regularly given to participate in international conferences, courses and workshops.

There are not enough words to thank Professor Peter C. Smith. Thank you for having supervised my work. I am entirely grateful not only for receiving such extremely useful technical advice, but also, and probably more importantly, for your support at all levels since the beginning of this marathon. I will never forget the encouragement, patience, guidance, and support in creating the best conditions for me to pursue my project.

The words are even more scarce when thanking Professor Marisa Miraldo. I have discovered, in this long process, someone that shares, not only the passion and reason why I am here, but also the things that are most important in life. And those things are truly quite far removed from doing research or getting a PhD.

I am grateful to Professor Pedro Pita Barros for first introducing me to Health Economics, for having launched the seed and encouraged its growth during all this time, and for supporting me during all the steps of this long walk.

I am especially grateful to my PhD Viva examiners Professor Panos Kanavos and Professor James Barlow who offered me their insightful comments and constructive feedback that enriched this thesis.

Thank you Professor Lise Rochaix for all your kindness and support in the end of this long adventure.

I am also grateful to Professor Tommaso Valletti and Professor Carol Propper for the support given in the beginning of this process, and to my examiners in the first and late stage reviews - Professor Lesong Conteh, Professor Chris Chapman, Professor Tommaso Valletti, and Professor Rifat Atun - for their comments, suggestions and advice.

Dr Rowena Jacobs please receive my gratitude for having shown me a new world of applied microeconometrics. For your kindness, rigorous supervision and guidance on my decisions.

Thank you to the participants on the several conferences and seminars I have attended that shared with me interesting discussions and useful comments. I thank to the participants at internal seminars at Imperial College in 2012 and 2013, at the Centre for Health Policy at Imperial in 2014, and at Bocconi CERGAS in February 2014, and attendants at the London Industrial Organisation Day in January 2014, Royal Economic Society Job Market 2014, London Health Economics Group meeting in July 2014, $13th$ Portuguese Health Economics Conference 2013 in Braga, International Health Economics Association conference 2013 in Sydney, $2nd$, Spanish Economics Association Symposium 2013 in Santander, Health Economics Study Group 2012 in Oxford, and European Health Economics Conference 2012 in Zurich.

Thanks to all my friends that despite being far away have always been present in my life. Anoka, Davids, Pu, Natal, Ana Marisa, Sarita, you are always with me.

Thank you to Rita, Peter, Hae-Kyung, Tiago, Anupa, Zulma and Paola for the good moments we spent together and the long coffees and meals. Those were precious moments in this marathon. Di, Martolas, Pedrocas, Rafa and Sandra thank you for supporting me in the very first steps of this adventure.

Thank you to all my friends in Mozambique that were partly responsible for the motivation to come here, study this topic and spend four years of my life finding a way to do something useful. Lindo: obrigada!

Macauzito you are and you will always be an inspiration to me.

And because my inner-child is so dominant, I leave the best until the end. To my family, *um muito obrigada*! Thank you for the sacrifice of not having me around that often. For having always been there, whether agreeing or disagreeing with my choices. For being always my safe place when all goes wrong. For believing in me and in my strength to keep going, when I thought I couldn't. Thank you for having shown me such a diverse world through your eyes since I was a little girl. It is only possible for me to be here because you always nurtured and encouraged my passion for life and what is around me. Papi, Mammy, avó Ia, avô Pires, avó Jóia, avô Barrenho, avô Badalinho and avó Tide, thank you for your lessons and demonstrations of love. Thank you to Valila, Cláudia, Caca, Cristinol, Carlos e Cuca, Silvia, tios e tias, primos, and, especially, to my beautiful little girls Madalena, Daniela and Iris. You are the brightest side of my life! I dedicate this thesis to you!

A criança que ri na rua,
A música que vem no acaso,
A tela absurda, a estátua nua,
A bondade que não tem prazo -

Tudo isso excede este rigor
Que o raciocínio dá a tudo,
E tem qualquer coisa de amor,
Ainda que o amor seja mudo.

Fernando Pessoa. 4-10-1934. Poesias Inéditas.

Abbreviations

ATC Anatomical Therapeutic Chemical
CAS Chemical Abstracts Service registry
CD Communicable Diseases
COPD Chronic Obstructive Pulmonary Disease
CUAs Cost-Utility Analysis
CROs Contract Research Organisation
DALYs Disability-Adjusted Life Years
EMA European Medicines Agency
EMAP Emerging Markets and Asia-Pacific
EphMRA European Pharmaceutical Market Research Association
EU European Union
EudraCT European Clinical Trials Directive
FDA (US) Food and Drug Administration
FTC Federal Trade Commission
GARD Genetic and Rare Diseases Information Centre
G7 Group of 7
GBD Global Burden of Disease Study
GDP Gross Domestic Product
HIC High-income countries
HIV/AIDS Human Immunodeficiency Virus
HHI Herfindahl-Hirschman Index
IHME Institute of Health Metrics and Evaluation
ICD(ICD-10) International Statistical Classification of Diseases and Related Health Problems
ICTRIP International Clinical Trials Registry Platform
IHD Ischaemic Heart Disease
IP Intellectual Porperty
LIC Low-income countries
LMIC Low- and middle-income countries
LRI Lower Respiratory Infections
OECD Organisation for Economic Co-operation and Development
PH Proportional Hazards
MDD Major Depressive Disorder
NCD Non-Comunicable Diseases
NCEs New Chemical Entities

NHS National Health System
NIH (US) National Institute of Health
NME New Molecular Entities
NTD Neglected Tropical Diseases
PPP Purchasing Power Parity
R&D Research and Development
TRIPS Agreement on Trade-Related Aspects of Intellectual Property Rights
UK United Kingdom
USA United States of America
WHO World Health Organisation
YLD Years of Life Lost due to time lived in disability
YLL Years of Life Lost

Chapter 1

Introduction

The economic impact of the pharmaceutical industry is incontestable. The industry generated more than $500 billion in sales in 2010, and the total drug bill across OECD countries in 2009 accounted for approximately 19% of total health care spending (OECD, 2011). The last two decades of sustained expenditure growth signal a continuous increase in future spending in pharmaceuticals, given the "increased drug insurance benefit coverage" across the world (Berndt, 2002).

As a consequence, health care payers are more aware of the need for cheaper treatments than ever. Payers are thus increasingly eager to impose stricter regulations on drug price, market entry and reimbursement, decreasing the profitability of pharmaceutical companies. The squeeze on profits, together with the high levels of risk associated with the nature of the R&D investments, can potentially reduce innovation incentives, possibly, reducing worldwide access to novel treatments in the future.

The pharmaceutical sector, one of the most research-intensive industries, has experienced a slowdown in the number of drugs in the pipeline for potential market launch. Indeed, the R&D process is risky, time-consuming and costly. Many factors affect the performance of this process, including the stricter policy framework that regulates the sector. Around 90% of drug candidates do not successfully complete the mid-stage of drug discovery (Paul et al., 2010; Mestre-Ferrandiz et al., 2012). This contributes to the escalation of the R&D costs[1]. Given the economic and health related impact associated with pharmaceutical innovation and the need for access to new therapies in the future, the slowdown in innovation is of tremendous importance to the society.

The slowdown in innovation does not equally affect all disease areas associated with high burden of disease. Indeed, this slowdown more strongly affects disease areas that are less profitable, riskier or scientifically more complex. Therefore, it is fundamental for health systems to design policies susceptible of promoting innovation in disease areas of substantial public health importance but for which there is either under-investment by the industry or where innovation is difficult. These policies must strike the right balance between promoting the development of drugs with substantial health improvement that are affordable for health systems, and allowing sufficient rents to the industry to incentivise R&D investment.

Achieving this balance is not a trivial exercise. The innovation process is complex and lengthy and involves the interaction of multiple stakeholders. Understanding the complexity of the pharmaceutical R&D process, often considered and analysed as a "black box", is a first step into the design of those

[1]Some estimates suggest that the process from the early steps of basic research until a product is licensed and ready to be marketed can take approximately 14 years (Paul et al., 2010). Most recent estimates find an average cost of over $1bn to bring a new compound to the market (DiMasi and Grabowski, 2007).

policies.

The aim of this thesis is to contribute to this debate by exploring the nature of the R&D process, assessing the factors affecting innovation across disease areas and its equity implications. We propose to measure and characterise failure within all stages of the global pharmaceutical R&D activity, as well as the implications for equity of access to drugs and the future availability of new therapies across the different disease areas.

We discuss the existing literature that fails to analyse these points in Chapter 2. We survey the literature on the determinants of pharmaceutical innovation and critically appraise the evidence on the factors that influence the scientific discovery and development of new therapies. The chapter identifies gaps in the current literature that are addressed in the subsequent chapters.

We explore how the existing literature lacks theoretical and empirical analysis that consider simultaneous interactions between demand-side, supply-side and policy-shaping factors affecting the innovation process, taking the R&D process thoroughly and considering the equity and welfare implications of current policies to the pharmaceutical innovation. These are indeed the issues to which this thesis makes several contributions explained in detail in Chapter 2.

We depart from the existing literature in three significant ways. The major contribution is the analysis of failure for all stages of the R&D process, making use of a unique global panel dataset built by merging data on lifecycle of innovation processes in the industry with global health data. Secondly, we are the first to assess, at global level, the impact of innovation on equity of access to new drug therapies. Thirdly, we have used methodological approaches that model the dynamic nature of R&D decisions to forecast drug availability in the coming decades.

The remainder of this thesis is organised as follows. In Chapter 3, we describe in detail the data and empirical methods that are an important contribution of this thesis. In Chapter 4, we analyse the factors associated with failure of projects during the different stages of R&D. In particular, we assess how competition and alliances between firms correlate with the likelihood of project failure considering the duration of the R&D projects, the specificities of the product being developed, and the characteristics of the relevant market. In Chapter 5, we develop a dynamic micro-simulation model forecasting the stock of new drug launches in the market in the coming decades in the absence of further policy interventions. In Chapter 6, we assess the inequalities of access to new therapies, by measuring inequalities in global drug R&D activity and market launches in terms of population health needs in the last two decades. Finally, in Chapter 7 we present and discuss the main results of the thesis, debating the potential policy implications and future avenues for research.

Chapter 2

The determinants of pharmaceutical innovation: a critical review of the literature

I. Introduction

Population health has improved considerably over the past century. Pharmaceutical innovation has been one of the key factors at the heart of this improvement. New drugs have substantially contributed to better healthcare and prevention, resulting in significant improvements in life expectancy and quality of life (Acemoglu and Johnson, 2006). Nevertheless, innovation still falls short of societal needs. Especially, many health needs such as neglected diseases endemic in low- and middle-income countries, chronic and mental health conditions in high-income countries remain un-addressed (Kremer, 2002; Pecoul et al., 1999; Trouiller et al., 2002).

Even though the pharmaceutical industry has a proven track record in the development of innovative treatments, it has been reported that pharmaceutical innovation was experiencing a productivity crisis by the end of 2009 (Munos, 2009; Paul et al., 2010; Dhankhar, 2012). This was reflected in the declining number of new chemical entities (NCEs) approved and simultaneous escalating R&D costs by the end of 2010 (PWC, 2011b).

Indeed, drug innovation is a long complex and costly research intensive process with many factors impacting on its performance. It has been reported that it takes around 14 years from the early steps in basic research until a product is licensed and brought to the market (Paul et al., 2010). Estimates per approved biopharmaceutical product find that the "average capitalised pre-tax costs bringing a new compound up to the point of initial US marketing approval" are over $1bn (DiMasi and Grabowski, 2007). Furthermore, not all investment translates into successful R&D projects. Ninety percent of drug candidates do not go successfully through Phase II trials (Paul et al., 2010).

In this chapter, to further explore the determinants of innovation within the pharmaceutical industry, we propose to assess the literature on pharmaceutical innovation by exploring theoretical and empirical contributions in the literature. This review focuses on both the determinants of successful innovation and the factors affecting innovation expressed in terms of R&D effort and investment by big and small firms, spin-offs and start-ups, research centres, and academic departments.

In line with the Oslo Manual from OECD innovation involves "new or significantly improved characteristics of a product [, process or method of production,] to existing innovations. Because of this, innovation necessarily involves a certain degree of novelty (Manual, 2005). This degree of novelty is intrinsically related to the "concept of new to the market and new to the world" translating the idea of "whether or not a certain innovation has already been implemented by other firms, or whether the firm

is the first in the market or industry or worldwide to have implemented it" (Manual, 2005).

Those innovations that are new to the market are considered radical or disruptive. These are the innovations that are "new to the world for all markets and industries, domestic and international" and contrast with the incremental innovations, those that are significant improvements of the existing products (Manual, 2005). Both types, radical and incremental, are necessarily present in the R&D process in the pharmaceutical industry and are part of this literature review and the analyses performed in this thesis.

Furthermore, innovation activities include three types, that are namely: i) the ones that are successful in having resulted in an implementation; ii) the ongoing, work in progress, innovations that are uncertain implementation; and, iii) the abandoned ones before the implementation Manual (2005). These three dimensions are entirely considered in this review when assessing the determinants of such intricate and complex pharmaceutical R&D process. These are, indeed, the focus of the analysis performed in Chapter 4 and Chapter 5.

As we shall see, although the existing literature suggests numerous factors that explain how the performance achieved in the drug innovation is determined, there are many flaws and gaps in the literature. These stem from a variety of sources. Some are related to the approaches chosen to consider different factors affecting R&D performance. Others are methodological issues regarding the measurement of innovation performance, the econometric robustness of the estimation strategies, or even the confounders used to control for many different factors that compete to explain innovation. These problems are key areas of further research that will be the focus of this thesis.

This review is organised as follows. In the following section we describe the method and scope of the review. In section III we present the framework of our analysis, whereas section IV provides a careful examination of the theoretical contributions found in the literature. Section V assesses the empirical evidence in detail. Section VI presents the discussion of the key messages from the review as well as the concluding remarks.

II. Method and scope of the review

We perform a narrative review in line with standard practice in Economics that aims to identify the determinants of drug innovation. We have conducted a search in July 2011 and included the following databases: PubMed from 1966 to 2011, EconLit from 1968 to 2011 (via EBSCO), and Web of Knowledge from 1900 to 2011. We have also identified grey literature, such as the reports from the Office of Health Economics, and other publications through cross-referencing with the screened publications resulting from the search. We describe below the search strategy, selection and inclusion criteria, and search results of the search.

The review is concerned primarily with the determinants of innovation within the biopharmaceutical industry, by exploring theoretical and empirical contributions of existing literature. We are, in this perspective, interested in innovative results generated by the R&D process. This way, we are particularly focused in 'useful' R&D, either incremental innovations or scientific breakthroughs. Because of the long R&D process, with early stages' knowledge serving as input to next ones, innovation is presented in multiple modes along the drug pipeline: molecules, new chemical combinations, tissues, and final products,

such as new drugs, combinations of existing drugs, vaccines and other biologic compounds. Therefore, successful drugs reaching the market approval phase are just a part of the innovation created.

As the literature uses different proxies to measure innovation, the use of a strict definition would needlessly restrict the scope and findings of this review. Therefore, we have opted for using a Boolean search applying criteria that combined a series of terms to identify publications to the defined scope. The keywords used in the search can be found in Figure 2.1.

Table 2.1. Keywords used in the Boolean search

Publications were identified matching the following keywords as part of the full text of the articles:
(a) ("R&D" OR "research and development" OR "innovation") AND (field TX all field)
(b) (Pharma* OR drug* OR medicin*) AND (field TX all field)
(c) (determinant* OR explanat* OR factor* OR cause*) AND (field TX all field)
(d) (data OR model OR theor*) (field TX all field)

A. Selection and inclusion criteria of the search

Two groups of criteria were defined and used in the two-stage selection strategy we carried out. In the first stage, the first set of four criteria included studies:

1. Related to determinants of biopharmaceutical research and development activity
2. Published in peer-reviewed scientific journals
3. Written in English
4. Original articles excluding comments, reviews, letters, editorials and other similar papers

Study titles, abstracts and publication type were sequentially evaluated against these four inclusion criteria (and duplicates were excluded). Articles meeting all four inclusion criteria were retrieved. The eligibility was limited to published studies in English. Although this may lead to a risk of publication bias, it is unlikely that any major contribution to this enquire is not reported in the published literature. Moreover, studies must bring conceptual and empirical validation to the analysis of pharmaceutical R&D determinants. For this reason, comments, reviews, opinions and editorials are not contemplated in this review.

After being shortlisted in the first stage, full texts of all the articles meeting the first set of inclusion criteria were further examined. In a second stage, selected studies were classified according to whether they explored the determinants of biopharmaceutical research and development from a theoretical perspective or through an empirical approach. The second set of inclusion criteria included:

5. Theoretical studies: original and significant contribution

6. Empirical studies: methodological robustness, use of econometric techniques, exclusion of purely descriptive statistics studies

The check of eligibility, in the second stage, meant that only papers that offered an original and significant theoretical contribution were selected. These offer new theoretical insight (rather than being simple applications of existing theories) and that contribute to understand how the different demand, supply and policy factors affect R&D activity and innovation performance in the pharmaceutical industry.

Regarding empirical studies, only articles employing a methodology supported by econometric techniques were retrieved. Studies were excluded if they performed an analysis of descriptive statistics, trends or average numbers.

B. Results from the search

Figure 2.2 shows the PRISMA flow chart (Moher et al., 2010) of information through the different phases of the review. The search identified 1680 articles and six duplicates were removed. Titles and abstracts of 1674 unique records were sequentially analysed against the first set of eligibility criteria. The selection at the first stage resulted in 1622 exclusions. Most of the excluded studies did not refer to the object of this study, i.e. they did not examine the determinants of R&D in the biopharmaceutical industry (1593). Most of the PubMed papers excluded on this basis examined the biochemical characterization of new chemical entities (NCEs) prior to the development of a new drug, whereas most of EconLit exclusions were related to non-pharmaceutical studies or articles on the political point of view and economics of innovation. The examination of the other three eligibility criteria resulted in the exclusion of 21 papers for not being published in peer-reviewed journals, five for being written in languages other than English, and three studies that were not original. The second stage analysis resulted in 30 exclusions. Sixteen articles were not considered given the lack of original contribution to the theoretical modelling of the determinants of R&D in the biopharmaceutical sector. For the most part, these articles asserted normative statements and points of view when analysing factors impacting R&D decision-making. Additionally, 14 empirical studies presented only trends and average numbers analysis. Ultimately 22 articles were selected for full assessment. Of these, and based on the stages of screening and eligibility checks, five papers offered major theoretical contributions with their respective empirical validation. The remaining 19 were studies assessing empirical evidence alone, and although some explored their theoretical background or implications, they did so in a limited way and always within the context of one of the five theoretical contributions previously identified.

By cross-referencing these 22 articles, we have identified 24 additional articles for the narrative review. The following critical appraisal of the literature focuses on the 46 articles.

III. What determines R&D? An overview framework

Since the seminal contribution of Schumpeter's visions to the economics of innovation (Schumpeter, 1942), the literature has attempted to disentangle the role of the different factors that drive innovation and technological progress. In particular, the literature as focused on: firm characteristics (size, technological experience, location, among others) and market power; regulatory environment (such as intellectual property (IP) rights, industrial policy in the context of international trade); the nature of the R&D process (e.g. input spillovers, scale of R&D programs, collaboration between innovators, scientific and clinical knowledge) and demand-side factors.

For the analysis of the impact of the several factors on pharmaceutical innovation activity, we have developed a framework that clusters factors into three broad categories: (1) demand-side factors; (2) supply-side factors; and (3) policy-shaping factors. This overview framework largely contributes to the analysis of the interactions between the different factors that mostly have been studied separately, and contributes to the identification of the gaps of the existing literature. We use this framework to review the theoretical contributions and empirical evidence. Each factor within each category will affect innovation either directly or indirectly through effects on factors in other categories. Not controlling for these direct and indirect effects will potentially lead to misleading results. The three categories should be therefore thought of as complementary and interconnected, as Figure 2.1 suggests. The comments on the literature will account for these links.

A. Demand-side factors

Underlying the demand-side explanation is the belief that the potential for innovation responds to the aggregate demand for medicines (drugs, vaccines, and other biologic products) because the existing demand determines the size of the market and, therefore, potential profitability and returns to R&D investment.

Demand can be briefly characterised by the epidemiological profile of the population, and affordability. The epidemiological distribution of disease prevalence and disease severity would ideally characterise the medical unmet need of the populations, and for that reason, determine the demand for health care. However, there are limitations on the measurement of global need that includes all diseases, especially those that affect the poor countries and analyses may suffer from this measurement bias. Moreover, affordability is itself directly related to drug prices, income and reimbursement from insurers. All these define the profit expected to flow from successful innovation and the knowledge needed to fulfil tomorrow's scientific challenges. Consequently, firms' decisions to allocate resources to R&D programs across diseases internalize today's and future expectations about the volume and value of demand, defined by disease prevalence, and willingness- and ability-to-pay for medicines.

However, in the health care sector, a complete examination of the demand requires a multi-agent perspective: patients, health care providers, governments, insurance companies, and health care organisations are potential consumers. In countries where the national health system is the main provider and payer of healthcare, governments demand a large proportion of drugs and decide on the reimbursement level to the general public, with direct effects on the affordability of patients' drugs. On the other hand,

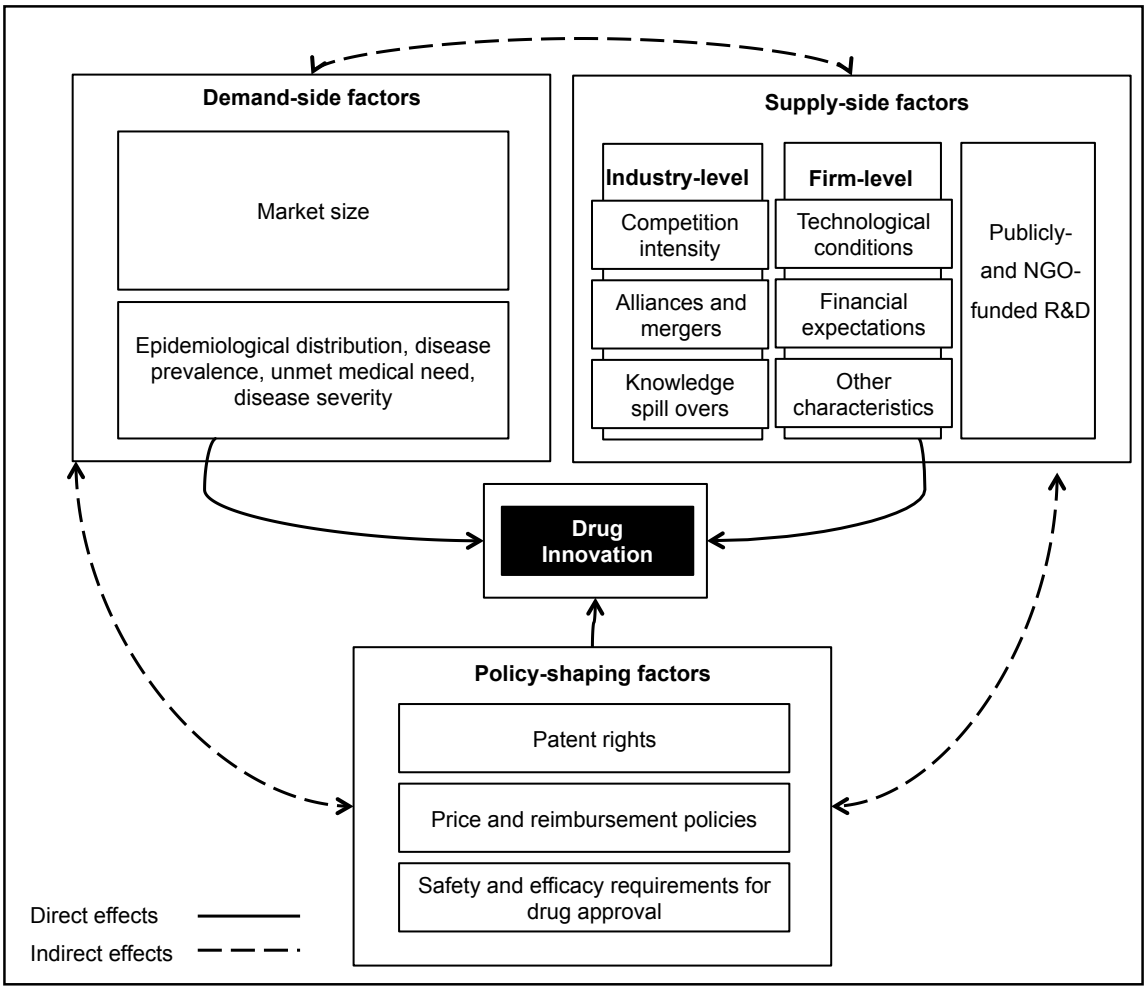


Figure 2.1. Framework of complementary factors explaining R&D allocation

in countries where the private sector plays a central role in the delivering and funding of healthcare, insurance companies and managed care organisations have implemented drug prescription lists to exercise bargaining power over the pharmaceutical industry to get cheaper prices, and contain moral hazard behaviour.

Lastly, demand affects and is affected by the policy context and regulatory framework, with potential strong implications to the dynamics of the market competition and allocation of R&D resources. Conditions that cause high disease burden to population health drive the need for new public health policies that improve access to new treatments. This need changes the interplay between health systems and innovators by shaping the incentive structure for innovation. On the one hand, increased need puts public budgets under pressure, leading to cost-containment efforts that can potentially reduce the prospects of profitability of new innovators. Many of these policies have indeed been targeted at fostering competition between branded and generic drugs, restricting drug prices and reimbursement levels with strong implications for the size and value of a product in the market. On the other hand, unmet health needs drive health systems to implement policies to facilitate drug innovation such as priority-drug approval

processes on areas as cancer and rare diseases (e.g. The Orphan Drug Act[1]), less restrictive rules on IP (such as patent flexibility and compulsory licensing as the TRIPS[2] Agreement contemplates), as well as public money available for R&D public-private partnerships.

B. Supply-side factors

On the supply-side, two levels of forces can affect innovation: firm-individual level characteristics, and industrial dynamics.

On the one hand, it is plausible to presume that innovation performance heterogeneity arises from differences in firms' characteristics and strategies. Different types of innovators compose the biopharmaceutical industry: biotechnology firms, laboratories, big-multinational pharmaceutical companies, small-specialised firms, laboratories, private-public consortia, and the general scientific community, placed in teaching hospitals or research centres. They perform differently accordingly to their individual characteristics as well as the capability and financial constraints they face. In particular, performance is likely to be affected by technological conditions, such as experience on disease-specific versus broad R&D programs, size, distance from inputs, alliances with other firms, access to external capital and internal funds.

On the other hand, supply-side factors relate to the industrial dynamics and the way organisations compete and/or collaborate in the market. The degree of competition and scope of collaboration between brand and generic firms, manufacturers, the scientific community, and laboratories may impose entry barriers and other costs, whilst facilitating alliances, mergers and knowledge spillovers across firms. All these factors impact on the variety and quantity of medicines available in the market, and impose constraints on the access to medicines.

Additionally, if evaluated as a pipeline, the R&D activity is potentially affected by other industries. Biotechnology firms, laboratories and academia play an important role at the level of basic scientific discovery and development of specialised therapies with molecules, tissues and biological compounds, benefiting disease-specific drug development. For that reason, the funding directed to life sciences and biopharmaceutical industry from publicly funded institutions and non-governmental organisations (NGO) affect the pharmaceutical innovation activity. Public money directed to biomedical and pharmaceutical R&D generates potential social and private benefits. Universities, research centres and public-private partnerships potentially impact on technological clusters and knowledge spillovers within the industry. Laboratories, teaching hospitals, and clinical testing centres may constitute factor endowments to R&D performance.

[1]The Orphan Drug Act [hereinafter ODA] of 1983 is an act to amend the Federal Food, Drug, and Cosmetic Act in the United States to facilitate the development and commercialisation of drugs to treat rare diseases and conditions, termed *orphan drugs*, Orphan Drug Act, Pub. L. No. 9714, 96 Stat. 2049 (1983), codified at 21 U.S.C.,xx 360aa60dd (2000)

[2]Trade Related Aspects of Intellectual Property Rights Agreement [hereinafter TRIPS Agreement] is an international agreement administered by the World Trade Organisation (WTO), that sets minimum standards for many forms of intellectual property regulation applicable to nationals of WTO members, Agreement on Trade-Related Aspects of Intellectual Property Rights, 15 Apr. 1994, Marrakesh Agreement Establishing the World Trade Organization, Annex 1C, Legal Instruments esults of the Uruguay Round vol. 31, 33 I.L.M. 81 (1994).

C. Policy-shaping factors

Multiple policies interplay within health systems and may influence the incentives for pharmaceutical innovation. These policies can be clustered in three broad areas. The first relates to drug approval requirements to ensure safety and clinical effectiveness. The second concerns patenting, including marketing exclusivity arrangements and fast-track regulations, in the way it affects market and R&D competition. Finally, the policies regarding price and reimbursement of medicines. These include cost-containment policies, health technology assessment, and competition policies between branded and generics, as well as reference pricing between countries in the way they affect RD incentives.

Such policy and regulatory factors impact on the demand and supply of medicines. The pharmaceutical sector is heavily regulated with non-negligible cost implications for the R&D processes: tighter clinical safety standards require more data on efficacy and cost-effectiveness of medicines that impose higher RD costs; patenting and license approvals impact directly on generics versus non-generics competition and price-setting; price and reimbursement policies determine buyers' purchasing power and firms' profitability.

A framework for the analysis of the factors that affect innovation

These factors above described can be brought together in a single equation that summarises the forces that drive innovation:

$$Innovation = f(D, S, P) \tag{2.1}$$

where D represents the demand-side factors, S expresses the supply-side factors, and P evokes the policy-shaping factors on innovation activity. Interconnections between these three explanations (illustrated in Figure 2.1) raise methodological questions when defining this function. Simultaneous effects bring additional modelling challenges of overcoming collinearity, endogeneity and reverse causality.

With this framework as a background, the sections that follow critically examine the existing literature, highlighting the potential methodological challenges in the analysis of the determinants of innovation that arise from this holistic perspective.

IV. Conceptual and theoretical contributions

Methodological considerations

The first challenge encountered in this review regards the inconsistency found in the literature at a conceptual level in the definition of innovation. Indeed, the terms innovation and R&D are used interchangeably, despite being two different concepts. Similarly, the concepts intensity, effort and performance may constitute different approaches when referring to innovation or R&D. The disputable conceptual framework gives room for the use of heterogeneous and less uniform use of performance measurements of innovation activity, as we will explore in section V.

Literature review

In this section, we provide a detailed analysis of the theoretical contributions following the framework of analysis presented in section III.

A. Demand-side factors

Three studies explore the effect of aggregate demand on innovation and investigate the role of market size on the composition of biomedical R&D using disease prevalence as a proxy for demand (Acemoglu and Linn, 2004; Lichtenberg, 2005; Bhattacharya and Packalen, 2011). Theoretical predictions from the three studies indicate that greater R&D effort is associated with more lucrative therapeutic areas (Acemoglu and Linn, 2004; Lichtenberg, 2005; Bhattacharya and Packalen, 2011). Optimal R&D effort is found to be intrinsically related to anticipated profitable market segments (Acemoglu and Linn, 2004). Furthermore, even at the academic level, when disease prevalence is used as a proxy for market size, basic scientific funded research is also found to respond to changes at the market level (Bhattacharya and Packalen, 2011).

The studies, however, are not entirely comparable.

Two studies have limitations by not incorporating supply-side and policy-shaping factors. Lichtenberg (2005), on the one hand, assumes that increases in the disease burden affect only the marginal social benefit but not the marginal cost of innovation, while Bhattacharya and Packalen (2011) model the socially optimal allocation of research according to the disease prevalence and the distribution of funding granted to the different diseases.

The study by Acemoglu and Linn (2004), however, simultaneously models demand and supply of drugs. They assume a competitive monopolistic model with product differentiation which allows for different R&D strategies and innovation rates across disease areas (Acemoglu and Linn, 2004). The study models exogenous demographic and income changes by age group in the United States in order to address two methodological challenges: (i) reverse causality problems at income and demand level (endogeneity); and (ii) cross-correlation with the past (autocorrelation).

B. Supply-side factors

Two studies investigated theoretically whether the profitability of market segments, price levels, and financial constraints are main determinants of innovation direction (Grabowski and Vernon, 2000a; Giaccotto et al., 2005). Both studies find a positive effect of expected returns and cash flows on R&D intensity (Grabowski and Vernon, 2000a; Giaccotto et al., 2005), reinforcing the expected profitability of the R&D projects a crucial factor in firms' R&D decisions. Most of the empirical studies use these theoretical contributions as a baseline.

C. Policy-shaping factors and regulatory framework

The impact of policies on innovation is assessed by controlling for supply-side determinants only in one study (Giaccotto et al., 2005). The predicted effect of two dummy-type policy changes are: (i) ambiguous effects of changes in drug price competition regulation[3] on firms' expected returns and cash flow levels; and (ii) negative effects of tighter drug approval rules[4] on firms' profitability. Ambiguous effects of the Waxman-Hatch Act are anticipated, given the lower entry barriers created for generic products and increased patent lives for branded ones. Regarding the later, stronger Food and Drug Administration (FDA) requirements for clinical testing are suggested to negatively affect R&D intensity levels due to additional R&D costs.

Table 2.2. Gaps of the existing theoretical literature

Publications addressing the three category of factors	**existing evidence**
Demand-side factors including supply-side factors	existing literature
Demand-side factors including policy-shaping factors	ambiguous
Supply-side factors including demand-side factors	ambiguous
Supply-side factors including policy-shaping factors	existing literature
Policy-shaping factors including demand-side factors	ambiguous
Policy-shaping factors including supply-side factors	ambiguous

Table 2.2 identifies the gaps in the theoretical literature by listing the types of studies following the three clusters and the main approach taken in the study (e.g. studies with demand-side factors including supply-side factors are studies which have as main focus the demand-side factors and potentially may include supply-side factors, whereas studies with supply-side factors including demand-side factors are studies that focus primarily on supply-side factors but that contemplate demand-side factors in the analysis).

To summarise, the theoretical literature confirms market profitability as a significant determinant of the current R&D effort. Also, the literature provides insight into the incentives underpinning academic R&D that are also determined by demand and research funding. This supports the consolidated idea of absence of incentives to pursue R&D projects targeted at markets and diseases with relatively low expected future returns.

However there are substantial gaps in the literature (see Table 2.2), in particular there are relatively few studies that simultaneously assess the role of demand, supply and policy factors in incentivising innovation. Furthermore, there is a lack of theoretical insight on the effects of policy instruments such

[3]It refers to the Waxman-Hatch Act Introduced in 1984, the Drug Price Competition and Patent term Restoration Act that extended the effective patent life of a new drug by a maximum of 5 years. Drug Price Competition and Patent Term Restoration Act of 1984, Pub. L. No. 98-417, 98 Stat. 1585 (codified at 15 U.S.C. xx 68b-68c, 70b (1994); 21 U.S.C. xx 301 note, 355, 360cc; 28 U.S.C. xx 2201 (1994); 35 U.S.C. xx 156, 271, 282 (1994)).)

[4]It refers to the Kefauver-Harry amendment is The Food, Drug and Cosmetic Act of 1962 instituted the regulatory requirements to bring a pharmaceutical product to market, through FDA approval.

as clinical guidelines, drug prescription list, price regulation, reimbursement policies, and international agreements on pharmaceutical innovation (Table 2.2).

V. Empirical evidence

Methodological considerations

The empirical studies analyse the determinants of pharmaceutical R&D in seven individual countries (US, Canada, Sweden, Denmark, Japan, Netherlands and France). However, many of them present an international examination, performing analysis at OECD and EU levels.

The studies are, however, not comparable since they use different levels of data aggregation and, because of that, cannot give definite conclusions about the impact of the different factors on pharmaceutical innovation. The data used varies significantly. In general, three types of data are used: (i) cross-country data; (ii) aggregated market-level data at research project, drug or research phase level; (iii) micro-data at the firm level, both by aggregation and disaggregation at either drug or molecule level. Moreover, data sources vary with the purpose of the studies: those that investigate determinants of R&D linked to firms' characteristics such as financial situation or knowledge spillovers, commonly use voluntary and self-reported information provided by firms or agency institutions (for example, Henderson and Cockburn (1996) or Arundel and Kabla (1998)). Studies with an aggregate market-level drug and research phase analysis use private information compiled by pharmaceutical consultancy firms (for instance, Danzon et al. (2005a) or Lichtenberg (2005)).

How to measure innovation?

Outcome variables also present an important source of heterogeneity. Studies lack uniformity when measuring innovation effort and R&D productivity. Most of the studies focus on the quantification of R&D intensity, proxied either by the level of R&D investment or the number of scientific publications and patents. Table 2.3 describes and groups the variety of outcome variables considered by the different studies.

[Table 2.3 here]

This raises issues regarding the comparability of these studies. In principle, studies that use measures of R&D intensity, such as R&D expenditure per sales or R&D investment do not accurately reflect its productivity. Also, the use of the number of scientific articles as a proxy for innovation performance suffers from disguised publication bias and could potentially be distorted by academic interests, thus not capturing private sector performance as a whole.

Additionally, the use of patents as a measure for scientific discovery can also be of concern. Indeed, using the number of patents biases the analysis towards successful compounds that effectively pass the drug approval and licensing requirements. Other factors have an impact on successful patenting. We could speculate how much of the propensity rate of being patented is explained by the value of the invention itself and by the firm specificities and competition dynamics. The same rational applies to drug launch

measures. The portions of FDA approvals or new drugs launched do not comprehensively capture the performance of a long, complex and multi-inputted R&D process. They may be biased towards the more profitable and less competition-deterred technologies.

To overcome such limitations, three studies (Danzon et al., 2005a; Abrantes-Metz et al., 2004; Aharonson et al., 2007) consider the probability of success rates across R&D phases, by evaluating what happens at molecular level (NCEs). This way, the bias from marketing, patenting or publishing is minimised.

Literature review

In the section that follows we provide a detailed analysis of the empirical evidence following the framework of analysis presented in section V.

A. Demand-side factors: market size and epidemiological distribution

Market size

The existing literature unanimously confirms the theoretical predictions of the aggregate market size as a significant determinant of R&D direction (Acemoglu and Linn, 2004; Lichtenberg, 2005; Civan and Maloney, 2006; Bhattacharya and Packalen, 2011; Dubois et al., 2011). Furthermore, this effect is significantly dependent on the income distribution (Acemoglu and Linn, 2004; Lichtenberg, 2005; Civan and Maloney, 2006; Bhattacharya and Packalen, 2011). On top of that, one of these studies finds the US market as a worldwide R&D driver (Civan and Maloney, 2006). But the analysis in this study is limited by the use of mortality and person-years lost due to disease proxies for potential market size. Consequently, the study does not consider conditions such as chronic and neglected tropical diseases with high comorbidity levels and low mortality.

With the epidemiological distribution being the most commonly used proxy for market size, problems of endogeneity and reverse causality between health and wealth arise. Some authors employ cancer incidence (Lichtenberg, 2007); and burden of disease measured in terms of DALYs across diseases (Lichtenberg, 2005) but do not tackle the arising methodological issues. Others go one step further using exogenous demographic (Acemoglu and Linn, 2004) and epidemiological shocks (Bhattacharya and Packalen, 2011) to deal with the endogenous effect of wealth on health and demand for drug innovation. However, they are not necessarily equally able to measure R&D performance: Acemoglu and Linn (2004) uses both the number of NCEs and number of FDA approvals, while Bhattacharya and Packalen (2011) subscribes the publication bias when considering the number of scientific articles per disease.

Acemoglu and Linn (2004) explores partially the multi-agent function of demand, illustrated in Figure 2.1. To explain the effect of the multi-agent function on drug innovation, this study uses US drug prescription data on ambulatory and private consultancy services to give a measure of market size from the healthcare system's point of view.

Nevertheless, demand-side factors are limited by not controlling for supply-side and policy-shaping factors, covering biases from reverse causality, endogeneity and confounding effects.

B. Supply-side factors

With the exception of a few cases below mentioned, no supply-side factors are incorporated in the analyses of the studies reviewed. Policy-shaping effects are also far from being explored, and inappropriate methodological strategies show how logistic regressions or Ordinary Least Squares are unable to disentangle policy from supply or demand effects as possible determinants of innovation. We provide a detailed analysis of each factor within the supply-side category considered by the literature, namely industrial and firm level factors.

The existing literature exhaustively explores the impact of forces such as technological specialisation and experience of firms, several other firm's characteristics, and access to capital in drug innovation.

Firm technological conditions

While two studies find evidence of a negative effect of diversification of knowledge across diseases on phase-specific success rates (Danzon et al., 2005a; Aharonson et al., 2007), two other studies suggest evidence of a positive effect of spillovers across firms with related research projects on patents granted by disease-type (Henderson and Cockburn, 1996) and the existence of economies of scope within a firm's R&D portfolio (Henderson and Cockburn, 1996). These results are not, however, entirely comparable since they use different measures to proxy innovation: patents granted by disease area (Henderson and Cockburn, 1996) do not necessarily measure the same as the phase-specific success rates (Danzon et al., 2005a; Aharonson et al., 2007).

Additionally, knowledge stock is regarded to benefit innovation creation. More experienced firms generally present a higher number of patents and R&D success rate, whether experience is measured by looking at retrospective R&D expenditures on sales (Alexander et al., 1995; Arundel and Kabla, 1998), patent stock (Ulku, 2007), firms' overall experience (Danzon et al., 2005a; Henderson and Cockburn, 1996; Pisano, 1990) or therapeutic category-specific experience (Danzon et al., 2005a). Several firm-level and industrial-level covariates are considered here, but they commonly lack policy relevant considerations with potential impact on industrial clusters, IP issues and public R&D spending. This may complicate the comparability of results amongst countries with different policy settings at these levels.

Lastly, Danzon et al. (2005a) find that experience is particularly relevant at the early stages of research (Phase I) for smaller firms, but shows to be not significantly important on clinical trials to large firms. This study presents an extensive analysis by considering data on 1910 compounds, developed by 900 US firms in the period between 1988 and 2000. They analyse, across time, the role of alliances on R&D phase-specific success rates by exploring a multilevel analysis at compound and firm level.

Firm specific characteristics

Geography and nationality are explored as determinants of the number of patented compounds (Alexander et al., 1995; Arundel and Kabla, 1998; Hirai et al., 2010; Pisano, 1990). Positive effects are found to be related to firms headquartered in dynamic markets such as Germany (Alexander et al., 1995) and the US (Arundel and Kabla, 1998). Non-Japanese firms are shown to have higher entry barriers to launch drugs in the Japanese market (Hirai et al., 2010). However, one study reports no R&D productivity differences in firms based in the US and Europe (Pammolli et al., 2011).

Furthermore, firm's age is found to have a negative impact on the number of patented compounds (Kim et al., 2009).

Two studies report that firm size has a negative effect on the number of patented compound (Alexander et al., 1995; Kim et al., 2009), whilst three other studies find that large firms are more successful in originating technical change (Vernon and Gusen, 1974) and in launching new drugs in the market (Abrantes-Metz et al., 2004; Carpenter and Turenne, 2001).

These studies, however, differ in terms of robustness of the analysis. Three of them present more robust analyses, given the covariates considered to control for inventor's characteristics, firms' time-varying features (Kim et al., 2009), therapeutic area specificities and drug approval requirements (Hirai et al., 2010; Abrantes-Metz et al., 2004). However, they do not incorporate demand-side factors and policy-shaping factors. For this reason, the effect of firm's size is rather unclear. The impact of the firm size on the R&D performance might be contaminated by the effects of the demand-side and policy factors.

Firm financial expectations

The literature seems to provide homogeneous evidence when supporting the theoretical predictions of the importance of expected returns and availability of internal funds in determining innovation performance (Grabowski and Vernon, 2000b; Giaccotto et al., 2005; Mahlich and Roediger-Schluga, 2006; Vernon et al., 2010; Mahlich and Roediger-Schluga, 2006; Vernon, 2005; Jensen, 1987). Capital intensity, whether measured in spending or capital-labour ratio, also contributes to better performance at patenting level (Kim et al., 2009; Boasson et al., 2005). The data used, however, suffers from potential self-report and sample bias: the studies report privately owned information of a small sample of firms, not greater than 14.

Additionality, only one study addresses possible endogeneity issues related to access to capital and more innovation. Vernon (2004) consider non-pharmaceutical cashflows as an instrument for accessibility to capital but lacks robustness in measuring innovation as it considers aggregated figures of R&D investment at the firm-level.

Apart from firm characteristics, aggregate forces at the industrial level determine how firms compete and collaborate in the innovation activity.

Alliances and mergers

The literature is not conclusive when evaluating the effect of alliances and mergers on innovation. Three studies demonstrate significant positive effects of alliance status on phase-specific R&D success rates (Danzon et al., 2005a), drug launch times (Hirai et al., 2010) and firm's internal productivity (Higgins and Rodriguez, 2006), whilst a study concludes on prejudicial effects of alliances with academia on R&D success rates (Aharonson et al., 2007).

The divergence in conclusions may derive from data comparability: the later is a firm-level study, while the former three studies refer to compound-level analysis; also, different countries are analysed in these studies: US, Japan/US/EU, and Canada, respectively. Likewise, control variables diverge significantly across studies: some are more focused on technological conditions of the R&D process, such as the proportion of clinical trials pursued, or quality of compounds tested (Danzon et al., 2005a) and therapeutic area

specificities (Hirai et al., 2010); while others focus on firm characteristics, access to supportive resources and the intensity of competition (Aharonson et al., 2007; Hirai et al., 2010; Higgins and Rodriguez, 2006).

Furthermore, two other studies suggest possible adverse effects of large horizontal mergers on the R&D investment and availability of new drugs (Comanor and Scherer, 2013; Graves and Langowitz, 1993).

Knowledge spillovers

Two studies find significant evidence of geographical distance (as a measure of technological distance) as a determinant of innovation performance: geographic distance between the firm and the nearest neighbour (Boasson et al., 2005), and competitors' output in the same research area (Henderson and Cockburn, 1994) show positive effects upon the number of patents. Nevertheless, the former provides a more detailed analysis per disease-specific research programme (Henderson and Cockburn, 1994).

Lastly, the difference between private-private versus public-private spillovers is also explored (Boasson et al., 2005; Cockburn and Henderson, 1998; Furman et al., 2006). While geographical and technological distance are found to have a significantly positive spillover effect in R&D productivity between private firms, no significant results are found regarding public-private partnerships.

Competition intensity

Intensity of competition is mostly regarded by the empirical literature as a significant determinant of innovation and it is closely linked with the IP rights' system and the competition policy in a country. Two types of competition have been suggested by the literature to affect innovation: market competition, exerted by the incumbents, and pressure from peers in the R&D process that race to innovate in a specific target.

In what concerns market competition, the studies are divergent in terms of data sources, methodology and estimation strategies, and they also differ in the conclusions about the effect of competition on innovation. Most of the studies suggest a significant positive effect (Alexander et al., 1995; Arundel and Kabla, 1998; Grabowski and Vernon, 2000b; Danzon et al., 2007; Giaccotto et al., 2005; Mahlich and Roediger-Schluga, 2006; Aharonson et al., 2007), whilst Kyle (2006) finds that market competition is negatively correlated with the likelihood of entry, and that this effect is higher when the number of competitors is low. Indeed, the impact of drugs longer established in the market seems to be greater than that of more recently introduced drugs.

Despite that, some of these studies make conclusions using poor measures for innovation productivity when evaluating the effect of industrial margins on R&D investment per sales (Grabowski and Vernon, 2000b; Giaccotto et al., 2005; Mahlich and Roediger-Schluga, 2006) and foreign sales (as percentage of total sales) (Giaccotto et al., 2005). A cross-industry analysis for 604 Dutch and French firms was performed to evaluate the effect of exports to USA and Japan (Arundel and Kabla, 1998). A more robust longitudinal approach is used by Aharonson et al. (2007) to evaluate the entry of Canadian firms in the sector over 9 years. However, the study is focused on measuring the reverse causality of R&D intensity geographical rings co-location on firms' entrance. Additionally, one study finds positive effect of firm global market share on patent application propensity rates (Alexander et al., 1995).

Moreover, inconclusive results are found in a study that considers the ratio of drug prices between

Japan and US to explain the median time lag of drug approvals between Japan and the US/EU (Hirai et al., 2010). With apparently solid data at drug level, this study takes a more comprehensive analysis by controlling for firms' demographic characteristics and some regulatory effects, such as category of drug approval process (standard, priority, and orphan status). Also, drug approval requirements are explored by controlling for the use of a bridging strategy by Japanese firms[5]. Such covariates have potentially strong effects on faster and costless drug launches, and may constitute a main difference between US/EU and Japanese markets.

Kyle (2006) presents the most robust study when analysing the effect of market competition on innovation. The author analyses all drugs developed in the 28 largest pharmaceutical markets between 1980 and 2000 and finds that several of the characteristics of entrants and incumbents are positively associated with the time-to-entry in the G7 markets. Kyle (2006) uses duration models to model time-to-entry when accounting for country-specific demand, therapeutic area specificities and the collaboration behaviour between firms.

Moreover, the effect of competition from R&D peers is found, in contrast to market competition, to positively affect drug innovation in two studies. Indeed, Cockburn and Henderson (1994) and Kyle (2006) have found that the race to innovate is associated with a higher R&D productivity between rivals (Cockburn and Henderson, 1994) and a positive effect on the likelihood of entry (Kyle, 2006).

Additionally, a phase-specific approach is explored by a third study which finds inconclusive results (Danzon et al., 2005a). By constructing a Herfindahl-Hirschman Index (HHI) on the proportion of a firm's compounds in each therapeutic category, the authors intend to capture a scale-scope effect with potential beneficial spillovers amongst firms, across therapeutic groups. HHI is found to have a positive effect on the number of success compounds in the Phase I trials but an inverse effect on the number of compounds in the late stages of research (Phase II and III) (Danzon et al., 2005a). This study uses R&D project-level across OECD countries, which allow the exploration of firm, projects and therapeutic class fixed effects simultaneously.

Finally, the existing literature has not adequately addressed the effect of parallel trade on innovation and has been reporting unclear consequences on price convergence across countries. One study reports that entry by parallel traders resulted in price reduction by originators in Sweden (Ganslandt and Maskus, 2004), whereas three other studies suggest that parallel trade imports have neither resulted in price convergence across countries nor lower prices (Kanavos and Costa-Font, 2005; Kanavos and Vandoros, 2010; Kyle et al., 2008).

With the exception of one study that speculates on the effect of declining drug prices (ensured by import policies) on the elasticity of drug development (Lichtenberg, 2007), the consequences of parallel trade on innovation, access to medicines and social welfare have not been addressed in the literature and require further research.

In the next section we examine the various levels of market competition that derive from IP system and patent rights that regulate the pharmaceutical industry.

[5]this consists in the use of data originated from Randomised Control Trials for Phase III carried out in other countries than where the drug application is being processed.

C. Policy-shaping factors: drug approval, patenting, price regulation and trade policy

A few number of studies account for the impact of policy settings on innovation. Some evidence is given at four levels of policy context: patent rights, price and reimbursement policies, and safety and efficacy requirements for drug approval.

Patent rights

The literature reports various effects on the role of patent rights on innovation in the pharmaceutical industry, and the issues raised by differences across countries in terms of IP rights.

When looking at the overall effect of patenting on R&D effort, one study finds no significant effect of sales on the R&D intensity (Giaccotto et al., 2005), whereas two other studies suggest positive benefits of stronger IP regulation and appropriation returns from licensing on the number of drugs and patent application propensity scores (Arundel and Kabla, 1998; Civan and Maloney, 2006). Stronger results using a quasi-experiment methodology with a case-control approach are drawn from a study that evaluates a Danish patenting policy in the academic sector applied to Danish scientists (Valentin and Jensen, 2007). A substitution effect between domestic nationals and non-nationals inventors is found to occur by avoiding stricter IP rights of the Danish scientists.

Furthermore, Kyle (2006) has shown that intermolecular competition affects the likelihood of entry. Competition pressure from products established in the market and enjoying from the patent status is negatively correlated with the likelihood of entry, and that this effect is higher when the number of competitors is low.

Developed versus developing world

Two studies have looked to the effect of the TRIPS agreement on the R&D effort. The TRIPS agreement increased the levels of patent protection around the world but has not been found to affect R&D investment in the diseases that are most prevalent in developing countries (Kyle and McGahan, 2012; Scherer and Watal, 2002). The existing literature proposes policy changes that encourage ramsey pricing and contain parallel exports from low-income countries (Scherer and Watal, 2002), or a regime in which firms could opt for patents in either developed or developing countries (Lanjouw and Cockburn, 2001).

Price and reimbursement policies

The optimal design of price regulation is complex and has not been fully addressed by the literature. The available literature does provide some evidence on the general trade-off between stricter price regulation and innovation (Hirai et al., 2010; Giaccotto et al., 2005; Vernon, 2005), but the results lack robustness as they do not consider time trends, and other demand and policy effects.

Across the different pricing and reimbursement policies, there is some evidence on the effect of external reference pricing and launch of new drugs (Danzon et al., 2005c; Kyle, 2007; Lanjouw, 2005; Danzon and Epstein, 2012). The external reference pricing compares the price of the new drug to the prices of products that are bio-equivalent in other countries. As a comparator, it uses the mean, median or minimum price of the same drug in a designated set of countries. The literature has suggested that the effects of price differentials across countries are associated with the convergence of firms' pricing strategies across

countries, and potential launch delays and non-launches on smaller markets and low-income countries (Danzon et al., 2005c; Kyle, 2007; Lanjouw, 2005; Danzon and Epstein, 2012).

Indeed, and because of the existence of global price differentials, an arbritage opportunity is given by the fact that countries with relatively low prices are allowed to resell drugs to countries where prices are high, without authorisation of the firm that owns the IP rights over the product.

Finally, no investigation on the impact on innovation of reimbursement policies, clinical guidelines or lists of prescription drugs is referred to in the existing studies.

Safety and efficacy requirements for drug approval

Several studies have shown that countries with stricter drug approval criteria experience higher R&D costs and higher delays in launch of new drugs (Wiggins, 1981, 1983; Baily, 1972; DiMasi et al., 2003; Grabowski et al., 1978; Vernon et al., 2010; Hirai et al., 2010; Giaccotto et al., 2005), although significant variation is reported across therapeutic categories. For instance, Danzon et al. (2005b) shows that both regulatory requirements and competition have contributed to the exit of vaccine manufacturers, as the duration of trials in order to detect adverse events has increased considerably.

At the same time, more relaxed rules at drug approval level, such as fast-track approval policies for important drugs and the use of comparable data used in regulatory processes in other countries, are reported to be associated with faster launch of drugs and less costly R&D processes (Hirai et al., 2010; Dranove and Meltzer, 1994; Yin, 2008).

However, the existing literature insufficiently assigns a causal relationship when assessing the impact of these regulations on market launch of new drugs and in terms of social welfare across time. Many other factors potentially contribute to the rising R&D costs and the decrease in the number of new drugs being launched in the market. These include changes in the R&D landscape provoked by firm's investment decisions, population health needs, payers' willingness-to-pay and other regulatory frameworks at the patenting level.

VI. Discussion and final remarks

Results from the theoretical contributions are, to a large extent, validated by empirical evidence. The literature finds robust evidence on the effect of several determinants of innovation: at an aggregate level, epidemiological and aggregate income distributions in more dynamic economies as proxies for market size (Acemoglu and Linn, 2004; Lichtenberg, 2005; Bhattacharya and Packalen, 2011; Dubois et al., 2011); at the industrial level, the intensity of competition between pharmaceutical companies in the R&D process (Kyle, 2006; Cockburn and Henderson, 1994), the presence of economies of scale and technological specialisation (Cockburn and Henderson, 1994; Danzon et al., 2007; Aharonson et al., 2007; Boasson et al., 2005) and at the firm level, the future commercial profitability and cash flows (Grabowski and Vernon, 2000b; Giaccotto et al., 2005; Mahlich and Roediger-Schluga, 2006; Aharonson et al., 2007; Giaccotto et al., 2005; Vernon et al., 2010), and firm characteristics such as size (Alexander et al., 1995; Kim et al., 2009; Abrantes-Metz et al., 2004; Carpenter and Turenne, 2001), age (Kim et al., 2009), location and nationality (Alexander et al., 1995; Arundel and Kabla, 1998; Hirai et al., 2010; Pisano, 1990).

We can denote these on Equation 2.2 based on the evidence above as:

$$Innovation_j = f(D(m_j), S(c_j, t_{ij,}, f_i, x_i)) \tag{2.2}$$

where Demand D is a function of market size per disease j denoted by m_j, Supply S is a function of competition intensity between firms within a disease-specific category j represented as c_{jt}, technological conditions such as scale and scope within each disease category j of firm i given by t_{ij}, financial expectations of each firm i which invests in disease-category j denoted by f_{ij}, and firm-specific characteristics x_i, such as size and location. Nonetheless, there are missing Policy-shaping factors P in the equation revealing a lack of consolidated effects of policies on innovation. Indeed, the search has not focused on the policy aspects of innovation and, because of that, it has only identified a few studies with policy considerations.

In other cases, findings of empirical studies have been inconclusive. This is specially the case for the effect of market competition and intensity of alliances (private-private and public-private), mergers, and technological distance in same therapeutic areas.

Furthermore, the current literature is scarce in providing a theoretical approach on simultaneous interactions between demand-side, supply-side and policy-shaping factors, as Figure 2.1 illustrates. Here we present three limitations found in the existing literature addressed by contributions made in this thesis.

The first issue refers to the lack of a dynamic analysis of demand on R&D decisions made by the industry. On the one hand, buyers' decisions are affected by the competition dynamics and supply-side conditions. On the other hand, the volume of demand and expected returns of industry are affected by (i) aggregate changes in the epidemiological profiles of the population, and (ii) the organisation of healthcare systems with potential long-term decisions on R&D investments and public policy implications. Finally, there is need to consider the interactions between demand and supply when assessing the impact of policies on pharmaceutical innovation. The existing empirical studies do not help to disentangle the policy effects from industrial trends, demand behaviour and firm characteristics. Following from this lack in the literature, we aim at developing a dynamic model that simulates the future R&D landscape by considering the impact of demand factors, industrial factors and firms' characteristics on R&D decisions, and that can further be applied to evaluate the impact of policy changes on the R&D landscape and availability of new therapies (Chapter 5).

In particular, further research is needed on the impact of drug approval requirements, patent rights, price regulation and clinical guideline policies on pharmaceutical innovation. Policy frameworks, such as health technology assessment and price regulation, have been implemented to contain an escalation of drug spending, by imposing guidelines for health professionals and by restricting patient choice at the retail level. However, little is known about the effect of these policy measures on pharmaceutical innovation. Healthcare systems, such as that of the UK, have implemented cost-effectiveness orientations on drug prescription lists to the National Health System. At the same time, the current coalition government is currently changing reimbursement and market access policies with potential significant consequences on R&D investment allocation in the UK, with potential spillover effects on the whole industry (Claxton et al., 2008). In the US, on the other hand, insurers have fixed drug formularies to restrict patient choices

and contain moral hazard but results remain unclear (Huskamp et al., 2000, 2003; Berndt, 2001). These policies have potentially a strong impact on the intertemporal trade-off between innovation and access to value-for-money medicines, not only within the countries where the policy changes have been set, but also in other jurisdictions, given the relevance of these markets for the pharmaceutical industry.

The second issue relates to the need to analyse the R&D process thoroughly, when assessing the factors that are impacting on the performance of the industry at innovation level. The innovation process in the pharmaceutical industry is long and shows high levels of heterogeneity across the different R&D stages. Most of the studies do not consider such level of heterogeneity when measuring R&D productivity and, because of that, lack understanding on what factors are imposing stronger influence on the performance of the innovation activity in the industry. By developing robust methodologies that account for the heterogeneity and time as important dimensions of the R&D process, we give further insight on the factors associated with project failure across the different stages of the R&D process in order to help designing policies that promote efficiency of the R&D process (Chapter 4).

Finally, there is insufficient research on the welfare implications of current policies to the pharmaceutical innovation. The evaluation of current policies and the design of better incentive mechanisms that promote innovation need to consider welfare and population health consequences. The literature is poor when including such consequences in a global perspective. It is true that to further discuss global health issues in the developing world we need intrinsically to address what happens in the most dynamic pharmaceutical markets, at demand, supply and policy level. However, the pharmaceutical innovation is characterised by being a multinational activity, causing significant spillovers across companies and regulatory jurisdictions. Because of that, it is needed a global perspective on the innovation activity of the industry, and their implications to the global population health. We propose to assess the inequalities of the access to new therapies, by measuring inequalities in global drug R&D activity in terms of global health needs in the last two decades (Chapter 6). This may inform the policy evaluation on the welfare and health implications, resulting from the current R&D process.

VII. Appendix: Tables

Table 2.3. Heterogeneity of outcome variables in the literature

Outcome variables to measure innovation performance		**Existing studies**
Intensity	R&D expenditure, in terms of sales	(Grabowski and Vernon 2000; Vernon 2004; Giaccotto, Santerre et al. 2005; Vernon 2005; Mahlich and Roediger-Schluga 2006; Vernon, Golec et al. 2009)
	Post-merger R&D investment	(Danzon, Epstein et al. 2007)
	Number of disease-specific scientific publications	(Bhattacharya and Packalen 2011)
	Market entry of firms	(Aharonson, Baum et al. 2007)
Probability of discovery	Number of New Chemical Entities (NCEs)	(Acemoglu and Linn 2004)
	Phase-specific success rates	(Danzon et al. 2005; Abrantes-Metz et al. 2004)
Launch of medicines	Number of new drugs launched	(Acemoglu and Linn 2004; Lichtenberg 2005)
	Probability of market launch	(Kyle 2006)
	FDA approvals	(Acemoglu and Linn 2004)
	Number of disease-specific drugs in the pipeline	(Civan and Maloney 2006)
	Number of existing disease-specific treatments, sold in the USA	(Lichtenberg 2005)
	Number of chemotherapy regimens	(Lichtenberg 2007)
	Number of "priority-review drug"	(Lichtenberg 2005)
Patents	Number of drug patents	(Boasson and MacPherson 2001; Kim, Lee et al. 2009)
	Number of patented compounds in the pipeline to be market registered	(Alexander, Flynn et al. 1995)
	Number of "important" drug patents (in US, Europe and Japan jurisdictions)	(Henderson and Cockburn 1996)
	Number of inventions patented	(Valentin and Jensen 2007)
	Patent application propensity rates	(Arundel and Kabla 1998)
	Successful patent applications	(Ulku 2007)

VIII. Appendix: Figures

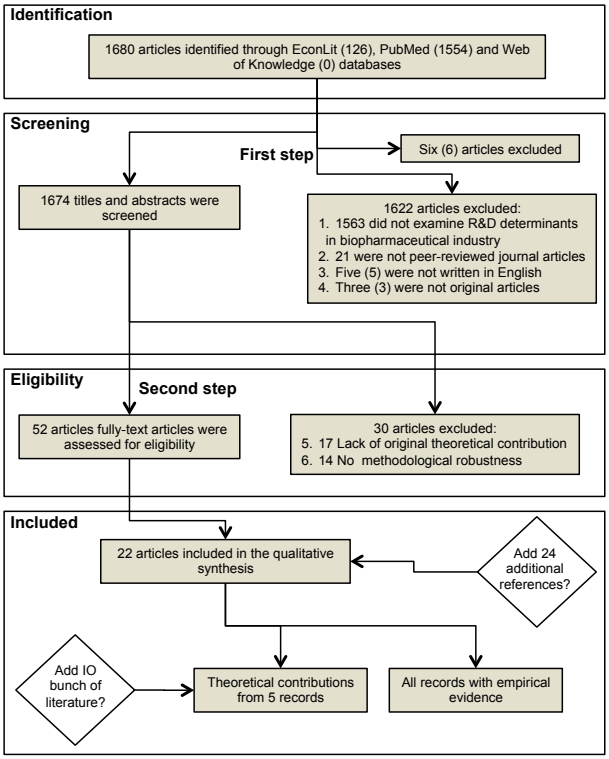


Figure 2.2. Preferred reporting Items for Systematic reviews and Meta-Analyses (PRISMA) flowchart

Chapter 3

Data and empirical methods of this thesis

In view of the fact that this thesis consists of a series of empirical essays on global pharmaceutical innovation, we have devoted this chapter to describe in detail the data and empirical methods that are an important contribution of this thesis. Indeed, the empirical work performed in the following chapters relies on a unique dataset that has required a large commitment on the management of several data sources, including its extraction, cleaning, merging and shaping. This large investment has been crucial to the econometric and quantitative analysis performed. In the following section, we have described the data sources and types of information used, including issues regarding the quality of data. In section II we have outlined the several assumptions considered in the construction of the dataset and application to the methods used in the different analyses. Finally, in section III we have presented the caveats and limitations drawn from the data and methods used. A more detailed outline and discussion of the empirical methods is presented in each chapter.

I. Description of data sources

We have built a unique panel dataset by merging five different sources of information that characterise globally pharmaceutical R&D activity and health outcomes. These are namely: (1) IMS R&D Focus from IMS Health (IMShealth, 2012); (2) Global Burden of Disease (GBD) study from the Institute of Health Metrics and Evaluation (IHME) (IHME, 2014); (3) ICD-10 codes from International Statistical Classification of Diseases and Related Health Problems (ICD) from WHO (WHO, 2014); (4) ranking of top pharmaceutical companies from ScripIntelligence (ScripIntelligence, 2014) and Fortune (Fortune, 2014); and (5) macroeconomic data from World Bank (WorldBank, 2014). We describe below the five data sources, the type of information provided and the issues regarding the quality of data involved.

Pharmaceutical R&D activity from IMS R&D Focus

IMS R&D Focus (IMShealth, 2012) is our major dataset and contains information on global pharmaceutical R&D activity from 1980 until 2012 (date of extraction). It is updated weekly and is typically used by pharmaceutical companies to monitor R&D activities of the competitors. The information is compiled from patent and regulatory filings, presentations at medical conferences, press releases, and information disclosed to financial analysts.

The dataset contains 34 051 pharmaceutical projects across all countries developing new compounds with information on the starting and ending dates of each R&D stage, namely early discovery, phase 1, phase 2, phase 3, pre-registration and registration, market launch and failure when it occurs. The dataset provides detailed information at six levels: (1) product (i.e. product name, trade name, generic name, laboratory code, CAS (Chemical Abstracts Service registry) number, substance origin, among

others); (2) drug indication/action fields (Anatomical Therapeutic Classification (ATC) code, therapeutic indication(s) (amongst 493), pharmacological activity, and the mode of administration), (3) companies involved in the R&D process of each project (including parent companies, academic research centres, hospitals, philanthropic organisations, and joint-ventures); (4) phase development history (namely the chronological record of R&D phases, the country assigned to each R&D phase, and the latest phase of development); and, (5) patent/franchise information (i.e. patentee and patentee nationality, the licensor and licensee).

Projects of new compounds consist of new molecular entities (NME) including small molecules, monoclonal antibodies, proteins, gene therapies, vaccines and immunotherapies, as well as fixed combination products, biosimilars, in vivo imaging agents, and specialized delivery systems (IMS, 2011). Therefore, 'me-too' drugs or other chemical generics are not recorded in this dataset and are not part of our analysis. Moreover, the compound projects can scientifically target several therapeutic indications simultaneously and consequently present several ATC codes.

Despite being a very rich dataset standardly used in research and for benchmark industry studies, there are potentially several issues surrounding the representativeness of the dataset with regards to R&D activity. The potential data quality issues relate to sample selection of the compound projects recorded and under-reporting of negative results of clinical trials. Data quality issues are plausible to exist due to three reasons. First, there tends to be secrecy at early stages of discovery when innovators, for strategic reasons, have an incentive to hide information about the novelties being tested (Scotchmer, 2004). This ceases when innovators have the incentive to protect their discoveries through patenting and information becomes public (Cohen and Levinthal, 1989; Scotchmer, 2004). This trade-off between the incentives of disclosure of information is more pronounced in the discovery and pre-clinical stage before any legal requirement imposes the registry of the ongoing investigational studies. Therefore, our data might fail to fully measure activity at early stages of the R&D process.

Second, and related to the above, some projects are firstly observed in the database at an R&D stage different from basic discovery and we do not observe the entire lifetime of the project. This could be partially related with the incentives of disclosure of information discussed above as well as the nature of the pre-clinical research that entails more scientific uncertainty and, sometimes, is not indication-specific. This is one of the reasons for the limited literature on success probability for preclinical stages (Paul et al., 2010). Therefore, we require the use of appropriate modelling options that account for left-censored data in the context of the duration models and the semi-parametric proportional hazard models.

Finally, the legal requirements of public registry of the ongoing investigational and clinical trials are relatively recent. The US Food and Drug Administration Modernization Act of 1997 (FDAMA or Public Act 105-115 (FDA, 2015b)) amended FDA Cosmetic Act to create a public information resource "ClinicalTrials.gov" (ClinicalTrials.gov, 2015) to track drug efficacy studies that was only made mandatory with the FDA Amendment Act (FDAAA) of 2007 (FDA, 2015a). In Europe, trial registration in the European Register (EudraCT) has been implemented since 2004 within the scope of the EU Clinical Trials Directive (EMA, 2015). Consequently, under-reporting of R&D activity data recorded by the industry is not uniform across the last 32 years.

While we can use appropriate methodologies (such as specification models that consider the use of left- and right-censored data (Chapter 4 and Chapter 5)) to mitigate the issues with left-censored data, there is little we can do to deal with under-reporting of R&D activity. However, we expect the data to be more affected by the problems of disclosure of information before the new regulations take place and

at the early stages of the R&D process. We will discuss in each relevant analysis how this caveat could influence our results.

Global Burden of Disease (GBD) study from IHME

The GBD study (IHME, 2014) contains up-to-date evidence on levels and trends for cause-specific mortality and disease burden readily available for 187 countries and facilitates studies that require cross-country comparisons. We detail more information about these measures in Chapter 6. The GBD study estimates yearly deaths and DALYs between 1990 and 2010 for a list of 291 diseases and injuries (IHME, 2009). However, the group of causes of death that relates to "external causes" for car accidents, falls, injuries, war, conflicts and intentional self-harm conditions were excluded from the dataset as these are not typically addressed by pharmaceutical R&D activity.

IHME uses the International Classification of Diseases (ICD) (WHO, 2014)[1] coding system to produce disease level estimates of global health outcomes. With this coding we merge global health outcomes with pharmaceutical R&D activity. The merged dataset of R&D activity with global health outcomes supports the analysis performed in Chapter 6 when measuring the dispersion of global R&D activity in terms of unmet health need.

International Statistical Classification of Diseases and Related Health Problems (ICD)

ICD data (WHO, 2014) contains codes for diseases, symptoms, and external causes of injury medically classified by the WHO. The coding set entails more than 14,400 different codes and allows tracking many new diagnoses. We have used ICD database[2], to merge the corresponding therapeutic indications of the compound projects in the IMS R&D Focus dataset with the GBD data. However, this correspondence is not always unique, given that each specific therapeutic indication may target more than one disease. This reflects the multi-usage of a drug in several health conditions and this fact imposes assumptions in the construction and management of the dataset that we will discuss in subsection II of this Chapter.

Fortune and ScripIntelligence

Fortune (Fortune, 2014) and ScripIntelligence (ScripIntelligence, 2014) are used to identify the ranking of the top100 pharmaceutical companies (in terms of revenues and profits) by Fortune in 2007 and by ScripIntelligence in 2011 (Fortune, July 2009; ScripIntelligence, 2013).

World Bank economic indicators

World Bank (WorldBank, 2014) provides us with a long time-series (1980-2014) of country level population and GDP data (US$ 2005 PPP).

II. Assumptions on the construction of the datasets

A. Unit of analysis - Innovation at therapeutic indication level

Since compounds often target several indications simultaneously, we have considered all therapeutic indications targeted. Therefore, our unit of analysis results to be at compound-indication project level.

[1] The ICD is the international standard classification of medical conditions for all general epidemiological, health management and clinical purposes. This coding system provides the basis for the compilation of national mortality and morbidity statistics by WHO member states.

[2] version 2010 available in http://apps.who.int/classifications/icd10/browse/2010/en

Moreover, the same compound project might be in the R&D pipeline in several countries simultaneously. This happens most likely when firms run multi-centre clinical trials. Because of this, we have considered all countries where the same compound-project has been developed as one observation. Consequently, our dataset expands to 65,845 country-compound-indication projects. Each country-compound-indication project is assigned to one of the 17 anatomical main groups (for example, dermatological conditions) and one of the 199 more specific classes (such as anti-psoriasis treatments) using the Anatomical Therapeutic Chemical (ATC) classification system used by the European Pharmaceutical Market Research Association EphMRA (2013) and that is standardly used in datasets such as the IMS R&D Focus to classify the different compounds. This is the basis for the definition of the relevant market that we explain in subsection C.

B. The construction of the datasets

Time-to-event data panel *(Chapter 4 and Chapter 5)*

We have constructed a time-to-event panel data that supports the analysis of Chapter 4 and Chapter 5 (for the estimation of the parameters that inform the micro-simulation model) by merging the five different datasets described in section I. This time-to-event panel data contains information for each compound-indication project for the time (in years) spent within each R&D phase (state duration), and the dates of the occurrence of all events, including transition between states, market launch or failure.

Each observation is a country-compound-indication-year project, reflecting the country, stage of development in the R&D pipeline and the time spent (duration) in that stage. For each R&D stage, a project is 'at risk' of failure from the year of its first entry into that R&D stage. If failure happens, then the project ceases being considered at risk in our analysis. Also, if a project moves to the subsequent R&D stage, then it ceases being considered within the risk set for that R&D phase.

To perform this analysis we have considered solely projects that do not present any biological component. Industry reports (PWC, 2011a) show crucial differences between the R&D process of non-biologics and biologics. Also, comparison between these must be cautious given the differences in sample sizes, production costs, development times and regulatory framework (DiMasi and Grabowski, 2007). The original 18 252 NCEs expand to 81 142 compound-indication year observations.

We have further information about the variables used in this analysis including failure, duration, market level, firm level and project level characteristics of each observation that we describe in section D and we discuss the assumptions taken for the definition of the product market in section C.

R&D portfolio in 2011 *(Chapter 5)*

The baseline dataset that is used as the base year for the simulations of Chapter 5 corresponds to the R&D portfolio in 2011 (that consists of the last full year extracted). The total number of projects in 2011 is 728 consisting of 667 NCEs and 61 biologicals.

We have defined market as the region-compound indication where a specific product will be at risk of being launched. For the therapeutic classification, and as used in Chapter 4, we classify the projects making use of the therapeutic categories presented by the EPhRMA[3] $3rd$ level ATC Classification.

For the definition of region, we have grouped countries into five main regional markets worldwide. We have done these for three main reasons: i) firms' strategy on market launch tends to be at regional level due to the similarities of health systems within regions and their strict and specific regulatory requirements;

[3]European Pharmaceutical Market Research Association

ii) countries that are part of each region are shown to be similar in the relevant characteristics, namely regulatory environment, IP rights, and technological conditions; iii) considering countries rather than regions would entail larger amount of missing information given that in many countries there is no R&D activity in many therapeutic areas. The five different regions are respectively: (1) USA; (2) the EMA members, which include the 28 European Union Member States as well as countries that belong to the European Economic Area (including Norway, Iceland and Liechtenstein), Canada, and Switzerland; (3) Japan; (4) Emerging Markets and Asia-Pacific (EMAP), including Brazil, India, Russia (and former Soviet Union), China, South Korea, Mexico, Nigeria, Israel, Vietnam and Taiwan ; and (5) the rest of the world.

Global R&D and global health need in 1990 and 2010 ***(Chapter 6)***

The merging of IHME and IMS R&D Focus data consists of linking therapeutic indication(s) of the compound projects to a disease using the ICD-10 code as we explain above. The final dataset includes a total of 59,301 country-compound-indication projects after matching non-missing therapeutic indications and ICD-10 codes (dropping 6,544 (0.099%) records with missing therapeutic indication). This dataset aggregates information at disease-level on the R&D activity, global health need and affordability. We have information about the variables used to measure these three dimensions in section D.

C. Product market definition

With the ATC classification system, products are classified into groups at five different therapeutic levels. The first level of the code indicates the anatomical main group. The second level of the code indicates the therapeutic main group. The third level of the code indicates the therapeutic/pharmacological subgroup. The fourth level of the code indicates the chemical/therapeutic/pharmacological subgroup. Finally, the fifth level of the code indicates the chemical substance.

We have used ATC code as the basis for market definition since it has been regularly used by competition authorities in the US and Europe (Backhaus, 2012) as well as in academic research (see, for example, Kyle (2006) and Backhaus (2012)). Within ATC the ATC 3rd level is the one commonly used in the definition of relevant market. ATC 3rd level identifies the therapeutic/pharmacological subgroup of medicines with the same therapeutic properties for a given disease or family of diseases (Backhaus, 2012) and indicates one from the total 262 therapeutic pharmacological subgroups to define the relevant market. This implies that medicines clustered in other therapeutic classes are not considered to be substitutes.

D. Variables

A complete list of observable characteristics of the compound-indication projects can be found in Table 4.1. They include the variables of occurrence of events, duration of the compound projects and three other levels of variables that characterise the compound projects that are respectively: i) market-level characteristics; ii) firm-level characteristics; and iii) project-level characteristics.

[Table 4.1 here]

R&D activity

Events: failure, transition to the next R&D stage and market launch

Failure occurs whenever the R&D activity for a project is interrupted. While we denote it as a failure there are many possible reasons why projects are abandoned. Failure of innovation may be due to a

combination of regulatory pressures, scientific/clinical non-achievements, or even strategic decisions of firms that withdrawal the project. Failure is the central topic of the analysis performed in Chapters 4 and 5, and therefore requires the construction of variables that identify the occurrence of failure. We have constructed a variable *failure* which identifies the occurrence of failure of a project that is R&D stage-specific. We have assumed that failure is a permanent condition that prevents future progression to any subsequent R&D phase (i.e. there is no *resurrection*). Given this, we have four variables that take the value '1' if failure happens in one of the phases $\{discovery, Phase1, Phase2, Phase3\}$, respectively $\{f_{discovery}, f_{Phase1}, f_{Phase2}, f_{Phase3}\}$, and the value for '0' to the non-occurrence of such an event (nothing happens, i.e. the project remains in the same R&D stage as in the previous period).

The successful progression to the next R&D phase is mutually exclusive of the occurrence of failure and is captured by the construction of four other R&D stage-specific variables. That means that for each of the R&D phases $\{discovery, Phase1, Phase2, Phase3\}$ we constructed a variable *success* that takes the value '2' if there is a successful progression to the next R&D stage, respectively $\{s_{discovery}, s_{Phase1}, s_{Phase2}, s_{Phase3}\}$. For the particular case of the value '2' in phase 3 indicates a successful transition of the drug project to the market. We have assumed that there is no regression in the R&D progress, i.e., a project cannot revert back from Phase 3 to Phase 2 of clinical trials, for instance. This is a plausible assumption given the strong level of regulation in place at each R&D phase as well as the nature of the process itself.

The above implies that our data is inherently right-censored since we do not necessarily observe the occurrence of an event. It may be the case, and that happens surely for the compound projects in the R&D pipeline that are current R&D investments, that a project neither fails nor progresses to the next R&D stage. The *'nothing happens'* scenario is captured by the value '0' in the variable *failure* and it is one of the reasons for using duration models in the analyses performed in Chapters 4 and 5.

R&D activity: total and market launches

For the analysis performed in Chapter 6, we have measured pharmaceutical R&D activity over two dimensions: (i) the overall R&D activity accomplished, by counting the disease-specific total number of projects (as done in the existing literature, for instance, Civan and Maloney (2006)); and (ii) the successful R&D activity, i.e. projects that are translated into new medicines (*disease-specific market launches*), by counting the disease-specific market launches (as in Acemoglu and Linn (2004) and Lichtenberg (2005)). With these, our analysis accounts for both the dispersion of total R&D activity and the dispersion at market launch level.

Duration

We have constructed a variable *duration* that captures the time dimension of the R&D process and is equal to the length of development of a compound project. The duration is equal to the time elapsed between the starting of the project (since it is observed in the dataset) and the occurrence of one of two definitive events: failure and progression to the next R&D stage (market launch in case the project is in Phase 3 trials). In the first case, and for the compound projects that fail to succeed in the R&D process, we observe the time elapsed until failure occurs (time-to-failure), whilst for the second case we observe the duration in the different R&D stages (i.e. the sum of the durations in the different R&D stages) in case progression in the R&D process occurs. For the projects we do not observe any of these events, the duration is equal to the time elapsed between the starting of the project in the dataset and the last year of data (2012). The duration is a critical piece when modelling hazard functions, as we do in Chapters 4 and 5, since it translates the risk of occurrence of an event of interest per unit of time.

Market-level characteristics

The empirical analyses performed in Chapters 4 and 5 consider the effect of market-level characteristics at three levels, namely: (i) competition; (ii) market size and affordability; and (iii) country-specific regulatory and technological environment.

Competition

Since we are interested in the effect of competition in the risk of failure of compound projects, we have constructed two levels of competition variables building on the existing literature (as outlined in Chapter 2 and explored empirically in Chapter 4): (i) competition in the final product market, and (ii) competition within the R&D process. Product market competition for each year is measured by: (i) the number of new drugs, i.e. established in the market in the last five years (*new drugs*) (ii) the number of old drugs, i.e. established in the market for more than five years for each relevant market (*old drugs*). We have used a five year cut-off point since it corresponds to the number of years of market exclusivity granted by FDA when extending the patent for innovators. Competition in the R&D process for each year is measured by the number of potential entrants in the same market (*potential competitors*), i.e., the number of projects being developed for the same market in each calendar year.

Market size and affordability

A measure of market size is incorporated in the model specifications of Chapters 4 and 5, and is used to proxy affordability in Chapter 6 when assessing R&D inequalities. This measure of market size and affordability is constructed by merging IMS Health R&D with country-level World Bank data on GDP per capita ($US 2005 PPP) (*GDPpc*) and population data from 1980 to 2012 (*population*).

Moreover, and for the inequality analysis performed in Chapter 6, we have measured country-disease specific health outcomes to proxy health need and we constructed a measure of national affordability. To do this we have merged the recently released 1990-2010 GBD data at disease-country level from IHME to gather information on the disability-adjusted life years *(DALYs)* and mortality levels *(deaths)* and rates for each disease at national level. To develop a measure of *affordability*, we have constructed a measure of country specific ability-to-pay by combining country specific health need (*'DALYs','deaths'*) and a measure of country specific income, using World Bank statistics on national Gross Domestic Product per capita *(GDPpc)*.

Country-specific regulatory and technological environment

The country-specific fixed effects capture systematic differences across countries such as differences in the regulatory and policy framework, and specificities of the technological environment. For this we have constructed country-specific dummies that identify the country where the compound project has been developed (*targetcountry*). Moreover, time trends have been considered in the analysis performed in Chapter 4 and time dummies have been constructed to control for time differences in the R&D process and likelihood of failure of R&D projects *(time dummies)*.

Firm-level characteristics

We have controlled for the nature of the parties involved in the R&D process using two variables: i) the participation of a big firm in the R&D project by characterising a big firm as one of the TOP100 firms *(big firm)*, using Fortune's ranking in 2007 and ScripIntelligence's ranking in 2011 to identify top100 Pharmaceutical companies (in terms of revenues and profits) (Fortune 2009; SCRIP Intelligence 2013); and (ii) the participation of academia, namely universities, hospitals, research centres, philanthropic and other non-profit organisations and public initiatives in the R&D process *(academic participant)*.

Project-level characteristics

Projects have been characterised at two levels, namely: (i) intensity of alliances; and (ii) therapeutic specificities of the R&D projects.

To characterise alliances, we have measured alliances by considering the logarithm (to consider non-linearities on the effect of the number of collaborators) number of firms collaborating in the R&D project *(intensity of alliances)*. We have constructed therapeutic indication dummies (*therapeutic specificities*) to control for systematic differences between therapeutic classes, and different technological and scientific specific conditions in each therapeutic category that could influence effort and the probability of failure (Mestre-Ferrandiz et al., 2012).

III. Implications of data limitations

There are mainly four caveats driven by the quality of our data.

First, data is not necessarily representative of the whole R&D activity due to lack of disclosure by firms and potential selection bias in terms of the type of projects being recorded. Indeed, as discussed above, it is plausible that for strategic reasons firms keep secrecy at early stages of discovery in order to hide information from competitors. This could imply under-representation of number of projects developed as well as under-recording of failure. If this is likely to exist it is plausible to presume that it is more pronounced in early stages of discovery before the incentive to patent happens. Also this caveat is more likely to affect the early years of our data corresponding to the years in which there was no compulsory regulation on the publication of data on clinical trials (before 2004 in Europe and 2007 in the USA).

The under-reporting of activity might also lead to selection bias. Indeed, in early discovery, firms are less likely to disclose information about compounds with a high likelihood of success since secrecy is important at this stage. While firms could also have a strategic incentive to non-disclose failure of projects it could be argued that the incentive for non-disclosure is more likely to happen for successful projects. Indeed, there have been over the years a number of initiatives in which pharmaceutical companies disclose failure projects[4].

For this reason, the recorded failures may be systematically different in nature from the ones that succeed, and this nature cannot be captured by the observable characteristics of the projects (for instance, we do not have information about the level of complexity of each project). This means that unobserved characteristics of the projects may be correlated with the observables, and the estimates in Chapters 4 and 5 may present some bias. Another issue relates to the left-censored nature of the data, related with late recording in the IMS dataset and presents incomplete information about transition between the R&D stages. While this issue should be taken into consideration in the interpretation of the results, we have adopted proper econometric methodologies to mitigate the issues arising in left-censored data discussed in Chapter 4.

Secondly, and because we have considered the different indications targeted by a compound as a different project, there is a potential estimation bias of failure. This is particularly true if multiple indications can be contained in one drug (and so one indication would be enough to characterise the project in terms of complexity and market), and we might be inaccurately estimating all the effects of variables of interest on failure. However, we believe that compound projects targeting simultaneously several indications are operating in different markets. In that case, our analysis makes sense to be

[4]GSK initiative is a good example: http://www.thesgc.org/about/what$_i$s$_t$he$_s$gc.

indication-specific.

On the other hand, if the same compound project is being developed in more than one country, and because we are considering these as separate projects, we might be over-estimating failure. Indeed, we might be under-estimating the positive network spillovers that these type of projects might generate given their multinational nature. Nevertheless, there are pretty strong documented locational effects on R&D that are captured by the country where R&D is taking place which supports our country specific analysis.

Thirdly, the use of country-specific measures (gdppc, population and country-specific died effects) might not suffice to measure market size for new therapies. Ideally, we would have market size proxied by some measure of affordability constructed with global disease burden weighted by ability-to-pay at country level. However, there is no time-series data for country-specific disease burden. The only information we have available for the time period of the analysis (1980-2012) is country-cause-specific mortality. However, mortality is a very crude proxy for health need. Mortality would over-represent conditions that affect low- and middle-income countries, and would under estimate need related to non-fatal conditions (such as chronic conditions) that are top drivers of burden of disease in high-income countries. Since most profitable markets are those disease areas that affect quality-of-life of the most developed countries[5] using mortality would entail a misrepresentation of the true market size. Moreover, there is no disease burden data at indication level. We would definitely misrepresent the market size of conditions by using a disease-grouped level estimate. Also, definitely there is lack of information about rare diseases and other under-represented conditions. The use of disease burden data to proxy market size would be biased towards more prevalent and most profitable diseases.

Therefore, we have used GDP per capita as a measure for buying power of the relevant market. We do not only consider the GDP per capita of all countries where the drug has been launched, but also the countries where the R&D process has taken place to have a measure of the markets where failure is more likely to happen. The location of the R&D activity translates most likely the country to which the therapy is aiming at being launched, since our data shows very low proportion of relocation of projects.

It is plausible to consider country-specific measures at R&D level because of several reasons. First, the location of the R&D and the first market launch (and before CROs have become relevant in the global R&D context) seems to be strategically chosen to recoup the investment made. Even if companies launch products in multiple markets it seems plausible to presume that the first market they choose is the most profitable one to allow them recouping the investment made. Also, and in the EU context in particular, given the inter-linkages of price regulations and strategic sequential launching we can expect that companies launch drugs in markets where they can price at higher levels. Despite the harmonisation at regulation level between the US, Europe and Japan, these price regulations are country specific, and associated with patenting, regulation of clinical trials and market registration. Secondly, the public funds targeting scientific and investigational discoveries in life sciences are national and dependent on domestic R&D policies. The National Institute of Health in the US, and Wellcome Trust and Cancer Research in the UK are examples of nationally designed institutes with a large impact on the domestic research capacity. This results in R&D investments being affected by country specific economic and technological contexts.

Fourth, while this is widely used in policy, regulation and academic research, product market definition

[5]in 2013 ScriptIntelligence reports the four drugs with largest sales in the US market: 1st is antipsychotic; 2nd Nexium is used to treat symptoms of gastroesophageal reflux disease; 3rd Humira (adalimumab) is used to treat rheumatoid arthritis, juvenile idiopathic arthritis, psoriatic arthritis, ankylosing spondylitis, and plaque psoriasis; 4th Crestor "statins", reduces levels of "bad" cholesterol.

using within ATC codes might misrepresent the market where a specific compound is operating. Indeed, there are two potencial caveats arising from this market definition. The first relates to the fact that there are products clustered simultaneously in several ATC-3 codes because they can treat several diseases. A second caveat arises from the fact that medicines clustered in the same ATC-3 code are not always competing since medicines potentially present a high level of differentiation and target very different and unrelated conditions[6]. While acknowledging these caveats, the lack of research on the definition of relevant markets in the pharmaceutical sector combined with the common usage of this definition renders it sensible its usage in our analysis.

Indeed, the definition of the relevant market in pharmaceutical products requires further research (Mihaescu and Rudholm, 2013) however given the existing literature and the practice of competition authorities we assume the same ATC3 code level for the market definition.

Besides these limitations, this thesis provides insightful analyses explored into detail in the following Chapters. Moreover, it gives room to an avenue of future research seeding from these analyses and that can be improved with additional data at firm and project level. We explore these in detail in each of the following Chapters.

[6]An anecdotal example from competition law reports the use of crude market definitions by the European Commission to assess market investigation of Teva/Barr (Case No COMP/M.5295 - TEVA / BARR REGULATION (EC No 139/2004) about several products including Tamoxifen. Tamoxifen is mainly used in the treatment of post-menopausal breast cancer (ATC L02), and also to treat infertility in women with anovulatory disorders (ATC G03), gynecomastia (abnormally large mammary glands in men) (ATC G,L), bipolar disorder (ATC N05) and angiogenesis (development of new blood vessels) (ATC C1X)

IV. Appendix: Tables

Table 3.1. Definition of variables across the datasets

Variable	Definition	Frequency	Chapter
failure	$\{f_{discovery}, f_{Phase1}, f_{Phase2}, f_{Phase3}\} = 1$ if occurrence of failure	R&D stage-specific compound project	Chapters 4 and 5
success	$\{s_{discovery}, s_{Phase1}, s_{Phase2}, s_{Phase3}\} = 2$ if there is a successful progression to the next R&D stage	R&D stage-specific compound project	Chapters 4 and 5
total disease-specific R&D activity	number of disease-specific compound-projects	drug compound project	Chapter 6
disease-specific market launches	number of disease-specific market launches	drug compound project	Chapter 6
duration	Length of time elapsed from the start of the compound project	Therapeutic class-year	Chapters 4 and 5
new drugs	Count of drugs in the market launched for less than 5 years	Therapeutic class-year	Chapters 4 and 5
old drugs	Count of drugs in the market launched for more than 5 years	Therapeutic class-year	Chapters 4 and 5
potential competitors	Count of R&D projects in the same market	Therapeutic class-year	Chapters 4 and 5
population, total	Population in 10s of millions	Country	Chapters 4 and 5
GDPpc	GDP per capita in US$1000s (constant 2000 US$)	Country	Chapters 4, 5 and 6
DALYs	disease-country disability-adjusted life years	Country	Chapter 6
deaths	disease-country mortality	Country	Chapter 6
affordability	combination of national health need (*DALYs,deaths*) and a measure of national income, *(GDPpc)*	Country	Chapter 6
targetcountry	Country-specific regulatory and technological environment where R&D is based	Country	Chapters 4 and 5
time dummies	time dummies	year	Chapters 4 and 5
big firm	Project has at least one TOP100 firm collaborator (dummy variable)	drug compound project	Chapters 4 and 5
Academia participation	Project has at least one academia collaborator (dummy variable)	drug compound project	Chapters 4 and 5
intensity of alliances	Log of total number of collaborators	Therapeutic class-Year	Chapters 4 and 5
therapeutic specificities	Therapeutic class that the project targets	Drug compound project	Chapters 4 and 5

Chapter 4

The determinants of failure in drug development: a duration analysis

I. Background

Product innovation in the pharmaceutical industry is costly, risky and time-consuming. With a decreased number of NCEs being discovered, and research and development (R&D) costs representing a high proportion of the revenues of the industry, the pharmaceutical industry is facing unprecedented challenges to its R&D model (PWC, 2011a,b; Paul et al., 2010). The industry's profitability and growth prospects are also under pressure as the finance of healthcare systems comes under increasingly intense scrutiny. Two of the key determinants of drug R&D activity costs are success rates and development times. Given the long, uncertain and multi-stage process of developing a new drug, understanding failure rates is key to better understand pharmaceutical industry performance, the magnitude of the long-term investments involved in R&D, and improving investment activity in the future (see, for instance DiMasi et al. (2003) or Mestre-Ferrandiz et al. (2012)).

The multi-stage nature of the drug R&D process is characterised by the regulatory criteria established by policy makers to ensure safe, efficacious and accessible drugs for consumers. These regulations imply that the successful completion of each development stage requires different amounts of resources, diversified scientific knowledge, and distinct competences from firms. The stages of R&D are therefore very heterogeneous in duration, scope, investment requirements and probability of success (Mestre-Ferrandiz et al., 2012).

Despite the evident heterogeneity across the different stages, the vast majority of the literature in this area tends to consider the R&D process as a "black box" when evaluating the determinants of failure of research projects. We believe, however, that it is important to unpack the R&D process by analysing the key success factors of each stage of the development process in order to design more focused policies and incentives that foster successful R&D. In this chapter, we measure the association of the probability of failure in the different R&D phases with the potential factors that can affect innovation (as outlined in Chapter 2 and briefly summarised below) with special focus on the role of competition and alliances. We do so by modelling the probability of failure of a project in each stage of the R&D process as a function of its R&D history considering the timing, duration and country of that stage.

Literature on the determinants of pharmaceutical innovation

There are three strands of literature that relate to our analysis. The first consolidates the following as determinants of innovation: at an aggregate level, epidemiological and income distributions in more dynamic economies matter for innovation given the extent to which they determine market size (see, for instance, Acemoglu and Linn (2004) or Lichtenberg (2005)); at the industrial level, the intensity of competition between pharmaceutical companies (see Giaccotto et al. (2005), Grabowski and Moe (2008) or

Mahlich and Roediger-Schluga (2006)), the presence of economies of scale and technological specialisation (Henderson and Cockburn, 1996), and alliances between firms (Aharonson et al., 2007); and at the firm level, the future commercial profitability of the drug project (Giaccotto et al., 2005; Grabowski and Vernon, 2000a; Vernon et al., 2010), and firm characteristics such as size (Alexander et al., 1995), age (Kim et al., 2009), location and nationality (see, for instance, Hirai et al. (2010)).

In all this literature, the process of R&D process is considered a "black box" and the different stages of the R&D process are not assessed separately.

A second strand of literature focuses on the measurement of success rates for the various stages of the R&D process using a diversity of methods and datasets (Adams and Brantner, 2006, 2010; DiMasi et al., 2010, 2003, 1991; Kola and Landis, 2004), with only one study providing detailed information on discovery success rates (Paul et al., 2010). These contributions show that drug success rates differ across the different stages of the R&D process and that the failure rates of the clinical stages have been increasing over time. For example, the most comparable studies report success rates for Phase 1 to be 71% (DiMasi et al., 2003) and 65% (DiMasi et al., 2010); for Phase 2 to be 44% (DiMasi et al., 2003) and 40% (DiMasi et al., 2010); and for Phase 3 to be 69% (DiMasi et al., 2003) and 64% (DiMasi et al., 2010).

Even though these studies provide valuable insight regarding the heterogeneity of the R&D process they do not address the factors associated with the success of each phase of the process.

The determinants of success in the different stages of the R&D process are assessed by a third strand of the literature that accounts for the project history and characteristics in the analysis of phase-specific success rates.

Danzon et al. (2005c) analyse the effect of alliances and firm experience on the phase-specific probability of success of projects for 1,910 compounds developed by US biopharmaceutical firms between 1988 and 2000. They find evidence of diminishing returns of firm experience in late stages of the R&D process; diseconomies of scale in Phase 3; a positive effect of alliances on the probability of success in Phases 2 and 3; and evidence of knowledge spillovers across firms in Phase 1.

Kyle (2006) analyses all drugs developed in the 28 largest pharmaceutical markets between 1980 and 2000 and finds that several of the characteristics of entrants and incumbents are positively associated with the time-to-entry in the G7 markets. When accounting for country-specific demand factors, competition appears to be negatively correlated with the likelihood of entry. Indeed, the impact of drugs that have been for longer in the market seems to be greater than that of more recently introduced drugs. Also, markets with many drugs being developed in the pipeline experience more entry.

Pammolli et al. (2011) examine the association between phase-specific R&D productivity and portfolio composition and regional location of R&D investments using more than 28,000 compounds investigated since 1990 in the US and Europe. They find that lower probability of success is associated with reorientation of R&D investments to riskier and highly uncertain therapeutic areas. Also, no productivity gap is found between companies based in the United States and Europe. They also present the most recent estimates of failure rates for Phase 1, Phase 2 and Phase 3 that are 25%, 52% , and 29% , respectively.

Finally, the Federal Trade Commission (FTC) uses information of all drugs that initiated the registration process through the US Food and Drug Administration for the first time between 1989 and 2002 to analyse the association of drug's observable characteristics (such as therapeutic group, route of administration and originator size) with its pathway through the three stages of clinical trials (Abrantes-Metz et al., 2004). The authors find that the duration of the R&D process has decreased from 1995 to 2002; drugs with longer durations are less likely to succeed, as well as drugs developed by smaller firms.

Our analysis relates to these contributions in that we i) measure failure rates in each stage of the R&D process and ii) analyse the characteristics associated with project failure across the different stages of the R&D process by estimating how phase-specific R&D failure rates correlate with competition and alliances. However, our analysis departs from these studies in two significant ways.

First, in contrast with some earlier contributions, we use two different semi-parametric proportional hazard models to estimate the impact of market structure and alliances on the phase-specific failure rates by considering the history of the R&D process, and the duration of the drug projects in each phase of the process. In line with Kyle (2006) and Abrantes-Metz et al. (2004), we believe this methodology fits more closely the dynamic and lengthy nature of the R&D process than the logistic regression models used by Danzon et al. (2005c) and Pammolli et al. (2011), which do not consider project duration as a potentially relevant part of the failure process.

Secondly, we use a much richer dataset with global data from 1980 to 2012 across all therapeutic areas, and we analyse the influence of competition and alliances on the failures rates in each phase of the entire R&D process from discovery to market launch. Kyle (2006) focuses on the market conditions that affect the probability of launching a drug into the market whereas Abrantes-Metz et al. (2004) and Pammolli et al. (2011) assess the role of drug observable characteristics (i.e., therapy category, route of administration) and company size and location on the success of transition in clinical phases. Danzon et al. (2005c) focuses on the role of alliances and firm experience on the success of clinical trials.

We are primarily interested in modelling the association between the probability of failure of the projects in any R&D phase and industrial level determinants, namely competition and alliances. The nature and timing of competition and alliances between firms, manufacturers, scientific community, laboratories and academia may foster or hinder innovation. Indeed, measures of competition and alliances have been widely used to explain success and launch of new drugs in the market and are, therefore, likely to also influence the other stages of the R&D process.

A further step, however not part of the scope of this chapter, would be to find causal relationships between these variables. However, this analysis is out outside the scope of this Chapter given the data limitations related with sample selection and few information at firm and project level.

The role of competition

Economic theory has explored the unstable relationship between competition and innovation (Aghion et al., 2005; Scherer, 1967). On the one hand, competition may increase firms' incentives to innovate in order to "escape competition" ((Aghion et al., 2005), p. 3) and maximise the future expected profits. On the other hand, competition may exert an extra pressure on firms and discourage innovation (Aghion and Howitt, 1992; Romer, 1990).

The evidence explores this ambiguous relationship between market competition and innovation. Some studies find a significant positive effect of competition on drug innovation (Aharonson et al., 2007; Alexander et al., 1995; Giaccotto et al., 2005; Arundel and Kabla, 1998; Grabowski and Vernon, 2000a; Mahlich and Roediger-Schluga, 2006), whereas one study reports a negative significant effect of market competition on drug time-to-entry in the market (Kyle, 2006).

These contributions focus on two types of competition: i) competition in the final product market and ii) competition within the R&D process. Competition in the final product market is proxied by sales outside a company's headquarters' country (as percentage of total sales) (Arundel and Kabla, 1998; Giaccotto et al., 2005), firms' global market share (Alexander et al., 1995) and number of drugs established in the market (Kyle, 2006). Competition within the R&D process is proxied by industrial margins on

R&D investment per sales (see for instance Giaccotto et al. (2005) or Grabowski and Vernon (2000a)), and the number of drugs launched anywhere in the world in the same market (Kyle, 2006).

To the best of our knowledge, Kyle (2006) is the only contribution that explores simultaneously competition in the final product market as well as competition within the R&D process. The study demonstrates that competition within the R&D process stimulates entry, whereas competition in the product market has a negative significant impact on entry. Competition from drugs longer established in the market appears to a have a greater impact than that from more recently introduced ones. Also, markets with many potential competitors (number of drugs launched anywhere in the world) experience more entry.

Following Kyle's results on the role of competition on market launch, we hypothesize that both types of competition can also impact success of the different stages of the R&D process (Kyle, 2006).

Our hypothesis is that if competition within the R&D process influences market entry (i.e. the transition from Phase 3 to the market) then it is plausible to presume that it could also have a significant impact on the strategic decisions within the R&D process and, in particular, the decision to abandon a drug project. By testing this hypothesis we expect to identify important and significant differences of the effect of competition in different stages of the R&D process. In assessing the role of competition, we assume that current projects in a given market take market structure, as well as competitors' strategies within the R&D process, as given and compete simultaneously in time t.

The role of alliances

Though not conclusive, the literature suggests an important effect of alliances (private/public) in the drug R&D productivity. Two studies demonstrate significant positive effects of alliance on phase-specific R&D success rates (Danzon et al., 2005c) and drug launch times (Hirai et al., 2010), whilst a study concludes on negative effects of alliances with academia on R&D success rates (Aharonson et al., 2007).

To the best of our knowledge, only Danzon et al. (2005c) demonstrate a positive effect of alliances on phase-specific probability of success of projects. In particular, they show that alliances have a positive effect on the probability of success in Phase 2 and Phase 3. Building on this literature we will investigate the role of alliances on the failure of R&D projects at each stage of the R&D process. We expect a positive effect of alliances on R&D success, which may be offset by a negative impact of some types of alliances, i.e. with a public institution/university as some literature suggests (Aharonson et al., 2007).

The remainder of this chapter is organised as follows. In section II we present the specification model and estimation strategy. In section III we describe the data. In section IV we discuss the descriptive statistics and non-parametric analysis, and we present the main results from the semi-parametric analysis. Finally, in section V we provide a discussion of the results and conclusions.

II. Specification model and estimation strategy

Failure expresses the opposite situation of keeping open the option of investment in the future on a particular R&D project, and there are many possible reasons why failure happens. Failure of innovation may be due to a combination of regulatory pressures, scientific/clinical non-achievements, or even strategic decisions of firms that withdrawal the project. There are therefore technical and economic risks that impact on the likelihood of failure of a R&D project (Pennings and Sereno, 2011).

Studying the option of deferring a decision of keeping investing or abandoning a project is part of the nature of R&D investments. These investment decisions involve a substantial degree of uncertainty

about the future and an enormous level of irreversibility. This means that the timing of investment is critical under these circumstances and represents one the main dimensions of R&D decisions, impacting on development times and R&D costs (Palmer and Smith, 2000; Dixit, 1994).

Duration models focus on the analysis of time duration and the occurrence of events to statistically infer on the relationship between some factors and the probability of non-occurrence (survival) of a certain event. We use duration models to model R&D failure and to account for the dynamic nature of sequencing R&D process. In each R&D stage a project, which represents a compound for a particular indication, is at risk of failure since the first year the project entered in that R&D stage and it ceases being at risk of failure if one of two things happens: (i) it transits to the next R&D stage, including the market launch (anywhere in the world); (ii) it is discontinued by the firm. We assume that once failed the project may not be reactivated by the firm.

The probability of failure of a new drug component in the short interval of time dt after t, can be represented by the hazard function $h(t)$ (Lancaster, 1992) given by:

$$h(t) = P(\text{failure at time } t \mid \text{R\&D until time } t) = \frac{P(\text{failure at time } t)}{P(\text{R\&D until time } t)} \tag{4.1}$$

The hazard function $h(t)$ can be rewritten as

$$h(t) = \lim_{dt \to 0} \frac{P(t \leq T < t + dt) \mid T \geq t)}{dt} = \frac{P(t \leq T < t + dt) \mid T \geq t)}{P(T \geq t)} \tag{4.2}$$

If we represent the duration distribution function as $P(T < t) = F(t)$[1], where $t \geq 0$ at point in time t, and letting the probability density function to be $f(t)$, then the hazard function $h(t)$ at time t is given by:

$$h(t) = \frac{F(t + dt) - F(t))}{(1 - F(t)} = \lim_{dt \to 0} \frac{F(t + \delta t) - F(t))}{dt} \cdot \frac{1}{1 - F(t)} = \\ = F\prime(t) \cdot \frac{1}{1 - F(t)} = \frac{f(t)}{1 - F(t)} \tag{4.6}$$

This is the hazard rate of failure, and represents the instantaneous rate of failure per unit of time at t, conditional on the fact that the project has been in development up to time t. The hazard function can

[1] In the context of duration models,

$$F(t) = 1 - S(t) \tag{4.3}$$

where $S(t)$ represents the survivor function. $S(t)$ gives the fraction of projects that stayed at least t years in the R&D process. It can be written in terms of an integral involving the hazard function t equals the exponential of the negative integral of the hazard function between on the interval $[0, t]$:

$$S(t) = e^{\int_0^t h(u)\delta u}$$

Similarly, the hazard function $h(t)$ can be written in terms of a derivative involving the survivor function:

$$h(t) = -\frac{\frac{\delta S(t)}{dt}}{S(t)} \tag{4.5}$$

See T. Lancaster, 1992, pp. 6-10.

be rewritten as a function of X systematic observable characteristics of our interest:

$$h(t, X) = h_0(t)\theta(X) \tag{4.7}$$

where $h(t, X)$ is a function of $\theta(X)$, and X represents a set of relevant observable characteristics, that vary across calendar time. This enables us to model the association between failure rates and X covariates of interest, in our case, competition and alliances. We are interested in modelling the relationship between competition and alliances and the rate at which a project fails the R&D process after t, given that the project did not fail before t. In order to do so, we model the failure rates h_i^j from state i to state j, with j = failure and $i \in \{d, p1, p2, p3\}$, where d denotes discovery, $p1$ denotes Phase 1, $p2$ represents Phase 2, and $p3$ denotes Phase 3 trials.

The advantages of separately modelling the phases are that the covariates of interest may be more important in some phases than others, that some covariates change at the beginning of each phase, and that the quality of the data may differ throughout the different stages. For example, clinical trials conducted in patients must be registered in most national regulatory bodies, whereas data regarding at pre-clinical stages may be somewhat self-selected by companies that choose to share information. This means that we also may expect differences in the quality of the data across the different R&D stages as we have discussed in Chapter 3.

Given the supportive literature we expect a positive effect of potential market size on drug innovation, provided that we also control for market size. We seek to control for country-specific regulatory characteristics that, among others, capture systematic differences in regulatory and policy framework, and specificities of the technological environment. Moreover, given that the firm size is reported to have mixed effects on R&D productivity we also control for the size of company (Abrantes-Metz et al., 2004).

We adopt two modelling strategies: (i) the single risk hazard model; and (ii) the independent competing risks model. For both of them, we use the Proportional Hazard (PH) (Van den Berg, 2001) specification as the estimation strategy, with the assumption that there is no unobserved heterogeneity ($\tau = 1$).

Single risk hazard model

In the single risk hazard model, we consider the transition to success or failure as the process of interest:

$$h_i^j(t, X, \tau) = h_0(t)\theta(X)\tau = h_0(t)\theta(X) \tag{4.8}$$

where $h_i^j(t, X)$ is the hazard rate for failure from state i, where $i \in \{d, p1, p2, p3\}$, and X a set of relevant observable characteristics, that vary across calendar time. Also, $h_0(t)$ denotes the baseline hazard and $\theta(X)$ the systematic part of the hazard. The hazard function is allowed to differ across projects through the systematic part $\theta(X)$. This means that the population of projects is assumed to be homogeneous with respect to the systematic factors that affect the distribution of T. $\theta(X)$ gives the shape of the hazard function for any given project and can be specified as:

$$\theta(X) = \exp(X\prime\beta) \tag{4.9}$$

It is possible to consistently estimate β in the exponential part of the model, even though the baseline hazard function $h_0(t)$ is left unspecified. This ensures that the fitted model will always give estimated

hazards that are non-negative. The interpretation of the coefficient of β is that it measures the effect on the log hazard of a unit change in the value of X at time t. The PH specification model allows for a non-parametric baseline hazard $h_0(t)$. The latter is a function representing the duration dependence through which the probability of failure changes with the elapsed duration of one unit of time t.

We are interested in estimating $\theta(X)$, i.e. the systematic part of $h_i^j(t, X)$. $h_i^j(t, X)$ measures the instantaneous rate of failure of the projects active at time t that fail in the short interval from t to $t + dt$, in a large population of projects that are homogeneous with respect to X. The β parameters are estimated consistently by maximization of a partial likelihood function that does not depend on the baseline hazard function, which can be estimated non-parametrically (Lancaster, 1992). Further detail on this can be found in the VI.Appendix at the end of this Chapter.

Competing risks model

Even though the single risks model provides a good baseline analysis, the nature of the R&D process is such that a project in a given state can either remain in that state, move on to the next state or be abandoned. The possibility of the project progressing to the next R&D phase impedes the occurrence of failure and can be considered a competing event. Progression to the next R&D phase is not considered as a censoring event (such as censoring due to loss to follow-up or no event at all). For example, consider a project in discovery showing a progression to Phase 1 after three years. The single risk hazard model considers this project as being at risk of failing, even though it succeeds in progressing to the next R&D phase. In this model progression to Phase 1 is indistinguishable from loss of follow-up in discovery, and then considered censoring. In reality, though, a proportion of the projects that are considered censored in the first model, have progressed to another R&D phase (*competing event*), and therefore, should not be considered at risk of failing in that specific R&D phase. The single risk hazard model described above fails to mirror accurately this more realistic formulation of the R&D process. To address these issues, we have considered as a second modelling strategy, the competing risks model.

In the competing risks model, we consider two possible, mutually exclusive, destination states for each R&D project: failure and progression to the subsequent R&D phase. In other words, observations are simultaneously exposed to several competing risks. This model imposes two assumptions. First, that failure is a permanent condition that prevents future progression to any subsequent R&D phase (there is no *resurrection*). Secondly, that we do not observe regression in the R&D progress, i.e., a project cannot revert back from Phase 3 to Phase 2 of clinical trials. This is a plausible assumption given the strong level of regulation in place at each R&D phase.

Let T_j denote the time to the event of interest (failure), T_k, denote the time to the competing event (transition to subsequent R&D phase), and T_c the time to no event. Then the observed time-to-event T is given by

$$T = min\{T_c, T_j, T_k\} \tag{4.10}$$

Because we only observe one (the first) event, and so the minimum T, the joint distribution of $\{T_c, T_j, T_k\}$ cannot be identified by the data. Therefore, the probability of failure in $t + dt$ is given by:

$$\begin{aligned} h(t) = P(\text{failure at time } t \mid t+ dt) = \\ = P(t \leq T_j < t + dt \mid \text{survival to t and all other } \{T_c, T_k\} \geq t + dt) \end{aligned} \tag{4.11}$$

Formally, we model the transition rates $\theta_i^j(t_j \mid X, \tau_j)$ from state i to state j, with j= failure and $i \in \{d, p1, p2, p3\}$, and the transition rates $\theta_i^k(t_k \mid X, \tau_k)$ from state i to state k, with $k \neq i, j$ and $k \in \{p1, p2, p3, m\}$, being the subsequent R&D stage after stage i. The total number of projects that remain in the R&D pipeline at t which depart to one of the two destinations is given by:

$$\theta_i(t, X, \tau_j) = \theta_i^j + \theta_i^k \tag{4.12}$$

which gives us the sum of transition intensities over both destination states - failure j and subsequent R&D phase k. From there, we can calculate the contribution of each destination stage to the hazard function.

Analogously to the single risks model, we model transition rates with the MPH specification:

$$\theta_i = \begin{cases} \lambda_j(t_j) \times \theta_{0,j}(X)\tau_j & \text{, if } j \text{ happens} \\ \lambda_k(t_k) \times \theta_{0,k}(X)\tau_k & \text{, if } k \text{ happens} \end{cases} \tag{4.13}$$

Where X stands for the set of observed project characteristics that differ across calendar time, and $\{t_j, t_k\}$ the unobserved project characteristics. Conditional on X, the variables t_j and t_k are assumed to be dependent only if τ_j and τ_k are dependent. So, in the case of independence of τ_j, τ_k, the model reduces to two unrelated ordinary PH models of t_j and t_k where the baseline transition rates $\lambda_j(t_j)$ and $\lambda_k(t_k)$ are left unspecified.

We considered two specifications of the baseline hazard in the competing risks model: i) the first assumes that the baseline hazard for both types of risks (failure and progressing to next R&D phase) is identical; ii) the second assumes proportionality of both baseline hazards. The advantage of using additional information about the competing risk comes at a price, in the form of the assumptions needed to consistently estimate the β parameters. First, we assume that both risks are independent, after controlling for observed characteristics (Cameron and Trivedi, 2005). When assuming state independency and mutually exclusivity of the destination states, we can estimate β by maximising the overall log likelihood of the two events parts. Details about the specification of the log-likelihood can be found in the VI.Appendix at the end of this Chapter. Secondly, we are assuming that $\{t_j, t_k\}$ are project-specific effects and distributed independently of the regressors (exogeneity). Finally, the effects of the covariates X are assumed to be proportional (Van den Berg, 2001).

We run several specification tests after choosing which specifications are economically relevant, and those that minimize the Akaike criterion (Akaike, 1974) (See Appendix for more details). This criterion statistic is commonly used to compare the quality of different models and/or models with different numbers of parameters by assessing the trade-off between the goodness of fit and the complexity of the model. We estimate the goodness-of-fit and test for the proportionality assumption, which is a central assumption for our methodology (See Appendix). We use the *linktest* that tests the proportionality-hazard assumption[2] by interacting time with the covariates and verify that the effects of these interacted variables are not

[2]The proportional assumption is vital to the interpretation and use of a fitted proportional hazards model. The proportional hazards model has a log-hazard function of the form

$$\ln h(t, X, \beta) = \ln h_0(t) + X\prime\beta \tag{4.14}$$

It assumes that a plot of the log-hazard over time would produce two continuous curves, one for $X = 0$, that would be equal to $\ln[h_0(t)]$, and the other for $X = 1$, which is $\ln h_0(t) + \beta$. The difference between these two curves at any point in time are β, regardless of the shape of the baseline hazard function (Hosmer et al., 2011; Cleves et al., 2010).

different from zero. We expect that the effects are not different from zero because the proportionality-hazards assumption states that the effects do not change with time except in ways that we already parametrized (with the semi-parametric function of the baseline hazard). This is the nucleus of the proportional hazard diagnostics (Hosmer et al., 2011) (See Appendix).

We also check for data outliers when evaluating the fit of the model. We use the method of the *efficient score residuals* to identify observations with disproportionate influence on the fit of the model and unusual configuration of covariates (Hosmer et al., 2011) .

III. Data and variables

To perform this analysis we have built a unique time-to-event panel dataset that is extensively described in Chapter 3. We have merged IMS Health R&D Focus of pharmaceutical compound projects with World Bank data on country level macroeconomic data. We also use the Fortune ranking in 2007 and ScripIntelligence in 2011 to identify top100 Pharmaceutical companies (in terms of revenues and profits) (Fortune, July 2009; ScripIntelligence, 2013).

The IMS Health R&D Focus contains information on global pharmaceutical R&D activity from 1980 until 2012. The dataset contains all compound projects across all countries and therapeutic areas with information on the starting and ending dates of each R&D stage, namely early discovery, Phase 1, Phase 2, Phase 3, and market/registration (states). Additionally, the dataset contains information for each project for the time (in years) spent within each R&D phase (state duration), including the time elapsed if and until a firm abandons a project (which we label *failure*), and the dates of transition between states.

Compound projects are broadly defined to include small molecules, monoclonal antibodies, proteins, gene therapies, vaccines and immunotherapies, as well as fixed combination products, biosimilars, in vivo imaging agents, and specialized delivery systems (IMS Health, 2011). For this analysis we have considered only the projects that do not present any biological component. Industry reports (PWC, 2011a) show crucial differences between the R&D process of non-biologics and biologics. Also, comparison between these must be cautious given the differences in sample sizes, production costs, development times and regulatory framework (DiMasi and Grabowski, 2007).

Since compounds often target several therapeutic indications simultaneously, we have considered all indications targeted by each compound. Moreover, the same compound project might be in the R&D pipeline simultaneously in several countries. Because of this, we have considered all countries where the same compound-project has been developed.

This results in the fact that the effective unit of analysis is a compound-indication project, reflecting the therapeutic indication, the country, stage of development in the R&D pipeline and the time spent in that stage. For each R&D stage, a project is 'at risk' of failure from the year of its first entry into that R&D stage. If failure happens, then the project ceases being considered at risk in our analysis. Also, if a project moves to the subsequent R&D stage, then it ceases being considered within the risk set for that R&D phase.

Moreover, each observation is assigned one of the total 262 therapeutic pharmacological subgroups (third level of ATC code) to define the relevant market. To define the relevant market, and similarly to the competition authorities and other papers in the literature (see, for example, Kyle (2006)) we use the third level of the ATC code defined by the European Pharmaceutical Market Research Association (EphMRA, 2013).

We have a total of 18 252 projects of NCEs that expand to 81 142 country-compound-indication year observations.

Finally, as explained in detail in Chapter 3, some projects are first observed in the database at a R&D stage different from basic discovery. This is related with the disclosure of information about the pre-clinical research. We accommodate this issue by using appropriate modelling options in our semi-parametric models to account for the left-censored data as we have explained in the section II.

We have further information about the variables used in this analysis including failure, duration, market level, firm level and project level characteristics of each observation that we describe in the next section.

Variables

A complete list of variables labels and description can be found in Table 4.1. Our dependent variable captures the occurrence of a failure for each project, conditional on the R&D stage. Because we are separately modelling the R&D phases, four dependent variables are constructed with the value of 1 if failure happens in one of the phases respectively $\{f_{discovery}, f_{Phase1}, f_{Phase2}, f_{Phase3}\}$, and the value for '0' to the non-occurrence of such an event (nothing happens, i.e. the project remains in the same R&D stage as in the previous period).

The successful progression to the next R&D phase is mutually exclusive of the occurrence of failure and is captured by the construction of four other R&D stage-specific variables. That means that for each of the R&D phases $\{discovery, Phase1, Phase2, Phase3\}$ we constructed a variable *success* that takes the value '2' if there is a successful progression to the next R&D stage, respectively $\{s_{discovery}, s_{Phase1}, s_{Phase2}, s_{Phase3}\}$. For the particular case of the value '2' in phase 3 indicates a successful transition of the drug project to the market. We have assumed that there is no regression in the R&D progress, i.e., a project cannot revert back from Phase 3 to Phase 2 of clinical trials, for instance. This is a plausible assumption given the strong level of regulation in place at each R&D phase as well as the nature of the process itself.

[Table 4.1 here]

We have considered several explanatory variables. In particular, to proxy industrial forces, we include competition in the final product market, competition within the R&D process, intensity and type of alliances. To measure market size, we consider population and GDP per capita. Finally, we have also included country fixed effects to control for regulatory and technological environment country-specific characteristics.

As described in Chapter 3, market is defined using the third level of the ATC code that characterises a compound-indication project, which indicates one from the total 262 therapeutic pharmacological sub-groups to define the relevant market, similarly to other papers in the literature (see, for example, Kyle (2006)) and the competition authorities in the US and Europe (Backhaus, 2012).

To measure competition we have followed Kyle (2006) by considering competition in the final product market and within the R&D process. In particular, market competition for each year is measured by: (i) the number of *new drugs*, i.e. established in the market in the last five years (ii) the number of *old drugs*, i.e. established in the market for more than five years for each relevant market. The five-year period captures the exclusivity period that a NCE is granted by FDA that protects it from new competition in the marketplace[3]. This is because we believe that the market exclusivity affects firms' R&D decisions since it determines the period in which the incumbents have been granted protection to set higher prices.

[3]New Drug Product Exclusivity provided by the Food, Drug and Cosmetic Act under section 505(c)(3)(E) and 505(j)(5)(F), also known as the Hatch-Waxman exclusivity amendments

After that protected period is expired, candidates in pipeline will revise their potential market profitability since there are no longer exclusivity rights for price setting.

Competition in the R&D process for each year is measured by the number of potential entrants in the same market, i.e., the number of projects being developed for the same market in each calendar year (*potential competitors*).

To measure alliances we construct three variables to characterise the intensity and type of alliances at project-level. Namely we consider: (i) the log number of firms collaborating in the R&D project; (ii) the participation of a big firm in the R&D project by characterising a big firm as one of the TOP100 firms (*Big firm*); and (iii) the participation of academia in the R&D process (*Academia participant*).

To proxy market size one would, in principle, use global pharmaceutical sales data by broad therapeutic area or disease level incidence rates as discussed in Chapter 3. However, global pharmaceutical sales data is prohibitively costly. And, disease incidence levels are difficult to find across all therapeutic areas and countries for the time span considered in this analysis. Therefore, and following Kyle (2006), we have used population size and GDP per capita (GDPpc) at country level from World Bank data as proxies for demand. We have considered the GDP per capita (*GDPpc*) and population (*Population*) of the country in which the R&D process takes place. This is a plausible assumption, given the R&D distribution: concentrated in high-income countries (US, EU and Japan account for more than 90% of the total projects) in all R&D phases, and the almost negligible levels of relocation of the projects between R&D phases (1.1% maximum of projects are relocated in our data). We also explore non-linearity in population and GDP per capita to account for decreasing returns to scale.

We further control for other covariates that can influence the R&D process. In particular, we control for two relevant observable time-invariant attributes that have been used in the R&D literature that characterise systematic differences between the projects, namely: (i) the target therapeutic class (therapeutic class); and (ii) the home country for the R&D of the project (*targetcountry*).

By considering systematic differences between therapeutic classes, we allow for different technological and scientific specific conditions in each therapeutic category that could influence effort and the probability of failure (Mestre-Ferrandiz et al., 2012).

The country-specific characteristics capture, among others, systematic differences in regulatory and policy framework, and specificities of the technological environment.

IV. Results

A. Descriptive statistics and nonparametric analysis

Our sample consists of 18,252 projects, 4,230 of which have failed. Table 4.2 summarises the descriptive statistics across the years between 1980 and 2012 for failures (i.e. compound-indication projects that were disrupted or abandoned) and successes (i.e. compound-indication projects that successfully progressed to the next R&D phase), duration of the projects, competition, alliances and market size proxies.

[Table 4.2 here]

The data is consistent with the most recent estimates on failure rates by Pammolli et al. (2011). It shows a relatively higher proportion of failure (29.9%) for projects in the preliminary stage of discovery than projects at later stages of the process. This proportion decreases to around 14% in Phase 1 and Phase 2 of clinical trials. And, around 10% of the projects that are in Phase 3 trials fail to be launched in

the market. These numbers vary over time: higher annual percentage rates of failure in the 1980s, mainly driven by the unsuccessful experience of the non-US projects, and a declining trend in the late decade of 2000 (Figure 4.1). This may be partially due to more "births" of new projects: Figure 4.2 shows relatively higher annual percentage rates of start-ups between 1980 and mid-1990s driven by the number of new projects in the US.

[Figure 4.1 and Figure 4.2 here]

Also, when looking across ATCs, we observe that general systemic anti-infectives (ATC code J) account for roughly 26.3% of the total cases of failure in the entire R&D process. Moreover, looking closely at more refined 3rd ATC level categorisation, three ATCs concentrate the highest proportions of total failures: (i) antineoplastics (ATC code L: Antineoplastic and immonumodulating agents) account for more than 4.6% of total R&D failures; (ii) other central nervous system drugs account for roughly 2% of total R&D failures (ATC code N: Nervous system); and (iii) antivirals for systemic use represent more than 1.8% of total R&D failures (ATC code J: general anti-infectives) (Table 4.3).

[Table 4.3 here]

The mean duration of failures (5.82 years in discovery, 4.78 years in Phase 1, 5.52 years in Phase 2 and 5.59 years in Phase 3) is more than twice the mean duration of successes (3.66, 2.22, 3.08, and 2.70 respectively). Also, failures face on average fiercer competition than the average of successes: this is more pronounced when looking at the mean number of new drugs in the market and competitors in the R&D process. Moreover, failures experience on average lower degree of alliances in all phases of the R&D process except in Phase 3 (1.43 against 1.69 firms in discovery, 1.43 against 2.08 firms in Phase 1, 1.65 against 2.46 firms in Phase 2 and 2.46 against 1.78 in Phase 3). The participation of a big firm and academia is relatively higher in failures in early stages of development, when comparing to the successes. The population and GDP per capita of countries with more failures are, on average, higher than those with more successes.

We run log-rank tests to test for the differences in the relative survival experiences of distinct groups that can be constructed by looking at different levels of each covariate. The *logrank test statistic* compares estimates of the hazard functions of these different groups at each observed event time (Cleves et al., 2010). For instance, it compares the true hazard function of projects facing no market competition with the true hazard function of projects facing competition at some level (for example, 1 competitor). Our results suggest that we can reject the null hypothesis that assumes no difference between the true survivor functions for the different groups of drug projects that face different intensities of competition, as well as intensities and type of alliances. Results from the non-parametric analysis support many of the descriptive statistics results and anticipate our conclusions. The several Kaplan-Meier (KM) survivor functions show survival estimates of projects that face different levels of competition and experience different levels of alliances, across the R&D phases.

The KM estimates suggest that projects facing higher levels of market competition fail more quickly in the discovery and Phase 1 stages when competing with newer or drugs longer established in the market (Figure 4.3). Also, projects tend to fail less in the discovery stage when facing more competition within the R&D process (Figure 4.4). Results are less clear when looking at the other R&D stages.

[Figure 4.3 and Figure 4.4 here]

The survivor functions of all R&D phases are clearly inconclusive when depicting the survival experience of projects with different intensities and types of alliances. Projects with more than two collaborators

fail more quickly but not as quick as the projects that are developed solely by one company or by a large number of collaborators (Figure 4.5). This suggests that there is an "optimal" number of collaborators. Moreover, projects in late stages of clinical trials survive more when involve alliances with at least one big company (Figure 4.6). Also, academic partners seem to linked with more successful projects in pre-clinical stages of development.

[Figure 4.5 and Figure 4.6 here]

These data and results form the basis for the semi-parametric analysis that follows.

B. Semi-parametric analysis

According to the *efficient score residuals* analysis criterion, we identify 17 drug projects from a total of 18,270 projects, with disproportionate influence on the fit of the model. These 17 drug projects correspond to 11 antineoplastic and immunomodulating agents, four drugs targeting the nervous system, one project relating to anti-infectives and one for the musculoskeletal system. These correspond to 235 project-year observations from a total of 102,935 project-year observations. We have excluded these from our analysis. However, we have run the final analysis without excluding these observations and results do not change qualitatively when incorporating the omitted observations.

We run several specifications allowing for all possible combinations of time-invariant and time-varying characteristics. The results for the several specifications remain qualitatively the same and are available from the authors upon request. All models include year dummies and three levels of covariates we are interested in exploring: (i) competition proxies (final product market competition from new and old drugs and competition within the R&D process from potential competitors); (ii) alliances (*log of the number of firms collaborating*, *Big firm*, and *academia*); and market size (proxied by *population* and *GDPpc*, and nonlinearities of both).

We discuss the results of a specification model that introduces an interaction term between the *log* number of firms and the presence of a big firm in the alliance, controlling for year and therapeutic level fixed effects. This model does not fail the *linktest* and respects the proportionality assumption.

Table 4.4 reports estimation results of single risk hazard model for all R&D phases, whilst Table 4.5 shows estimation results of the competing risks model. Two specifications of the baseline hazard are considered in the competing risks model: the first assumes that the baseline hazard for both types of risks is identical; the second assumes proportionality of both baseline hazards. The results are robust across the different models of the baseline hazard. However for presentation purposes we present the results of the second specification of the hazard. All model specifications and robustness checks are available upon request.

[Table 4.4 and Table 4.5 here]

Competition

The results of the single risk model show that more competition from *new drugs* established in the market is associated with a significantly higher risk of failure of projects before they reach Phase 3 of clinical trials. This result is robust across all specifications except in two cases: when omitting the number of old incumbents or the number of potential entrants. The effect seems to weaken when the project reaches Phase 3 of clinical trials. The results of the competing risks model are analogous to these; however they are only statistically significant in the discovery stage. This seems to suggest that pressure

from young incumbents is not significantly different in influencing failure or success after the project passes the discovery stage.

Experiencing competition from drugs that have been longer established in the market (*old drugs*) seems not to be correlated with the risk of failure of projects across the R&D process. This is a robust result across the single and competing risks model in all specifications.

Exposure to *potential entrants* in the market (competition within the R&D process) is associated with lower risk of failure of drug projects in discovery and Phase 1 of clinical trials. This result is robust to all model specifications and both modelling strategies. The same result is also found for Phase 2 but only for the single risks model. Competition within the R&D process is not significantly associated with failure or success in late stages of development.

Alliances

We have found weak evidence of a link between the level of alliances and drug R&D failure. Our results demonstrate that increasing the number of collaborators in the R&D process is associated with lower risk of failure in Phase 2 in the single risks model. Still referring the results of the single risks model, there is some analogous evidence for discovery and Phase 1 though not significant across all specifications. In the competing risks model however the significance is not robust across all specifications.

This result is in line with the findings from the literature that report ambiguous results regarding the role of alliances on R&D productivity (Danzon et al., 2005a, 2007; Hirai et al., 2010). In particular, Danzon et al. (2005a) find the greater the number of firms collaborating in the project the higher the probability of success of the projects only in Phases 2 and 3.

Ambiguous effects are found when considering the type of alliances. In the single risks model, alliances with a big firm are related to higher risk of failure in discovery but lower risk of failure in Phase 3 clinical trials. However in the competing risks model, alliances with a big firm decrease the risk of failure in both phases.

Finally the introduction of the interaction term between the *log* number of firms and the dummy that captures participation of a *big firm* reveals a nonlinear effect of alliances in decreasing the risk of failure in Phase 2.

With respect to alliances with academia, in both the single and competing risk models, research partnerships with academia are associated with a higher risk of failure in discovery, and lower risk of failure in Phase 3 for the single risks model.

Market size

When looking at the effect of market size in the likelihood of failure, results are qualitatively similar in both modelling strategies. However, and as in Kyle (2006), neither *population* nor *GDPpc* are significantly associated with failure. The market size of the home country of the project is significantly associated with its likelihood of failure.

However, when including a non-linear effect of population, by introducing population squared to the specification model, the results show significantly lower risk of failure in discovery stage in more populated countries.

The results also show some non-linearity when introducing GDP per capita (GDP per capita square) as a proxy for affordability at country level. This is more evident when controlling for country-specific fixed effects. Projects being developed in richer countries present a higher risk of failure in discovery stage than projects developed in poorer countries. This is indicated by the negative coefficient of GDP

per capita squared. These results confirm evidence from the existing literature in this field (for instance see Acemoglu and Linn (2004) or Dubois et al. (2011)).

Additionally, these results are in line with the regional innovation paradox referred to in the innovation and economic growth literature: there seems to be an apparent contradiction between the comparatively greater need to spend on innovation in poor regions and their relatively lower capacity to invest in innovation related activities, compared to more advanced regions (Barro and Sala-i Martin, 1992; Nelson, 1996; Oughton et al., 2002).

These results may be capturing two effects: the relevance of the US on the worldwide R&D activity, and the saturation of the market in higher income countries. The first is related to the disproportionate proportion of projects started and failed in the US. The second effect may be related to the fact that rewards to R&D investment are likely to be higher in richer countries, so it may be more worthwhile taking the risk of investment in richer rather than in poorer countries.

As a final remark, our results suggest significant effect of competition and alliances on the rate of failure in two particular phases: discovery and Phase 2. These appear to be two crucial stages of development where there are systematic differences between failure and progression to next R&D phase, as also shown by the competing risks model.

V. Discussion and conclusions

This chapter seeks to measure the association of competition and alliances with the probability of failure in the different R&D phases using semi-parametric duration models to model global R&D data.

Three important results emerge from our analysis. The first is that the determinants of failure differ across the different phases of the R&D process. In particular, the advocated role of competition and alliances as platforms for successful innovation is not verified across all stages of the research process. Consequently, failure is heterogeneous in the different phases and therapeutic areas. Because of that there are areas needing more investment since projects targeting these areas are more likely to fail. However, in order to provide more detailed policy recommendations one would need to disentangle the strategic from the non-strategic abandonment of projects to better identify the key factors associated with failure that are policy amenable.

Secondly, we show competition to be significantly associated with failure in discovery and Phase 2 in two ways. Indeed, we find that the probability of failure of a project declines in the number of *potential competitors* (competition within the R&D). This may be due to the fact that the number of projects in a therapeutic area signals a target where there is potential for innovation, or unaddressed health needs and scientific challenges that motivate further R&D investment. Also, this effect may suggest the positive pressure to become the first to get patented and launch in the market to recoup the investment in R&D (the "escaping competition" mentioned by Aghion et al. (2005)). This is expected to be more pronounced in the discovery stage. This result suggests a possible policy agenda to boost competition at science level through programmes that encourage innovation in early discovery and pre-clinical stages. Moreover, firms might be, not only, benefiting from positive externalities, but also cooperating strategically to differentiate themselves and therefore strategically abandoning projects in the pipeline in order to focus on areas with reduced competition in the market.

Furthermore, we find that the risk of failure increases in the number of new drugs in the market (market competition), i.e. the drugs that are fully enjoying the patent status. Our result is specific to

Phase 2 clinical trials. If prospect profitability signaled by the number of competitors in the market is low then it seems natural that firms are keener in abandoning the projects if Phase 2 clinical trials reveal that the drugs do not offer a substantial advancement when benchmarked in terms of incremental effectiveness with existing drugs. This result emphasizes the importance of the choice of the relevant comparator in cost effectiveness and analysis and health technology assessment.

Even though novel, both findings seem to be aligned with Kyle's results that, by focusing only on market launch, shows that competition from new incumbents reduces the probability of launching a new drug in the market; on the other hand, drugs with more potential competitors are more likely to experience entry (Kyle, 2006).

In our analysis competition does not play any role in project abandonment in Phase 3 clinical trials, but it is associated with failure in discovery and Phase 2. These two stages indicate two important milestones for the R&D process. Failing to successfully complete the discovery stage implies that the drug does not satisfactorily pass the "first toxicity dose" levels required to support administration to a human. This is largely a scientific issue. On the other hand, Phase 2 trials are increasingly a large financial commitment. Indeed, given the increased regulatory requirements in Phase 3 trials, firms have increasingly expanded the number of individuals in Phase 2 trials in order to predict at an earlier stage whether it is commercially viable to proceed to Phase 3 (Scannell et al., 2012). Failing to complete Phase 2 trials can reveal at an earlier stage not only the lack of drug efficacy, particularly for therapeutic areas with animal models of efficacy that are hardly predictive such as oncology (Kola and Landis, 2004), but also, insufficient commercial differentiation from existing drugs in the market (Arrowsmith and Miller, 2013). Since progressing to Phase 3 implies a substantial financial commitment associated with large-scale clinical safety and efficacy studies required for the "launch decision" (Mestre-Ferrandiz et al., 2012) Phase 2 trials can be strategically used to unveil important information regarding the likelihood of success of Phase 3 trials .

The third result suggests a mixed association between alliances and the probability of failure of the projects, confirming the ambiguous findings of the literature.

At first sight, our results suggest that the qualitative effect of the number of collaborators on failure seems to change across the different specifications and the different phases of the R&D process. However, controlling for the firm size, a closer analysis clearly shows that in Phases 2 and 3: (i) the risk of failure decreases in the number of collaborators; (ii) the participation of at least one big firm company in the research project is associated with a lower rate of failure of drug projects, particularly in the transition from Phase 3 clinical trials to market launch; and (iii) that this participation presents increasing rate of failure when extending the alliances protocol to a greater number of participants, provided that we interact the log number of alliances with the participation of a big firm. These results suggest the positive effect of policies to foster collaboration between private sector and academia in the last stages of development, and programmes for product launch for academic projects.

These findings are more evident in the scaling up of clinical trials (Phase 3). At this stage of the R&D process innovators must, not only focus on product development and clinical trials, but also on a series of functional activities that ultimately lead to a successful launch, including: scaling up manufacturing, logistics and distribution processes, marketing effort, regulatory compliance, among others. Moreover, there is a higher financial effort and access to capital required for Phase 3 trials and going through the market authorisation process. The transition to the market is therefore more likely to be successful with the alliances with big firms that have established these capabilities over a long period of time. This result

is consistent with findings reported in the literature (Abrantes-Metz et al., 2004; Mestre-Ferrandiz et al., 2012).

When looking at the partnerships with academia, we find that alliances with academia are associated with increased risk of failure of projects in discovery. This may be due to the fact that academia is normally engaged in exploratory research of riskier targets and consequently areas in which it is harder to innovate, or even not commercially appealing, normally funded by public money. For example, there is evidence of academia being involved in various projects on the potential role of genomics in fostering drug discovery in many tropical diseases (Gardner et al., 2002; Ridley et al., 2006; Rosamond and Allsop, 2000). Other possible reasons may be linked to the documented divergences of perspectives and interests of academia and industry researchers that hinder successful collaborative effort (Siegel et al., 2003). These divergences arise on a trade-off between disclosure and secrecy of knowledge in the process of patenting structure in drug discovery (Rhoten and Powell, 2007), and unveil the conflicting objectives, work environments and scientific methodologies of both parties (Munos, 2009; Perkmann and Walsh, 2007; Murray and Stern, 2007). Also, firms might be acting strategically to benefit from positive knowledge spillovers from publicly-funded R&D projects (Cockburn and Henderson, 1998), while accepting that the academic partner leads the R&D process and assumes an important share of the risks.

Moreover, firms might have a higher incentive to strategically not reveal information about failure in R&D projects (Reinganum, 1981; Dasgupta and Stiglitz, 1980; Barzel, 1968), compared with academia, particularly in early discovery when targets being tested do not need compulsory public registration. Because of this, policy implications drawn from these results must accommodate the fact that failure might be under-reported by the industry and academia is more likely to invest in riskier and scientifically more complex projects.

Moreover, the change in signs across R&D phases when evaluating the role of alliances on failure could be related to the nature of the alliance that is not captured in our models. There might be attributes of the alliances not available in our data that may be masking important characteristics of the projects and/or nature and process of the alliances that can contribute to the likelihood of the failure of the R&D project.

There are several caveats of our analysis driven by the quality of our data and the lack of detailed data at firm level and project level as discussed in Chapter 3.

First, there is a potential problem for selection bias: projects that fail may be systematically different in nature from the ones that succeed, which means that unobserved characteristics of the drug projects may be correlated with the level and intensity of competition, different type of alliances or even market size. This means that we might be inaccurately estimating the true effect of competition and alliances on failure if projects targeting the same therapeutic indication differ substantially in terms of scientific complexity and expected return on investment.

This relates to the problem of asymmetry of information in early stages of discovery discussed in Chapter 3. There is a trade-off between the incentives to disclose information in the R&D process and to hide it from rivals. This is particularly pronounced in the discovery and pre-clinical stage, and before the legal requirements to register ongoing investigational studies has been imposed since 2004 in Europe and 2007 in the US. For this reason, the results for the discovery stage might suffer from this limitation. This sheds light on the need to design policies that promote the publication of negative clinical trials to reduce the publication bias that affects firmsecisions on R&D investments.

Secondly, we are not modelling failure as a function of strategic behaviour of the firm and its competi-

tors. Projects may clearly fail to complete a specific R&D phase due to several reasons. Failure may be a combination of rejection by the market approval bodies, strategic withdrawal by the firm of certain R&D investments, merger and/or acquisition by a competitor, or even scientific or clinical failure in targeting a disease. With our data we cannot separately identify the reasons behind project abandonment. Such analysis would require firm and project level data that is unavailable in our dataset. In particular, in our data, the majority of the projects are owned by several companies and we cannot identify and measure the role and effort of each firm in the R&D process. This fact restricts the use of firm-level fixed effects as a means to incorporate the strategic behaviour of the firms in our analysis.

Thirdly, there is a potential risk of mis-estimating failure given the assumption on the replication of the compound observations according to the number of indications targeted by each compound. This is particularly true if multiple indications can be contained in one drug. In that case, we might be inaccurately estimating all the effects. However, compound projects with several indications can be seen as several compound projects running in parallel because they target different markets. In that case, the analysis of the role of alliances and competition on failure shall be indication specific. On the other hand, if the same compound project is being developed in more than a country, drugs with a higher probability of success might be those for which there are multiple countries. In that case, we might be underestimating the likelihood of failure of such type of compound projects. Make it country-specific undermines the multinational perspective of the project, which might create network spillovers greater than the sum of the individual countries. Nevertheless, there are pretty strong documented locational effects on R&D that are captured by the country where R&D is taking place.

Finally, the use of country-specific measures (gdppc, population and country-specific died effects) might underestimate the true market size for new therapies. Ideally, we would have market size proxied by some measure of affordability constructed with global disease burden weighted by ability-to-pay. However, this information is not readily available for a time-series and the GDP per capita gives a proxy for the total affordability of a country to medicines.

Addressing these caveats requires data at firm and project level to be more readily available for research. Many of these issues would be potentially solved with more information about pharmaceutical companies, such as R&D investments, financial accounts, their patenting and licensing activities, and the clinical and financial risk of their R&D portfolio. This information is not available which restricts our analysis. Despite these caveats we retain our analysis as a novel and relevant insight into the nature of the R&D process. This piece of work may contribute to the policy debate on the presumed role of competition and alliances as platforms for successful innovation.

VI. Appendix: Tables

Table 4.1. Definition of variables

Variable	**Definition**	**Frequency**
Competition from new drugs	Count of drugs in the market launched for less than 5 years	Therapeutic class-year
Competition from old drugs	Count of drugs in the market launched for more than 5 years	Therapeutic class-year
Potential competitors	Count of R&D projects in the same market	Therapeutic class-year
Intensity of alliances	Log of total number of collaborators	Therapeutic class-year
Big firm participation	Project has at least one TOP100 firm collaborator (dummy variable)	Drug project
Academia participation	Project has at least one academia collaborator (dummy variable)	Drug project
Population, total	Population in 10s of millions	Country
GDP *per capita*	GDP per capita in US$1000s (constant 2000 US$)	Country
Regulatory forces	Country where R&D is based	R&D phase-year
Therapeutic specificities	Therapeutic class that the project targets	Drug project

Table 4.2. Descriptive statistics

Variables(mean)	**Discovery**			**Phase 1**			**Phase 2**			**Phase 3**		
	Total	Failures	Successes	Total	Failures	Successes	Total	Failures	Successes	Total	Failures	Successes
Duration	3.48	5.82	3.66	2.48	4.78	2.22	2.7	5.52	3.08	2.24	5.59	2.7
New drugs	6.06	13.61	8.04	7.21	18.28	10.18	6.69	17.72	10.37	7.59	21.91	7.65
Old drugs	0.88	2	1.21	0.95	2.22	1.28	1.02	1.86	1.34	1.07	2.07	1
Potential entrants	11.02	22.79	14.21	13.08	30.67	17.98	12.37	30.6	18.37	13.7	39.42	13.78
Population	1.7E+09	2.0E+09	1.8E+09	1.7E+09	1.9E+09	1.8E+09	1.7E+09	1.9E+09	1.8E+09	1.9E+09	2.3E+09	1.8E+09
GDP *per capita*	29058.1	31988.5	30052.2	29882.4	32233.5	30570.5	30734.4	32398.7	31483.5	30503.2	32320	30037.7
Intensity of alliances	1.65	1.43	1.69	2.05	1.43	2.08	2.24	1.65	2.46	2.87	2.46	1.78
Big firm	0.59	0.56	0.54	0.66	0.64	0.66	0.68	0.67	0.73	0.76	0.69	0.53
Academia	0.12	0.13	0.11	0.09	0.06	0.08	0.08	0.08	0.08	0.08	0.07	0.12
Observations	54600			11109			9644			5789		
Drug projects	10952	3277		3257	444		2446	361		1597	148	
Failures (%)		29.90%			13.60%			14.80%			9%	
Therapeutic classes	78											
Years covered		1980-2012										

Table 4.3. R&D activity and failure across therapeutic classes

	Total number of R&D projects	Proportion of failures as a proportion of ATC-specific failures	Proportion of failures as a proportion of total failures
A: Alimentary tract and metabolism	1583		0.203
Stomatologicals, mouth preparations, medicinal dentifrices etc	44	0.25	0.0006
Drugs used in diabetes	645	0.203	0.0072
Vitamins	5	0	0
Anabolics, systemic	3	0	0
Appetite stimulants	7	0.143	0.0001
Other alimentary tract and metabolism products	124	0.129	0.0009
Antacids, antiflatulents and anti-ulcerants	128	0.203	0.0014
Functional gastro-intestinal disorder drugs	133	0.128	0.0009
Antiemetics and antinauseants	66	0.167	0.0006
Cholagogues and hepatic protectors	48	0.146	0.0004
Laxatives	7	0.143	0.0001
Antidiarrhoeals, oral electrolyte replacers and intestinal anti -inflammatories	187	0.156	0.0016
Antiobesity preparations, excluding dietetics	181	0.387	0.0038
Digestives, including digestive enzymes	5	0.2	0.0001
B: Blood and Blood forming organs	653		0.247
Antithrombotic agents	402	0.279	0.0061
Blood coagulation system, other products	113	0.23	0.0014
Anti-anaemic preparations	98	0.143	0.0008
All other haematological agents	40	0.225	0.0005
C: Cardiovascular system	1767		0.239
Cardiac therapy	538	0.251	0.0074
Lipid-regulating/anti-atheroma preparations	302	0.265	0.0044
Cardiovascular multitherapy combination products	3	0.334	0.0001
Antihypertensives	151	0.212	0.0018

Diuretics	40	0.3	0.0007
Cerebral and peripheral vasotherapeutics	141	0.234	0.0018
Antivaricosis/anti-haemorrhoidal preparations	3	0	0
Other cardiovascular products	208	0.274	0.0031
Beta-blocking agents	44	0.114	0.0003
Calcium antagonists	137	0.182	0.0014
Agents acting on the renin-angiotensin system	200	0.21	0.0023
D: Dermatologicals	631		0.17
Antifungals, dermatological	39	0.051	0.0001
Anti-acne preparations	65	0.138	0.0005
Other dermatological preparations	132	0.227	16437
Wound healing agents	110	0.164	0.001
Anti-pruritics, inc. topical antihistamines, anaesthetics, etc	21	0.19	0.0002
Nonsteroidal products for inflammatory skin disorders	188	0.17	0.0018
Topical antibacterials and antivirals	57	0.14	0.0004
Topical corticosteroids	19	0.263	0.0003
G: Genitourinary system and sex hormones	668		0.166
Gynaecological anti-infectives	11	0.182	0.0001
Other gynaecologicals	106	0.189	0.0011
Sex hormones and products with similar desired effects, systemic action only	198	0.217	0.0024
Urologicals	353	0.133	0.0026
H: Systemic hormonal preparations (exc. sex hormones)	142		0.169
Pituitary and hypothalamic hormones	32	0.125	0.0002
Systemic corticosteroids	6	0	0
Thyroid therapy	5	0.2	0.0001
Other hormones	99	0.192	0.001
J: General anti-infectives (systemic)	2626		0.263
Systemic antibacterials	795	0.258	0.0112
Systemic agents for fungal infections	181	0.309	0.0031
Antimycobacterials	27	0.223	0.0003

Antivirals for systemic use	1080	0.306	0.0181
Sera and gamma-globulin	56	0.161	0.0005
Vaccines	379	0.071	0.0015
Other anti-infectives	108	0.528	0.0031
L: Antineoplastic and immunomodulating agents	4578		0.236
Antineoplastics	3484	0.243	0.0463
Cytostatic hormone therapy	137	0.212	0.0016
Immunostimulating agents	359	0.12	0.0024
Immunosuppressants	598	0.269	0.0088
M: Musculoskeletal system	970		0.252
Anti-inflammatory and anti-rheumatic products	600	0.272	0.0089
Topical anti-rheumatics	19	0	0
Muscle relaxants	37	0.162	0.0003
Anti-gout preparations	33	0.03	0.0001
Other drugs for disorders of the musculo-skeletal system	281	0.263	0.0041
N: Nervous system	3184		0.25
Anaesthetics	48	0.21	0.0005
Analgesics	557	0.26	0.0079
Anti-epileptics	158	0.19	0.0016
Anti-parkinson drugs	166	0.188	0.0017
Psycholeptics	520	0.25	0.0071
Psychoanaleptics excluding anti-obesity preparations	434	0.182	0.0043
Other cns drugs	1301	0.284	0.0202
P: Parasitology	143		0.217
Antiprotozoals and anthelmintics	141	0.22	0.0017
Ectoparasiticides, including scabicides, insecticides and repellents	2	0	0
R: Respiratory system	928		0.209
Nasal preparations	87	0.08	0.0004
Anti-asthma and copd products	575	0.206	0.0065
Cough and cold preparations	23	0.174	0.0002

Systemic antihistamines	73	0.233	0.0009
Other respiratory system products	170	0.282	0.0026
S: Sensory organs	379		0.135
Ophthalmologicals	362	0.141	0.0028
Otologicals	17	0	0

Table 4.4. Parameter estimation results: Single Risks Model

	Discovery	**Phase1**	**Phase2**	**Phase3**
Competition from new drugs	1.008***	1.012**	1.018***	1.005
	(-0.00152)	(-0.00497)	(-0.0057)	(-0.00796)
Competition from old drugs	1	0.999	0.999	1.006
	(-0.00201)	(-0.00617)	(-0.00716)	(-0.00818)
Potential competitors	0.996***	0.994**	0.990***	0.998
	(-0.000841)	(-0.00276)	(-0.00312)	(-0.00449)
Population, total	1	1	1	1
	(-3.71E-10)	(-1.05E-09)	(-1.17E-11)	(-1.83E-11)
Population, total squared	1	1	1	1
	(-4.55E-19)	(-1.19E-18)	(-1.36E-18)	(-1.61E-18)
GDP *per capita*	1	1	1	1
	(-2.22E-7)	(-6.54E-7)	(-4.07E-7)	(-7.26E-7)
GDP *per capita* squared	1	1	1	1
	(-3.83E-10)	(-1.15E-09)	(-7.83E-12)	(-1.34E-11)
Intensity of alliances	1.041	0.965	0.471***	0.85
	(-0.0841)	(-0.248)	(-0.11)	(-0.318)
Big firm participation	1.282***	1.107	0.871	0.457***
	(-0.0583)	(-0.147)	(-0.139)	(-0.131)
Academia participation	1.160**	0.928	0.977	0.465*
	(-0.0781)	(-0.207)	(-0.255)	(-0.215)
Big firm#Intensity of alliances	0.807**	0.718	1.37	1.577
	(-0.0743)	(-0.198)	(-0.381)	(-0.679)
Fixed effects				
Year dummies	Yes	Yes	Yes	Yes
Therapeutic class	Yes	Yes	Yes	Yes
Targetcountry	No	No	No	No
#Observations	41502	6990	5995	2817
#Projects	7811	1932	1437	857
#Failures	2856	344	276	102
Log likelihood	-20917.4	-1965.3	-1534.7	-480
Akaike's_criterion	41914.8	4014.6	3167.4	1018
Goodness-of-fit	0.344	0.769	0.134	0.986
Proportionality assumption	0.6726	0.4404	1	0.094

Notes. Exponentiated coefficients; Standard errors in parentheses.
* $p < .10$, ** $p < .05$, *** $p < .01$
The results refer to the reference case country USA, year 1990, and therapeutic category Sensory organs.

Table 4.5. Parameter estimation results: Competing Risks Model

	Discovery	**Phase1**	**Phase2**	**Phase3**
Competition from new drugs	1.004***	1.005**	1.005	1.001
	(-0.000976)	(-0.00217)	(-0.00277)	(-0.0075)
Competition from old drugs	1	1.001	1.002	1.001
	(-0.00133)	(-0.0027)	(-0.00398)	(-0.00766)
Potential competitors	0.998***	0.997**	0.998	0.999
	(-0.000533)	(-0.0011)	(-0.00153)	(-0.00432)
Population, total	1.000***	1	1	1
	(-2.14E-10)	(-3.67E-10)	(-4.81E-10)	(-1.71E-09)
Population, total squared	1	1	1	1
	(-2.45E-19)	(-3.59E-19)	(-5.45E-19)	(-1.57E-18)
GDP *per capita*	1	1	1	1
	(-0.0000097)	(-0.0000173)	(-0.0000182)	(-0.000076)
GDP *per capita* squared	1	1	1	1
	(-1.75E-10)	(-3.25E-10)	(-3.42E-10)	(-1.38E-09)
Intensity of alliances	0.949	1.12	0.826*	0.85
	(-0.0436)	(-0.0888)	(-0.106)	(-0.0887)
Big firm participation	0.912***	0.948	0.995	0.610*
	(-0.0255)	(-0.0511)	(-0.0742)	(-0.17)
Academia participation	1.132***	0.896	1.031	0.946
	(-0.0403)	(-0.0758)	(-0.0975)	(-0.367)
Big firm#Intensity of alliances	1.162***	1.063	1.454***	1.389
	(-0.0662)	(-0.11)	(-0.174)	(-0.524)
Transition to next R&D phase	3.78E-11	5.22E-23	4.97E-14	6.69E-21
	()	()	()	()
Fixed effects				
Year dummies	Yes	Yes	Yes	Yes
Therapeutic area	Yes	Yes	Yes	Yes
Targetcountry	No	No	No	No
#Observations	87788	14925	12865	5627
#Projects	9129	2390	1856	1125
#Failures	7314	1741	1197	117
Log likelihood	-57104	-11466.6	-7456.9	-553.4
Akaike's_criterion	114303.9	23035.3	15007.8	1188.9

Notes. Exponentiated coefficients; Standard errors in parentheses.
* $p < .10$, ** $p < .05$, *** $p < .01$
The results refer to the reference case country USA, year 1990, and therapeutic category Sensory organs.

Estimation strategies

Partial likelihood

In the context of Cox proportional hazard model, we can estimate the relationship between the hazard rate and explanatory variables without having to make any assumptions about the shape of the baseline hazard function. This is called a semi-parametric model. When recalling single risks model specification in Equation (4.8):

$$h_i^j(t, X, \tau_j) = h_0(t) \exp(X\prime\beta) = h_0(t)\theta(X) \tag{4.15}$$

And using proportional hazard assumption together with other assumptions, it is possible to estimate consistently β using partial likelihood method of estimation, rather than maximum likelihood. Partial likelihood is used when there is no full information on the form for the joint data distribution (Allison, 1984).

What differs from maximum likelihood is that instead of individuals or projects, we are interested to model the occurrence of ordered (according to duration time) events i.

We are interested to model the probability distribution of the duration of abandoned projects T, for any particular drug project and regard t_i as a realisation of the random variable T_i for a project with characteristics of each drug project in the sampleLancaster (1992). The sample Partial Likelihood is given by:

$$PL = \prod_{i=1}^{S} \mathcal{L}_i \tag{4.16}$$

$\mathcal{L}_i = P$(project d has event at $t = t_i$ conditional on being in the risk set at $t = t_i$)

To work out this probability, we need to use the rules of conditional probability together with the fact that *f(t)=h(t)S(t)*, and so the probability that an event occurs in the tiny interval $[t, t + \delta t)$ is $f(t)dt = h(t)S(t)\delta t$. Considering the illustrative dataset in the following table:

Table 4.6. Illustrative dataset

Drug project $\#d$	Duration T_d	Event i
1	2	1
2	4	2
3	9	(no event ensored)
4	11	3
5	13	4
6	14	(no event ensored)

Consider the event $i = 4$ with risk set $d = \{5, 6\}$. We can define

$$\begin{cases} A = P(\text{event experienced by } d{=}5 \text{ and not } d{=}6) = [h_5(13)S_5(13)\delta t][S_6(13)] \\ B = P(\text{event experienced by } d{=}6 \text{ and not } d{=}5) = [h_6(13)S_6(13)\delta t][S_5(13)] \end{cases} \tag{4.17}$$

The probability of either A or B using the standard conditional probability formula is equal to

$$\mathcal{L}_5 = \frac{A}{A + B} = \frac{h_5(13)}{h_5(13) + h_6(13)} \tag{4.18}$$

The survivor function terms cancel out. With this, we can derive all the other $\mathcal{L}_s$. For example, $\mathcal{L}_1 = \frac{h_1(2)}{(h_1(2)+h_2(2)h_6(2)}$.

All projects are in the risk set for the first event. Now let us apply PH model specification, and we have:

$$\mathcal{L}_5 = \frac{h_5(13)}{h_5(13) + h_6(13)} = \frac{h_5(13)\theta_5}{h_5(13)\theta_5 + h_6(13)\theta_6)} = \frac{\theta_5}{\theta_5 + \theta_6} \tag{4.19}$$

The baseline hazard contributions cancel out. Similarly,

$$\mathcal{L}_1 = \frac{\theta_1}{\theta_1 + \theta_2 + \ldots + \theta_6} \tag{4.20}$$

And so on for all events. Given each $\mathcal{L}_i$ expression, we can construct the complete PL expression for the whole sample of events, and then maximise it to derive β. As said before, the baseline hazard function is completely unspecified (Jenkins, 2005; Lancaster, 1992). Also, to note that each $\mathcal{L}_s$ expression does not depend on the precise survival time at which the s^th event occurs, but only the order of events affects the PL expression.

We also highlight the fact that the PH assumption implies that the hazard function for two different projects has the same shape, differing only by a constant multiplicative scaling factor that does not vary with survival time. This assumption may be tested.

Moreover, just to remember that we incorporate time-varying covariates. PL estimates the information at each event time. This means that covariates are only valuateduring the estimation at event times, and so it does not matter what happens to their values in between.

Log-likelihood

In the case of competing risks model, when we consider two possible destination (risk), the overall model likelihood value is the sum of the likelihood values for each of the destination-specific models, θ_i^j and θ_i^k, recalling Equation 4.13. The log-likelihood for the whole sample is the sum of this expression over all individual records in the sample.

The log-likelihood is given by

$$ln\mathcal{L} = \{\delta^j[ln\ \theta_i^j] + lnS_j\} + \{\delta^k[ln\ \theta_i^k] + ln\ S_k(T)\} \tag{4.21}$$

Two main assumptions are taken to use this estimation method (Lancaster, 1992).

State independency. The chances of making transition from the current state do not depend on transition history prior to entry to the current state.

To estimate destination-specific hazard rates, there is a weak identification assumption to hold: risks independence. This implies that $h(t) = \sum_{i=j,k} \theta_i(t)$ or even $h(t;X) = \sum_{i=j,k} \theta_i(t;X)$.

That is, $h(t \mid X) = \theta_i^j(t_j \mid X) + \theta_i^k(t_k \mid X)$ i.e. the hazard rate for transition to any destination is the sum of the destination-specific hazard rates, controlling for X observables. Once failure occurs, the failure to destination j has probability $\frac{\theta_i^j(t)}{\theta_i^k(t)}$.

Independence also means that the survivor function for transition to any destinations can be factored into a product of destination-specific functions:

$$S(t) = \exp[-\int_0^t h(u)\,\mathrm{d}u] = \exp[-\int_0^t [\theta_i^j(u) + \theta_i^k(u)]\,\mathrm{d}u] = \\ = \exp[-\int_0^t [\theta_i^j(u)]\,\mathrm{d}u] \times \exp[-\int_0^t [\theta_i^k(u)]\,\mathrm{d}u] = S_j(t)S_k(t) \tag{4.22}$$

Destinations are mutually exclusive and exhaust the possible destinations. The individual sample likelihood contribution in the independent competing risk model with two destinations is of three types:

A.1 $\mathcal{L}^j$**=transition to** j**, where** $\mathcal{L}^j = f_j(T)S_k(T)$

A.2 $\mathcal{L}^k$**=transition to** k**, where** $\mathcal{L}^k = f_k(T)S_j(T)$

A.3 $\mathcal{L}^C$**=censored spell, where** $\mathcal{L}^C = S(T) = S_j(T)S_k(T)$

$\mathcal{L}^j$, the likelihood contribution summarises the chances of a transition to j combined with no transition to k, and vice versa in the $\mathcal{L}^k$ case. The destination-specific censoring indicators:

$$\delta^j = \begin{cases} 1 & \text{, if transition to } j \\ 0 & \text{, otherwise (either exit to } k \text{ or censored)} \end{cases} \tag{4.23}$$

The overall contribution from the individual to the likelihood, $\mathcal{L}$, is

$$\mathcal{L} = (\mathcal{L})^{\delta^j}(\mathcal{L}^k)^{\delta^k}(\mathcal{L}^C)^{1-\delta^j-\delta^k} =$$

$$= [f_j(T)S_k(T)]^{\delta^j}[f_k(T)S_j(T)]^{\delta^k}[S_j(T)S_k(T)]^{\delta 1-\delta^j-\delta^k} =$$

$$= \left[\frac{f_j(T)}{S_j(T)}\right]^{\delta^j} S_j(T) \left[\frac{f_k(T)}{S_k(T}\right]^{\delta^k} S_k(T) =$$

$$= [h_j]^{\delta^j} S_j(T)[h_k]^{\delta^k} S_k(T) = \{[h_j]^{\delta^j} S_j(T)\}\{[h_k]^{\delta^k} S_k(T)\} \quad (4.24)$$

We can maximise the overall (log) likelihood by maximising the two component parts separately. The overall model likelihood value is the sum of the likelihood values for each of the destination-specific models. The log-likelihood for the whole sample is the sum of this expression over all projects in the sample.

$$ln\ \mathcal{L} = \{\delta^j[ln\ \theta_i^j] + lnS_j\ \}\ +\ \{\delta^k[ln\ \theta_i^k]\ + ln\ S_k(T)\} \quad (4.25)$$

Akaike's criteria

Akaike's (AIC) criteria: for each model specification the value is computed as: $AIC = -2 \times$ (log-likelihood) $+ 2(p + 1 + s)$, where p denotes the number of covariates in the model and $s = 1$ for the Weibull and log-logistic models (Akaike, 1974; Hosmer et al., 2011).

Testing goodness-of-fit and proportionality assumption

As a specification test and a measure of goodness-of-fit, we use the link test which basically regresses on and, where now the original model regressors are omitted. It tests whether the coefficient of is zero under the null hypothesis (Cleves et al., 2010; Hosmer et al., 2011).

We use the link test to interact a function of time on time-varying variables and test whether their coefficients are zero under the null hypothesis (Hosmer et al., 2011). We carry out the log-rank test that assumes under the null hypothesis that the different group hazards functions are similar, where the groups are defined by the different level of covariate X (Cleves et al., 2010).

The preferred method of performing this analysis is to compare the estimated parameter $\hat{\beta_X}$ obtained from the full data with the estimated parameters $\hat{\beta_X^i}$ obtained by fitting the model to the $n-1$ observations remaining after the i^{th} observation is removed. If $\hat{\beta_X} - \hat{\beta_X^i}$ is close to zero, then the i^{th} observation has little influence on the estimate (Cleves et al., 2010).

VII. Appendix: Figures

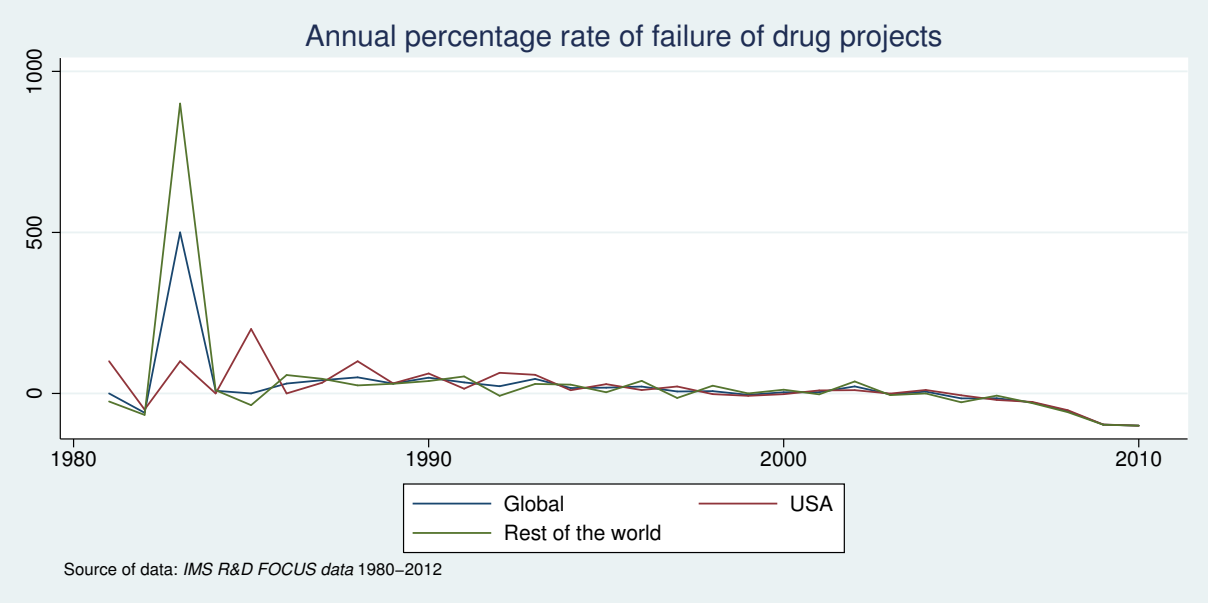


Figure 4.1. Annual percentage rate of failure of drug projects

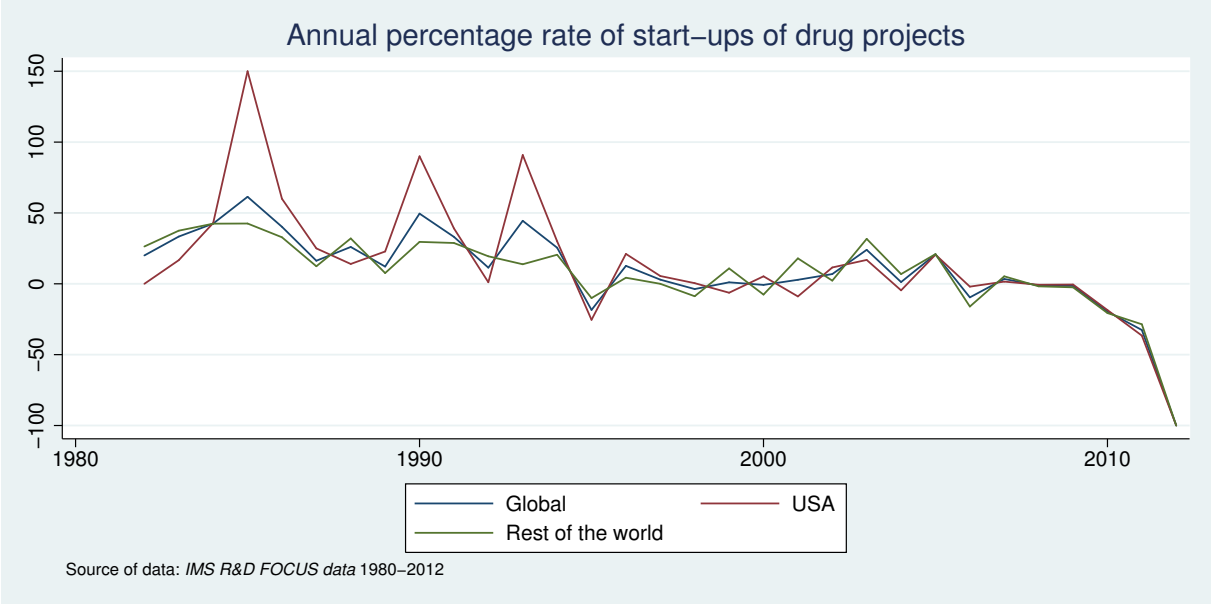


Figure 4.2. Annual percentage rate of start-ups of drug projects

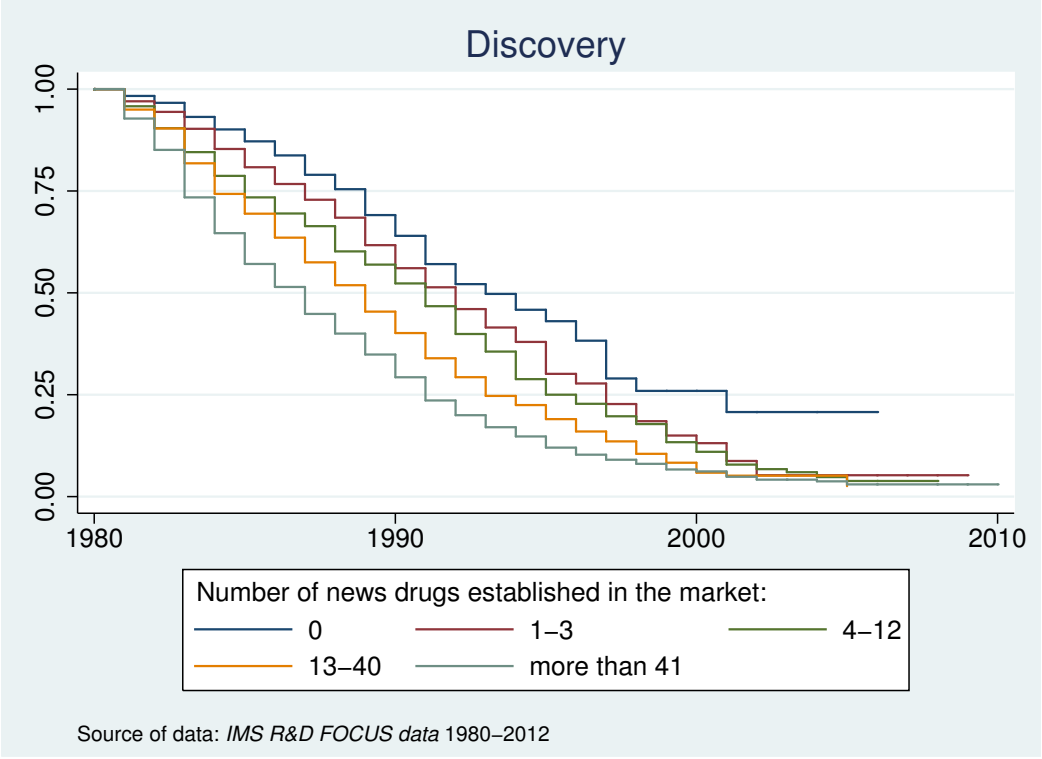


Figure 4.3. Kaplan-Meier survival function: market competition from young drugs in discovery

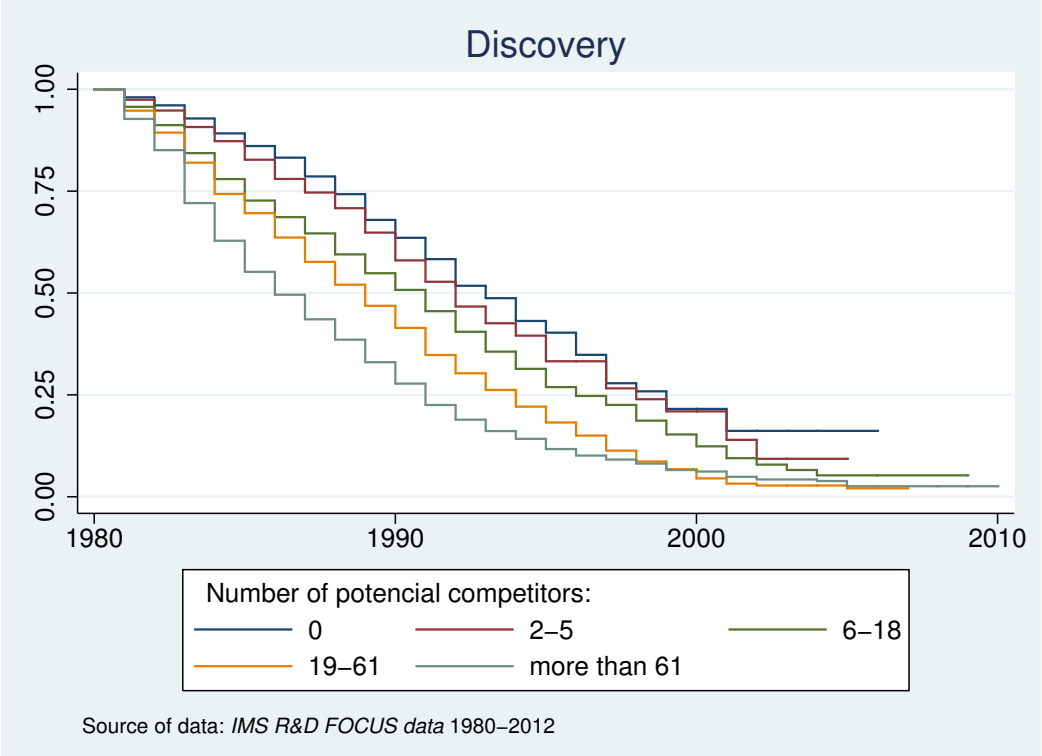


Figure 4.4. Kaplan-Meier survival function: competition from potential competitors in discovery

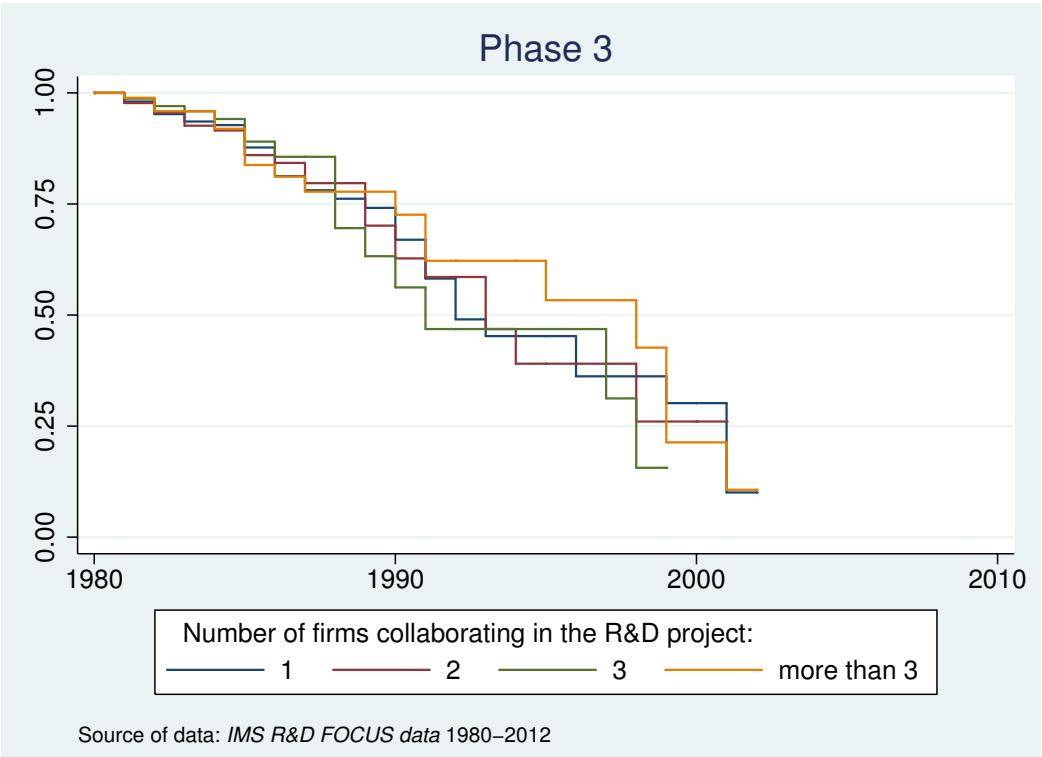


Figure 4.5. Kaplan-Meier survival function: intensity of alliances in Phase 3 trials

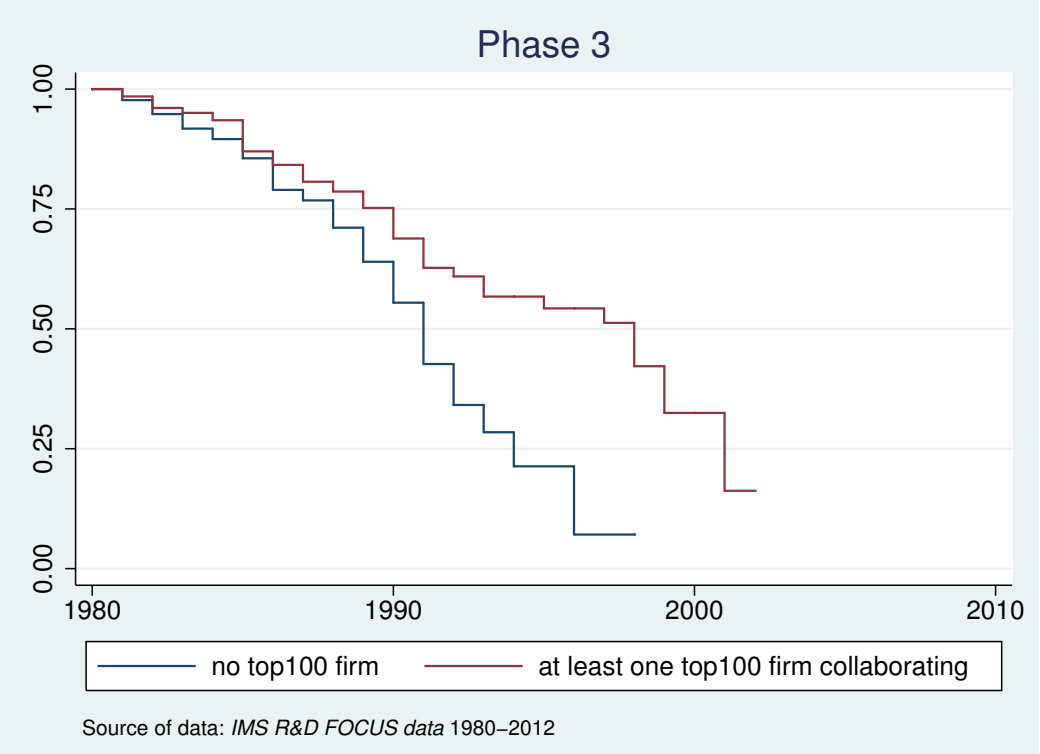


Figure 4.6. Kaplan-Meier survival function: participation of big firms in Phase 3 trials

Chapter 5

Micro-simulation of future pharmaceutical R&D landscape

I. Introduction

Despite the substantial increase in R&D spending in recent decades, evidence suggests a declining number of approvals of new molecules developed, amounting from a total of 52 in 1996, to around 20 in 2007, of new molecular entities and new biologics approved (PWC, 2011b; Bouvier, 2007; Paul et al., 2010; Strauss, 2010; Reuters, 2009), in spite of sustained increase in R&D spending.

Indeed, development costs are estimated to have increased yearly by 30% in the last two decades due to increased development times (Reuters, 2009) and higher failure rates amongst R&D phases (DiMasi, 2002; Paul et al., 2010). Furthermore, there are challenges regarding the true societal value of such investment. Most new drugs brought to the market are variants of formulations or combinations of existing therapies ("follow-on drugs"), with few out of the 25 new medicines discovered per year being in effect novel in modulating targets and disease pathways (EY, 2010; Paul et al., 2010).

This context raises issues of efficiency of allocation of resources as well as equity of access to new therapies that can offer better quality-of-life for conditions driving the global burden of disease. As Chapter 6 will explore in more detail, the pharmaceutical R&D investment has targeted the most prevalent health needs of the richest countries. The disincentives to invest in disease areas that affect the poorest and, thus financially less attractive markets, and the strategic disengagement from the pharmaceutical industry to pursue new avenues of discovery in some of the most disabling conditions, put in danger the prospect of access to new, better and cost-effective therapies in a large number of therapeutic areas.

The pressure from national health systems, insurance companies and other payers of health care to access more cost-effective drugs contrasts, to a large extent, with the societal interests to encouraging investment in under-invested research areas. Indeed, health systems have been notably confronted with trade-offs in promoting the development of drugs with substantial health improvement that are affordable, while allowing sufficient rents to the industry to incentivise R&D investment.

In the last three decades, there have been profound changes in the design of the incentive mechanisms in the top global regional innovators, namely, the USA, Europe and Japan to accelerate the innovation process and deliver novel therapies to the market (Dranove and Meltzer, 1994). These reforms have been implemented to grant priority review to therapies targeted at serious life-threatening diseases, rare and neglected conditions and those that pose a public health threat[1] and are found difficult to develop.

In spite of these reforms, innovation productivity crisis still persists in many therapeutic and disease areas such as vaccines, drug resistant infections, rare conditions and neurological disorders. To stimulate

[1] namely the Orphan Drugs Act in 1983, Fast Track in 1988, Priority Reviews in 1992, Accelerated Approval in 1993, FDA Modernization Acts in 1997 and 2007, TRIPS agreement in 1994 and lately, the Safety and Innovation Act in 2012. A full list of legislative and regulatory events can be found in Hwang et al. (2014).

innovation in these areas, new forms of incentive mechanisms are currently being introduced in health systems (for instance, value based-pricing in the UK, risk-sharing in Australia), however, there is still uncertainty regarding their impact on the future landscape of the R&D pipeline and on innovation.

The existing literature has focused on optimal project selection strategies in the pharmaceutical firm R&D portfolio by using mathematical programming and discrete event system simulation techniques to model uncertainty in the R&D pipeline (Subramanian et al., 2003; Rajapakse et al., 2005; Rogers et al., 2002). Even though these approaches explore in detail the stochastic optimisation problem around decisions given its resource-constraints and performance-oriented targets, the studies lack in general an economic perspective on the factors that impact R&D success, failing to model the role that market conditions and project characteristics play on the R&D strategies by the pharmaceutical industry.

Therefore, the aim of this study is to forecast the stock of new therapies to be launched in the market between 2011 and 2030, by exploring the dynamic effect of market conditions and project characteristics on the likelihood of market launch of new therapies. Throughout this analysis, it is assumed an absence of new R&D policies to the existing ones, meaning this analysis captures the "do nothing" scenario on what concerns the effect of policy and regulation on the R&D landscape.

The remainder of the paper is organised in three sections. The next section describes the model used to incorporate market launch, failure and births in the dynamics of the R&D pipeline. The covariates and parameter estimation are also discussed in full detail in this section. In section II we explain the dynamic micro-simulation model and the main results. In section III we discuss the results of the micro-simulation analysis. Finally, in section IV we conclude with a discussion and interpretation of the results.

II. Methodology

The aim of this study is to simulate the number of new therapies that will be launched in the market from 2011 until 2030 for a specific therapeutic indication and in a specific locational market under a *ceteris paribus* assumption of no further policy intervention. We have used historical data to calibrate the micro-simulation model and to infer about the likelihood of both events failure and market launch, and to forecast the stock of new drugs in the near future.

The ceteris paribus assumption should not be regarded as a limitation. It is indeed one of the many possible future scenarios for the simulations but given that it makes use of historical data we believe it is the most evidence-based scenario we could have chosen. Because of this, it imposes fewer assumptions on the uncertainty around the likelihood of events we are interested to model and simulate. Many other scenarios can be analysed in future research but we believe this is the obvious starting point for the analysis.

We have defined market as the region-therapeutic area where a specific product will be at risk of being launched. For the therapeutic classification, and as used in Chapter 4, we classify the projects amongst the therapeutic categories presented by the EPhRMA[2] third level Anatomical Therapeutic Chemical (ATC) Classification.

For the definition of region, given that firms approach the market from a regional perspective due to the strict and specific regulatory requirements, and because of the larger amount of missing information for countries where there is no R&D in many therapeutic areas, we have grouped countries into five main regional markets around the world. The countries that are part of each region are shown to be similar

[2]European Pharmaceutical Market Research Association

in terms of regulation environment, IP rights, and technological conditions. The five different regions are respectively: (1) USA; (2) the EMA members, which include the 28 European Union Member States as well as the European Economic Area (including Norway, Iceland and Liechtenstein), Canada, and Switzerland; (3) Japan; (4) Emerging Markets and Asia-Pacific (EMAP), including Brazil, India, Russia (and former Soviet Union), China, South Korea, Mexico, Nigeria, Israel, Vietnam and Taiwan ; and (5) the rest of the world.

We simulate the stock of new market entries N_i of category i, given the stock of projects in the global R&D portfolio in 2011, i.e., $P_{il,2011}$. For every calendar year t the sum of new therapies, both New Chemical Entities (NCEs) and biologics of therapeutic type i in region l is given by:

$$N_{i,t} = \sum \sum_{l,t} NCEs_{il,t} + \sum \sum_{l,t} biologics_{il,t} \tag{5.1}$$

Let $P_{il,t}$ denote the portfolio of projects in region l at the end of year t (net of births, failures and market launches occurring in year t), it can be represented by:

$$P_{il,t} = P_{il,t-1} + B_{il,t} - D_{il,t} - M_{il,t} \tag{5.2}$$

with $P_{il,t}$ denoting the past portfolio of projects of type i in region l at the end of year $t-1$, $B_{il,t}$ is the number of births of type i occurring in region l during the year t, $D_{il,t}$ is the number of deaths of type i occurring in region l at the end of year t, and finally, $M_{il,t}$ the number of product market of type i launched in region l during the year t.

Therefore, a new cohort of drug projects is added to the starting R&D portfolio (at the start of 2011) and during the year of 2011, a simulated proportion of projects will leave the R&D portfolio due to failure and market launch. Analogously, to the projects that survive in the R&D pipeline, a new cohort of newborn projects is added in the following year ($t = 2012$) and, similarly, some projects will leave the R&D pipeline caused by failure or market launch. This process is repeated every year until $t = 2030$.

In the next section, we specify failure and entry as functions of factors which impact on the life of pharmaceutical projects. We describe these factors by reviewing the economics and management literature on entry and failure, and we present the estimation strategy to obtain parameters that will inform the micro-simulation model on the probabilities of occurrence of these two events of interest, i.e., market launch and failure. In section B we specify the birth function that models the birth process of projects in the future. In section C we describe the dynamic micro-simulation model. Finally, in section D we give details about the data and variables used to measure these factors.

A. Market entry and failure as competing events in the dynamics of the R&D pipeline

Market entry is a function of the carrying stock of projects in the R&D pipeline (that we consider here to be equal to the stock of projects in the previous year), i.e. $P_{il,t-1}$, the transition probability of market entry of region l of type i at time t, i.e. $p^{hm}_{il,t}(X)$, and a vector of factors X. The proportion of drug projects that are launched in the market at time t is a stochastic process that is given by:

$$M(X) = M_{il,t}(p^{hm}(X), P_{t-1}) = p^{hm}_{il,t}(X) \times P_{il,t-1} \tag{5.3}$$

Similarly, the number of failures $D(X)$ is a function of the stock of projects in the R&D pipeline at $t-1$, $P_{il,t-1}$, the likelihood of failure of a project of therapeutic type i in region l at time t, $p^{hd}_{il,t}(X)$, and

a vector of factors X. The number of projects in the pipeline that die out at time t is given by:

$$D(X) = D_{il,t}(p^{hd}(X), P_{t-1}) = p^{hd}_{il,t}(X) \times P_{il,t-1} \tag{5.4}$$

These two events of interest - market entry and failure - are modelled as competing risks for the progression of new drugs in pipeline. We can construct a competing risks model with one transient state (staying alive in R&D pipeline) and two absorbing states (market entry and failure). These two absorbing states prevent further transitions to occur to other states. We are assuming here that the transient state "in R&D pipeline" includes the four sequential stages of drug development (Discovery, Phase 1, Phase 2, Phase 3) as we are particularly interested in modelling the risk of market launch, that imposes that all the intermediate R&D stages have occurred.

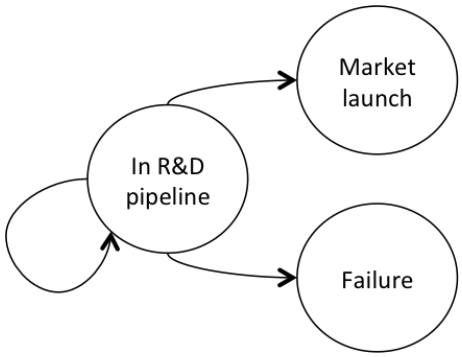


Figure 5.1. Illustration of the competing risks model

These multi-state processes are fully described by the transition intensities, i.e., the cause-specific hazards between states. Let $K = (K(t), t \geq 0, t \in T)$ be a time stochastic process with three possible states $S = \{0, 1, 2\}$, with $K(t)$ being the state occupied at time $t \geq 0$ and T is time interval $\in$ [0,τ]. The transition intensity or hazard rate, conditional on observable characteristics X, between state h and state j is equal to

$$\theta^{hj}(t|X) = \lim_{\delta t \to 0} \frac{Pr^{hj}(t, t+\delta t)}{\delta t} \tag{5.5}$$

Where $h,\ j \in S,\ t \in T$, and the transition probability that is going to inform our simulation is represented by

$$Pr^{hj}(t, t+\delta t|X) = pr\{K(t) = j|X(s), \mathcal{F}_{t-}\} \tag{5.6}$$

where $s \leq t$ with $s,\ t \in T$ and $\mathcal{F}_{t-}$ is a σ-algebra[3] generated by the history of process over [0,t), including state visited and time of transitions. The cumulative incidence function for jth competing risks F_j is defined as the following this expression:

$$F_j(t|X) = pr\{T \leq t, J = j\} = \int_0^t \theta^{hj}(u; \mathcal{F}_{t-}) \times \exp\left(\int_0^u \sum_{j=1}^{J} \theta^{hj}(\nu|X)\, \mathrm{d}\nu\right) \mathrm{d}u \tag{5.7}$$

[3]If $\{A1, A2, A3$, is a countable partition of X, then the collection of all unions of sets in the partition (including the empty set) is a σ-algebra.

This gives the proportion of projects at time t that left the R&D pipeline from cause j, accounting for the fact that projects can leave the pipeline because of two reasons (failure and market launch). This is done by integrating the transition intensity θ^{hj} that are modelled as a function of observable factors X. These observable factors are part of the function of market launch and of the function of failure and they are described in detail in section D. They contribute to estimate the likelihood of market launch and failure of a project.

With the estimation of the likelihood of market launch and failure of a project we predict the proportion of projects that leave the R&D pipeline between a short time interval. The transition probability from pipeline to the market of a project k belonging to the therapeutic class i and firm-type f, in region l at time t, is given by $p^{hm}(t|X)$, and the competing transition probability to failure is given by $p^{hd}(t|X)$, and they can be written as:

$$\begin{cases} p^{hm}(t|X) = f(Y_{il,t-1}, Z_{ifl,t-1}, W_{ki,t-1}) \\ p^{hd}(t|X) = g(Y_{il,t-1}, Z_{ifl,t-1}, W_{ki,t-1}) \end{cases} \tag{5.8}$$

Where $Y_{il,t-1}$ are the set of market-level characteristics, $Z_{jil,t-1}$ are the firm-level characteristics and $W_{ki,t-1}$ represent the project-level characteristics for $t-1$.

After estimating the transition probabilities, the resulting transition matrix for each R&D project is a square positive matrix with the same number of states at t and $t+1$, where $0 \leq \hat{p}_{hj} \leq 1$, $j = \{m, d\}$ that can be written as follows:

$$\begin{pmatrix} \hat{p}^{hh} & \hat{p}^{mh} & \hat{p}^{dh} \\ \hat{p}^{hm} & \hat{p}^{mm} & \hat{p}^{dm} \\ \hat{p}^{hd} & \hat{p}^{md} & \hat{p}^{dd} \end{pmatrix} = \begin{pmatrix} \hat{p}^{hh} & 0 & 0 \\ \hat{p}^{hm} & 1 & 0 \\ \hat{p}^{hd} & 0 & 1 \end{pmatrix} \tag{5.9}$$

Where $\hat{p}^{mm} = \hat{p}^{dd} = 1$ and $\hat{p}^{mh} = \hat{p}^{md} = \hat{p}^{dh} = \hat{p}^{dm} = 0$, due to the absorbing nature of failure and market entry states.

For the prediction of the parameters included in this matrix we need a flexible parametric model specification. This is due to two reasons. First, flexible parametric estimation does not impose restrictions on the shape of the underlying hazard function. The flexible parametric specification approach to competing risks models has been proved to be statistically identifiable even if both risks have identical predetermined variables, delivering the advantage of unrestricted correlated across the stochastic disturbances (Han and Hausman, 1990). Secondly, because the aim of this estimation is to inform the micro-simulation model on the transition probabilities, we need to estimate a model that allows us to predict the effect of time (duration) on the transition intensities. And this can only be done with a modelling approach that allows for time-dependency on the estimated hazard function (Briggs et al., 2006).

We use a spline-based parametric proportional hazard (PH) model to allow for turning points across time given the fact that the estimated hazard functions not necessarily respond to time in the same way. This allows for time-dependency in the effect of a covariate on hazard rates through the life course of a project. The approach taken by Royston and Lambert (2011) is to model the logarithm of the baseline cumulative hazard function as a natural cubic spline function of log time. A detailed explanation of this

model is presented in detail in the appendix V. The estimation results are in line with the results found in Chapter 4 and are briefly explained in the appendix V of this Chapter.

The next section presents in detail the assumptions to model birth of projects in the simulation model.

B. The birth function

Alongside with the market entries and failures, a new cohort of projects $B_{il,t}$ of type i in region l is born and added to the pipeline every calendar year t.

Since we assume that new projects are born at random, the expected number of new projects $B_{il,t}$ of type i in region l at time t follows a Poisson distribution[4]. This implies that the probability of a given number of births occurring in a calendar year can be expressed by a statistical distribution with a known average rate of birth, which is independent of the time elapsed since the last birth had occurred. Therefore we are ale to predict, with a Poisson distribution, the (random) degree of spread of births around a known average rate.

To calculate the average birth rate we have considered the historical information (from 1980 until 2010) that characterises the process of generation of new projects in the R&D pipeline. We have separated the analysis between biologics from chemically synthesized projects (here defined as NCEs), and we have computed the historical average μ_{il} of new projects in early discovery of each type i in region l for the last 30 years.

Therefore, if a birth process $b_{il}(X)$ follows a Poisson model with average rate $\{\mu_{il}^{NCEs}, \mu_{il}^{bio}\}$ for projects of type i in region l, the expected number of births for NCEs and biologics $\{B_{il}^{NCEs}, B_{il}^{bio}\}$ for region l and therapeutic area i satisfies the following equations

$$\begin{cases} B_{il}^{NCEs} = E_{il}^{NCEs}[p|X] = e^{b_{il}^{NCEs}(X)} \text{ if } NCEs \\ B_{il}^{bio} = E_{il}^{bio}[p|X] = e^{b_{il}^{bio}(X)} \text{ if } biological \end{cases} \tag{5.11}$$

There are three underlying assumptions with this method.

First, the historical average computed from the last three decades reflects the long run birth rate of R&D projects started in early discovery. This means that the levels of newborn projects, across therapeutic areas and across regions, show the firms' response to expected economic profits and losses

[4]The Poisson distribution may be derived by considering an interval, in time, space or otherwise, in which events happen at random with a known average number μ. The interval is divided in n subintervals $I_1, ...I_n$ of equal size, such that $n > \mu$. The probability that an event will fall in the subinterval I_k is for each k equal to $\frac{\mu}{n}$, and the occurrence of an event in I_k may be approximately considered to be a Bernoulli trial (Lancaster, 1992; Cameron and Trivedi, 2005). The birth process of drug projects is a random variable that follows a Poisson distribution with parameter $\mu > 0$, if, for $n = 0, 1, 2$, the probability mass function (pdf) of B is given by:

$$f(n;\mu) = pr\{b = n\} = \frac{\mu^b e^{-\mu}}{b!} \tag{5.10}$$

Where e is the base of the natural logarithm and $n!$ is the factorial of n, and μ the parameter that characterises the expected number of births and its variance. The Poisson distribution specifies how likely it is that the count of births will be 3, or 5, or 10, or any other number, during one period of observation. This expresses the probability of a given number of births occurring in one year if these births occur with a known average rate μ and independently of the time since the last event. The historical average birth rate is going to inform on the stochastic process of generation of new projects of $b(x)$, i.e. $b(x) \sim Pois(\mu)$, where $\mu_{ij} = \sum_{il,t} \frac{b_{il,t}}{B_{il}}$, $t = t_1, ..., T$. Given this, the birth of new projects can be simulated through a Monte-Carlo process by considering $b(x)$.

related to the project's expected likelihood of market entry and/or failure. Since we are estimating these probabilities with the full data available for the global R&D portfolio, we have opted to use similar timespan (1980-2010) for the computation of the historical average rates of both NCEs and biologics $\{\mu_{il}^{NCEs}, \mu_{il}^{bio}\}$.

Secondly, that the average rate is justifiable to be computed for both cases of NCEs and biologics. Biologics are distinct from chemically synthesised pharmaceutical products and face different regulation pathways in some jurisdictions. Since we have not find consistent evidence of differences of in project births between NCEs and biologics, we have computed $\{\mu_{il}^{bio}\}$ with the same full data timespan to calculate the average rate of birth for both biologics and NCEs.

Thirdly, as the literature suggests (for instance, as shown by Mestre-Ferrandiz et al. (2012)) there are substantial differences in the R&D effort across regions, and therapeutic areas that characterise the process of generation of new projects. This is the basis for opting for computing market-level (therapeutic-level and regional-level) average rates $\{\mu_{il}^{NCEs}, \mu_{il}^{bio}\}$ for projects of type i from region l.

However, the newborn projects need to be fully characterised to be modelled as functions of market entry and failure, as described in section II. We explain that in detail in the subsection D.

The subsection C describes the micro-simulation model that combines dynamically and over time the processes of market entry, failure and birth of pharmaceutical projects.

C. The dynamic micro-simulation model

There is established literature on the dynamic micro-simulation models that project population samples over time, and model life course events under various policy rules. These processes are based on the creation of transition probabilities to update individual observable characteristics over time, incorporating Monte Carlo simulation methods to generate a stochastic process that determines individual transition probabilities (Willekens, 2009).

We use Monte Carlo simulation methods to establish the actual incidence of the transitions of market launch and failure in consecutive periods of time.

The occurrence and timing of each event are established by comparing, in a first instance, the predicted probability of survival, $\hat{p} = \{\hat{p}^{no\,event}\}$, with a random number, r, drawn from an uniform distribution $U \sim U \in [0, 1]$. There are two possible outcomes: 1) the project survives to the occurrence of any event, i.e. the project remains in the R&D pipeline, if $r \leq \hat{p}^{no\,event}$; or 2) there is the possibility of occurrence of an event if $r \geq \hat{p}^{no\,event}$.

If that is the case that $r \geq \hat{p}^{no\,event}$, then the type of event occurring is established by comparing the set of predicted probabilities for both transition probabilities $\hat{p} = \{\hat{p}^{hm}, \hat{p}^{hd}\}$ with another random number, u, drawn from an uniform distribution $U \sim U \in [0, 1]$. There are two possible outcomes: 1) the transition to market occurs if $u \leq \hat{p}^{hm}$; or, 2) the project dies, otherwise.

Figure 5.2 illustrates the simulation process. The baseline portfolio (in January 2011), that is modified by the transitions incurred by projects at the end of time period t, becomes the input dataset for the following time period, i.e. $t + 1$. This is repeated until the entire simulation cycle is complete (last time period is reached in $t = 2030$).

Given the model specifications and the parameter estimates, we can specify the following functions that characterise the lifetime of projects and the stock of new therapies launched in the market in the coming years:

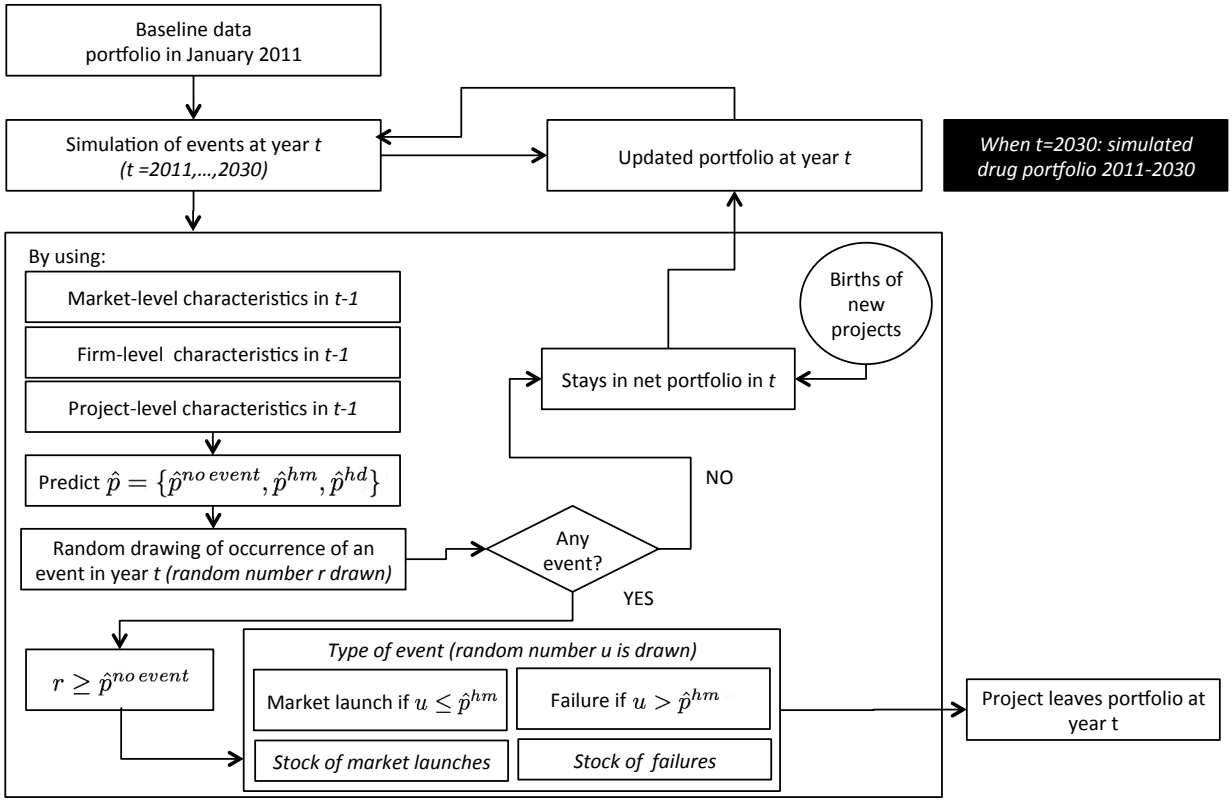


Figure 5.2. Illustration of the micro-simulation model

$$\begin{cases} \hat{N}_{i,t} = \sum \sum_{l,t} \hat{M}_{il,t} \\ \hat{M}_{il,t}(X) = \hat{p}^{hm}_{il,t}(X) \times \hat{P}_{il,t-1} \\ \hat{D}_{il,t}(X) = \hat{p}^{hd}_{il,t}(X) \times \hat{P}_{il,t-1} \\ \hat{P}_{il,t-1} = \hat{P}_{il,t-2} + \hat{B}_{il,t-1} - \hat{D}_{il,t-1} - \hat{M}_{il,t-1} \end{cases} \quad \begin{cases} \hat{N}_{i,t} = \sum \sum_{l,t} \hat{M}_{il,t} \\ \hat{M}_{il,t}(X) = \hat{p}^{hm}_{il,t}(X) \times \hat{P}_{il,t-1} \\ \hat{D}_{i,l,t}(X) = \hat{p}^{hd}_{il,t}(X) \times \hat{P}_{il,t-1} \\ \hat{P}_{il,t-1} = (1 - \hat{p}^{no\,event}(X)) \times \hat{P}_{il,t-2} + \hat{B}_{il,t-1} \end{cases} \tag{5.12}$$

Where $\hat{D}_{il,t-1} = \hat{p}^{hd}_{il,t}(X) \times \hat{P}_{il,t-2}$ and $\hat{M}_{il,t-1} = \hat{p}^{hm}_{il,t}(X) \times \hat{P}_{il,t-2}$, representing respectively the expected number of failures of type i in region l at time $t-1$, and the expected number of market launches of type i in region l at time $t-1$, as functions of the expected portfolio of projects in $t-2$, $P_{il,t-2}$. $\hat{B}_{il,t-1}$ is the expected number of births of project type i in region l during $t-1$. Given this, we predict the number of new therapies of type i at time t, $\hat{N}_{i,t}$, by summing up the market launches of drugs of type i at time t, either being of chemical (NCEs) or biological composition, across the different regions and therapeutic areas.

We detail in the next section the definition of market and the measurement of the covariates used for the estimation of the transition probabilities and the simulation process.

D. Data and variables

Economic literature suggests that market entry is a function of market size, the level of competition, and the fixed costs associated with product launch. Empirical findings prove that the number of entries increases with the size of the market but at a decreasing rate (Bresnahan et al., 1987; Bresnahan and Reiss, 1991, 1990; Scott Morton, 1999).

However, and as already discussed in Chapter 4, the literature is scarce in analysing the factors influencing the exit and, especially, the disruption of R&D. Chapter 4 of this thesis is the first attempt to measure the association of competition and alliances with the probability of failure across the different R&D phases. Three groups of factors influencing market launch and failure of R&D projects are discussed in Chapter 2 and can be briefly outlined here. They are respectively: i) market-level characteristics; ii) firm-level characteristics; and iii) project-level characteristics.

In this section, we discuss these factors that significantly affect the decision to launch a new drug or abandon a drug project. For the measurement of the market-level, firm-level and project-level characteristics, we have followed the same strategy as presented in Chapter 4. These covariates are dynamically updated at every calendar year in the simulation process.

Market-level characteristics

We incorporate in this analysis the size of the market and competition forces as these have been found factors to affect significantly the R&D process, as discussed in Chapter 4. Indeed the literature argues that on one hand, there is a concentration of R&D investment in more lucrative therapeutic areas (as the literature has suggested, for instance see Acemoglu and Linn (2004)). However, those markets, on the other hand, also present a higher risk of failure, specifically when projects are in early stages of discovery. Similarly to Chapter 4, we have used a fixed effect at regional level to proxy market size for the jurisdiction where the R&D process takes place. To avoid including more complexity in the simulation model, we have opted to not include forecasts on the GDP per capita and population, and proxy the market size by controlling for the fixed effects at regional level.

Furthermore, the literature has shown evidence of a two-sided effect of competition on R&D: first, the negative effect of competition from incumbents on entry of new medicines targeting the same therapeutic classification; secondly, the positive effect of competition from peers in pipeline on entry, as shown in Chapter 4 and, for example, in Berndt et al. (1996). To proxy competition we have considered the two measures presented in Chapter 4: (1) competition at the final product market, and (2) competition from peers, i.e., within the R&D process. Specifically, final product market competition at time t is measured by counting: (i) the number of drugs established in the market in the last five years before $t-1$ (*new drugs*); and, (ii) the number of drugs launched in the market previous to $t-6$ (*old drugs*). Moreover, competition in the R&D process at time t is measured by counting the number of projects being developed for the same market in $t-1$ (*potential competitors*).

In addition to size and competition forces in the market, the regulatory environment has been suggested to significantly affect drug prices (for example Danzon and Chao (2000)) and entry costs (for instance Djankov et al. (2002)) and, ultimately, to affect market launch of new drugs (Danzon et al., 2005c; Lanjouw, 2005; Kyle, 2007). Countries with stricter entry regulation and relatively little price regulation, such as the United States and the United Kingdom, show highly concentrated domestic industries whose

products diffuse more extensively to foreign markets (Thomas, 1994). Given that, we have included a regional fixed effect to account for the effect of the regulatory and technological environment on market entry and failure of projects.

For the market-level characteristics of the newborn projects, we fully characterise the size of the market and the regulatory environment by identifying the region where they are born. The competition forces, i.e., the number of *young drugs*, *old drugs*, and *potential competitors* are characterised dynamically by the microsimulation model. A project being born at time t faces the number of *young drugs* that a project of type i in region l has faced in time $t-1$. The same dynamic process happens to feed *old drugs*, and *potential competitors*.

Firm-level characteristics

The literature suggests that firm-level characteristics and product-level characteristics matter when explaining the speed at which a new drug is approved, since the costs of gaining regulatory approval differ across firms and types of drugs within a regulatory environment (Dranove and Meltzer, 1994; Carpenter, 2004).

The size and type of firm is suggested to significantly impact on market entry and the likelihood of failure of a R&D project (as presented in Chapter 4 and, for instance, in Abrantes-Metz et al. (2004); Cockburn and Henderson (1998, 1994)). Following the Chapter 4, we have measured the size of the firm by identifying the participation of a TOP100 firm in the R&D project (*big firm*), and the participation of an academic/public institution in the R&D project (*academia participant*). The newborn projects have been attributed an average level of participation of a big firm and academia of projects of type i from region l, for the two types of products, i.e., NCEs and biologics.

Project-level characteristics

Chapter 4 and the existing literature specify therapeutic novelty, composition of the product (biological vs chemical), therapeutic indication, duration of the project and collaboration intensity as relevant factors impacting entry and failure of a drug project (Abrantes-Metz et al., 2004; Kyle, 2006; Mestre-Ferrandiz et al., 2012).

We have measured alliances, technological specificities and duration of the project using the same approach as in Chapter 4.

To measure collaborative intensity of the project (*alliances*), we have calculated the *log* number of firms collaborating in each R&D project. We have also included an interaction term which considers the log number of firms collaborating in each R&D project with the presence of a big firm. The newborn projects have been attributed an average level of log number of firms collaborating of projects of type i in region l, for the two types of products, i.e., NCEs and biologics.

We have used a fixed effect at therapeutic area level to control for different technological and scientific specific conditions in each therapeutic category that influence effort and the probability of failure and market launch of a particular project.

Finally, we have included in the analysis the duration of the project (duration) by ageing the project one year if the project survives and stays in the R&D pipeline to the following year of the microsimulation. For the newborn projects we have been attributed a one-year duration as initial duration.

III. Results

In this section we discuss the historical trends for the launch of new therapies in the five regional markets, and report the descriptive statistics for the results of the simulation. In section B we examine the results of the microsimulation analysis. The results for the parameter estimation that support the simulation model can be found in the Appendix V.

A. Descriptive statistics

Figure 5.3 presents the historical trends for the (very first) launch of, respectively, NCEs and biologics in the five regions. For both types of therapies, USA and EMA lead the number of new launches. For NCEs, and with the exception of the trends between Japan and EMAP, the differences in trends across the five regions are statistically significant. For the biologics, the only statistical significant differences happen to be between the USA and EMA, and the EMAP and the rest of the world.

[Figure 5.3 here]

With the exception of EMA's trend for NCEs, there is a slight decreasing trend of the number of market launches in the years preceding 2010 across all regions for both types of products. However, similar downward trends are found in past periods (Figure 5.3) shedding some light on the debate about productivity crisis or natural inflection points in the business cycle of the pharmaceutical industry (Cockburn, 2007; Munos, 2009).

With regards to NCEs, there are clearly two important periods for the industry in the USA and EMA markets: (i) the period between 1992 and 2005, showing a slight uprising trend on the number of new therapies marketed; and (ii) the period between 2005 and 2010 that shows a steeper increasing trend. Moreover, the Japanese market shows a different pattern following the year of 1995, when there is a decreasing number of new therapies being launched. In contrast, the emerging markets and the rest of the world and other markets show an uprising trend since 2000.

With respect to the biologics, there are three distinctive periods: (i) the period from 1992 to 2001, showing a steep uprising trend for the number of market launches; (ii) the period between 2001 and 2004, revealing a drastic fall in the number of new therapies launched; and, (iii) the years following 2004 with a uprising trend (even though slightly flatter than the trend shown in the nineties). Interestingly, trends in Japan exhibit a small variation across time in the market launch of biologics, as they show persistent low numbers during the observed period. This may be related to the fact that, historically, firms have chosen to first place their products in the European/North American markets (IMS, 2010; IMAP, 2011). It is interesting the exponential uprising trend of new biological therapies being launched in the emerging economies after 2000, as the industry reports have been stating (IMS, 2010; IMAP, 2011). The regional differences are, however, non-significant across regions, with the exception of the difference between the USA and EMA (p-value=0.0542).

When looking at the historical trends across the main therapeutic groups, there are clearly two dominant groups showing a rapid increase in the number of new therapies over time: ATC category J, general anti-infectives, and ATC category L, antineoplastic and immuno-modulating agents. These two ATC categories dominate the number of new therapies in the market for the NCEs (Figure 5.4) and biologics (Figure 5.5).

[Figure 5.4 and Figure 5.5 here]

The US and EMA markets lead the market launch of new therapies launched in these two categories. Figure 5.6 and Figure 5.7 show the historical trends for the NCEs for both regions (with a p-value=0.0000 between both regions). Interestingly, and for the rest of the regions (Japan (Figure 5.8), EMAP (Figure 5.9) and Others (Figure 5.10)), there is a larger number of therapies being launched in the market targeting the group of the general anti-infectives. Japan presents significantly higher number of new therapies than the USA (p-value=0.0116) and EMA (p-value=0.0592).

[Figures 5.6 - 5.10 here]

This scenario changes when looking at the market of new biologics. The USA and EMA regional markets have experienced in the last two decades a larger proportion of new biologics for anti-infectives compared with the the number of antineoplastic drugs and immunomodulating agents. The Japanese and EMAP markets show a less consistent trend of new launches on these ATC categories (Figures from Figure 5.15 to Figure 5.19 showing these trends can be found in the technical appendix IX).

[Figure 5.15 - Figure 5.19 here]

B. Micro-simulation results

We run the model specification presented in section A for NCEs and biologics separately (the results can be found in the technical appendix V), and we have run 200 repetitions of the Monte-Carlo simulation. Given the large volume of results, we present in this section aggregated summaries and statistical tests at regional and main therapeutic level (ATC-2 level), and a few detailed cases that illustrate the results of the simulation analysis. This is because the differences across ATC-3 level can be captured at ATC-2 level. All detailed results are available upon request.

The model includes three levels of covariates that are dynamically updated with the simulation process. These three level of covariates, that are explained in section D, are respectively: (i) market-level characteristics, including proxies for market competition (*new drugs*, *old drugs* and *potential competitors*), market size (regional fixed effects) and regulatory and technological environment characteristics (regional fixed effects); (ii) firm-level characteristics, including a dummy for firm size (*big firm*) and the type of organisation that is investing in the R&D project (*Academia participant*); and finally, (iii) project-level characteristics that include a measure of the intensity of alliances of the project (log *number of firms*), an interaction term between the intensity of alliance and the participation of a big firm in the project (*big firm* * log nr of firms), the *duration* of the project, and therapeutic area fixed effects.

We present in appendix V the descriptive statistics and corresponding t-tests for mean-comparison testing of the simulated number of new therapies, births, duration of projects, probability of no-event, transition probability to market and transition probability to failure. Tables 5.2 and 5.3 present, respectively, the regional descriptive statistics and the t-tests for NCEs (Tables 5.4 and Table 5.5 for biologics), whereas Tables 5.7 and 5.9 present, respectively, the descriptive statistics at therapeutic level and corresponding t-tests for NCEs (Table 5.8 and Table 5.10 for biologics).

The results of the micro-simulation suggest a clear continuation of the global leadership of the USA, followed by EMA, on the market launches of NCEs and biologics. The USA secures a relatively significantly higher number of market launches of NCEs (913.38 and 776.49 with p-value = 0.0303) and biologics (699.16 and 629.06 with p-value = 0.0324) than the follower, EMA.

At the same time, the results for both of these regions show consistently comparable numbers for the R&D activity in NCEs and biologics, as shown in Table 5.3 and Table 5.5. There are no statistical significant differences between both regions in what relates to the number of new projects (33.90 versus 31.45 with p-value = 0.8644 for NCEs; and, for biologics 1.26 versus 1.31 with p-value = 0.7975). The same happens to what regards the duration of projects in pipeline (9.79 years versus 9.35 years with p-value = 0.1782 for NCEs; and, for biologics 9.78 years versus 9.42 years with p-value = 0.4654), and probabilities of market launch (0.4 versus 0.39 with p-value = 0.2513 for NCEs; and, for biologics 0.37 versus 0.37 with p-value = 0.7975) and failure (0.49 versus 0.50 with p-value = 0.1823 for NCEs; and, for biologics 0.53 versus 0.53 with p-value = 0.5738). Therefore, both regions are perceived to have similar R&D landscapes, translating the past trends of similar levels of R&D investment concentrated in these regional areas (IMS, 2010; IMAP, 2011).

On the other hand, the regional markets of EMAP and Japan show relatively similar future trends for the market launches for NCEs, with similar average number of new projects being born (202 in Japan and 199 in EMAP, which are not statistically different from each other (p-value=0.9813)), probabilities of market launch (0.346 and 0.351, respectively, with p-value = 0.7078) and duration of projects (of approximately 7.5 years and 7.7 years, respectively, and p-value = 0.6958). The sales growth rates of the pharmaceutical industry in emerging economies like China, Brazil, Russia, India, and South Korea have been reporting an explosion and redefinition of the global pharmaceutical business (IMS, 2010).

On top of these regional disparities, there are persistent differences across the therapeutic categories.

For the market for NCEs, the results suggest that the USA will unequivocally lead the market launches in drugs targeting the nervous system (ATC N, Figure 5.11k). There are significant favourable differences from EMA market (2nd) on the number of market launches (41.85 versus 27.57 with p-value = 0.0124), duration in the R&D pipeline (12.6 years versus 9.9 years with p-value = 0.0027), probability of market launch (0.45 versus 0.41 with p-value = 0.0107) and probability of failure (0.49 versus 0.51 with p-value = 0.0308). There is also some evidence of the dominance of the US on the R&D activity on drugs targeting the musculo-skeletal system (ATC M, Figure 5.11i). The number of market launches in the US is significantly higher than those launched in the EMA market (27.041 versus 20.562 with p-value = 0.0997), and the US presents a more mature R&D pipeline (9.55 years versus 9.0 years with p-value = 0.014).

Furthermore, the EMAP regional market will affect greatly the global market for antineoplastic and immunomodulating agents (Figure 5.11h). The emerging markets are expected to register a significantly promising number of more than 80 new anticancer chemical therapies per year, contrasting with 58 in EMA, 75 in the US, 32 in Japan and 16 in the rest of the world (even though the differences are not shown to be statistically significant with corresponding p-values of 0.7428, 0.9597, 0.3011 and 0.191). The results suggest a "catching-up" phenomenon at the level of anti-cancer drugs between emerging markets and the global leaders, USA and EMA. This is reinforced by the non-significant differences for the probability of failure across regional markets (0.493 against 0.492 in the US (p-value=0.9064), 0.502 in EMA (p-value=0.2516), and 0.506 in Japan (p-value=0.1805)). As discussed above, this "catching-up" phenomenon is in line with the most recent forecasts from the industry reports for the emerging economies (IMS, 2010).

[Figure 5.11 here]

The market for biologics shows, however, to be regionally more concentrated. The USA, EMA and Japan will continue to play the main role in most of the therapeutic areas, particularly in those areas that target the blood and blood forming organs (ATC B, Figure 5.12b), dermatological conditions (ATC

D, Figure 5.12d), systemic hormonal preparations (ATC H, Figure 5.12f), respiratory systems (ATC R, Figure 5.12l), and sensory organs (ATC S, Figure 5.12m). EMAP and the rest of the world show no R&D activity targeting these therapeutic groups.

Furthermore, the US will lead the number of new biologic therapies being placed in the global market for four main therapeutic groups. These are: alimentary and tract metabolism (ATC A, Figure 5.12b), blood and blood forming organs (ATC B, Figure 5.12b), genitourinary system and sex hormones (ATC G, Figure 5.12e) and nervous system (ATC N, Figure 5.12k). The number of new therapies is significantly higher in the US market than in the second-best placed market (for ATC A, 29.4 versus 20.4 in EMA with p-value = 0.0118; for ATC B, 12.54 versus 8.29 in EMA with p-value = 0.0038; for ATC G, 6.66 against 2.49 in Japan with p-value = 0.0984; and, finally for ATC N, 26.65 versus 19.52 with p-value = 0.0496).

[Figure 5.12 here]

These aggregate figures disguise some dispersion found in the results at a more refined level. We present here two cases that depict such variation within main therapeutic groups. For instance, by looking at specific products, for instance, for the specific case of antineoplastics (ATC code L: antineoplastic and immunomodulating agents) and vaccines (ATC code J: anti-infectives for systemic use), the results suggest a clear dominance of other regions on the future production and market launch of new therapies.

The case of antineoplastics

Antineoplastics act to prevent, inhibit or halt the development of tumors or malignant cells, and are used to treat, for instance, metastatic cancers and other forms of cancer. Some of these chemotherapy drugs are also used to treat other conditions, including multiple sclerosis, Crohn's disease, psoriasis, psoriatic arthritis, systemic lupus erythematosus, rheumatoid arthritis, and scleroderma.

The results of the micro-simulation show that there is a substantial expected number of new antineoplastic agents with chemical composition being placed in the EMAP regional market (Figure 5.13(d)). In 2020, the best-scenario forecast for the number of new therapies placed in the USA (Figure 5.13(a)) is around ten times smaller than the number of new therapies expected to be placed in the EMAP market (Figure 5.13(d)). Even though the USA, EMA and Japan show higher probability of market entry and lower probability of failure, there is a much greater number of newborn projects in EMAP compared with those regional markets (yearly average of 23 versus 1.3 in EMA, 2.9 in Japan, 1.1 in the USA, and 1.3 in the rest of the world).

[Figure 5.13 here]

The case of vaccines

Vaccines are biological preparations that improve immunity to a particular disease. Vaccines typically contain an agent that resembles a disease-causing microorganism and are often made from weakened or killed forms of the microbe[5].

When looking at the production and market availability of vaccines in the future, the results suggest that the pure biological production of vaccines will quite steadily increase to around six vaccines to be

[5]Vaccines can be prophylactic (example: to prevent or ameliorate the effects of a future infection by any natural or "wild" pathogen), or therapeutic (e.g., vaccines against cancer are also being investigated).

available in each regional market. However, the results clearly suggest more promising numbers when looking at the vaccines produced with mixed chemicals and other non-biological components[6]. By 2030, the regional market EMAP (Figure 5.14(d)) is expected to provide more than three-fold of the number of new vaccines in the market available in the rest of the world (Figure 5.14).

[Figure 5.14 here]

IV. Discussion and conclusions

This paper aims at forecasting the future global R&D landscape and the market launch of new therapies with both chemical and biological composition, assuming no major policy changes by governments or regulators. We do this by developing a micro-simulation model that incorporates three crucial events of the R&D lifecycle of drug projects: birth, failure and market entry. The micro-simulation model is dynamically updated over time and uses parameter estimates of a flexible parametric competing risks model. With this flexible parametric competing risks model we model the likelihood of market entry as a competing event of failure of projects as a function for market, firm and project-level characteristics, and time to the occurrence of any of those events.

Three main conclusions arise from this analysis.

First, the results suggest that the historical regional disparities are expected to persist and that the United States will continue leading the global market launches of new chemical and biologic therapies. This is shown by the significantly higher number of market launches in the US compared with those in the EMA market, that comes usually second in the global ranking. These results express the historical high-value of the US market: the US market in 2011 was valued 31% of the global pharmaceutical sales (PWC, 2011a).

Specifically, the United States are expected to drive global innovation in new therapies targeting the nervous system of both chemical and biologic composition, as there are favourable statistically significant differences from the EMA for this market on the number of market launches, duration in the R&D pipeline, probability of market launch and probability of failure. The same is true for new chemicals targeting the musculo-skeletal system, given the significantly higher number of market launches in the US and the more mature R&D portfolio (higher duration of projects). This is also the case for the markets for biologics, where the USA will be the established global innovator showing significantly higher number of market launches in therapies targeting the alimentary and tract metabolism and genitourinary systems, and new biologic components to produce blood and blood forming organs.

These facts are in line with the past trends and reinforce the business forecasts for the coming decades. By 2020, the industry reports that the US will represent 27% of the 1.6 trillion USD that the pharmaceutical market will globally value (PWC, 2011a; IMAP, 2011). The better performance of the US at the global market level derives from three reasons. First, the differences in demand growth between the two regional markets have led to a more attractive market in the United States. Pricing and reimbursement policies differ across these regions and EMA has been facing large pressures from national health systems to reduce health spending, having engaged in stricter price and reimbursement regulations, and stricter health technology assessment requirements (PWC, 2011a; IMS, 2010; Morton and Kyle, 2012). Second, the US competitive advantage in innovation relies on higher intensity of collaboration between firms and

[6]Beside the active vaccine itself, there are several other excipients that are commonly present in vaccine preparations, namely aluminium salts or gels, antibiotics, proteins, amino-acids and chemicals.

public research centres, especially in pre-clinical stages of discovery (Pammolli et al., 2001). These have flourished in a sheltered environment by public and philanthropic money devoted to science (Wadman, 2007; Odling-Smee, 2007) that has deepened the US competitive advantage over the "molecular biology revolution" (Pammolli et al., 2001). Finally, the US has been perceived as a more secure IP system that lowers uncertainties about patenting and licensing (PWC, 2011a). Evidence suggests that the US is a preferred destination of research, including by the European companies, resulting in a higher share of patents in new biotech fields (Pammolli et al., 2001). In fact, big firms, as Pfizer, have based in both markets and have strategically decided to relocate their R&D department from the UK to the US in 2011. Also, the growing economies, which are, nowadays, to some extent competing with the US for market share at global level, are seen as currently lacking financial and legal power to enforce the international IP system and reward innovation (PWC, 2011a; Lanjouw, 2005).

Secondly, this analysis suggests comparable R&D productivities between the United States and EMA. There are no structural differences in the market entry of NCEs and biologics between both regions in what relates to the duration of projects in pipeline, and corresponding probabilities of market launch and failure. This is in line with the evidence presented by Pammolli et al. (2011).

In fact, this might be revealing how these markets are similar at several levels. First, both US and Europe show identical health needs. As the latest disease burden numbers show (Murray et al., 2013), North America and Western Europe suffer from similar conditions, with ischaemic heart disease, trachea, bronchus and lung cancer, and chronic obstructive pulmonary disease, representing common main causes of mortality and disability. Secondly, the regulatory environment of both regions is very similar. The regulators for marketing approval in both regions, the European Medicines Agency (EMA) and the FDA, have gone through a harmonisation of requirements for marketing approval, enabling firms to prepare a common dossier for entry in both markets (Danzon, 2011; Morton and Kyle, 2012). The regulation gaps found in one region about orphan drugs, biologics and biosimilars have produced translational legislation on the other jurisdiction (Danzon, 2011; Morton and Kyle, 2012). Furthermore, there is an almost perfect overlap of the intellectual property rights environment in both regional markets (Morton and Kyle, 2012).

Thirdly, the emerging markets are expected to show noticeable achievements in bringing to the market promising number of new anticancer drugs, also used in a number of rare diseases, and other specific products such as vaccines, as a more refined analysis suggests. The emerging economies have been steadily gaining share at the expense of the US and the top five European countries (France, Germany, Italy, UK and Spain), accounting for close to 34 percent of global growth in 2009 (IMAP, 2011).

There are a number of reasons why this might be occurring. On the one hand, the emerging economies are becoming comparably more attractive than the well-established markets of Europe and the US. The industry reports forecast an explosion for the demand for medicines in growing economies that will, by 2020, represent more than 32% of the world global sales volume (PWC, 2011a). The great wealth of the richest, and an affluent "middle-class" in these countries, that present identical health needs to the developed world (Murray et al., 2013; PWC, 2011a; IMS, 2010), represent greater business opportunities, that contrast with the pessimistic economic environment in EMA (PWC, 2011a). Moreover, many of these countries are expanding the coverage of their national health systems, generating large market opportunities for the industry (Lancet, 2010; Evans and Etienne, 2010). On the other hand, the literature has been reporting a (re)localisation of R&D (mostly clinical trials) in emerging economies (Berndt et al., 2007; Thiers et al., 2008). This is regarded as a response to various economic, scientific and regulatory issues that are deterring successful innovation in the USA and in the EMA market. They

are namely related to issues around the regulatory approval process in these regions, namely regarding the comparability and extrapolation of safety and effectiveness data between biologics and chemicals, and a reference product (Dingermann et al., 2012; Kelly and Mir, 2009; Grabowski, 2008). Finally, the explosion of R&D activity in these markets might derive from the fact that firms seem to have intensified their global strategic behaviour in emerging economies. Global pharmaceutical players have acquired strategically some domestic manufacturing companies to penetrate these markets, accounting for 50% of the merger and acquisitions between 2008 and 2010 (IMAP, 2011).

There are several caveats to our analysis that relate to the availability and the quality of our data. First, projects start in a region and are assumed to develop, fail or enter in the market on that specific region, not taking into account the possible relocation of the project to another region, which might not be necessarily true. More detailed data at project level in terms of relocation could inform our model on this possibility. Second, we are relying on the last thirty years of data to give us an accurate picture of the dynamics and composition of the R&D pipeline. Without more years of information, we had to decide to use all the information we have available to carry this analysis. However, this means that possible future unpredictable (from the past trends and numbers) shocks in the pharmaceutical industry are not captured by this analysis. These shocks might be originated from new players entering in the market, changes in funding science, new regulatory settings that could change the nature and relocation of R&D or any other macroeconomic level changes.

Addressing these limitations requires more refined historical data at firm and project level and incorporation of random shocks in the analysis. Also, there are potential modeling extensions for this study to introduce unforeseeable uncertainty to the model. Despite these caveats and possible extensions, we retain our analysis as a relevant insight for the future of R&D landscape, and a useful tool that can be extended for the evaluation of policy changes in the global pharmaceutical industry. Many of these changes might come from challenges around the unmet global health needs, that have been driving the attention of public health policymakers, governments, the industry, and several private and public donors on funding more biomedical research and development.

V. Appendix: Tables

Spline-based parametric PH model

We introduce splines in our flexible parametric PH models to account for time-dependency in the effect of a covariate on hazard rates through the life course of a project. We take quite literally the lessons from Royston and Lambert (2011) to apply this model to our case. The approach taken by Royston and Lambert (2011) is to model the logarithm of the baseline cumulative hazard function as a natural cubic spline function of log time, so the cumulative general function of the baseline hazard $h_j(t_j)$ of the previous Chapter 4 is approximated by a spline function. For a hazard rate such as

$$\theta_{hj}(t_j|x) = \lambda_j(t_j) \times \theta_{0,j} \tag{5.13}$$

Royston and Lambert (2011) show that the PH spline model with fixed covariate vector x may be written as

$$\ln \theta^{hj}(t_j|x) = \ln \bigwedge_{0,j}(t) + \beta' x = s(z_t) + \beta' x \tag{5.14}$$

where $\theta^{hj}(t_j|x)$ represents the cumulative subhazard function of jth cause conditional on x observables. The ln baseline cumulative hazard function $\ln \bigwedge_{0,j}(t)$ can be written as a natural cubic spline function of log time, where the natural cubic line splines are defined as cubic splines constrained to be linear beyond boundary knots $\{k_{min}, k_{max}\}$. Such knots are placed at the extreme observed z-values. In addition, m internal knots $k_1 \cdots k_m$ are specified, with $k_1 > k_{min}$ and $k_m < k_{max}$. For three internal knots ($m = 3$), , Royston and Lambert (2011) suggest that the natural cubic spline may be written as

$$s(z) = \gamma_0 + \gamma_1 z + \gamma_2 \nu_1(z) + \gamma_3 \nu_2(z) + \gamma_4 \nu_3(z) \tag{5.15}$$

where the ith basis function is defined for $i = 1, 3$ as

$$\nu_j(z) = (z - k_i)_+^3 - \gamma_i (z - k_{min})_+^3 - (1 - \gamma_i)(z - k_{max})_+^3 \tag{5.16}$$

and

$$\gamma_i = \frac{(k_{max} - k_i)}{(k_{max} - k_{min})} \tag{5.17}$$

$$(z - a)_+^3 = \max\{0, (z - a)^3\} \tag{5.18}$$

As Royston and Lambert (2011) mention the "curve complexity is governed by the number of degrees of freedom, which equals $m + 1$", in this case four since we considered two knots that are placed in the ln time function at the 33 and 67 centiles of the distribution of the uncensored log survival times.

Results from parametric estimation

We run and discuss results of the model that was discussed in Chapter 4. The model includes three levels of covariates we are interested in exploring: (i) competition proxies (final product from new and old drugs and competition within the R&D process from potential competitors); (ii) alliances (log of the region-therapeutic area average number of firms collaborating, region-therapeutic area average of big

firm participation, region-therapeutic area average of academia participation, and interaction between the log number of firms and the presence of a big firm in the alliance); and, (iii) market size (proxied by region-level fixed effects). The model also controls for therapeutic level fixed effects. This model does not fail the linktest and respect the proportionality assumption.

We use a flexible parametric function to approximate the baseline hazard by using two modelling approaches. First, we consider two separate baselines, one for each of the causes - market launch and failure - to allow the shape of the baseline hazard to vary between these two types of events. Secondly, we include a non-linear effect of log time on the baseline hazards and their time-dependent effects. We have used 2 degrees of freedom (d.f.) for the baseline and only 2 d.f. for the time-dependent effects on the events of interest). The model failed to converge when using more than 2 d.f. for the time-dependent effects. This failure is likely to be due to there being a small number of projects in some markets with these events.

Table 5.1 reports estimation results for the subhazard rates of the semi-parametric competing risks model for NCEs and biologics. The results refer to the reference case region USA, year 1990, and therapeutic main group of otologicals and ophthalmologicals. The magnitude of the estimations is not different from estimation results for the hazard of failure in Chapter 4, in what regards to the covariates of interest competition, alliances and market size. However, individual parameters are almost impossible to interpret, and so we use predictions to infer about the effect of our characteristics of interest.

Table 5.1. Parameter estimation results: Competing risks model with spline function

Covariates	Parameter estimates	
	NCEs	Biologics
Market launch	1.342	0.65
	(-0.411)	(-0.354)
Failure	0.602*	0.257**
	(-0.185)	(-0.142)
Young drugs	1.003**	0.993
	(-0.00148)	(-0.0041)
Old drugs	0.996*	0.999
	(-0.00191)	(-0.0062)
Potential competitors	0.998**	1.003
	(-0.00077)	(-0.0022)
ln firms	0.946	1.025
	(-0.0663)	(-0.199)
Big firm	1.105*	1.12
	(-0.12)	(-0.121)
Academia participant	1.026	1.127
	(-0.0573)	(-0.169)
Big firm $*$ ln $firms$	0.742***	0.620**
	(-0.0558)	(-0.128)
Restricted cubic spline (rcs) function used for the baseline hazard rate		
_rcs_market1	3.138***	3.420***
	(-0.0566)	(-0.2)
_rcs_market2	1.119***	1.112***
	(-0.0101)	(-0.0307)
_rcs_failure1	3.846***	4.078***
	(-0.147)	(-0.407)
_rcs_failure2	0.992	0.901**
	(-0.0169)	(-0.0379)
Time-dependent effects		
_d_rcs_market1	3.138***	3.420***
	(-0.0566)	(-0.2)
_d_rcs_market2	1.119***	1.112***

	(-0.0101)	(-0.0307)
_d_rcs_failure1	3.846***	4.078***
	(-0.147)	(-0.407)
_d_rcs_failure2	0.992	0.901**
	(-0.0169)	(-0.0379)
	Fixed effects: region, therapeutic areas	

Notes. Exponentiated coefficients; Standard errors in parentheses.
* $p < .10$, ** $p < .05$, *** $p < .01$
The results refer to the reference case country USA,
year 1990, and category of otologicals and ophthalmologicals.
Therapeutic level parameter estimates available upon request.

Table 5.2. Simulated NCEs across regional markets - descriptive statistics

Region	NCEs		Births		Duration		Prob(no event)		Prob(market launch)		Prob(failure)	
	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd
USA	913.38	33.90	251.67	15.70	9.791372	1.741007	0.11097	0.029486	0.400754	0.038463	0.499816	0.006404
Japan	405.61	22.40	202.70	13.85	7.493668	0.955991	0.201202	0.055379	0.345777	0.024121	0.502513	0.006029
EMA	776.49	31.45	255.36	15.80	9.350002	1.930296	0.109543	0.037549	0.393139	0.043657	0.503527	0.009827
EMAP	419.13	25.45	198.95	15.85	7.701065	1.008116	0.135431	0.036781	0.35098	0.026472	0.499643	0.00537
Others	517.38	27.25	246.38	18	7.385184	0.505511	0.107732	0.032101	0.346431	0.016456	0.502914	0.008871

Notes. Average and standard deviations calculated from 200 Monte-Carlo simulations.

Table 5.3. Simulated NCEs across regional markets - Testing mean pairwise differences (t-tests)

NCEs								Births							
Region	Japan		EMA		EMAP		Others		Japan		EMA		EMAP		Others	
USA	0.0019	***	0.0303	**	0.0009	***	0.0405	**	0.3696		0.8644		0.6311		0.9324	
Japan			0.0009	***	0.8827		0.6008				0.1982		0.9813		0.6852	
EMA					0.0059	***	0.1707						0.5058		0.9091	
EMAP							0.7634								0.7768	
Duration									Prob(no event)							
Region	Japan		EMA		EMAP		Others		Japan		EMA		EMAP		Others	
USA	0.0005	***	0.1782		0.0001	***	0.0001	***	0	***	0.7777		0.0003	***	0.589	
Japan			0.0014	***	0.6958		0.6904				0	***	0	***	0	***
EMA					0.001	***	0.0008	***					0	***	0.7793	
EMAP							0.1905				0	***			0.0006	***
Prob(market launch)									Prob(failure)							
Region	Japan		EMA		EMAP		Others		Japan		EMA		EMAP		Others	
USA	0.0001	***	0.2513		0	***	0	***	0.3468		0.1823		0.9526		0.4103	
Japan			0.0003	***	0.7078		0.9203				0.8044		0.0439	**	0.8523	
EMA					0.0004	***	0.0002	***					0.4469		0.9013	
EMAP							0.48								0.0961	*

Notes. * $p <.10$, ** $p <.05$, *** $p <.01$

Table 5.4. Simulated Biologics across regional markets - descriptive statistics

Region	Biologics		Births		Duration		Prob(no event)		Prob(market launch)		Prob(failure)	
	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd
USA	699.16	29.23	1.26	1	9.777629	2.016145	0.18241	0.064694	0.373276	0.051612	0.529728	0.015211
Japan	563.34	50.05	1.14	1.15	7.511153	1.00361	0.165529	0.068542	0.315059	0.023823	0.532519	0.011875
EMA	629.06	19.20	1.31	1.2	9.419939	2.217615	0.169516	0.058415	0.367424	0.061007	0.531979	0.012952
EMAP	302.76	59.79	1.10	0.8	7.567262	0.977185	0.147579	0.066156	0.316199	0.024876	0.529573	0.010635
Others	350.27	45.15	1.78	0.7	7.334253	0.696551	0.207292	0.05392	0.311137	0.02533	0.539259	0.013113

Notes. Average and standard deviations calculated from 200 Monte-Carlo simulations.

Table 5.5. Simulated Biologics across regional markets - Testing mean pairwise differences (t-tests)

Biologics								Births							
Region	Japan		EMA		EMAP		Others	Japan		EMA		EMAP		Others	
USA	0.0094	***	0.0324	**	0.0144	**	0.0126 **	0.0003	***	0.7975		0.0013	***	0.0406	**
Japan			0.0066	***	0.0322	**	0.4512			0.0098	***	0.0584	*	0.3928	
EMA					0.011	**	0.0042 ***					0.0144	**	0.0042	***
EMAP							0.1324							0.1344	
Duration								Prob(no event)							
Region	Japan		EMA		EMAP		Others	Japan		EMA		EMAP		Others	
USA	0.0029	***	0.4654		0.0089	***	0.0018 ***	0.0677	*	0.2631		0.1966		0.0625	*
Japan			0.0369	**	0.6194		0.4346			0.3814		0.5585		0.0229	**
EMA					0.0561	*	0.0065 ***					0.4867		0.0425	**
EMAP							0.553							0.0093	***
Prob(market launch)								Prob(failure)							
Region	Japan		EMA		EMAP		Others	Japan		EMA		EMAP		Others	
USA	0.002	***	0.6451		0.0072	***	0.0013 ***	0.9825		0.5738		0.8454		0.1084	
Japan			0.032	**	0.546		0.2963			0.6196		0.9284		0.2953	
EMA					0.0528	*	0.004 ***					0.2995		0.442	
EMAP							0.4778							0.2292	

Notes. * $p <.10$, ** $p <.05$, *** $p <.01$

Table 5.6. Main therapeutic groups: Anatomical Therapeutic Chemical Classification

ATC code	Description
A	Alimentary tract and metabolism
B	Blood and blood forming organs
C	Cardiovascular system
D	Dermatologicals
G	Genito-urinary system and sex hormones
H	Systemic hormonal preparations, excluding sex hormones and insulins
J	Antiinfectives for systemic use
L	Antineoplastic and immunomodulating agents
M	Musculo-skeletal system
N	Nervous system
P	Antiparasitic products, insecticides and repellents
R	Respiratory system
S	Sensory organs
V	Various

Notes. EphRMA classification

Table 5.7. Simulated NCEs across regions and therapeutic groups - descriptive statistics

Therapeutic group	Region	NCEs		Births		Duration		Prob(no event)		Prob(market)		Prob(failure)	
		mean	sd	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd
A	USA	54.949	30.207	1.725	1.135	8.930	3.057	0.122	0.083	0.385	0.084	0.508	0.063
	Japan	30.037	20.807	1.477	0.750	7.167	2.928	0.193	0.111	0.339	0.105	0.504	0.100
	EMA	51.282	30.455	1.689	1.443	8.208	3.182	0.110	0.080	0.369	0.093	0.509	0.072
	EMAP	24.957	17.037	1.979	1.266	7.339	2.933	0.145	0.094	0.344	0.103	0.502	0.100
	Others	18.689	11.967	1.170	0.143	7.487	3.081	0.083	0.072	0.348	0.108	0.500	0.104
B	USA	19.498	11.225	1.812	0.758	9.659	2.584	0.118	0.054	0.399	0.062	0.504	0.031
	Japan	10.924	7.517	1.339	0.200	7.474	2.763	0.204	0.093	0.349	0.091	0.509	0.074
	EMA	17.112	9.876	1.614	0.497	9.460	3.144	0.109	0.047	0.394	0.072	0.504	0.040
	EMAP	6.124	4.089	1.442	0.402	7.831	2.824	0.162	0.087	0.355	0.097	0.499	0.091
	Others	4.303	3.098	1.113	0.162	6.749	2.915	0.116	0.081	0.326	0.115	0.494	0.122
C	USA	84.428	55.409	3.905	4.835	8.348	3.433	0.166	0.095	0.360	0.102	0.497	0.091
	Japan	32.001	23.080	1.907	1.779	7.393	2.873	0.253	0.109	0.342	0.100	0.502	0.095
	EMA	66.327	45.306	3.206	2.312	7.239	2.802	0.166	0.089	0.341	0.100	0.503	0.092
	EMAP	31.937	22.427	2.288	1.097	7.554	2.696	0.189	0.089	0.346	0.092	0.504	0.086
	Others	128.431	91.771	13.049	15.533	6.838	2.901	0.123	0.070	0.330	0.110	0.499	0.111
D	USA	24.962	14.714	1.390	0.346	8.490	3.270	0.067	0.058	0.370	0.097	0.502	0.083
	Japan	33.517	23.570	2.051	2.031	6.838	2.903	0.121	0.104	0.330	0.109	0.501	0.108
	EMA	33.048	19.055	1.567	1.178	8.370	2.893	0.057	0.051	0.375	0.085	0.508	0.065
	EMAP	7.670	5.379	1.224	0.202	6.878	2.906	0.066	0.067	0.332	0.108	0.502	0.104
	Others	15.621	10.315	1.423	0.429	7.411	2.659	0.061	0.059	0.347	0.093	0.500	0.090
G	USA	12.536	7.889	1.334	0.415	7.916	3.293	0.098	0.093	0.357	0.104	0.503	0.089
	Japan	4.483	2.990	1.060	0.048	7.660	2.691	0.090	0.095	0.348	0.095	0.495	0.100
	EMA	14.430	9.452	1.630	0.824	7.247	2.860	0.097	0.081	0.344	0.100	0.505	0.091
	EMAP	7.428	5.118	1.212	0.105	7.308	2.794	0.103	0.102	0.341	0.103	0.495	0.107
	Others	9.058	6.321	1.470	0.367	6.939	2.889	0.081	0.082	0.335	0.107	0.500	0.103
H	USA	12.982	8.283	1.911	1.063	7.953	3.321	0.127	0.085	0.358	0.109	0.500	0.101
	Japan	2.134	1.681	1.207	0.053	6.740	2.968	0.268	0.114	0.323	0.117	0.496	0.125
	EMA	10.759	6.175	1.313	0.224	8.321	3.082	0.150	0.077	0.373	0.085	0.515	0.045
	Others	12.034	8.536	2.984	1.122	7.198	2.761	0.120	0.074	0.338	0.103	0.497	0.105

J	USA	51.508	26.031	1.317	0.353	10.742	4.356	0.107	0.081	0.422	0.094	0.495	0.065
	Japan	8.930	6.714	1.253	0.245	6.756	2.897	0.248	0.131	0.325	0.115	0.496	0.122
	EMA	45.828	24.098	1.927	1.580	9.324	3.519	0.117	0.082	0.394	0.089	0.505	0.061
	EMAP	19.795	13.776	1.992	1.564	7.208	2.991	0.152	0.102	0.339	0.110	0.498	0.110
	Others	16.642	10.778	1.486	0.453	7.591	2.899	0.129	0.086	0.353	0.096	0.506	0.082
L	USA	75.787	37.121	1.482	0.505	12.134	4.058	0.096	0.066	0.449	0.075	0.492	0.048
	Japan	32.340	21.719	3.811	1.317	7.364	2.890	0.217	0.107	0.345	0.100	0.506	0.090
	EMA	58.446	29.090	2.533	1.365	10.787	3.472	0.100	0.060	0.427	0.070	0.502	0.038
	EMAP	80.132	53.522	9.335	8.215	8.637	4.048	0.132	0.094	0.370	0.116	0.493	0.102
	Others	16.309	8.858	1.326	0.206	8.031	3.907	0.103	0.082	0.359	0.121	0.498	0.110
M	USA	27.041	13.048	1.222	0.280	9.547	3.837	0.127	0.087	0.400	0.096	0.503	0.073
	Japan	15.670	11.194	2.022	1.593	6.941	2.855	0.230	0.123	0.331	0.109	0.499	0.111
	EMA	20.562	10.391	1.075	0.055	9.000	4.239	0.131	0.102	0.381	0.119	0.495	0.099
	EMAP	8.322	5.402	1.251	0.127	8.070	3.235	0.147	0.104	0.355	0.106	0.495	0.101
	Others	6.212	3.716	1.239	0.232	7.826	2.999	0.140	0.096	0.359	0.100	0.507	0.089
N	USA	41.845	20.738	1.111	0.115	12.578	3.351	0.097	0.054	0.455	0.058	0.489	0.041
	Japan	13.616	9.844	1.187	0.149	7.047	2.877	0.237	0.105	0.336	0.106	0.503	0.103
	EMA	27.565	14.296	1.185	0.183	9.958	3.034	0.126	0.073	0.406	0.071	0.510	0.040
	EMAP	9.878	6.239	1.074	0.069	7.736	2.938	0.153	0.078	0.356	0.098	0.505	0.089
	Others	15.912	10.633	1.247	0.168	7.300	2.803	0.123	0.068	0.346	0.096	0.507	0.080
P	USA	23.052	12.242	1.215	0.171	12.699	4.580	0.082	0.060	0.466	0.077	0.491	0.054
	Japan	2.681	1.605	1.029	0.012	8.872	2.600	0.175	0.075	0.387	0.065	0.517	0.022
	EMA	16.023	9.989	1.062	0.018	14.621	4.627	0.038	0.040	0.508	0.068	0.475	0.066
	EMAP	3.878	1.948	1.009	0.006	10.364	2.906	0.091	0.050	0.424	0.057	0.511	0.032
	Others	3.928	2.301	1.361	0.067	8.417	3.026	0.081	0.049	0.386	0.072	0.529	0.031
R	USA	15.913	8.891	1.094	0.108	8.298	3.496	0.157	0.115	0.370	0.099	0.505	0.073
	Japan	11.650	8.572	1.261	0.177	7.083	2.813	0.241	0.129	0.335	0.104	0.501	0.103
	EMA	19.490	10.070	1.219	0.230	8.479	4.153	0.151	0.118	0.369	0.113	0.502	0.087
	EMAP	8.253	6.079	1.099	0.100	6.752	2.901	0.178	0.126	0.325	0.115	0.495	0.122
	Others	9.829	7.292	1.735	0.896	6.765	2.915	0.173	0.105	0.325	0.115	0.495	0.122
S	USA	14.488	7.315	1.732	0.570	9.994	3.182	0.078	0.037	0.417	0.067	0.509	0.037
	Japan	5.862	3.361	1.025	0.014	10.080	2.293	0.139	0.055	0.405	0.056	0.504	0.024
	EMA	9.311	4.753	1.046	0.030	10.536	3.415	0.071	0.042	0.430	0.065	0.512	0.042

EMAP	2.281	1.673	1.151	0.035	6.738	2.940	0.106	0.064	0.325	0.116	0.495	0.124
Others	3.070	2.061	1.414	0.197	7.456	2.760	0.067	0.038	0.351	0.092	0.506	0.072

Notes. Average and standard deviations calculated from 200 Monte-Carlo simulations.

Table 5.8. Simulated Biologics across regions and therapeutic groups - descriptive statistics

Therapeutic group	Region	Biologics		Births		Duration		Prob(no event)		Prob(market)		Prob(failure)	
		mean	sd	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd
A	USA	29.347	16.084	8.865	0.072	9.239	3.393	0.202	0.147	0.365	0.095	0.544	0.070
	Japan	5.475	3.622	2.443	0.061	7.580	3.078	0.106	0.064	0.325	0.102	0.541	0.098
	EMA	20.443	11.802	6.857	0.070	8.619	3.469	0.171	0.109	0.353	0.098	0.544	0.076
	EMAP	4.890	3.254	2.116	0.026	7.716	3.157	0.130	0.086	0.329	0.105	0.537	0.098
	Others	4.503	3.090	2.122	0.026	7.657	3.067	0.186	0.107	0.323	0.102	0.544	0.098
B	USA	12.539	7.156	4.561	0.056	10.144	2.805	0.211	0.146	0.382	0.070	0.533	0.034
	Japan	4.111	2.925	2.174	0.041	7.783	2.799	0.288	0.163	0.325	0.091	0.553	0.045
	EMA	8.285	4.793	3.288	0.040	9.888	3.034	0.250	0.103	0.374	0.076	0.534	0.034
C	USA	21.744	12.783	7.410	0.139	10.282	3.541	0.170	0.110	0.385	0.087	0.527	0.068
	Japan	2.225	1.471	1.066	0.019	10.019	1.754	0.125	0.059	0.359	0.052	0.507	0.095
	EMA	25.217	17.427	12.519	0.157	8.365	3.278	0.186	0.114	0.339	0.103	0.539	0.080
	EMAP	2.602	1.917	1.364	0.046	7.327	2.761	0.143	0.063	0.315	0.093	0.545	0.057
	Others	16.600	13.787	10.644	0.256	6.764	2.969	0.314	0.124	0.286	0.116	0.534	0.134
D	USA	13.480	8.151	5.193	0.082	9.410	2.858	0.219	0.118	0.366	0.075	0.541	0.041
	Japan	4.581	3.439	2.496	0.055	7.009	3.043	0.218	0.159	0.309	0.101	0.542	0.106
	EMA	14.158	9.422	6.422	0.072	8.262	2.963	0.229	0.124	0.338	0.092	0.548	0.069
	Others	2.587	1.751	1.212	0.037	8.213	2.573	0.153	0.057	0.340	0.079	0.545	0.036
G	USA	6.657	4.482	3.170	0.037	8.236	3.315	0.318	0.131	0.331	0.102	0.549	0.084
	Japan	2.492	1.839	1.364	0.053	8.211	2.226	0.282	0.097	0.329	0.075	0.532	0.083
	EMA	1.790	1.443	1.066	0.021	6.740	2.962	0.220	0.104	0.290	0.115	0.529	0.133
	EMAP	1.909	1.387	1.015	0.009	8.422	2.211	0.251	0.089	0.335	0.070	0.533	0.074
H	USA	3.931	3.136	2.372	0.045	6.751	2.919	0.228	0.104	0.289	0.113	0.531	0.132
	Japan	1.793	1.402	1.033	0.014	6.754	2.959	0.171	0.088	0.293	0.115	0.527	0.132
	EMA	4.210	3.235	2.366	0.051	7.067	2.889	0.218	0.093	0.303	0.105	0.541	0.099
J	USA	44.747	22.917	7.811	0.058	10.594	4.300	0.155	0.132	0.396	0.107	0.524	0.079
	Japan	7.034	5.484	4.155	0.034	6.745	2.902	0.163	0.138	0.292	0.112	0.527	0.130
	EMA	30.939	16.336	7.623	0.057	9.627	4.041	0.160	0.125	0.374	0.107	0.534	0.077
	EMAP	3.547	2.773	2.123	0.018	6.742	2.918	0.178	0.151	0.292	0.113	0.528	0.131
	Others	6.011	4.591	3.540	0.067	6.824	2.970	0.211	0.176	0.297	0.110	0.530	0.132

L	USA	75.931	35.509	9.046	0.140	10.914	4.196	0.103	0.093	0.406	0.096	0.520	0.066
	Japan	6.996	4.883	3.456	0.042	7.243	3.045	0.091	0.093	0.315	0.106	0.536	0.109
	EMA	48.786	24.251	6.691	0.107	10.786	4.148	0.096	0.087	0.407	0.095	0.529	0.062
	EMAP	5.068	3.074	2.123	0.019	9.268	3.458	0.028	0.020	0.356	0.101	0.511	0.094
	Others	6.945	5.349	4.004	0.077	6.754	2.906	0.138	0.112	0.294	0.112	0.526	0.130
M	USA	14.645	7.385	2.611	0.063	9.407	4.600	0.162	0.122	0.368	0.132	0.529	0.112
	Japan	3.029	1.988	1.268	0.042	8.524	2.627	0.098	0.039	0.353	0.075	0.545	0.028
	EMA	10.360	5.536	2.104	0.022	9.675	4.597	0.157	0.115	0.371	0.129	0.523	0.107
	EMAP	2.021	1.592	1.174	0.035	6.744	2.957	0.152	0.082	0.293	0.114	0.527	0.132
	Others	5.315	3.596	2.471	0.045	7.900	3.169	0.202	0.107	0.328	0.105	0.543	0.099
N	USA	26.649	16.631	5.308	0.049	12.016	4.164	0.107	0.100	0.421	0.101	0.507	0.079
	Japan	4.257	3.277	2.395	0.057	6.760	2.935	0.096	0.098	0.297	0.113	0.524	0.130
	EMA	19.518	11.560	6.931	0.078	9.250	3.211	0.122	0.122	0.362	0.093	0.535	0.067
	EMAP	1.925	1.501	1.102	0.023	6.751	2.963	0.151	0.081	0.293	0.115	0.527	0.132
	Others	3.363	2.687	2.052	0.020	6.756	2.926	0.243	0.108	0.289	0.113	0.532	0.132
P	USA	17.349	9.724	1.101	0.024	14.307	4.764	0.074	0.089	0.487	0.095	0.497	0.094
	EMA	15.061	8.939	1.004	0.005	13.862	4.756	0.065	0.075	0.484	0.095	0.499	0.092
	Others	2.847	1.789	1.022	0.011	8.381	3.315	0.175	0.123	0.354	0.089	0.569	0.054
R	USA	6.014	4.681	3.469	0.044	6.959	2.892	0.227	0.098	0.298	0.108	0.538	0.111
	Japan	1.916	1.501	1.128	0.026	6.759	2.958	0.171	0.088	0.292	0.114	0.529	0.133
	EMA	5.818	4.419	3.224	0.043	6.985	2.960	0.224	0.106	0.301	0.107	0.542	0.110
	Others	2.267	1.845	1.369	0.054	6.758	2.970	0.243	0.110	0.289	0.115	0.531	0.133
S	USA	2.532	1.619	1.053	0.017	8.848	2.610	0.196	0.063	0.357	0.073	0.546	0.030
	Japan	1.832	1.451	1.064	0.023	6.747	2.954	0.177	0.089	0.292	0.115	0.528	0.133
	EMA	1.399	0.555	0.000	0.000	13.333	5.916	0.107	0.168	0.481	0.129	0.519	0.129

Notes. Average and standard deviations calculated from 200 Monte-Carlo simulations.

Table 5.9. Simulated NCEs across regions and therapeutic areas - Testing mean pairwise differences (t-tests)

Therapeutic group	Region	NCEs								Births							
		Japan		EMA		EMAP		Others		Japan		EMA		EMAP		Others	
A	USA	0.0146	**	0.8133		0.0197	**	0.0032	***	0.1774		0.8643		0.2033		0.0211	**
	Japan			0.0833	*	0.5822		0.0426	**			0.38		0.5378		0.0314	**
	EMA					0.1106		0.0062	***					0.309		0.0401	**
	EMAP							0.4496								0.3787	
B	USA	0.1264		0.6113		0.0136	**	0.0233	**	0.4159		0.7706		0.0517	*	0.033	**
	Japan			0.0455	**	0.2865		0.1167				0.3708		0.2589		0.1195	
	EMA					0.0801	*	0.6869						0.2641		0.0557	*
	EMAP							0.2795								0.7469	
C	USA	0.135		0.6113		0.1445		0.4623		0.1846		0.6952		0.1905		0.4101	
	Japan			0.1339		0.9962		0.2784				0.186		0.8795		0.2903	
	EMA					0.0559	*	0.4608						0.0597	*	0.4444	
	EMAP							0.2569								0.2596	
D	USA	0.5283		0.3968		0.0035	***	0.1104		0.2976		0.4784		0.0269	**	0.3648	
	Japan			0.9791		0.0654	*	0.2035				0.6272		0.057	*	0.1684	
	EMA					0.0101	**	0.0312	**					0.0166	**	0.0833	*
	EMAP							0.0452	**							0.0586	*
G	USA	0.0928	*	0.6292		0.0755	*	0.206		0.1862		0.3147		0.2237		0.4686	
	Japan			0.2331		0.2777		0.2878				0.2141		0.3063		0.307	
	EMA					0.2389		0.259						0.2117		0.2577	
	EMAP					0.3139		0.3599								0.3136	
H	USA	0.1871		0.805		0.102		0.9305		0.2485		0.6843		0.1359		0.9667	
	Japan			0.1986		0.391		0.1921				0.175		0.391		0.1952	
	EMA					0.1109		0.8649						0.065	*	0.5781	
J	USA	0.0496	**	0.594		0.1887		0.0596	*	0.0426	**	0.4107		0.8822		0.4609	
	Japan			0.0967	*	0.203		0.2044				0.181		0.281		0.3441	
	EMA					0.2854		0.1765						0.6579		0.3527	
	EMAP							0.7323								0.5744	
L	USA	0.4483		0.4868		0.9597		0.3261		0.0212	**	0.1215		0.2063		0.7157	
	Japan			0.4223		0.3011		0.0768				0.0306	**	0.3491		0.0586	*

	EMA					0.7428		0.25						0.2822		0.2595	
	EMAP							0.191								0.186	
M	USA	0.1182		0.0997	*	0.1199		0.0487	**	0.6824		0.3638		0.285		0.1512	
	Japan			0.389		0.3144		0.1217				0.5378		0.3005		0.1321	
	EMA					0.1463		0.0406	**					0.3512		0.1283	
	EMAP							0.5388								0.4396	
N	USA	0.0136	**	0.0124	**	0.0153	**	0.0203	*	0.6373		0.0998	*	0.0978	*	0.8286	
	Japan			0.0227	**	0.2954		0.1547				0.3795		0.0598	*	0.6217	
	EMA					0.0252	**	0.0537	*					0.0725	*	0.5315	
	EMAP							0.1939								0.0928	*
P	USA	0.3139		0.2092		0.2995		0.2988		0.4882		0.5033		0.4788		0.6394	
	Japan			0.5		0.5		0.5				0.5		0.5		0.5	
	EMA					0.5		0.5						0.5		0.5	
	EMAP							0.5								0.5	
R	USA	0.356		0.2968		0.3063		0.5157		0.1997		0.4364		0.376		0.9294	
	Japan			0.3008		0.2615		0.7416				0.8056		0.2124		0.6933	
	EMA					0.28		0.4025						0.1699		0.7325	
	EMAP							0.7371								0.7334	
S	USA	0.4642		0.5422		0.4437		0.4883		0.4039		0.425		0.4814		0.5556	
	Japan			0.3052		0.3894		0.553				0.5		0.517		0.7288	
	EMA					0.3498		0.4358						0.5495		0.7058	
	EMAP							0.5								0.5	
Therapeutic group	Region	Prob(no event)								Prob(market launch)							
		Japan		EMA		EMAP		Others		Japan		EMA		EMAP		Others	
A	USA	0.113		0.9404		0.1728		0.0292	**	0.016	**	0.7595		0.0029	***	0.0077	***
	Japan			0.031	**	0.0029	***	0.0014	***			0.0425	**	0.0348	**	0.1005	
	EMA					0.0979	*	0.0108	**					0.0016	***	0.0052	***
	EMAP							0.202								0.5321	
B	USA	0.012	**	0.6817		0.3846		0.0755	*	0.048	**	0.8611		0.0974	*	0.1122	
	Japan			0.025	**	0.0106	**	0.0132	**			0.1063		0.2009		0.1656	
	EMA					0.5466		0.2763						0.1062		0.0866	*
	EMAP							0.6102								0.9264	
C	USA	0.0077	***	0.916		0.1378		0.0116	**	0.1334		0.1982		0.0549	*	0.0025	***

	Japan			0.005	***	0.001	***	0.0014	***			0.3318		0.2053		0.0217	**
	EMA					0.1227		0.007	***					0.0949	*	0.0082	***
	EMAP							0.088	*							0.2756	
D	USA	0.0045	***	0.8547		0.1631		0.0715	*	0.8998		0.2532		0.0122	**	0.0669	*
	Japan			0.0033	***	0.0145	**	0.0012	***			0.0034	***	0.0113	**	0.1127	
	EMA					0.1846		0.0667	*					0.0027	***	0.0238	**
	EMAP							0.5								0.1368	
G	USA	0.4647		0.956		0.4831		0.2004		0.1357		0.6014		0.2739		0.2558	
	Japan			0.35		0.5989		0.7711				0.2053		0.3909		0.4178	
	EMA					0.484		0.144						0.354		0.3217	
	EMAP							0.1925								0.3849	
H	USA	0.5255		0.7802		0.1011		0.2165		0.186		0.9402		0.0624	*	0.2893	
	Japan			0.4894		0.391		0.8807				0.1632		0.391		0.3876	
	EMA					0.078	*	0.3066						0.06	*	0.2626	
J	USA	0.3768		0.2666		0.9268		0.3736		0.0239	**	0.0739	*	0.08	*	0.0466	**
	Japan			0.5511		0.2436		0.1896				0.0274	**	0.2654		0.2595	
	EMA					0.669		0.1728						0.0917	*	0.0595	*
	EMAP							0.4901								0.9024	
L	USA	0.0097	***	0.5143		0.0829	*	0.5495		0.0057	***	0.0476	**	0.0219	**	0.0185	**
	Japan			0.0043	***	0.0096	***	0.0025	***			0.0023	***	0.3834		0.5666	
	EMA					0.0726	*	0.6715						0.0517	*	0.0298	**
	EMAP							0.0263	**							0.194	
M	USA	0.2537		0.5999		0.3935		0.0975	*	0.1516		0.2725		0.109		0.0304	**
	Japan			0.2251		0.0933	*	0.0363	**			0.2491		0.4944		0.221	
	EMA					0.3506		0.0846	*					0.1401		0.0459	**
	EMAP							0.1254								0.3421	
N	USA	0.0117	**	0.0174	**	0.7187		0.6027		0.0187	**	0.0107	**	0.0179	**	0.0249	**
	Japan			0.0307	**	0.0151	**	0.001	***			0.0718	*	0.2683		0.0775	*
	EMA					0.2729		0.1322						0.033	**	0.0983	*
	EMAP							0.5441								0.2322	
P	USA	0.9736		0.4826		0.7452		0.6975		0.3157		0.4929		0.3678		0.3138	
	Japan			0.5		0.5		0.5				0.5		0.5		0.5	
	EMA					0.5		0.5						0.5		0.5	

	EMAP						0.5						0.5	
R	USA	0.0142	**	0.1206		0.4089		0.0025	***	0.1981	0.9323	0.2816	0.1483	
	Japan			0.0168	**	0.0534	*	0.0068	***		0.3292	0.3275	0.1538	
	EMA					0.5752		0.002	***			0.3161	0.1537	
	EMAP							0.0196	**				0.3763	
S	USA	0.0131	**	0.0307	**	0.0151	**	0.001	***	0.5315	0.1898	0.408	0.4441	
	Japan			0.2729		0.1322		0.5442			0.226	0.401	0.439	
	EMA					0.0122	**	0.1179				0.3873	0.4223	
	EMAP							0.0356	**				0.5	
Therapeutic group	Region	Duration								Prob(failure)				
		Japan		EMA		EMAP		Others		Japan	EMA	EMAP		Others	
A	USA	0.0009	***	0.0031	***	0.0201	**	0.0353	**	0.2885	0.5672	0.0151	**	0.0302	**
	Japan			0.1221		0.0007	***	0.0029	***		0.1397	0.0347	**	0.0756	*
	EMA					0.4997		0.0001	***			0.0058	***	0.0131	**
	EMAP							0	***					0.5925	
B	USA	0.044	**	0.0833	*	0.1927		0.235		0.6062	0.995	0.1868		0.1662	
	Japan			0.1618		0.0802	*	0.049	**		0.6945	0.1633		0.1671	
	EMA					0.8767		0.4074				0.1797		0.1821	
	EMAP							0.0241	**					0.9904	
C	USA	0.0359	**	0.0007	***	0.4432		0.2233		0.3969	0.1945	0.096	*	0.0152	**
	Japan			0.0121	**	0.1236		0.0055	***		0.3266	0.1759		0.0335	**
	EMA					0.213		0.0498	**			0.084	*	0.0128	**
	EMAP							0.2125						0.3316	
D	USA	0.0089	***	0.0318	**	0.0059	***	0.0113	**	0.3563	0.2992	0.0289	**	0.1568	
	Japan			0.1444		0.0012	***	0.0133	**		0.2275	0.0112	**	0.0739	*
	EMA					0.1207		0.0203	**			0.0096	***	0.0642	*
	EMAP							0.0059	***					0.1711	
G	USA	0.2109		0.1803		0.2514		0.4043		0.1795	0.6798	0.3547		0.372	
	Japan			0.5028		0.3898		0.2967			0.1684	0.3923		0.3737	
	EMA					0.3773		0.6133				0.3457		0.3621	
	EMAP							0.6132						0.4302	
H	USA	0.07	*	0.2362		0.1665		0.391		0.1819	0.9604	0.0577	*	0.3815	
	Japan			0.3892		0.0646	*	0.2097			0.1708	0.391		0.3916	

	EMA					0.1828		0.609					0.0577	*	0.3589	
J	USA	0.0627	*	0.0253	**	0.0222	**	0.2322		0.0677	*	0.1517		0.1592		0.1918	
	Japan			0.2481		0.075	*	0.0253	**			0.0644	*	0.3418		0.3172	
	EMA					0.8772		0.3388						0.1522		0.1728	
	EMAP							0.1622								0.9608	
L	USA	0.0197	**	0.0145	**	0.0127	**	0.332		0.3726		0.0091	***	0.9064		0.6636	
	Japan			0.4896		0.0574	*	0.024	**			0.7677		0.1805		0.2272	
	EMA					0.192		0.3111						0.2516		0.6824	
	EMAP							0.2178								0.4586	
M	USA	0.0842	*	0.014	**	0.2059		0.6861		0.3607		0.2537		0.18		0.0785	*
	Japan			0.2471		0.1297		0.0335	**			0.3994		0.3613		0.1832	
	EMA					0.2611		0.0019	***					0.1831		0.0802	*
	EMAP							0.3029								0.3977	
N	USA	0.0055	***	0.0027	***	0.0206	**	0.3321		0.4559		0.0308	**	0.0934	*	0.4857	
	Japan			0.066	*	0.0135	**	0.0297	**			0.3221		0.1824		0.4697	
	EMA					0.2862		0.14						0.0762	*	0.3441	
	EMAP							0.0551	*							0.1751	
P	USA	0.2173		0.1255		0.5		0.5		0.5552		0.5033		0.5479		0.5694	
	Japan			0.5		0.5		0.5				0.5		0.5		0.5	
	EMA					0.5		0.5						0.5		0.5	
	EMAP							0.5								0.5	
R	USA	0.2761		0.1538		0.3773		0.2919		0.5122		0.6706		0.3244		0.1575	
	Japan			0.1342		0.3224		0.1753				0.9012		0.3677		0.174	
	EMA					0.3763		0.8837						0.3235		0.1518	
	EMAP							0.72								0.3732	
S	USA	0.3761		0.4162		0.4897		0.3495		0.6336		0.7342		0.4807		0.4944	
	Japan			0.39		0.3599		0.3975				0.0058	***	0.4768		0.4908	
	EMA					0.5		0.5						0.4666		0.4804	
	EMAP							0.3975								0.5	

Notes. * $p < .10$, ** $p < .05$, *** $p < .01$

Table 5.10. Simulated Biologics across regions and therapeutic areas - Testing mean pairwise differences (t-tests)

	Region	Japan		EMA		EMAP		Others		Japan		EMA		EMAP		Others	
A	USA	0.0011	***	0.0118	**	0.0004	***	0.0129	**	0.0096	*	0.2646		0.0035	***	0.0031	***
	Japan			0.0089	***	0.8392		0.8659				0.0533	*	0.842		0.9	
	EMA					0.005	***	0.0557	*					0.031	**	0.0182	**
	EMAP							0.9423								0.9978	
B	USA	0.3708		0.0038	***					0.1315		0.3456					
	Japan			0.1896								0.3639					
	EMA																
C	USA	0.0046	***	0.7813		0.0021	***	0.765		0.0504	*	0.5344		0.0124	**	0.7364	
	Japan			0.1209		0.3739		0.4525				0.1929		0.3739		0.4315	
	EMA					0.0875	*	0.7254						0.147		0.9	
	EMAP							0.3739									
D	USA	0.0229	**	0.2851				0.0417	**	0.1461		0.351				0.0375	**
	Japan			0.1746				0.7015				0.5236				0.6276	
	EMA							0.3465								0.4024	
G	USA	0.0984	*	0.1896		0.0292	**			0.3328		0.1549		0.1837			
	Japan			0.868		0.4226						0.9005		0.4226			
	EMA					0.9739								0.9802			
	EMAP																
H	USA	0.5321		0.4242						0.5113		0.1738					
	Japan			0.4624								0.5132					
	EMA																
J	USA	0.0607	**	0.1215		0.0403	**	0.0581	*	0.0628	*	0.6931		0.0086	***	0.0597	*
	Japan			0.052	*	0.1835		0.6169				0.0445	**	0.1776		0.6013	
	EMA					0.0295	**	0.0538	*					0.0101	**	0.063	*
	EMAP							0.2766									
L	USA	0.2661		0.2964		0.2491		0.2656		0.2795		0.38		0.173		0.2716	
	Japan			0.2459		0.3624		0.4993				0.2256		0.2474		0.3632	
	EMA					0.2186		0.2451						0.082	*	0.2144	
	EMAP							0.3592								0.12	
M	USA	0.3658		0.4399		0.4986		0.5214		0.4523		0.021	**	0.465		0.833	

	Japan			0.3164		0.8746		0.5231				0.6057		0.9755		0.6034	
	EMA					0.5252		0.5767						0.5978		0.5979	
	EMAP							0.4244								0.34	
N	USA	0.0206	**	0.0446	**	0.018	**	0.0264	**	0.0352	**	0.3772		0.0121	**	0.0501	*
	Japan			0.024	**	0.3739		0.6679				0.054	*	0.3739		0.61	
	EMA					0.0173	**	0.044	**					0.0256	**	0.0777	*
	EMAP							0.4661								0.4268	
P	USA			0.5203				0.2988				0.5774				0.1756	
	EMA							0.9955								0.5	
R	USA	0.7259		0.3739		0.3756		0.3427		0.6169		0.2766		0.3627		0.21	
	Japan			0.1821		0.2851		0.1746				0.1825		0.1748		0.1746	
	EMA					0.7015		0.1982						0.2661		0.2964	
S	USA	0.391		0.2227						0.2459		0.3624					
	Japan			0.1486								0.2451					
	EMA																

Therapeutic group		Prob(no event)								Prob(market launch)							
	Region	Japan		EMA		EMAP		Others		Japan		EMA		EMAP		Others	
A	USA	0.0117	**	0.2887		0.0135	**	0.0308	**	0.0007	***	0.0621	*	0.0003	***	0.0011	***
	Japan			0.0287	**	0.83		0.6396				0.0178	**	0.986		0.9958	
	EMA					0.028	**	0.0654	*					0.0136	**	0.0353	**
	EMAP							0.7601								0.9869	
B	USA	0.2936		0.0265	**					0.0969	*	0.3971					
	Japan			0.5073								0.2179					
	EMA																
C	USA	0.0554	*	0.1912		0.012	**	0.0663	*	0.0046	***	0.0638	*	0.0001	***	0.0126	**
	Japan			0.0397	**	0.3739		0.6322				0.011	**	0.3739		0.8935	
	EMA					0.0098	***	0.0622	*					0.0002	***	0.0152	**
	EMAP							0.3739								0.3739	
D	USA	0.208		0.4489				0.0938	*	0.0673	*	0.323				0.0553	*
	Japan			0.386				0.5346				0.2208				0.6828	
	EMA							0.2963								0.3713	
G	USA	0.1719		0.1798		0.1393				0.113		0.1888		0.1189			
	Japan			0.8993		0.4226						0.9483		0.4226			

	EMA					0.9466								0.9416		
	EMAP															
H	USA	0.1064		0.0008	***					0.5086		0.4712				
	Japan			0.5								0.4509				
	EMA															
J	USA	0.4062		0.0108	**	0.1848		0.0644	*	0.0196	**	0.0206	**	0.0038	***	0.012	**
	Japan			0.1833		0.9423		0.0057	*			0.0297	**	0.1742		0.3769	
	EMA					0.1376		0.0883	*					0.0048	***	0.0187	**
	EMAP							0.0732	*							0.3362	
L	USA	0.0974	*	0.0866	*	0.9811		0.2862		0.2266		0.8774		0.1726		0.1792	
	Japan			0.2643		0.105		0.1097				0.2141		0.5159		0.349	
	EMA					0.8376		0.1965						0.1569		0.1706	
	EMAP							0.4432								0.6542	
M	USA	0.5211		0.0644	*	0.3695		0.5115		0.2984		0.3819		0.5058		0.4953	
	Japan			0.8644		0.3389		0.1054				0.2893		0.9416		0.4675	
	EMA					0.2863		0.1739						0.5045		0.4887	
	EMAP							0.3389								0.5081	
N	USA	0.1837		0.1584		0.8766		0.099	*	0.0045	***	0.0762	*	0.0045	***	0.0089	***
	Japan			0.3739		0.2521		0.0834	*			0.0064	***	0.3739		0.3818	
	EMA					0.4328		0.2313						0.0065	***	0.0371	**
	EMAP							0.5899								0.3825	
P	USA	0.1826		0.2715		0.7813		0.1206		0.1209		0.3739		0.4525		0.7254	
	EMA					0.3759		0.3427						0.1821		0.2851	
R	USA	0.1746		0.1982		0.7015		0.3465		0.7015		0.3465		0.391		0.2227	
	Japan			0.391		0.2227		0.2079				0.2079		0.1658		0.1486	
	EMA					0.2079		0.1486						0.1869		0.866	
	EMAP							0.1896								0.4226	
S	USA	0.868		0.4226						0.9739		0.4229					
	Japan			0.9739								0.1089					
	EMA																
Therapeutic group						Duration								Prob(failure)			
	Region	Japan		EMA		EMAP		Others		Japan		EMA		EMAP		Others	
A	USA	0.0003	***	0.0307	**	0.0001	***	0.0008	***	0.0027	***	0.1764		0.0025	***	0.0027	***

	Japan			0.0141	**	0.9794		0.9927				0.0331	**	0.9924		0.9969	
	EMA					0.021	**	0.465						0.0329	**	0.0328	**
	EMAP							0.833								0.9917	
B	USA	0.0615	*	0.3966						0.2085		0.3762					
	Japan			0.1531								0.4542					
	EMA																
C	USA	0.002	***	0.0936	*	0.0009	***	0.0151	**	0.0196	**	0.4788		0	***	0.0171	**
	Japan			0.0176	**	0.3739		0.8203				0.0127	**	0.3739		0.975	
	EMA					0.0006	***	0.0179	**					0	***	0.0136	**
	EMAP							0.3739								0.3739	
D	USA	0.0205	**	0.2338				0.0634	*	0.1979		0.4475				0.054	*
	Japan			0.1536				0.7154				0.3791				0.6408	
	EMA							0.36								0.3845	
G	USA	0.0058	***	0.0959	*	0.0743	*			0.177		0.1826		0.1781			
	Japan			0.082	*	0.0198	**					0.9979		0.4226			
	EMA					0.4226								0.9975			
	EMAP																
H	USA	0.5774		0.5774						0.4955		0.5232					
	Japan			0	***							0.4701					
	EMA																
J	USA	0.0282	**	0.028	**	0.0166	**	0.0237	**	0.0701	**	0.0838	*	0.0067	***	0.0279	**
	Japan			0.0043	***	0.0111	**	0.0286	**			0.0609	*	0.1712		0.3655	
	EMA					0.1726		0.3752						0.0065	***	0.0237	**
	EMAP															0.3466	
L	USA	0.0271	**	0.0804	*	0.1868		0.7873		0.4103		0.3604		0.1549		0.3765	
	Japan			0.2091		0.1532		0.1786				0.379		0.3354		0.4711	
	EMA					0.7352		0.3858						0.1461		0.3452	
	EMAP															0.3666	
M	USA	0.4355		0.7038		0.1911		0.4858		0.5217		0.455		0.4951		0.4877	
	Japan			0.5003		0.5018		0.154				0.5341		0.9891		0.4881	
	EMA					0.9216		0.45						0.4959		0.4789	
	EMAP							0.4997								0.4948	
N	USA	0.4999		0.4999		0.0045	***	0.0983	*	0.0186	**	0.1387		0.0185	**	0.0987	*

	Japan		0.0044	***	0.0028	***	0.0064	***		0.0141	**	0.3739		0.3699	
	EMA				0.3739		0.3749					0.014	**	0.06	*
	EMAP						0.0063	***						0.3692	
P	USA		0.1001				0.3659			0.4624				0.5	
	EMA						0.5231							0.1746	
R	USA	0.4244	0.5214		0.4986		0.5214		0.2656	0.2456		0.3615		0.4992	
	Japan		0.4986		0.3164		0.5756			0.2186		0.2451		0.3592	
	EMA				0.5767		0.5214					0.2815		0.2801	
	EMAP						0.6679							0.1616	
S	USA	0.6679	0.4661						0.6101	0.3133					
	Japan		0.1039							0.4399					
	EMA														

Notes. * $p <.10$, ** $p <.05$, *** $p <.01$

VI. Appendix: Figures

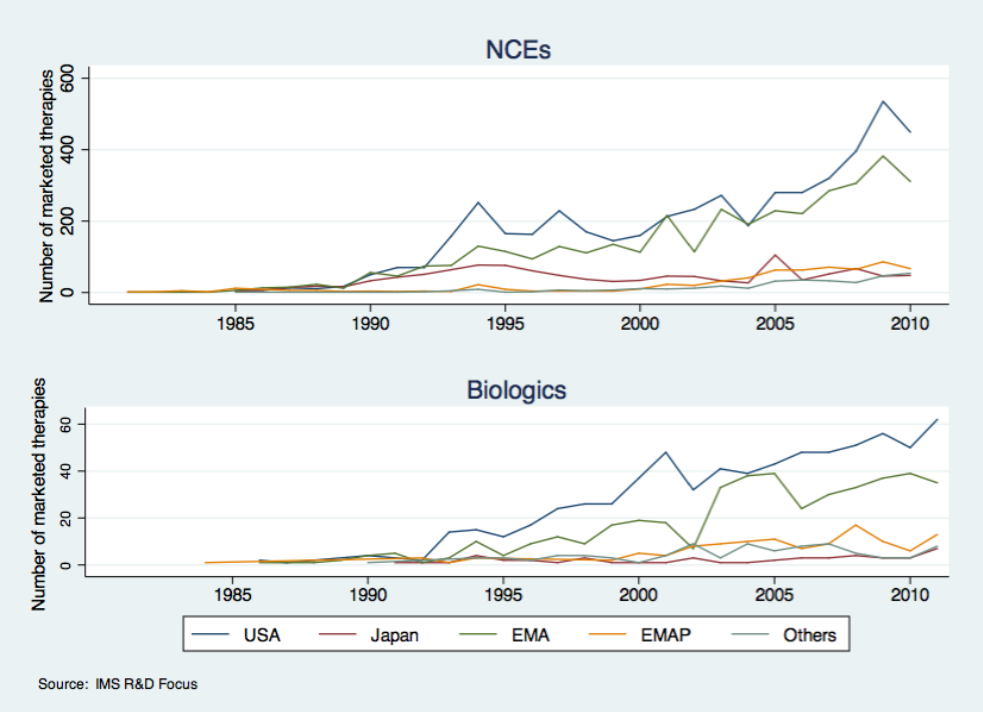


Figure 5.3. Historical regional trends in the number of market launches

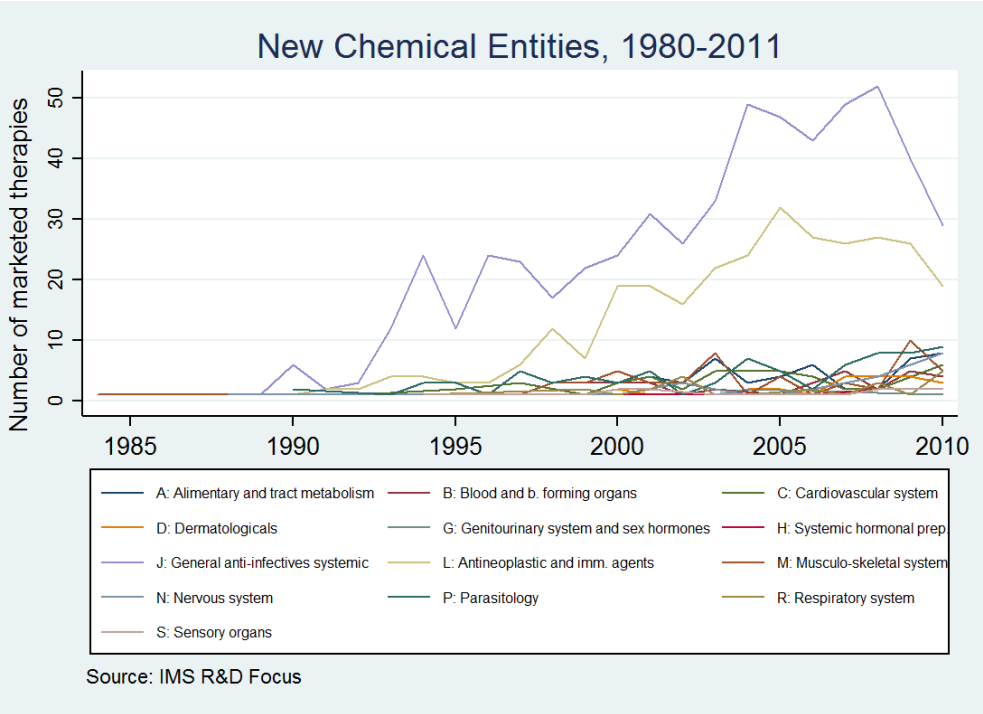


Figure 5.4. NCEs, historical trends in the number of market launches at therapeutic level

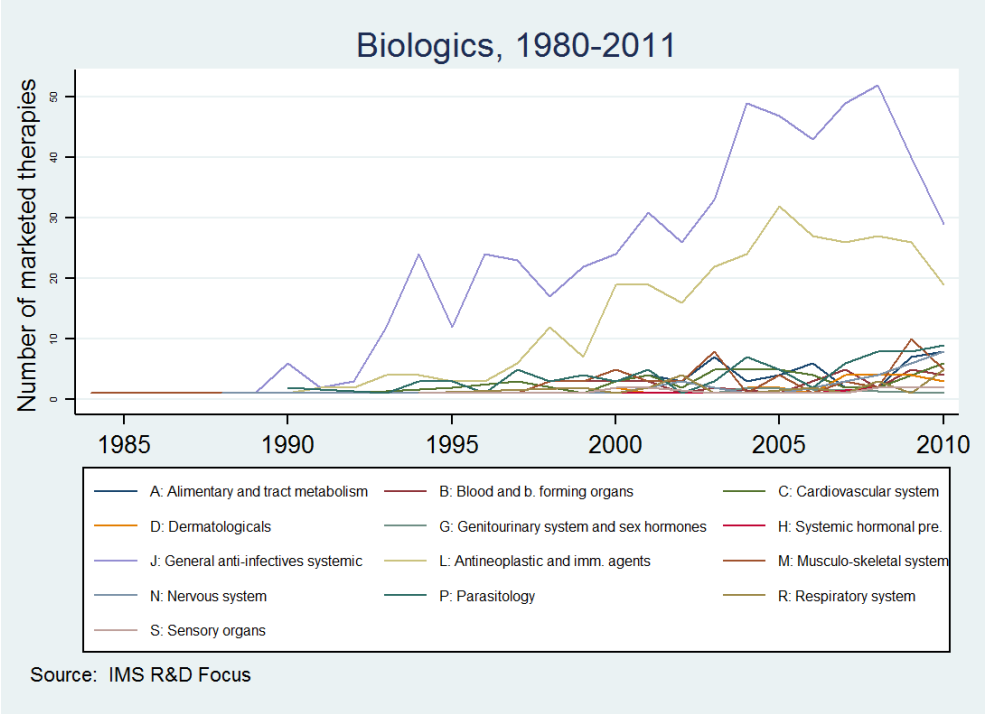


Figure 5.5. Biologics, historical trends in the number of market launches at therapeutic level

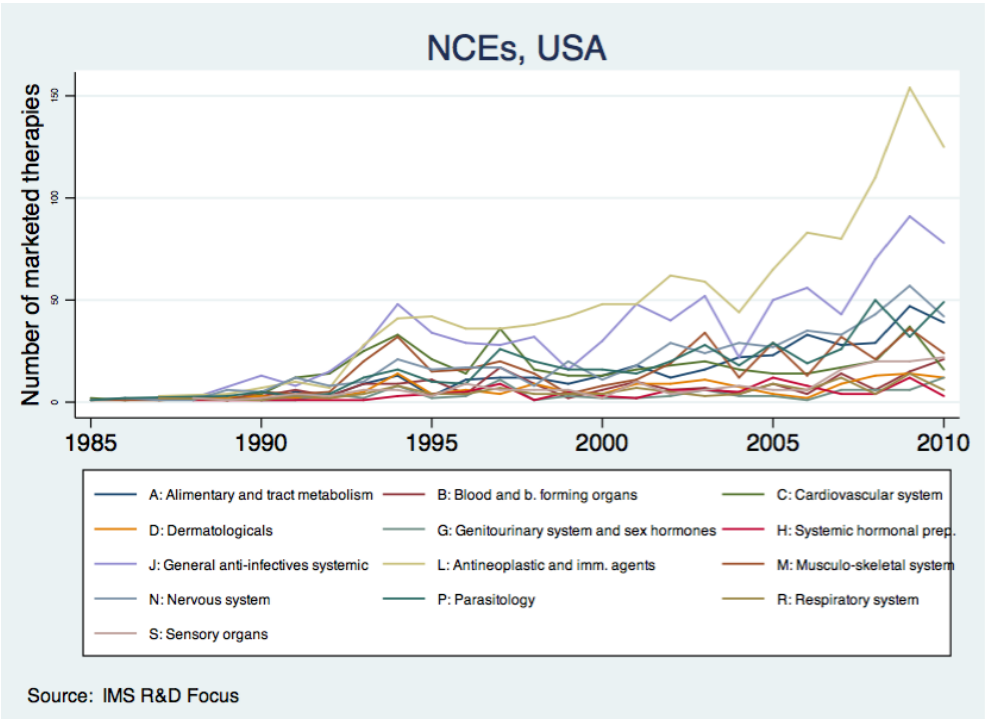


Figure 5.6. NCEs, historical trends in the number of market launches at therapeutic level in the USA

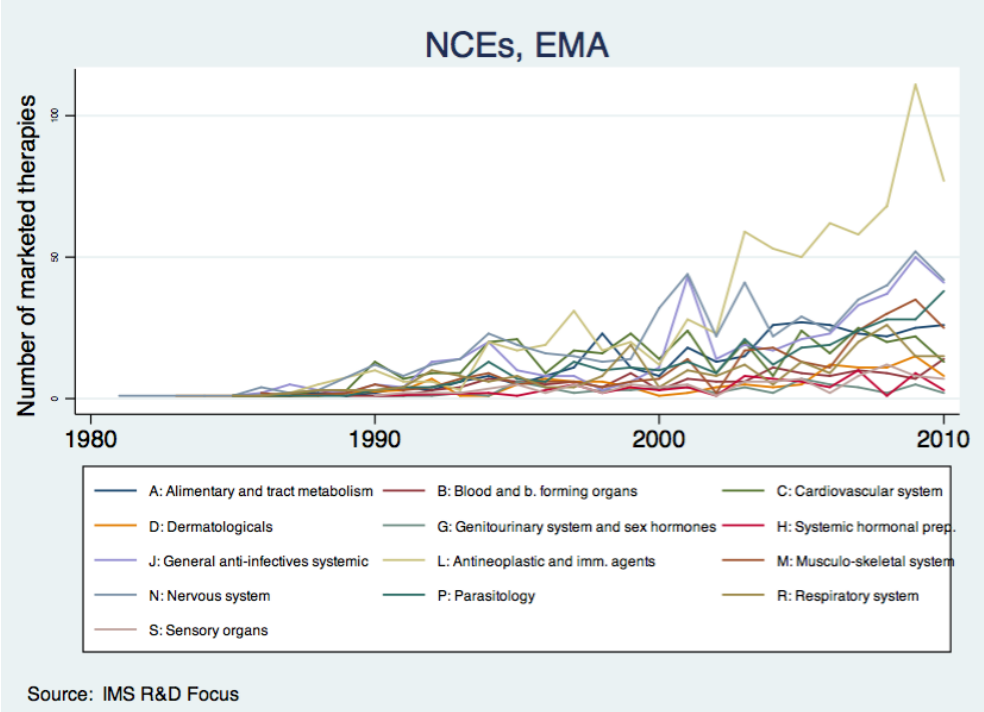


Figure 5.7. NCEs, historical trends in the number of market launches at therapeutic level in the EMA region

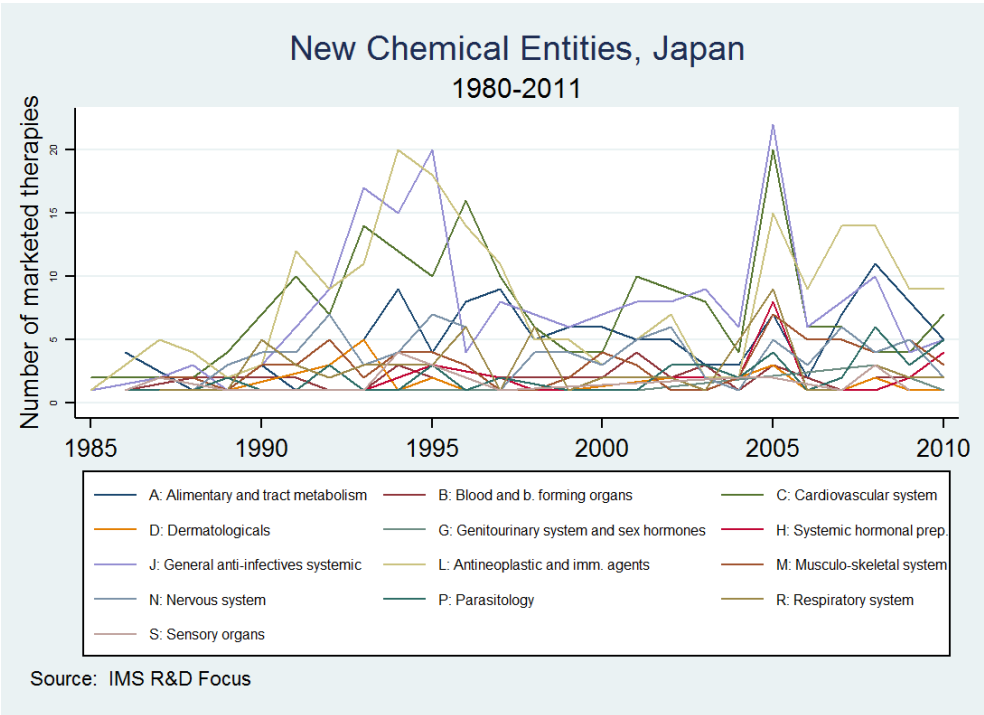


Figure 5.8. NCEs, historical trends in the number of market launches at therapeutic level in Japan

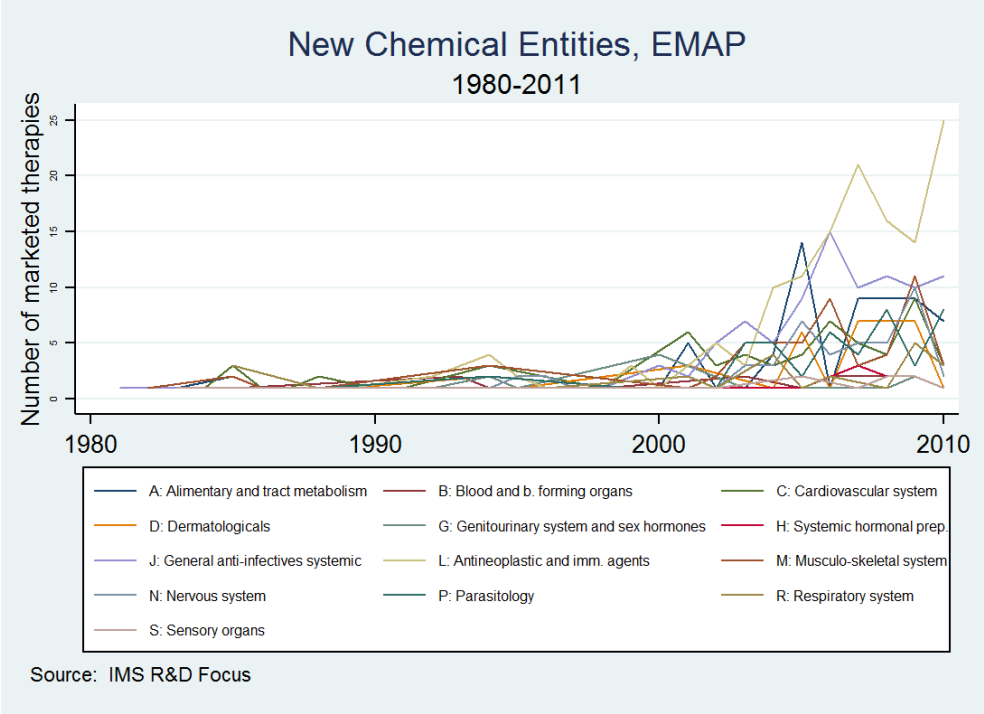


Figure 5.9. NCEs, historical trends in the number of market launches at therapeutic level in the EMAP region

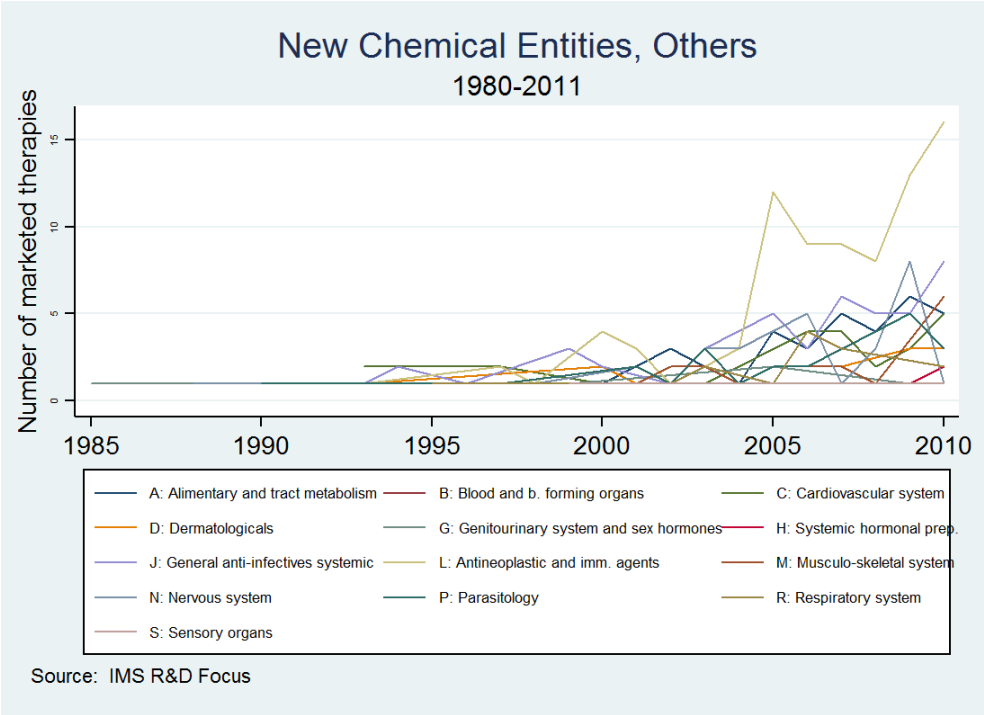


Figure 5.10. NCEs, historical trends in the number of market launches at therapeutic level in the rest of the world

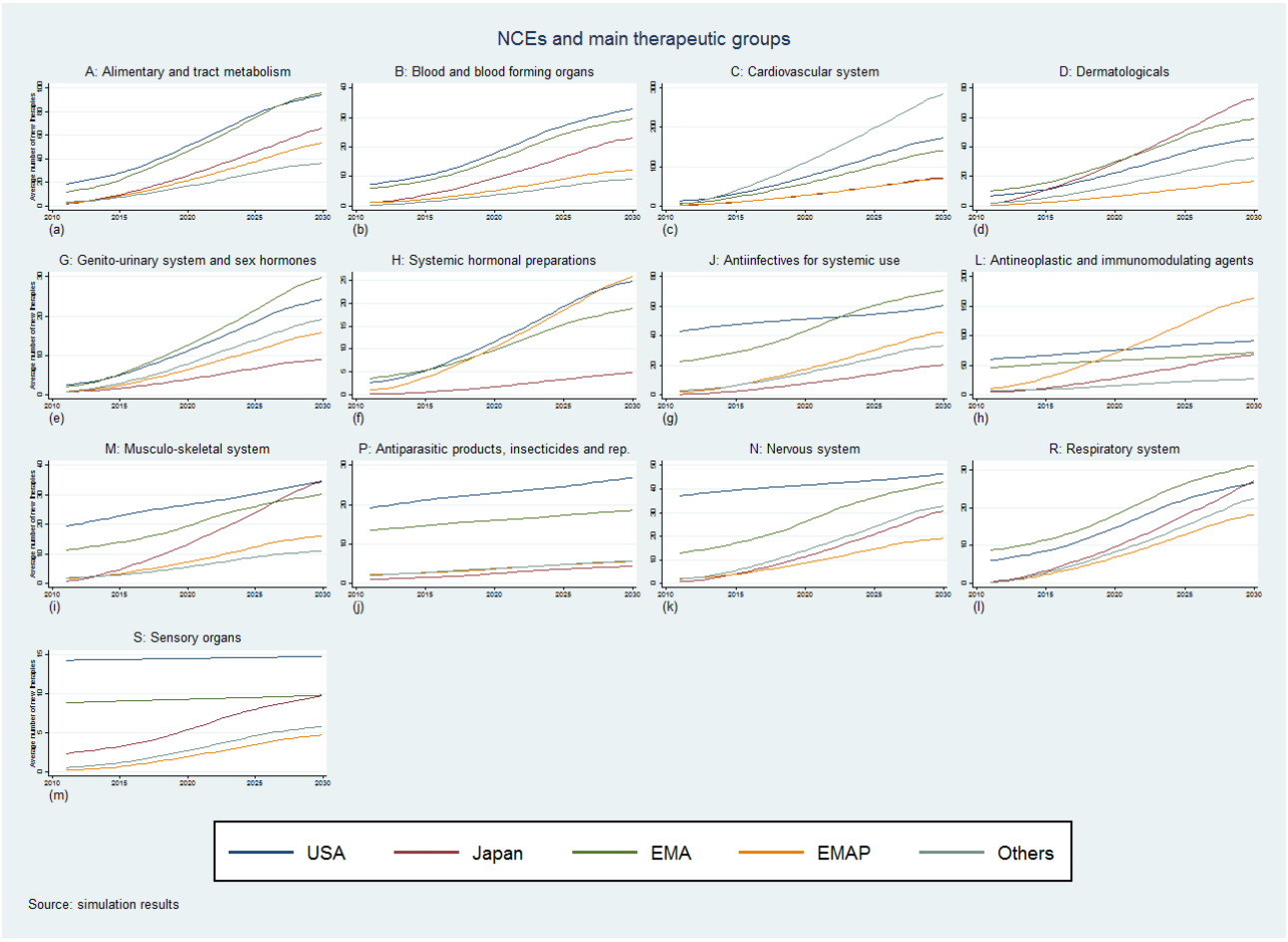


Figure 5.11. Simulated number of market launches for NCEs across regions

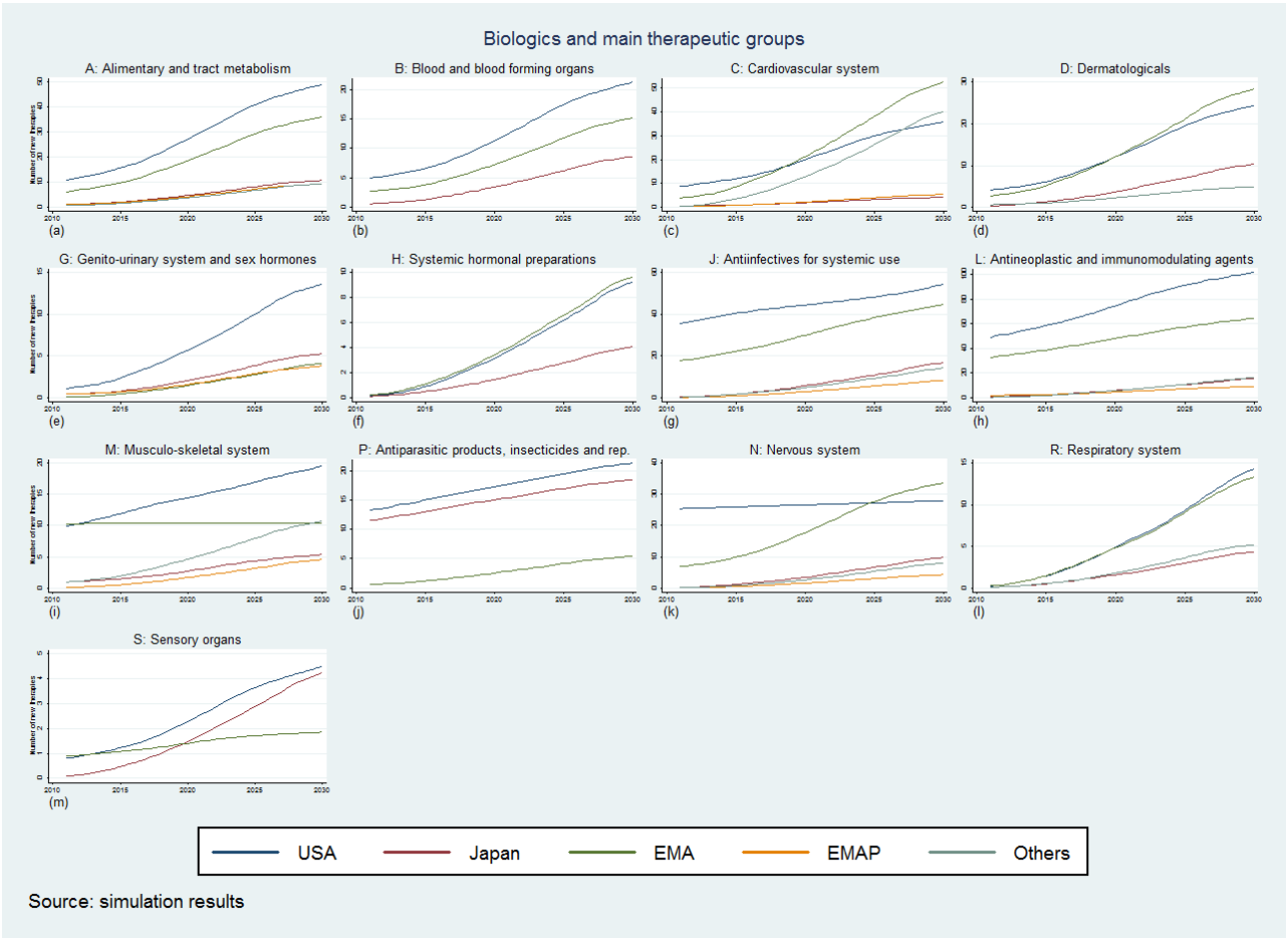
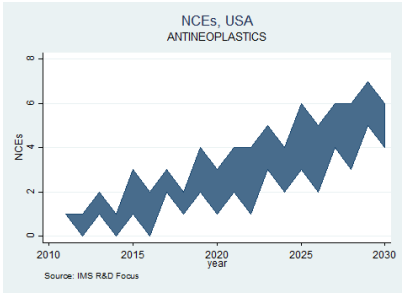
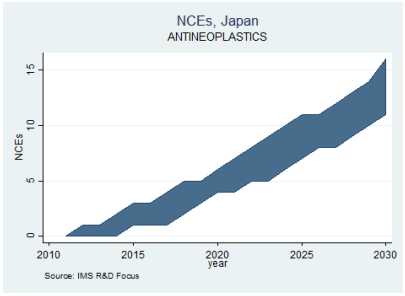


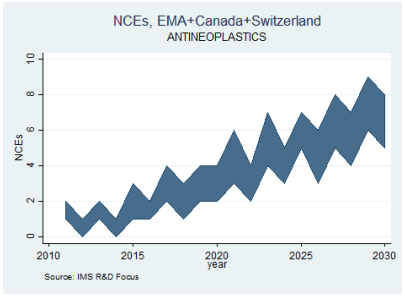
Figure 5.12. Simulated number of market launches for biologics across regions



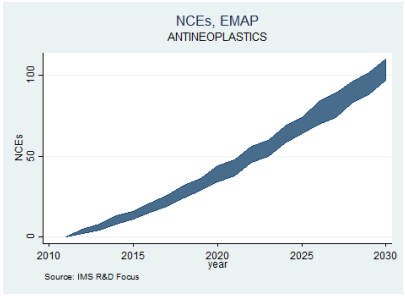
(a) USA



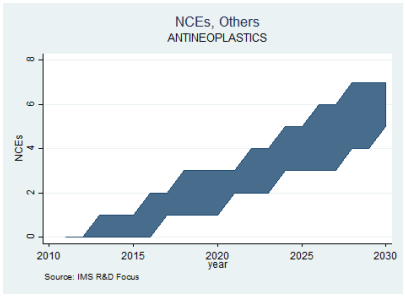
(b) Japan



(c) EMA+Canada+Switzerland

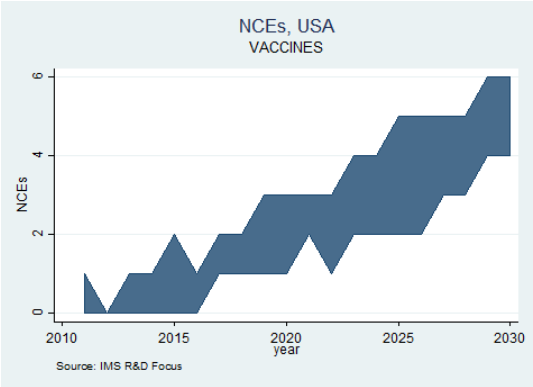


(d) EMAP

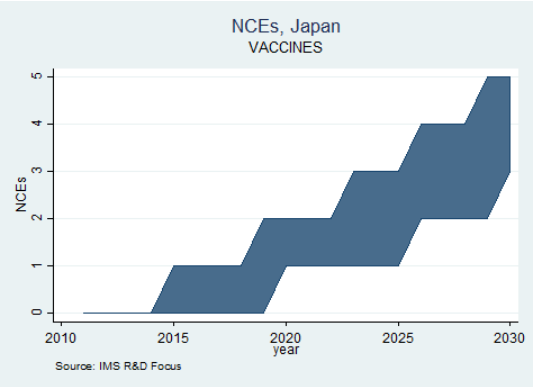


(e) Others

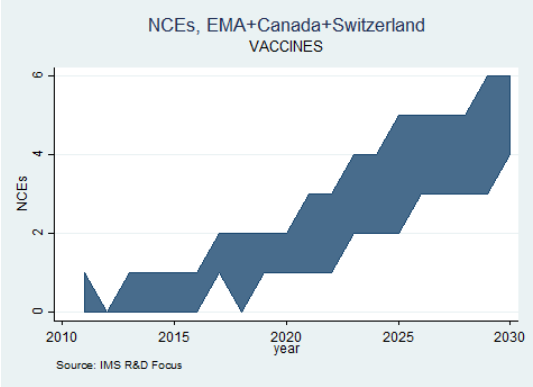
Figure 5.13. Simulated number of market launches for new antineoplastic agents



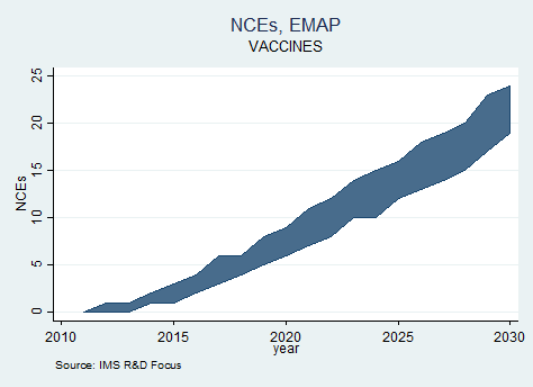
(a) USA



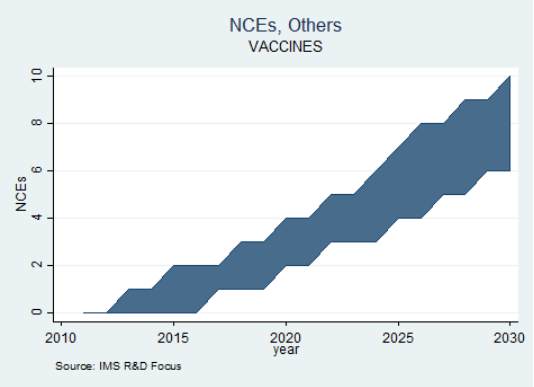
(b) Japan



(c) EMA+Canada+Switzerland



(d) EMAP



(e) Others

Figure 5.14. Simulated number of market launches of new vaccines across regions

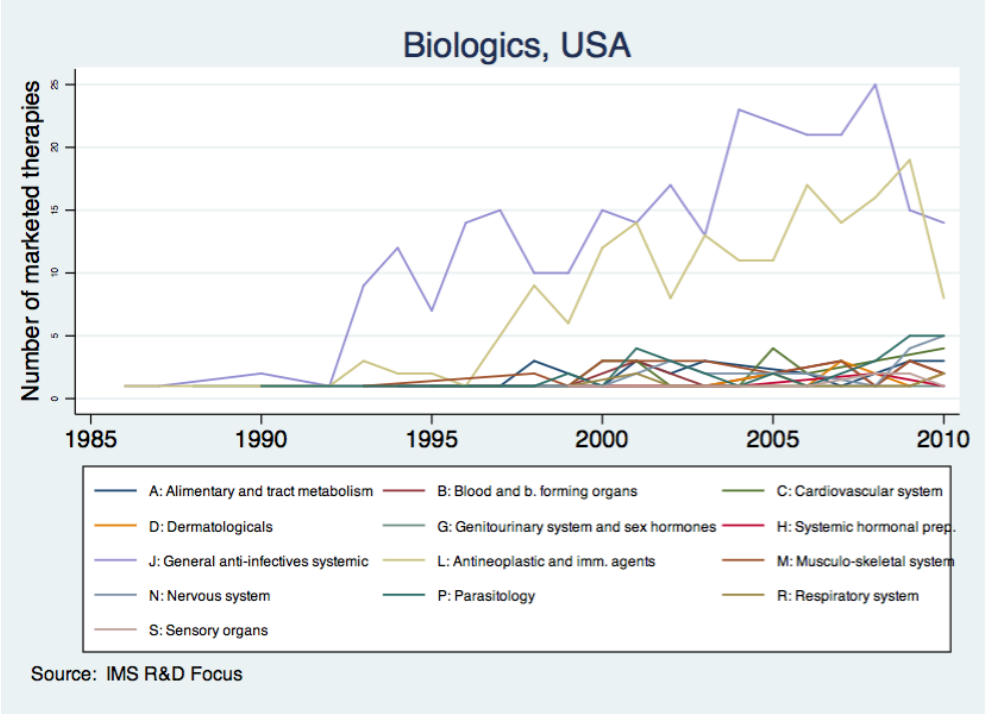


Figure 5.15. Biologics, historical trends in the number of market launches at therapeutic level in the USA

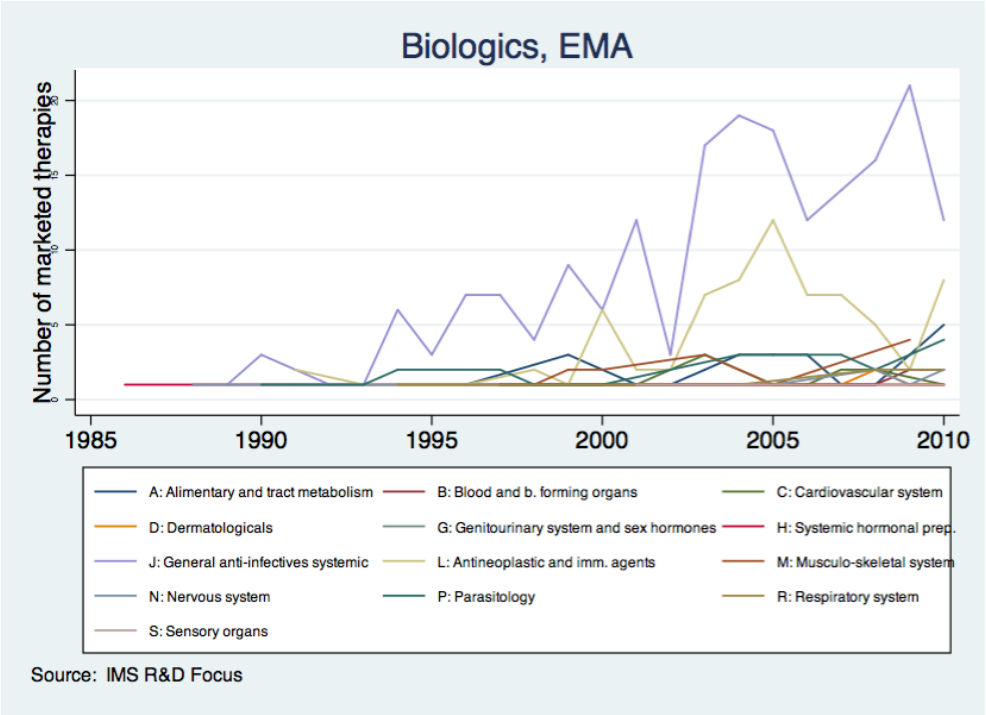


Figure 5.16. Biologics, historical trends in the number of market launches at therapeutic level in the EMA region

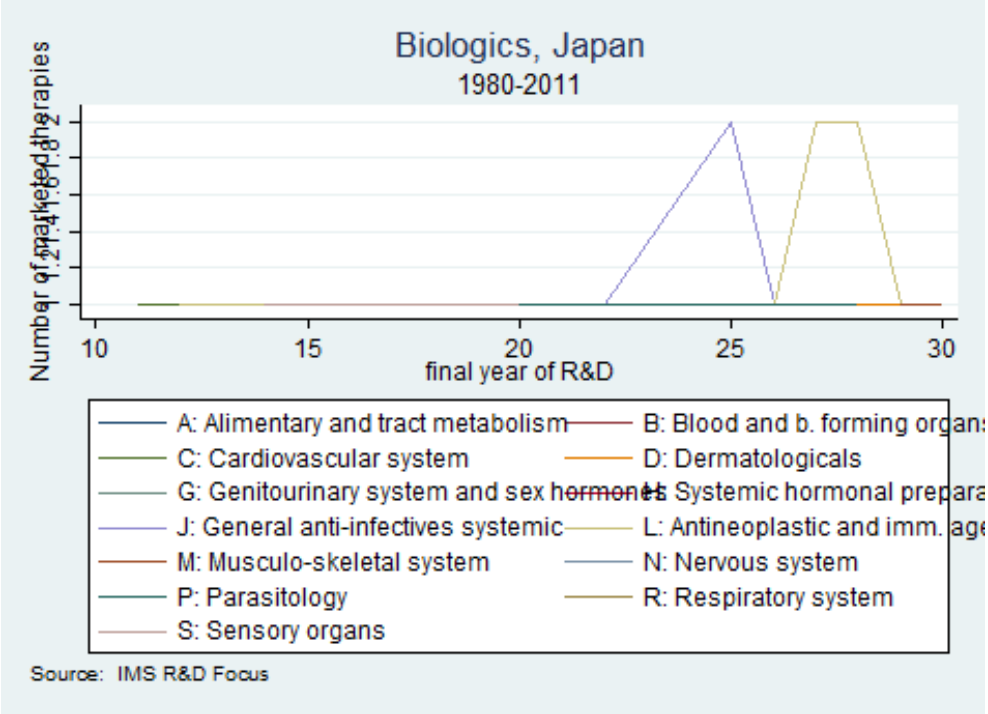


Figure 5.17. Biologics, historical trends in the number of market launches at therapeutic level in Japan

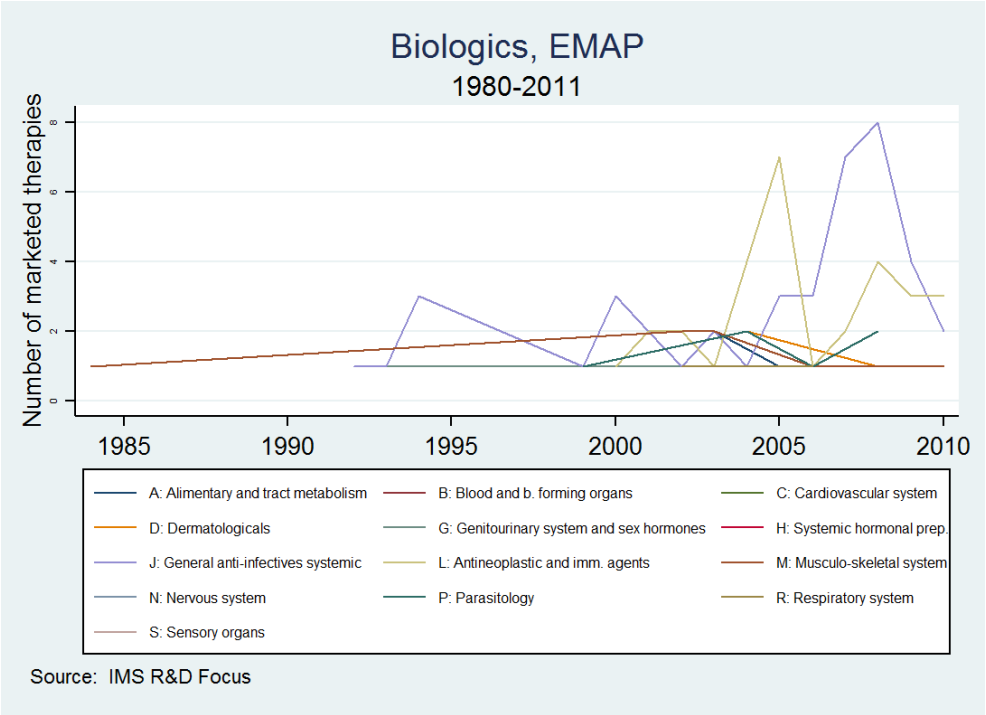


Figure 5.18. Biologics, historical trends in the number of market launches at therapeutic level in the EMAP region

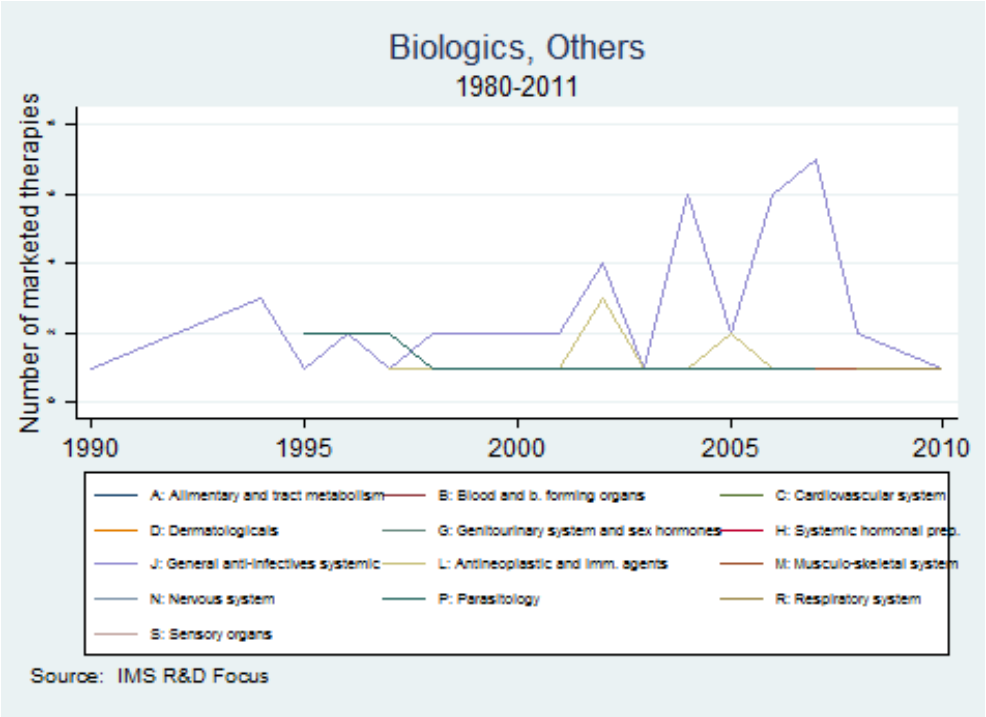


Figure 5.19. Biologics, historical trends in the number of market launches at therapeutic level in the rest of the world

Chapter 6

Measuring inequalities in pharmaceutical innovation in terms of unmet health need

I. Introduction

Pharmaceutical innovation has been concentrated in a small number of therapeutic categories. Since 2000, R&D investment has mainly targeted five therapeutic areas: (i) cancer, which accounts for more than half of the total industry's R&D spending (53.2%), followed by (ii) antibacterials and antimycotics (18.4%), (iii) treatments targeting the central nervous system (15.5%), (iv) therapies targeting metabolism conditions (8.8%) and, finally, (v) those to treat cardiovascular diseases (6.2%) (Pammolli et al., 2011).

Some of these therapeutic areas represent the most profitable markets for the industry (Acemoglu and Linn, 2004; Lichtenberg, 2005; Civan and Maloney, 2006; Bhattacharya and Packalen, 2011), as they target disease areas that affect mostly the high-income countries. The latest figures of global disease burden suggest an overlap of the industry R&D decisions and the main causes of disability in Europe, North America and Japan. They are respectively: cardiovascular and circulatory diseases, cancer, musculoskeletal diseases and mental disorders (Murray et al., 2013).

However, these are not necessarily the disease areas presenting the highest global health need. The majority of the world population lives in developing countries, for whom disease burden is mostly driven by infectious diseases, such as diarrhoeal diseases and lower respiratory infections, and neonatal conditions (Murray et al., 2013). However for these the market has sub-optimally delivered innovation (Trouiller et al., 2002). There are also diseases affecting the developed world for which there is lack of dynamic incentive mechanisms to promote the innovation. Some examples of neglected conditions are, namely, the thyroid disease (Lancet, 2012b), tuberculosis (Lancet, 2012c), viral, fungal and bacterial infections facing antimicrobial resistance (Smith and Coast, 2002; ECDC, 2013; DoH, 2012), and neurological diseases such as Parkison's disease and multiple sclerosis (Lancet, 2012a).

Therefore, there is a clear mismatch between the disease areas that have received more R&D investment from the pharmaceutical industry and global health needs.

This mismatch derives from two main reasons detailed in Chapter 4. The first reason relates to the intrinsic complexity of innovating in certain disease areas that consequently present higher level of failure in the discovery process. Therapies targeting these diseases are scientifically challenging, pose substantial safety and toxicity issues for the clinical development, and, compete with safe, cheaper, and readily available drugs established in the market for a long time.

The second reason relates to the strategic behaviour of firms when pursuing R&D investment decisions. These R&D strategies target areas that provide higher expected return on investment, larger market opportunities and lower uncertainty (DiMasi et al., 1991; Joglekar and Paterson, 1986; BCG, 2008; PWC, 2011b). This results in a lack of investment in new medicines targeting diseases prevalent in less profitable markets. These markets are unattractive to the industry because of lower willingness-to-pay and/or

affordability of the patients, or the low number of cases that demand a new medicine (Trouiller et al., 2002; Dukes, 2002; Drummond et al., 2007; Pecoul et al., 1999).

As a consequence, in the absence of optimal incentive mechanisms targeted at promoting innovation in riskier or less profitable disease areas, the current model has failed to deliver innovations in therapeutic areas with potentially large economic impact (Kremer, 2002; Kremer and Glennerster, 2004; Morton and Kyle, 2012).

It is then crucial to understand the extent to which the current R&D investments translate into health improvements and to assess who is benefiting from global innovation. It matters to understand whether investment strategies bring value-for-money to health systems or if there is a more efficient and more equitable allocation of resources that maximise the health and welfare benefits to populations. These concerns are not only pertinent to health systems but also to both private and public R&D investment initiatives, such as the Melinda and Bill Gates Foundation, Global Alliance for TB Drug Development, and International AIDS Vaccine Initiative, and to the pharmaceutical industry.

In order to inform policies to better promote R&D incentives in neglected areas and address these issues, it is critical to first identify the mismatches between pharmaceutical R&D strategies and the global health needs over time.

Therefore, in this Chapter we propose to measure inequalities in global drug R&D activity in terms of population health needs using measurement tools used in the vast literature in health inequalities.

This Chapter is organised in seven sections. The next section outlines the background literature. In section III we explain the methods used for the empirical analysis. In section IV we describe the data and variables used in the empirical analysis, and in section V we present the descriptive statistics. This is followed by a section that examines the results of concentration curves and concentration indices. Finally, section VII concludes with a discussion and interpretation of the results.

II. Background and literature

The meaning of "health need" is quite debatable and subject to considerable discussion in the literature. This debate is rooted in the various philosophical theories of social justice on the way in which welfare economists appraise the distribution of health benefits across populations (Culyer and Wagstaff, 1993; Wagstaff and Van Doorslaer, 2000; Williams and Cookson, 2000).

For policy purposes, however, it matters to define and measure "need" with a comparable and general health outcome indicator. It is desirable to have health outcome indicators that reflect the general health status of the individuals, and, at the same time, comprehensively characterise all health dimensions of a particular health condition (Gold et al., 1996).

Given this challenging task, there is substantial debate in the literature about what relevant measures should be used to assess health interventions from the perspective of the policy makers (as an example please see Drummond et al. (2005); Harris (2005); Claxton and Culyer (2007); Barker and Green (1996)) as the standardly used health outcome measures (avoidable deaths, health adjusted life years outcomes, number of deaths, amongst others) may not be desirable to use when assessing R&D priorities. Given this gap in the literature, we intend with this analysis to bring this debate closer to the issue of the usefulness of these measures to the distributional perspective of pharmaceutical R&D decisions.

To provide background to our study and address the gap of the literature, a narrative literature review was carried out related to ascertaining the studies that have measured dispersion of innovation in terms

of unmet global health need. A broad literature search was conducted in Medline, Embase and Cochrane Library and additional records were identified from grey and established literature discussing the topic namely the book byKremer and Glennerster (2004).

As we found, most of the literature explores anecdotally the idea of unequal distribution of biomedical innovation activity but fails to present measures of relative dispersion when referring to unmet health needs (Cicciò, 2004; Kremer and Glennerster, 2004; Finkelstein and Temin, 2008; Cohen et al., 2010; Fisk et al., 2011).

Nevertheless, there are three studies in the literature that measure the dispersion of innovation activity in terms of unmet health need (Catalá-López et al., 2011; Neumann et al., 2005; Viergever et al., 2013).

Two of them compare disease burden with the number of economic evaluations in Spain (Catalá-López et al., 2011) and in the US and other developed countries (Neumann et al., 2005).

Catalá-López et al. (2011) identify the most common disease causes (cardiovascular diseases, infectious diseases, malignant neoplasms and neuropsychiatric conditions) addressed by 477 economic evaluations of healthcare interventions published in Spain between 1983 and 2008. They find an unclear statistical relationship between the number of economic evaluations and the overall disease burden, whether disease burden is measured as the years of lost life, years lived with disability or disability-adjusted life years.

Neumann et al. (2005) examine 512 published cost-utility analyses (CUAs) in the US and in other developed countries, between 1976 and 2001, and compare rankings of the most common diseases covered by the CUAs to the rankings of US disease burden. They find no complete overlap between these two dimensions. They suggest that the leading causes of disease burden are mismatched by an over-representation of studies focusing on cerebrovascular diseases, diabetes, breast cancer and HIV/AIDS, and an under-representation of studies in depression and bipolar disorder, injuries, and substance abuse disorders.

A third study explores the mismatch between disease burden and the number of clinical trials recorded in the World Health Organisation (WHO) International Clinical Trials Registry Platform (ICTRIP) (Viergever et al., 2013). They find evidence of a large concentration of health R&D on non-communicable diseases (52.4 versus 7.4 trials in communicable, maternal, perinatal and nutritional conditions per million disability-adjusted life years (DALYs)), when exploring 2,381 records of trials.

Our study relates closely to these contributions in that it measures the dispersion of pharmaceutical innovation activity in terms of disease burden.

However, we depart from these contributions in several significant ways. First, we use global data on pharmaceutical R&D activity, and we merge it with global health data from the Global Burden of Disease Study (GBD) by the Institute of Health Metrics and Evaluation (IHME) (Murray et al., 2013). Secondly, we use tools commonly used in the inequalities literature (LeGrand, 1978; Wagstaff et al., 1989, 1991; O'Donnell et al., 2008) to quantify the mismatch between unmet health need and pharmaceutical R&D activity. Thirdly, we evaluate the dispersion of pharmaceutical R&D activity at three levels: i) unmet health need measured by disease burden, ii) unmet health need measured by mortality, and iii) willingness-to-pay/affordability. Lastly, we assess differences in this dispersion over time, i.e. between 1990 and 2010.

Based on the existing literature, we hypothesise that pharmaceutical R&D activity is concentrated on more commercially attractive, not necessarily more needy, disease markets. We test how dispersion in total R&D activity differs from the dispersion in the number of market launches, and how these relate to global health needs and affordability over time.

III. Methods for the measurement of inequalities

We use two tools that are referred to as the *"workhorse in health inequalities"* (Fleurbaey and Schokkaert, 2009): concentration curves and concentration indices. In the next three sections we discuss the methods used in the analysis and its application to the context of the pharmaceutical innovation. We do not intend to construct an extensive exposition of these methodologies. For a comprehensive discussion and guideline see Erreygers and Van Ourti (2011). In particular, in section A we discuss the concentration curve, statistical significance testing and its potential limitations in the context of pharmaceutical R&D activity, while in section B we outline the family of concentration indices discussed in the literature and applied to this context.

A. The concentration curve

We develop a framework of analysis that looks at the share of drug R&D activity r_i that corresponds to the cumulative proportion of two variables: (i) health need d_i, by disease i; and (ii) market attractiveness for disease i, m_i.

Diseases i are ranked according to need from the less needy to the neediest diseases. A concentration curve $L_r(d_i)$ represents the cumulative proportion of R&D activity targeted at disease i, and d_i captures the cumulative distribution of health need, that ranks diseases accordingly. The unit of analysis is the disease area, meaning that d_i, m_i and r_i are measured and grouped at disease level.

The representation of an equality line at 45-degrees highlights the idea of inequalities "in favor" of the less needy (or neediest) if the curve lies everywhere above (under) the equality line. With this, we can plot more than one curve in the same plane and perform comparisons over time.

However, graphical representation may be inconclusive since curves may cross the equality line several times (Lambert, 2001). This may also be true for simultaneous curves plotted against each other. Given that, we need robust decision rules to infer statistical dominance which are referred by the literature as dominance testing rules (O'Donnell et al., 2008).

Dominance testing rules

The statistical significance of the differences between curves, and between them and the equality line, can be inferred using dominance testing[1]. This feature is due to the relaxation of the restriction of increasing monoticity, contrary to the Lorenz curve representation (Van Doorslaer and Van Ourti, 2011). (O'Donnell et al., 2008). We will use two decision rules that have been theoretically presented and empirically used to infer statistically significant differences between concentration curves. One decision rule is named the *multiple comparison approach (mca)*. It considers a curve is dominant if all the points are statistically significantly different from those in another curve in one direction and with no statistically significant difference in the other. The second decision rule requires statistically significant difference between ordinates at all quantile points in order to assign dominance. This is consistent with the *intersection union principle* and reduces the probability of erroneously rejecting non-dominance at the cost of reducing the power of detecting dominance when true (O'Donnell et al., 2008)[2]. However, we have

[1] Owen Oonnell developed a dominance test that can be freely accessed through World Bank website: http://web.worldbank.org/ wbsite/external/topics/exthealthnutritionandpopulation/extpah/.

[2] Testing dominance when curves are dependent requires standard errors between ordinates allowing for dependence. A variance-covariance matrix between curves has been derived to allow for dependence between curves Bishop, J. A., K. V. Chow and J. P. Formby (1994). "Testing for marginal changes in income distributions with Lorenz and concentration curves."

only performed the *intersection union principle* because of limitations about number of observations. It should be remarked that our conclusions about dominance of concentration curves might be sensitive to the type of dominance test used. For this reason, the discussion of our results include the results from concentration curves and concentration indices.

B. The family of rank-dependent concentration indices

Concentration curves do not give a measure of the magnitude of inequality and dominance testing may be inconclusive. In particular, when curves cross the equality line the analysis becomes inconclusive. Therefore, a family of concentration indices has been developed to measure the magnitude of inequalities when the analysis performed with concentration curves is inconclusive.

The concentration index $I(r)$ can be generally defined as:

$$I(r) = f(r_i) \sum_{i=1}^{n} z_i r_i \text{ with} f(r_i) = (a_r, b_r, \mu_r, n) \tag{6.1}$$

where a_r and b_r correspond, respectively, to lower and upper bounds of the pharmaceutical R&D activity variable, n and μ_r denote the population and population mean of R&D activity r_i.

There are three main possible measures for concentration indices: (i) the Kakwani concentration index $C(r)$; (ii) the modified concentration index $\widehat{C}(r)$; and the Erreygers index $E(r)$.

The Kakwani index, $C(r)$, equals twice the area between the concentration curve and the diagonal, and reflects offsetting inequalities in different parts of the distribution (Wagstaff et al., 1991). It is represented as:

$$C(r) = \frac{2}{(n^2 \mu_r)} \sum_{i=1}^{n} \gamma_i r_i \tag{6.2}$$

This concentration index ranges from -1 to 1, with the sign of the ranking γ_i separating pro-needy from pro-less needy pharmaceutical R&D effort: positive values (those observations with ranks γ_i smaller than $\frac{(n+1)}{2}$) indicate inequalities in favour of needy disease groups. Similarly, in the case of the measure for market attractiveness, positive values indicate inequalities in favour of the more commercially attractive markets. This is given by $\sum_{i=1}^{n} \gamma_i r_i$ that denotes the rank-dependent weighted sum of all r_i levels. r_i is weighted by γ_i which ranks observations according to need variable d_i or variable market attractiveness m_i.

However, $C(r)$ imposes some restrictions on the properties of the variables that can be used in the analysis (Erreygers and Van Ourti, 2011). Not complying with these properties has profound implications for the use and interpretation of the concentration index. These are properties that relate to the level of independence of the numbers, cardinal invariance and mirror of the numbers that measure our variables of interest. These properties have three characteristics. First, that $C(r)$ can only give a measure of relative inequality if variables are measured on a ratio scale, as the index must remain unchanged regardless of possible transformations of the variables (Erreygers and Van Ourti, 2011). Secondly, when health variables are bounded (in many cases are left bounded) it is difficult to compare populations with different mean values that affect the bounds of the concentration index. Consequently, the concentration index must

International Economic Review: 479-488, Davidson, R. and J. Y. Duclos (1997). "Statistical Inference for the Measurement of the Incidence of Taxes and Transfers." Econometrica 65(6): 1453-1465.

reflect the lower and upper limits of the variables used. Finally, the degree of inequality of a variable is expected to mirror the same inequality of its opposite sign (for instance, health/ill health variable). This condition is only satisfied using unbounded variables (Erreygers and Van Ourti, 2011).

By construction, and as we will discuss further in section IV, our proxies for pharmaceutical R&D activity and health need variables are left-bounded and fixed-scaled/ratio-scaled, meaning that the measurement scale is unique with the zero point corresponding to a situation of complete absence (we are assuming non-negative R&D activity).

This implies that we need to use a transformed concentration index that considers the boundeness of variables (Clarke et al., 2002; Erreygers, 2009; Erreygers and Van Ourti, 2011). The modified concentration index $\widehat{C}(r)$ (Wagstaff, 2005) and Erreygers index $E(r)$ (Erreygers, 2009) are suggested by the literature as the appropriate indices when variables do not comply with the properties discussed above (Erreygers and Van Ourti, 2011). Given this, we devote particular attention to the results of $\widehat{C}(r)$ and $E(r)$ that can be written respectively as:

$$\widehat{C}(r) = \frac{2}{n^2(\mu_r - a_r)} \sum_{i=1}^{n} \gamma_i r_i \tag{6.3}$$

$$E(r) = \frac{8}{n^2(b_r - a_r)} \sum_{i=1}^{n} \gamma_i r_i \tag{6.4}$$

$\widehat{C}(r)$ incorporates upper b_r and lower a_r bounds of the measure of R&D activity r_i in the index calculation to remedy the issues with the bounds and measurement scale of the variable (Wagstaff, 2005). $E(r)$ presents a slightly modified version of the standard concentration index $\widehat{C}$ which satisfies all measurement properties, including the mirror property.

We present in section VI the results for the standard Kakwani concentration index $C(r)$, the modified concentration index $\widehat{C}(r)$ and the Erreygers index $E(r)$.

IV. Variables and data

We have fully described the dataset that supports this analysis in Chapter 3. As explained, three key variables at disease level are used in this analysis: (i) global pharmaceutical R&D activity, (ii) global health need and (iii) global market attractiveness. The following section describes the measures chosen to proxy global pharmaceutical R&D activity. In section B we detail the challenges and approaches to proxy global health need. In section C we discuss the construction of a measure for disease global market attractiveness to test the hypothesis of concentration of pharmaceutical R&D activity into more profitable markets. Finally, in section D we describe the data sources used for this analysis.

A. Global pharmaceutical R&D activity

Ideally, we would measure pharmaceutical R&D activity with yearly research expenditures at disease level and for each specific development phase of a project. However, pharmaceutical companies do not report R&D spending at such detailed level. Also, many of the firms are not publicly traded and do not disclose any financial information about their R&D spending.

Despite these limitations, and given the available information, we measure pharmaceutical R&D activity in two dimensions: (i) the overall R&D activity accomplished, by counting the disease-specific total number of projects (as used in existing literature, for instance, (Civan and Maloney, 2006)); and (ii) the successful R&D activity translated into new medicines, by counting the disease-specific market launches (as used by Acemoglu and Linn (2004) and Lichtenberg (2005)). With these, our analysis accounts for both the dispersion of total R&D activity and the dispersion at market launch level.

B. Global health need

The definition of health need imposes challenges, as exhaustively discussed in the literature (Culyer and Wagstaff, 1993; Wagstaff and Van Doorslaer, 2000; Williams and Cookson, 2000), and ultimately rests on value judgments (Gravelle et al., 2006).

In principle, it would be desirable to have measures of population health needs evaluated through health losses attributable to fatal and non-fatal diseases, injuries and associated risks (Lopez and Murray, 1998; Mathers et al., 2008). Preferably, this measure would aggregate clinical and patient-reported health in several domains thus accounting for all forms of disability (physical, psychological, mobility, functioning, and wellbeing).

To address this, a number of different standardized self-assessment measurement instruments have been developed to measure disability (Quality of Well-Being Scale, McMaster Health Index, EuroQoL EQ-5D, WHO DAS-II). However, these are criticised for several reasons, including being incomplete on the health dimensions contemplated, redundancy, prone to societal bias, and presenting huge practical constraints in data collection, which undermine cross-population comparability (Salomon et al., 2003).

Due to such limitations we use two measures of health outcomes that allow performing the analysis by considering the global nature of the pharmaceutical R&D activity.

As a first measure, we use cause-specific mortality to proxy health need. Mortality levels are readily available from IHME for 187 countries for 291 diseases and injuries for the years 1990 and 2010 (IHME, 2013). They have the advantage of objectively measuring the number of deaths caused by a certain condition across the globe, facilitating any study that requires cross-country comparisons.

This crude measure of health need presents, however, some drawbacks.

On the one hand, death registration is reported to be poor in some countries, with inaccurate definitions of the cause of death, raising issues about the harmonisation of the death coding for cross-country studies. Most recently, many efforts have been directed by the IHME initiative on the GBD study to use verbal autopsies as an additional source of information on causes of death in the least developing countries (IHME, 2013).

On the other hand, this is regarded as a one-dimensional measure that does not capture any non-fatal dimension that may impact on the health of populations. There are conditions that, not being mortal, persistently affect the quality of life of individuals and therefore ought to be captured in the analysis of the burden of disease. Therefore, measures such as disability-adjusted life years (DALYs) have been standardly used to quantify the burden of diseases, injuries and risk factors on human populations in international cost-effectiveness studies. DALYs are a general composite comparable measure that aggregate information at mortality and morbidity levels. DALYs are time-based measures that combine years of life lost (YLL) due to premature death, and years of life lost due to time lived in states of poor health or disability (YLD). One DALY can be interpreted as one lost year of "healthy" life, and the disease burden can be thought of as a measure of the gap between current health status and perfect

health. YLL are calculated using the number of deaths at each age multiplied by an age specific global standard life expectancy. YLD for a particular cause, in a particular time period, are estimated as the number of incident cases in that period multiplied by the product of the average duration of the disease and a weight factor[3] (Murray and Acharya, 1997; Gold et al., 1996).

DALYs have been however criticised for the economic and ethical assumptions that are made in the parameters used in the calculations. There are parameters related to the discounting factors, age weighting, and the social preferences assumed to estimate disability weights (Mathers et al., 2006; Gold et al., 1996; Barker and Green, 1996).

Despite this criticism, we have used DALYs for the readily available data for 187 countries for 291 diseases and injuries (IHME, 2013) and given the importance that DALYs play on the global health agenda and policy development.

C. Disease market attractiveness

To develop a measure of disease market attractiveness, we construct a measure of national *ability-to-pay* m_i for each disease market i by combining national health need and a measure of national income, using World Bank statistics on national Gross Domestic Product (GDP) per capita (measured in purchasing power parity terms at 2005 constant US$), and aggregating national ability-to-pay as:

$$m_{ij} = \sum_{ij} need_{ij} GDP_j \tag{6.5}$$

where $need_{ij} = \{deaths_{ij}, DALYs_{ij}\}$ denotes "need" for disease i in country j and GDP_j refers to the GDP per capita for country j. This measure assumes that all resources devoted to improving the health of a country are diverted to a single disease. This is unrealistic as resources are scarce and allocated according to policy decisions incorporated in national health programmes. However, we do not have information on the distribution of resources across diseases. We assume resources could be distributed in accordance to the health need related to a specific disease i of a particular country. Nonetheless, it is likely that the distribution of resources across diseases and health need are correlated. Therefore, we construct a second disease market attractiveness measure $\overline{m}_{ij}$ that is adjusted by the proportion of "need" of disease i on the total need of country j, $\overline{need_{ij}} = \sum_{ij} \frac{need_{ij}}{\sum_i need_i}$. This is expressed by:

$$\overline{m}_{ij} = \sum_{ij} \overline{need_{ij}} GDP_j \tag{6.6}$$

D. Sources of Information

We have built a unique dataset that matches data from three sources as explained in detail in Chapter 3: (i) IMS Health R&D Focus database that provides information on global pharmaceutical R&D activity;

[3]The weight factor reflects the severity of the disease on a scale from 0 (perfect health) to 1 (death) (weights used for the Global Burden Disease (GBD) study 2004 are listed in Annex Table 3.A6 of Mathers et al Mathers, C. D., A. D. Lopez and C. J. L. Murray (2006). "The burden of disease and mortality by condition: data, methods, and results for 2001." Global burden of disease and risk factors 1: 45-93). The GBD 2004 study used the following formula: $W = 0.1658Y^{e-0.04Y}$ where Y is the age in that year and W is the value assigned to it relative to an average value of 1. Future years were discounted at a 3% rate, so that a weighted year of life saved next year is worth 97% of a year of life saved this year.

(ii) the GBD study from IHME that includes the most recent global estimates on cause and country-specific DALYs and mortality levels for 1990 and 2010; and (iii) World Bank statistics on national GDP per capita (measured in purchasing power parity terms at 2005 constant US).

IMS Health R&D Focus database is updated weekly and is typically used by pharmaceutical companies to monitor R&D activities of the competitors. It provides a history of all global projects known to be in development from the early-1980s to the present, compiling information from patent and regulatory filings, presentations at medical conferences, press releases, and information disclosed to financial analysts (IMS, 2011). This dataset registers the progress of compound projects across R&D development phases (i.e., early discovery, Phase 1, Phase 2, Phase 3, and market/registration), and records the successful compounds, which followed an approval process by the regulatory body and reached the market, as well as the discontinued compounds that for several reasons failed to successfully go through the development process (i.e., technological failure due to toxicity and efficacy tests, or economic failure due to profitability issues or commercial risk).

Our data consists of 65,845 country-compound-indication projects at the date of extraction (October 2011) (please refer to Chapter 3 for complete explanation of the unit of analysis). Since compounds often target several indications in more than one country simultaneously simultaneously, we have considered all therapeutic indications targeted and all countries where the R&D is taking place, resulting that our unit of analysis is at country-compound-indication level. We used GBD data to inform estimates of burden of disease and mortality to proxy global population health need. This data is publicly available for 1990 and 2010 at disease-country level. IHME uses the International Classification of Diseases (ICD)[4] coding system to produce disease level estimates of burden and mortality for a list of 291 diseases and injuries (IHME, 2009). The group of causes of death that relates to "external causes" for car accidents, falls, injuries, war, conflicts and intentional self-harm conditions were excluded from this analysis as these are not typically addressed by pharmaceutical R&D activity.

Both datasets were matched by corresponding therapeutic indications to the ICD disease categories, using the ICD-10[5] code system and online medical dictionaries. The final dataset includes a total of 59,301 country-compound-indication projects after matching non-missing therapeutic indications and ICD-10 codes (dropping 6,544 (0.099%) records with missing therapeutic indication). With this match and using the categorisation of diseases and injuries used by the IHME, we measure the dispersion of global R&D activity in terms of unmet health need.

V. Descriptive statistics

In this section, and before introducing the results in section VI, we describe how the data on health need and pharmaceutical R&D activity is dispersed across the different diseases. We first start by describing how DALYs and mortality levels spread over the diseases. Then we examine the dispersion of pharmaceutical R&D activity across diseases.

Unmet health need across the globe

If we plot mortality across disease areas ranking the five main global causes of death for 2010, respectively:

[4]The ICD is the international standard classification of medical conditions for all general epidemiological, health management and clinical purposes. This coding system provides the basis for the compilation of national mortality and morbidity statistics by WHO member states.

[5]version 2010 available in http://apps.who.int/classifications/icd10/browse/2010/en

ischaemic heart disease (IHD), stroke, chronic obstructive pulmonary disease (COPD), lower respiratory infections and lung cancer. However, DALYs show that diahrrea and HIV/AIDS present more disability than COPD and lung cancer, representing two of the five main global causes of disease burden in 2010 (Table 6.1). These are findings of the most recent evidence on distribution of mortality and morbidity across the globe (Murray et al., 2013).

When looking at these figures at regional level, the GBD study finds large variations in the numbers and differences in terms of the causes of death and disability across high-income and low-income regions. In high-income countries, such as North America and Western Europe, non-communicable diseases are the major causes of death and disability, namely conditions related with the cardiovascular and circulatory system (e.g. ischaemic heart disease, COPD, cerebrovascular diseases), mental health disorders (e.g. major depressive disorder) and trachea, bronchus and lung cancers; whilst the low-income countries (namely in Southeast Asia and Sub-Saharan Africa) commonly suffer from infectious and parasitic diseases such as tuberculosis and lower respiratory infections. In the case of Sub-Saharan Africa, the three main causes of death are respectively: HIV/AIDS, Malaria and Diahrrea (Table 6.2).

[Table 6.1 and Table 6.2 here]

However, there are regions showing structural changes in the epidemiological profile in the last two decades. In Southeast Asia, Latin America and Caribbean, and North Africa and Middle East a larger proportion of disease burden is nowadays attributable to NCD traditionally prevalent in the high-income settings. In the Southeast Asia, diseases such as LRI, diarrheal diseases, tuberculosis and malaria that were prevalent in the region 1990 were replaced by cerebrovascular diseases, IHD and COPD as main causes of disease burden in 2010. In Latin America and Caribbean, diabetes emerges as the fifth cause of death and MDD the fourth cause of disease burden in 2010. The epidemiological transition to NCD is also clear in North Africa and Middle East, where IHD, MDD, stroke and low back pain prove to be the four main causes of disease burden in the region in 2010, contrasting with LRI and diarrheal diseases that were the first two causes of disease burden in the region in 1990 (Murray et al., 2013).

[Table 6.3 here]

Dispersion of R&D across diseases

R&D activity is markedly dispersed across diseases. Figure 6.1 suggests an inverse relationship between disease burden and pharmaceutical R&D activity, when taking a snapshot of the R&D landscape in 2010.

Conditions with larger disease burden and higher mortality (red diamonds in the figure) receive barely a few dozen pharmaceutical projects targeting such diseases (left-hand side figure). In 2010, for conditions with higher mortality such as hemorrhagic and other non-ischaemic stroke (3,043,326 and 62 843 000 DALYs per million people), trachea, bronchus, lung cancers (2,835,852 and 33 405 000 DALYs per million people), and diabetes mellitus (1,282,553 and 46 823 000 DALYs per million people), the total number of pharmaceutical projects was approximately 167, 370 and 174, respectively.

At the same time, conditions associated with higher R&D activity present relatively lower mortality and disease burden (green diamonds in the figure): other neoplasms are targeted in around 8,030 projects (16 615 000 DALYs per million of people and 608,658 deaths); other endocrine and immune disorders are targeted in approximately 3,388 projects, brain and nervous system cancers are targeted in around 2,100 projects, but present a negligible number of deaths and DALYs.

The same dispersion occurs for the number of new medicines that reach the market (right-hand side figure). Outliers such as other neoplasms with 2,929 new therapies, other mental and behavioural disorders

with 1,075, and other endocrine and immune disorders with 1,225 show incomparably low number of deaths and DALYs, in relation to the large numbers for mortality and disability presented by hemorrhagic and other non-ischaemic, trachea, bronchus, lung cancers, and diabetes mellitus.

At the end of the spectrum of dispersion of pharmaceutical R&D activity, there is a group of 53 disorders for which no novel therapies have been launched in the market in the last two decades, despite the level of R&D activity and the relatively high disease burden (full list presented in technical appendix V Table 6.4). In particular, no new therapies were developed to target preterm birth complications, congenital heart, rheumatic heart disease, polycystic ovarian syndrome and acute glomerulonephritis, and several neglected tropical diseases[6].

Finally, the dispersion of pharmaceutical R&D activity is also clearly shown in the concentration of R&D investments in fewer than 178 rare diseases. Amongst 6,800 rare and genetic diseases (including neglected tropical diseases) listed by the Genetic and Rare Diseases Information Centre (GARD) of the US National Institute of Health, data shows a concentration of R&D investment in approximately 178 disorders in the last two decades.

Less than twenty of those account for more than 52.5% of the total number of R&D projects targeting rare diseases[7] (list presented in Table 6.5). Among these, the conditions exhibiting the highest level of R&D activity are retinopathy, lupus nephritis, dengue fever and *li-fraumeni* syndrome, and melanoma. The dispersion illustrated by the descriptive statistics is clearly shown by the results for the concentration curves and concentration indices, as we present in the next section.

VI. Results from concentration curves and concentration indices

In the following section, we examine the concentration curves and concentration indices for the two proxies of R&D activity we use in this analysis: i) the total disease-specific number of R&D projects; and (ii) the disease-specific number of market launches.

We have performed three types of analysis. In the first we have measured the dispersion of drug innovation and R&D activity for the broad disease groups, i.e. CD and NCD (Section A). In the second we have used as unit of analysis disease subcategories by selecting the four disease subcategories causing more deaths and the four disease subcategories with higher disability in developed and developing countries (Section B).[8]. We follow the burden of disease cause classification list presented by the GBD study[9]. Additionally, and within the analysis for CD, we also present the special case of NTDs, given its public health relevance. Finally, we have looked at the evolution of inequalities between 1990 and 2010 for

[6]such as food-borne trematodiases, trichuriasis, ascariasis and yellow fever have no therapies.

[7]Rare diseases are generally considered to be diseases that affect fewer than 200,000 people. This definition was included by United States Congress in the Orphan Drug Act of 1983.

[8]For the developed countries: the top 4 causes of death are cardiovascular and circulatory diseases, neoplasms, chronic respiratory diseases, and diarrheal diseases and LRI; the top 4 causes of disease burden are cardiovascular and circulatory diseases, neoplasms, musculoskeletal disorders, and mental and behavioral disorders. For the developing countries: the top 4 causes of death are the same as the developed countries; the top 4 causes of disease burden are diahrreal diseases and LRI, neonatal disorders, cardiovascular and circulatory diseases, and nutritional deficiencies. (Murray et al., 2013)

[9]The GBD 2013 Study classifies disease and injury causes using a tree structure. The first level of disaggregation comprised three broad cause groups: Group I: Communicable, maternal, perinatal, and nutritional conditions; Group II: Noncommunicable diseases; and Group III: Injuries. Each Group is divided into major subcategories. For example, cardiovascular diseases and malignant neoplasms (cancers) are two major cause subcategories of Group II. Beyond this level, there are two further disaggregation levels. The major cause sub-categories are closely based on the chapters of the ICD, with a few significant differences (IHME, 2013)

both the broad disease groups and disease subcategories causing highest mortality and morbidity, and the particular case of NTDs.

We depict concentration curves that illustrate the cumulative percentage of R&D activity (y-axis) against the cumulative percentage of DALYs, mortality and market attractiveness (x-axis). Dominance testing (Table 6.6 and Table 6.7) and concentration indices (Table 6.8 and Table 6.9) are calculated accordingly, with special attention paid to the Kakwani $C(r)$, modified concentration index $\widehat{C}$ and Erreygers index E. Detailed results are presented below.

In section C we investigate possible statistically significant changes between 1990 and 2010.

A. The two broad disease groups

There is a general concentration of pharmaceutical activity towards more needy disease areas.

This can be clearly seen when splitting the analysis into CD and NCD. These results are in line with those found in the analysis of the dispersion of pharmaceutical R&D activity against DALYs and mortality levels. Figures for both groups of diseases illustrate a concentration of R&D activity towards more disabling/more fatal diseases over the period 1990-2010 - concentration curves for both years (1990 and 2010) lie under the equality line (Figure 6.2 and Figure 6.3).

[Figure 6.2 and Figure 6.3 here]

These figures also suggest different levels of inequality when comparing communicable diseases and non-communicable diseases. The pro-disabling/pro-fatal concentration of R&D activity is more pronounced in the group of communicable diseases than in the group of non-communicable diseases. These conclusions apply to both measures of R&D activity, i.e. for the: (i) cumulative share of the total disease-specific number of R&D projects (Figure 6.2), and the (ii) cumulative share of the disease-specific number of market launches (Figure 6.3).

Regarding affordability, as suggested by the literature, pharmaceutical R&D activity is concentrated on more commercially attractive disease groups (Acemoglu and Linn, 2004; Dubois et al., 2011; Lichtenberg, 2005). Interestingly, this inequitable distribution is only significantly different from the equality line when mortality is used as a measure of health need (Figure 6.10 and Figure 6.11). This means that R&D activity is more attracted by more pro-fatal than pro-needy diseases. Disability seems not to be significantly correlated with the dispersion of R&D and affordability.

[Figure 6.10 and Figure 6.11 here]

This is shown by the 45-degree-line dominating the concentration curves for both years, presenting larger concentration indices (Table 6.8). Indeed, these inequalities are more pronounced for Communicable diseases than for NCD, as suggested by the larger values of the concentration indices for the group of CD (Table 6.8).

[Table 6.8 here]

B. Disease subcategories

R&D activity has been concentrated on disease subgroups with higher mortality levels for the ranked top four causes of death. The positive concentration indices (Table 6.7) confirm an unequal distribution of R&D activity towards pro-fatal diseases, namely for the number of market launches. The exceptions

are the chronic respiratory diseases, diarrheal diseases, lower respiratory infections and other infectious diseases where insufficient observations are available to compute concentration indices. Concentration curves cross the equality line. The figures with the corresponding concentration curves that depict the distribution of pharmaceutical R&D activity across the TOP4 most global fatal diseases are provided in the technical appendix IX (Figure 6.6 and Figure 6.7).

[Table 6.6 and Table 6.7 here]

[Figure 6.6 and Figure 6.7 here]

The pro-needy distribution of R&D activity is also suggested by the concentration curves (Figure 6.4 and Figure 6.5) and concentration indices (Table 6.9) for the disease subgroups that are the four highest causes of disease burden[10].

[Figure 6.4 and Figure 6.5 here]

For the disease subcategories of cardiovascular and circulatory diseases (1st cause of disease burden in developed countries and 3rd in developing countries), neoplasms (2nd cause of disease burden in developed world), musculoskeletal disorders (3rd cause of disease burden in developed countries), and diarrheal, lower respiratory infections and other infectious disorders (1st cause of disease burden in developing countries), there is a clear concentration of R&D activity on conditions with higher disease burden, as the curves lie below the 45-degree-line (Figure 6.4 and Figure 6.5) and as confirmed by the positive concentration indices (Table 6.9). For example, within the disease subgroup of musculoskeletal disorders, there are more new therapies placed in the market for osteoarthritis compared with those for juvenile rheumatoid arthritis. The same result is found for the total number of pharmaceutical projects and can be seen in the technical appendix IX (Figure 6.8 and Figure 6.9).

[Table 6.8 and Table 6.9 here]

Concentration curves are unclear when assessing the direction of the inequality in two of the disease subcategories with the highest burden of disease: mental and behavioural disorders (4th cause of disease burden in developed countries) and neonatal disorders (2nd cause of disease burden in developing world) as they cross the equality line. However, the positive concentration indices suggest a marked pro-needy concentration of the R&D activity (Table 6.9). This is true for both the R&D activity measures of total number of projects and number of new therapies in the market. These results hold when considering the measure of affordability.

The particular case of Neglected Tropical Diseases[11]

Results suggest a clear concentration of innovation towards Neglected Tropical Diseases (NTD) with higher mortality, whether measuring innovation by the number of total projects or available therapies in

[10] the top four diseases with the highest burden of disease (measured in DALYs): musculoskeletal discarders and mental and behavioural disorders in the developed world; neonatal disorders and nutritional deficiency in the developing world (Murray et al., 2013).

[11] The NTD are a group of parasitic and bacterial diseases that cause substantial illness and are endemic amongst the poorest countries. According to the Centre for Disease Control and Prevention (CDC) and WHO, 17 diseases are part of this group, including: Buruli ulcer, Chagas disease, Cysticercosis, Dengue fever, Dracunculiasis, Echinococcosis, Fascioliasis, African Sleeping Sickness, Leishmaniasis, Leprosy, Lymphatic filariasis, Onchocerciasis, Rabies, Schistosomiasis, Soil-transmitted helminthes, Trachoma, and Yaws. For the purpose of this analysis we also included Malaria in the group of NTD.

the market. This is suggested by the dominance testing (Table 6.7) and the positive concentration indices (Table 6.7).

However, the results are inconclusive when using DALYs. This seems to be because of two reasons. First, mortality is a poor measure of need for the NTD affecting by far the developing world. The ranking of the NTD in terms of associated DALYs does not overlap with the ranking of NTD in terms of associated mortality. Lymphatic filariasis and Trichuriasis present 0% of global deaths but are responsible for 0.11% and 0.026% of global DALYs in 2010, respectively. Second, results from dominance tests (Table 6.7) suggest that there is not necessarily a concentration of drug innovation and R&D activity in more disabling diseases. This is expressed by the concentration curve (Figure 6.12) crossing the 45-degree line showing a shift of the dispersion of innovation and R&D activity from relatively more disabling conditions (Malaria, Leishmaniasis and Rabies) to relatively less disabling ones (Schistosomiasis and Filariasis). This shows that the pharmaceutical industry has been targeting significantly more fatal, but not necessarily more disabling NTD. Indeed, the industry has not been taking into account disability in the R&D decisions targeting NTD or that mortality might more accurately predict the way the industry looks at this market.

[Figure 6.12 here]

These results emphasise the need for further discussion about the R&D decisions to target NTDs that can deliver health improvements to populations.

C. What changed from 1990 to 2010?

When looking at the broad disease categories, for CD results suggest no significant changes over the last two decades on the dispersion of innovation and R&D activity against need as suggested by the dominance tests between the concentration curves of 1990 and 2010 (Table 6.6). These results hold when we measure need with both DALYs and deaths. For NCD, however, results suggest there has been an increase over time in the concentration of drug innovation and R&D activity towards more needy diseases (Table 6.6). This change over time is explained by the increased burden of disease associated with NCD in developing countries. Indeed, a closer look into the analysis reveals that the ranking of diseases according to disease burden has significantly changed between 1990 and 2010, with the top disease burden drivers in developing countries coinciding with those diseases that affect richer countries in 2010 (as illustrated in Table 6.2). Therefore, while the strategy of the industry might still be purely driven by financial considerations, the convergence in causes of mortality and morbidity between developing and developed countries is a probable explanation behind the increased concentration towards pro-needy conditions over time.

When looking at the subcategory analysis, there are no significant changes over time (Table 6.7) as confirmed by the non-statistically significant difference between the concentration curves depicting the dispersion of innovation and R&D activity between 1990 and 2010. This also holds for the specific case of the NTD.

VII. Discussion and conclusions

This paper is the first to measure inequalities in global drug R&D activity in terms of population health needs. We have examined the distributional dimension of pharmaceutical R&D activity across diseases using concentration curves and indices explored in the vast literature in health inequalities.

Three important conclusions are drawn from this analysis.

First, the mismatch between need and innovation and affordability shows an inequality of innovation in disease areas that have higher affordability and high mortality. The results are similar when assessing inequalities in terms of DALYs. That said, the additional information on morbidities incorporated in the measure of disease burden does not alter the results. This may imply that the industry does not account for disease burden in the R&D decisions or that mortality might more accurately predict the way the industry looks at the market.

The exception goes to the subcategory of diseases composed by the NTD to which conclusions change depending on the measurement of need considered. There is an inequality of innovation favouring the most fatal but not necessarily the most disabling NTD.

Second, inequalities have been unchanged in the last two decades. The significant redistribution of innovation found in the group of NCD towards more profitable and most fatal NCD is masking a relatively immutable dispersion of innovation between the nineties and 2010, when we consider the subcategories of CD and NCD.

Third, disparities are unevenly distributed across diseases. Disparities are larger in the group of CD that represent the biggest disease burden in LIC, and are relatively "more needy" conditions in the world panorama of health needs. In the last two decades, novel therapies were absent in 53 disorders, some of them presenting high levels of mortality. For example, preterm birth complications accounts for more than 77 thousand deaths; congenital heart anomalies for more than 225 thousand deaths and rheumatic heart disease for about 300 thousand deaths. Disorders with high burden of disease such as some musculoskeletal disorders, and several NTDs are still unresolved. This is because of the low ability to pay for new therapies in low-income countries that suffer from these conditions (Acemoglu and Linn, 2004; Kremer, 2002; Lichtenberg, 2005), and the already effective (and cheap) treatment alternatives available (schistosomiasis and soil transmitted helminths can be effectively controlled by drugs such as praziquantel)[12]. At the other end of the spectrum of the distribution of the R&D activity, cancer receives more than a half of global R&D investment. There are signals of a market expansion of these therapies in the developed world. This could be associated with the redefinition of health technology assessment rules for cancer therapy drugs in several European countries (Drummond and Mason, 2007), and with more private health insurances advertising the inclusion of cancer drugs on the reimbursement formularies.

In conclusion, these results show evidence of profound disparities of pharmaceutical R&D activity in terms of unmet health need across diseases. Further research is required to assess its determinants. These disparities could be associated with multiple factors including the peculiarities of the disease pathways and their clinical and scientific complexity, or the market conditions and strategic behaviour of pharmaceutical companies. It could also be the result of policies and interventions targeted at the sector. These policies comprise, among others, incentives to invest in research in certain diseases, through patenting, privileged and fastest regulatory process (e.g. the orphan drug act), purchase commitments (such as the case of pull-push commitment incentives for neglected diseases, cost-effectiveness rules for approval of drugs), and safety regulatory rules. It seems pertinent to bring the policy debate towards the urgency for the evaluation of the impact of these policies on the level of disparities acknowledged in the pharmaceutical R&D activity (please see Chapter 2 for an insightful discussion on these factors).

With all these conclusions it should be remarked that equal distribution of R&D activity is not necessarily desirable or expected. It is important to bear in mind that pharmaceuticals are a part of the

[12]Also, other external factors, namely sanitation conditions and hygiene conditions, are essential to the efficacy of new therapies targeting, for instance, tetanus and hookworm diseases . Others, like diphtheria, depend on the efficacy of public health interventions

health production function, where other determinants are similarly important for the provision of health. Nevertheless, examining discrepancies at this level helps to shed light on the discussion of how effective R&D decisions have been in promoting better health outcomes across the world, and whether current global public/private initiatives are focusing attention on health interventions that could generate the most impact.

There are several caveats to our analysis. It is out of the scope of this analysis to assess structural or causal relationships between the different dimensions assessed in this analysis. Further work on the determinants of drug R&D activity may help inform the policy debate on more effective incentive schemes to boost R&D in certain disease areas. A promising avenue of research is to consider the dynamic and sequencing nature of the drug development by using the past history of R&D stages. By controlling for R&D determinants extensively explored in the literature, this approach may help examining how different determinants may affect the performance of the R&D activity, with important equity and welfare implications.

As a last remark, it should be mentioned that among the caveats of this analysis, the measures we use to quantify health need are controversial and may raise some issues about their international comparability.

First, using mortality as a measure for health need assumes perfectly comparable death and diagnostic records across the globe. Second, it disregards the large disease burden of disabling conditions such as long-lasting conditions that include diabetes, mental health disorders, asthma, and epilepsy, as well as non-fatal but, nonetheless, hugely debilitating diseases such as human african trypanosomiasis, lymphatic filariasis and malaria.

Furthermore, composite measures like DALYs have not overcome the limitations around parameter assumptions in their calculation, including the value judgments around choices for discount rates, and the role of experts and local communities on the measurement of time preferences. Moreover, there is the need for careful consideration of the role of DALYs in relation to equity and distributional concerns. The use of the DALYs approach, in an attempt to target the use of resources to maximise the benefits gained, is likely to help those who can most benefit, those whose their conditions are quantifiable, over and above those who are in greatest need (Barker and Green, 1996). There is clear scope for further research on measurement of health outcomes and improvements to measures like DALYs.

It would make sense to consider more refined measures of need, for instance at therapeutic indication or sub-indication level to measure more accurately unmet need. However, and because global health outcomes data is constructed and measured at disease level, there is no such data available. In order to use so we would have to weigh the relative importance of each indication in explaining the burden of disease of a disease group. Because there is no such one-to-one relationship between diseases and indications (as the ICD-10 codes show), that attempt would always be arbitrary and so we have chosen to structure our analysis at disease level.

Finally, this analysis could be enriched by assessing inequality in other dimensions, namely access and distribution of therapies across regions and within countries.

VIII. Appendix: Tables

Table 6.1. IHME, Top5 disorders with higher global mortality and disease burden in 1990 and 2010

Global DALYs - TOP5 disorders 2010 GBD IHME study	
1990	2010
1. Lower respiratory infections	1. Ischaemic heart disease
2. Diahrrea	2. Lower respiratory infections
3. Preterm birth complications	3. Stroke
4. Ischaemic heart disease	4. Diarrhea
5. Stroke	5. HIV/AIDS
Global deaths - TOP5 disorders 2010 GBD IHME study	
1990	2010
1. Ischaemic heart disease	1. Ischaemic heart disease
2. Stroke	2. Stroke
3. Lower respiratory infections	3. COPD
4. COPD	4. Lower respiratory infections
5. Diahrrea	5. Lung cancer

Table 6.2. IHME, Regional Top5 disorders with higher global mortality and disease burden in 2010

High-income countries: DALYs 2010		
High-income Asia	High-income North America	Western Europe
1. Cerebrovascular diseases	1. Ischaemic heart disease	1. Low back pain
2. Low back pain	2. COPD	2. Ischaemic heart disease
3. Ischaemic heart disease	3. Low back pain	3. Cerebrovascular diseases
4. Other musculoskeletal diseases	4. Trachea, bronchus and lung cancers	4. Major depressive disorders
5. Self-harm	5. Major depressive disorders	5. Trachea, bronchus and lung cancers
High-income countries: Deaths 2010		
1. Cerebrovascular diseases	1. Ischaemic heart disease	1. Ischaemic heart disease
2. Ischaemic heart disease	2. Trachea, bronchus and lung cancers	2. Cerebrovascular diseases
3. Self-harm	3. Cerebrovascular diseases	3. Trachea, bronchus and lung cancers
4. Trachea, bronchus and lung cancers	4. COPD	4. Colon and rectum cancer
5. Lower respiratory infections	5. Road injury	5. COPD
Low-income countries: DALYs 2010		
Southeast Asia	Southern Sub-Saharan Africa	Central Sub-Saharan Africa
1. Cerebrovascular diseases	1. HIV/AIDS	1. Malaria
2. Tuberculosis	2. Lower respiratory infections	2. Diahrrea
3. Ischaemic heart disease	3. Diahrrea	3. Protein-energy malnutrition
4. Lower respiratory infections	4. Tuberculosis	4. Lower respiratory infections
5. COPD	5. Interpersonal violence	5. HIV/AIDS
Low-income countries: Deaths 2010		
1. Cerebrovascular diseases	1. HIV/AIDS	1. Malaria
2. Tuberculosis	2. Lower respiratory infections	2. Diahrrea
3. Lower respiratory infections	3. Diahrrea	3. Protein-energy malnutrition
4. Ischaemic heart disease	4. Interpersonal violence	4. Lower respiratory infections
5. Road injury	5. Tuberculosis	5. HIV/AIDS

Table 6.3. IHME, Regional Top4 disorders with higher global mortality and disease burden in 1990 and 2010

DALYs		Deaths	
1990	2010	1990	2010
North America			
1. IHD	1. IHD	1. IHD	1. IHD
2. TBLC	2. COPD	2. Stroke	2. TBLC
3. COPD	3. Low back pain	3. Lung cancer	3. Cerebrovascular diseases
4. MDD	4. TBLC	4. COPD	4. COPD
Southern Sub-Saharan Africa			
1. Diahrrea	1. HIV	1. LRI	1. HIV
2. LRI	2. LRI	2. Diahrrea	2. LRI
3. HIV	3. Diahrrea	3. Tuberculosis	3. Diahrrea
4. Tuberculosis	4. Tuberculosis	4. HIV	4. Interpersonal violence
Latin America and Caribbean			
1. Diahrrea	1. Disasters	1. IHD	1. IHD
2. LRI	2. IHD	2. Stroke	2. Stroke
3. Preterm birth complications	3. Violence	3. LRI	3. LRI
4. IHD	4. MDD	4. Diahrrea	4. Disaster
North Africa and Middle East			
1. Diahrrea	1. IHD	1. IHD	1. IHD
2. Congenital anomalies	2. MDD	2. Stroke	2. Stroke
3. IHD	3. Stroke	3. LRI	3. LRI
4. Preterm birth complications	4. Low back pain	4. Diahrrea	4. Road injury

Table 6.4. List of of conditions with zero novel therapies between 1990 and 2010

Main disease groups	diseases
Cardiovascular and circulatory diseases	Rheumatic heart disease, Endocarditis
Chronic respiratory diseases	Pneumoconiosis
Diabetes, urogenital, blood, and endocrine diseases	Genital prolapse, Polycystic ovarian syndrome, Acute glomerulonephritis, Chronic kidney disease due to hypertension, Chronic kidney disease due to diabetes mellitus, Uterine fibroids, Other hemoglobinopathies and hemolytic anemias, G6PD deficiency
Digestive diseases (except cirrhosis)	Paralytic ileus and intestinal obstruction without hernia, Appendicitis, Inguinal or femoral hernia
Mental and behavioral disorders	Idiopathic intellectual disability, Other drug use disorders
Musculoskeletal disorders Gout	Low back pain, Neck pain
Neglected tropical diseases and malaria	Food-borne trematodiases, Trichuriasis, Onchocerciasis, Trachoma, Ascariasis, Echinococcosis, Cysticercosis, Yellow fever, Hookworm disease
Neonatal disorders	Preterm birth complications, Other neonatal disorders
Neoplasms	Cancer of other part of pharynx and oropharynx, Mouth cancer, Nasopharynx cancer, Larynx cancer
Nutritional deficiencies	Other nutritional deficiencies, Vitamin A deficiency, Iodine deficiency, Protein-energy malnutrition
Other communicable, maternal, neonatal, and nutritional disorders	Sexually transmitted chlamydial diseases
Other n on-communicable diseases	Dental caries, Viral skin diseases, Sudden infant death syndrome, Fungal skin diseases, Congenital heart anomalies, Cellulitis, Edentulism, Neural tube defects, Abscess, impetigo, and other bacterial skin diseases, Scabies, Other sense organ diseases, Cleft lip and cleft palate, Down's syndrome, Refraction and accommodation disorders

Table 6.5. List of rare and genetic diseases with higher proportion of pharmaceutical R&D activity

Diseases	%R&D activity of total R&D in rare diseases
t-cell leukemia	0.019
scleroderma	0.019
crohn disease	0.020
allan-herndon-dudley syndrome	0.021
myelodysplastic syndrome	0.026
pulmonary fibrosis	0.026
autism	0.028
multiple sclerosis	0.028
barrett esophagus	0.035
intraepithelial neoplasia	0.035
hirsutism	0.036
psoriasis	0.036
melanoma	0.038
li-fraumeni syndrome	0.039
dengue fever	0.039
lupus nephritis	0.039
retinopathy	0.040
others	0.475

Table 6.6. Dominance tests I

	Total number of projects		Number of new therapies in the market	
	DALYs	Deaths	DALYs	Deaths
Communicable diseases				
1990 against 2010	non-dominance	non-dominance	non-dominance	non-dominance
1990 against 45-degree	non-dominance	non-dominance	non-dominance	non-dominance
2010 against 45-degree	non-dominance	non-dominance	non-dominance	non-dominance
Non-communicable diseases				
1990 against 2010	non-dominance	2010 dominates 1990	2010 dominates 1990	2010 dominates 1990
1990 against 45-degree	non-dominance	non-dominance	2010 dominates 1990	2010 dominates 1990
2010 against 45-degree	non-dominance	non-dominance	2010 dominates 1990	2010 dominates 1990

Note: results for intersection union principle (iup). Please check section 2.1 for more details.

Table 6.7. Dominance tests II

	Total number of projects		Number of new therapies in the market	
	DALYs	Deaths	DALYs	Deaths
Cardiovascular and circulatory disease				
1990 against 2010	non-dominance	non-dominance	non-dominance	non-dominance
1990 against 45-degree	45-degree	45-degree	45-degree	45-degree
2010 against 45-degree	45-degree	45-degree	45-degree	45-degree
Neoplasms				
1990 against 2010	non-dominance	non-dominance	non-dominance	non-dominance
1990 against 45-degree	45-degree	45-degree	45-degree	45-degree
2010 against 45-degree	45-degree	45-degree	45-degree	45-degree
Musculoskeletal disorders				
1990 against 2010	non-dominance	non-dominance	non-dominance	non-dominance
1990 against 45-degree	non-dominance	non-dominance	non-dominance	non-dominance
2010 against 45-degree	non-dominance	non-dominance	non-dominance	non-dominance
Mental and behavioral disorders				
1990 against 2010	non-dominance	non-dominance	non-dominance	non-dominance
1990 against 45-degree	non-dominance	non-dominance	non-dominance	non-dominance
2010 against 45-degree	non-dominance	non-dominance	non-dominance	non-dominance
Chronic Respiratory diseases				
1990 against 2010				
1990 against 45-degree				
2010 against 45-degree				
Diarr+LRI+OTH				
1990 against 2010	non-dominance	non-dominance	non-dominance	non-dominance
1990 against 45-degree	45-degree	45-degree	45-degree	45-degree
2010 against 45-degree	45-degree	45-degree	45-degree	45-degree
Neonatal disorders				
1990 against 2010	non-dominance	non-dominance	non-dominance	non-dominance
1990 against 45-degree	45-degree	45-degree	45-degree	45-degree
2010 against 45-degree	45-degree	45-degree	45-degree	45-degree
Nutritional deficiencies				

1990 against 2010				
1990 against 45-degree				
2010 against 45-degree				
Neglected tropical diseases + Malaria				
1990 against 2010	non-dominance	non-dominance	non-dominance	non-dominance
1990 against 45-degree	non-dominance	45-degree	non-dominance	45-degree
2010 against 45-degree	non-dominance	45-degree	non-dominance	45-degree

Note: results for intersection union principle (iup). Please check section 2.1 for more details.

Table 6.8. Concentration indices I

	Kawkani $C(r)$		Modified Wagstaff $\widehat{C}(r)$		Erreygers E	
	(1)	(2)	(1)	(2)	(1)	(2)
Communicable diseases						
DALYs 1990	0.221	0.230	0.229	0.235	0.006	0.003
DALYs 2010	0.219	0.228	0.227	0.233	0.006	0.003
Deaths 1990	0.213	0.224	0.220	0.228	0.006	0.003
Deaths 2010	0.213	0.224	0.220	0.228	0.006	0.003
$\bar{m}$ DALYs 1990	0.113	0.039	0.117	0.045	0.006	0.051
$\bar{m}$ DALYs 2010	0.107	0.037	0.111	0.042	0.006	0.052
$\bar{m}$ Deaths 1990	0.136	0.122	0.141	0.125	0.006	0.003
$\bar{m}$ Deaths 2010	0.125	0.112	0.129	0.115	0.006	0.003
Non-communicable diseases						
DALYs 1990	0.022	0.046	0.023	0.046	0.001	0.001
DALYs 2010	0.022	0.046	0.023	0.046	0.001	0.001
Deaths 1990	0.022	0.046	0.023	0.046	0.001	0.001
Deaths 2010	0.022	0.046	0.023	0.046	0.001	0.001
$\bar{m}$ DALYs 1990	-0.056	-0.055	-0.058	-0.056	0.001	0.001
$\bar{m}$ DALYs 2010	-0.051	-0.050	-0.053	-0.051	0.001	0.001
$\bar{m}$ Deaths 1990	-0.043	-0.042	-0.045	-0.043	0.001	0.001
$\bar{m}$ Deaths 2010	-0.034	-0.034	-0.035	-0.034	0.001	0.001

(1)Total number of pharmaceutical projects

(2) Number of new therapies in the market

Table 6.9. Concentration indices II

	Kawkani $C(r)$		Modified Wagstaff $\widehat{C}(r)$		Erreygers E	
	(1)	(2)	(1)	(2)	(1)	(2)
Cardiovascular and circulatory disease						
DALYs 1990	0.242	0.318	0.331	0.368	0.071	0.541
DALYS 2010	0.240	0.316	0.329	0.365	0.072	0.541
Deaths 1990	0.230	0.309	0.315	0.358	0.072	0.541
Deaths 2010	0.230	0.309	0.315	0.358	0.072	0.541
Neoplasms						
DALYs 1990	0.309	0.311	0.325	0.327	0.203	0.195
DALYS 2010	0.312	0.315	0.329	0.331	0.203	0.195
Deaths 1990	0.294	0.298	0.310	0.314	0.203	0.195
Deaths 2010	0.290	0.295	0.306	0.311	0.203	0.196
Musculoskeletal disorders						
DALYs 1990	0.215	0.320	0.467	0.433	0.035	0.954
DALYS 2010	0.215	0.320	0.467	0.433	0.035	0.954
Deaths 1990	0.342	0.467	0.744	0.633	0.014	0.943
Deaths 2010	0.342	0.467	0.744	0.633	0.014	0.943
Mental and behavioural disorders						
DALYs 1990	0.082	0.113	0.090	0.119	0.335	0.222
DALYS 2010	0.082	0.112	0.089	0.119	0.335	0.222
Deaths 1990	0.531	0.576	0.580	0.610	0.333	0.221
Deaths 2010	0.511	0.557	0.558	0.590	0.333	0.221
Chronic Respiratory diseases						
DALYs 1990	0.107	0.127	0.155	0.158	0.237	0.775
DALYS 2010	0.107	0.127	0.155	0.158	0.237	0.775
Deaths 1990	0.107	0.127	0.155	0.158	0.237	0.775
Deaths 2010	0.107	0.127	0.155	0.158	0.237	0.775
Diarr+LRI+OTH						
DALYs 1990	0.328	0.358	0.350	0.397	0.245	0.383
DALYS 2010	0.328	0.360	0.350	0.398	0.245	0.383
Deaths 1990	0.298	0.344	0.318	0.381	0.245	0.383

Deaths 2010	0.313	0.354	0.334	0.392	0.245	0.383
Neonatal disorders						
DALYs 1990	0.333	0.184	0.667	0.283	0.350	0.350
DALYS 2010	0.333	0.184	0.667	0.283	0.350	0.350
Deaths 1990	0.333	0.184	0.667	0.283	0.350	0.350
Deaths 2010	0.333	0.306	0.667	0.471	0.350	0.331
Nutritional deficiencies						
DALYs 1990						
DALYS 2010						
Deaths 1990						
Deaths 2010						
Neglected tropical diseases + Malaria						
DALYs 1990	0.054	0.092	0.066	0.099	0.566	0.301
DALYS 2010	0.089	0.136	0.109	0.148	0.565	0.301
Deaths 1990	0.149	0.193	0.181	0.209	0.564	0.300
Deaths 2010	0.193	0.241	0.236	0.261	0.563	0.300

(1)Total number of pharmaceutical projects

(2) Number of new therapies in the market

IX. Appendix: Figures

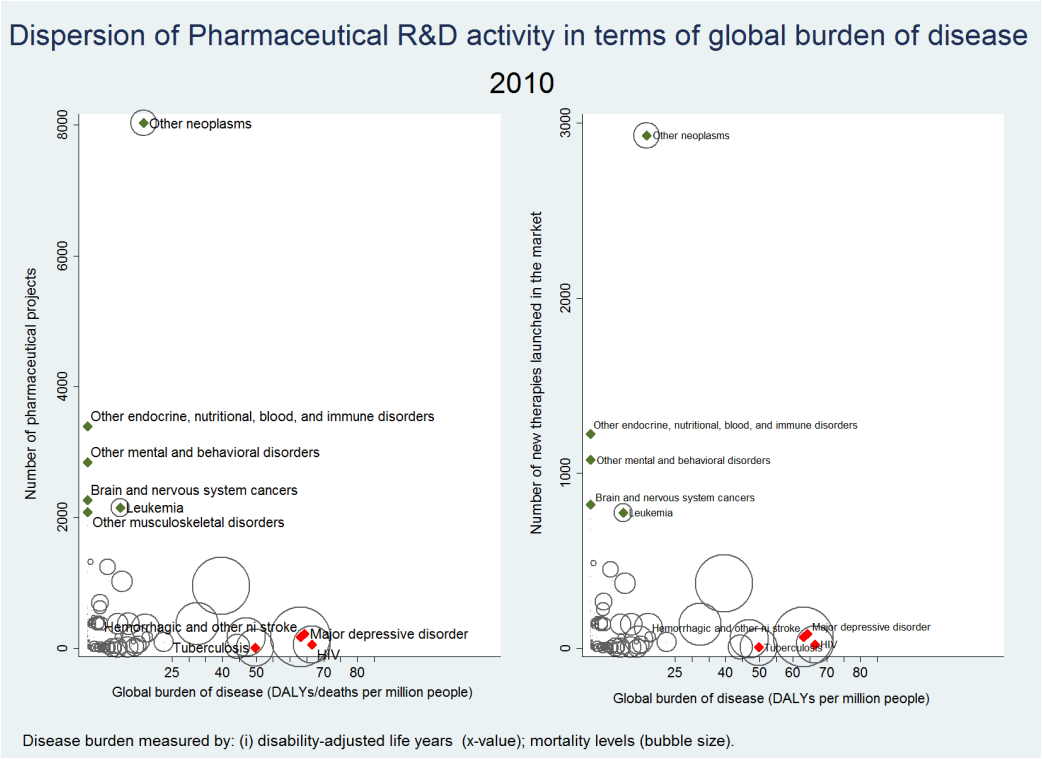


Figure 6.1. Dispersion of pharmaceutical R&D activity in terms of global disease burden

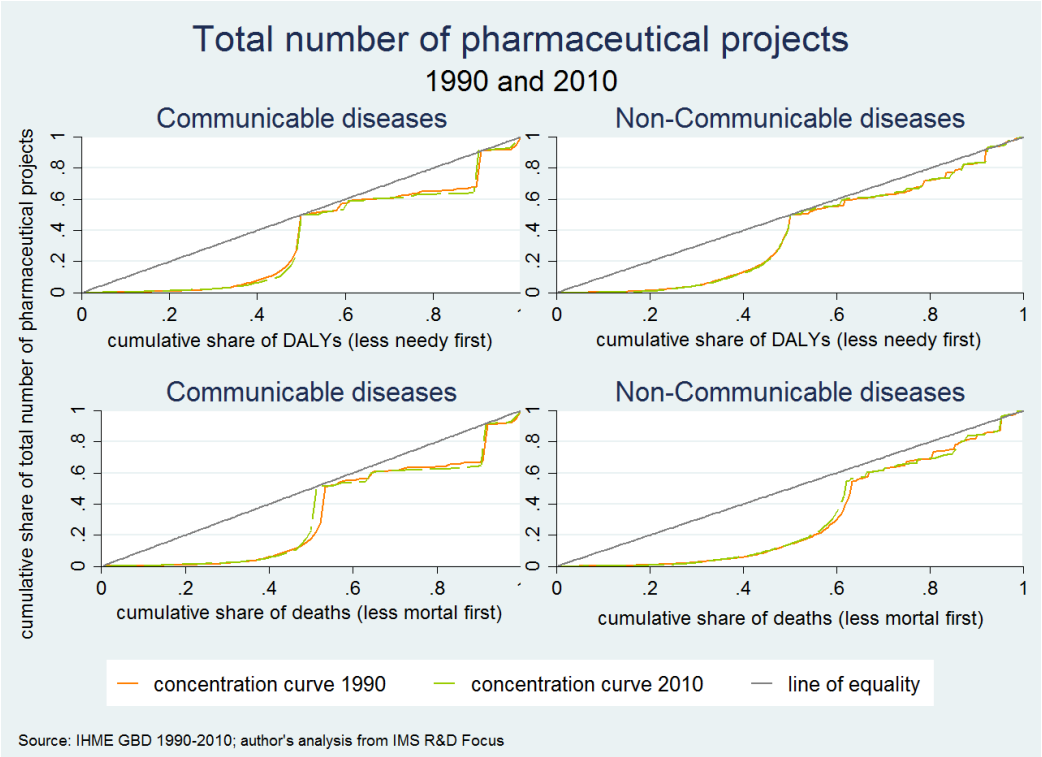


Figure 6.2. Total R&D activity and health need: communicable and non-communicable diseases

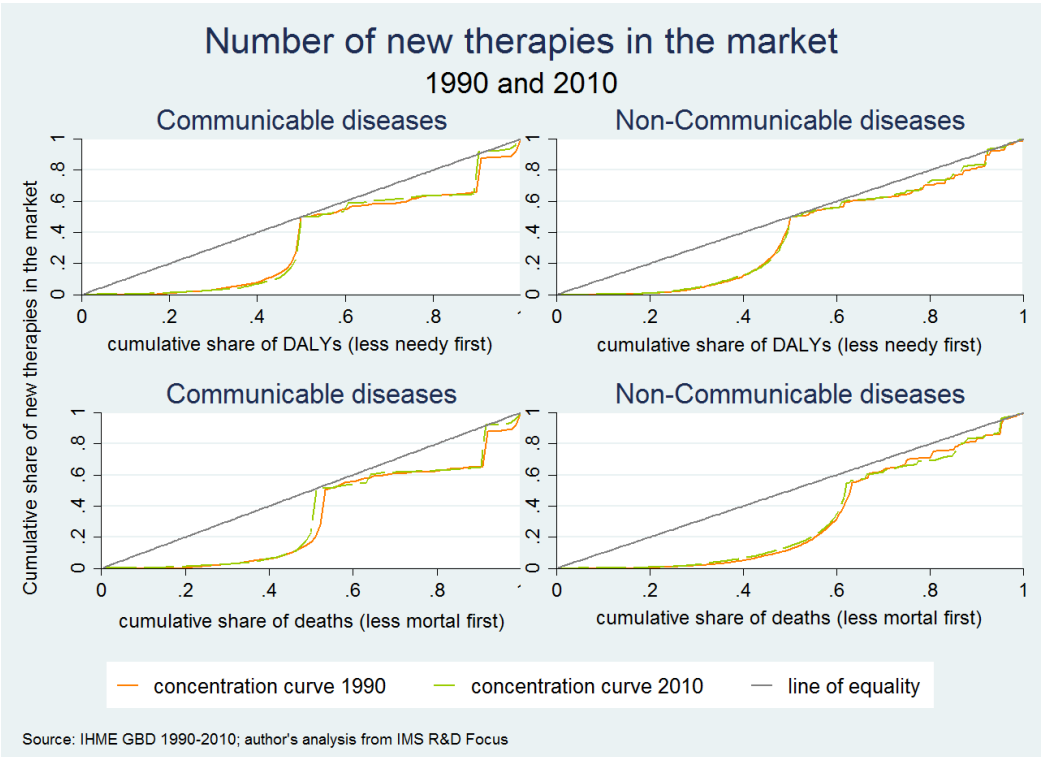


Figure 6.3. Market launches and health need: communicable and non-communicable diseases

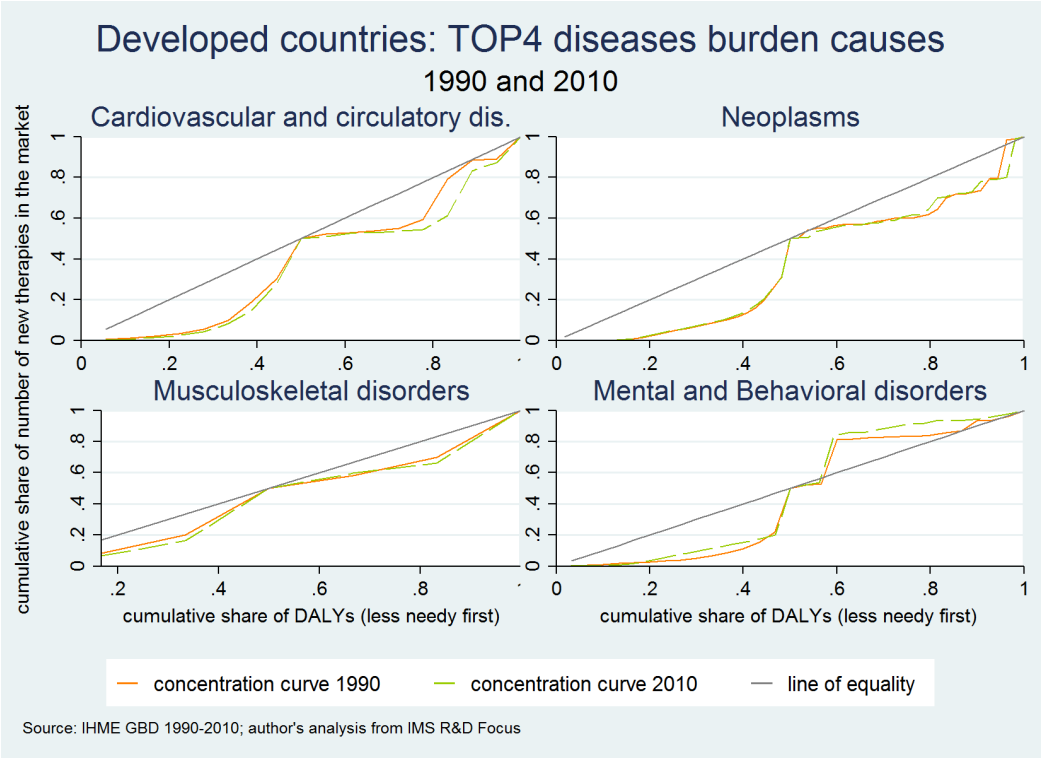


Figure 6.4. Developed countries TOP4 causes of disease burden and market launches

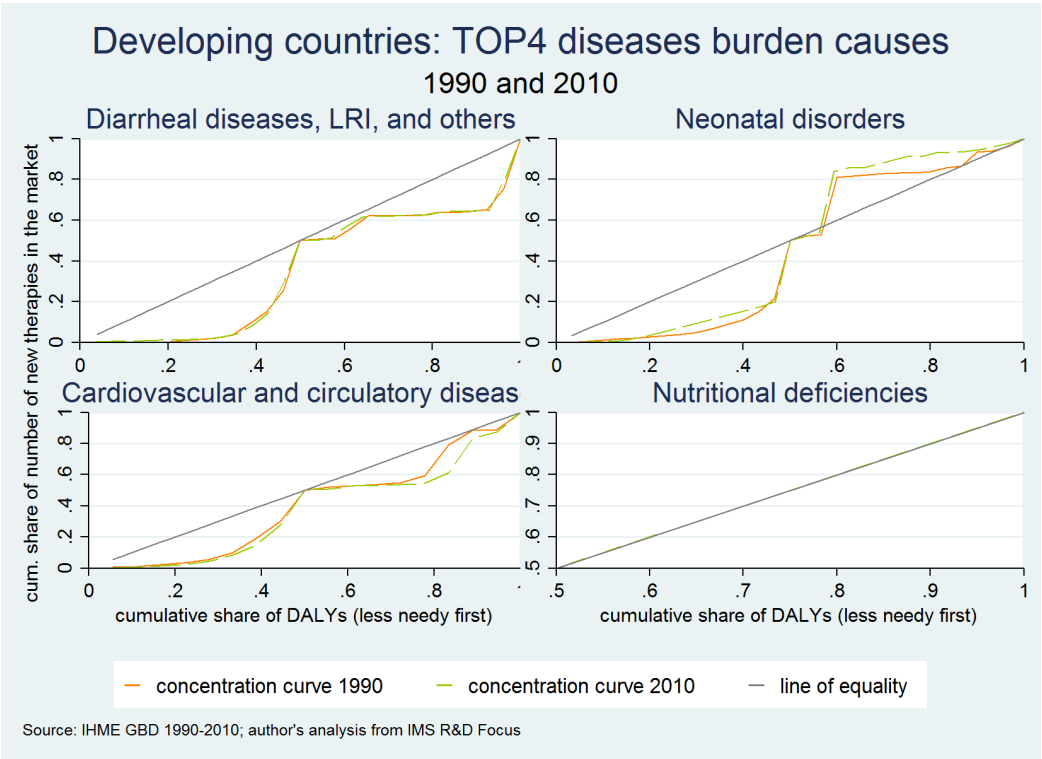


Figure 6.5. Developing countries TOP4 causes of disease burden and market launches

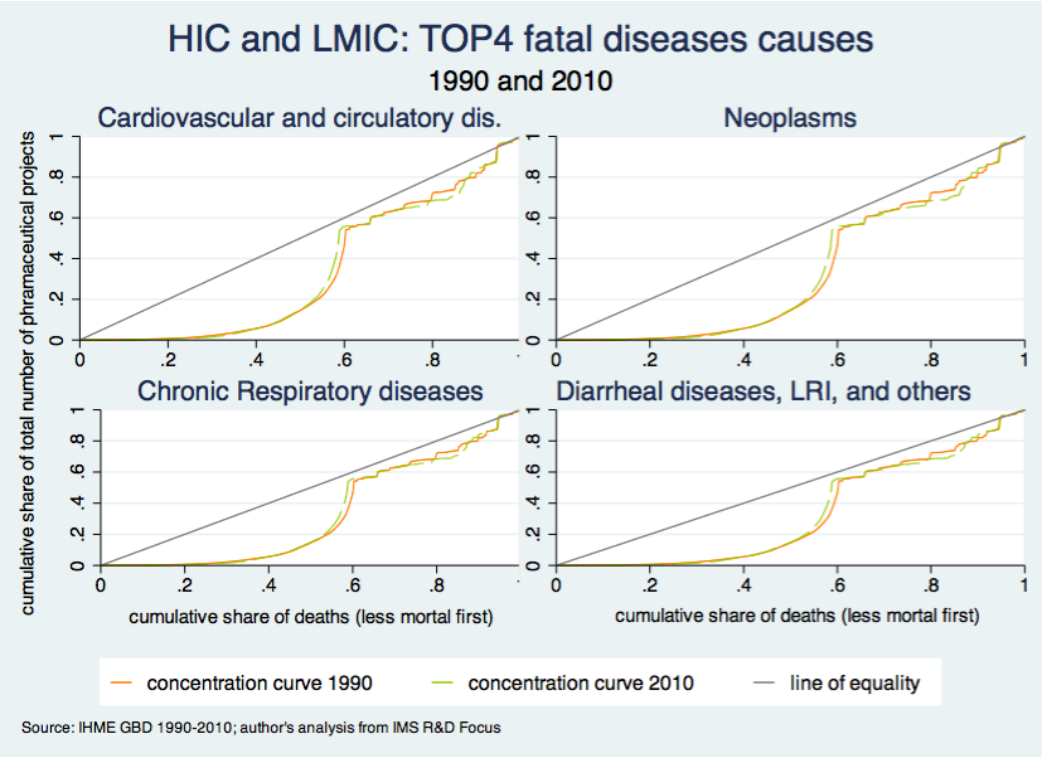


Figure 6.6. Developed and developing countries TOP4 causes of death and total R&D activity

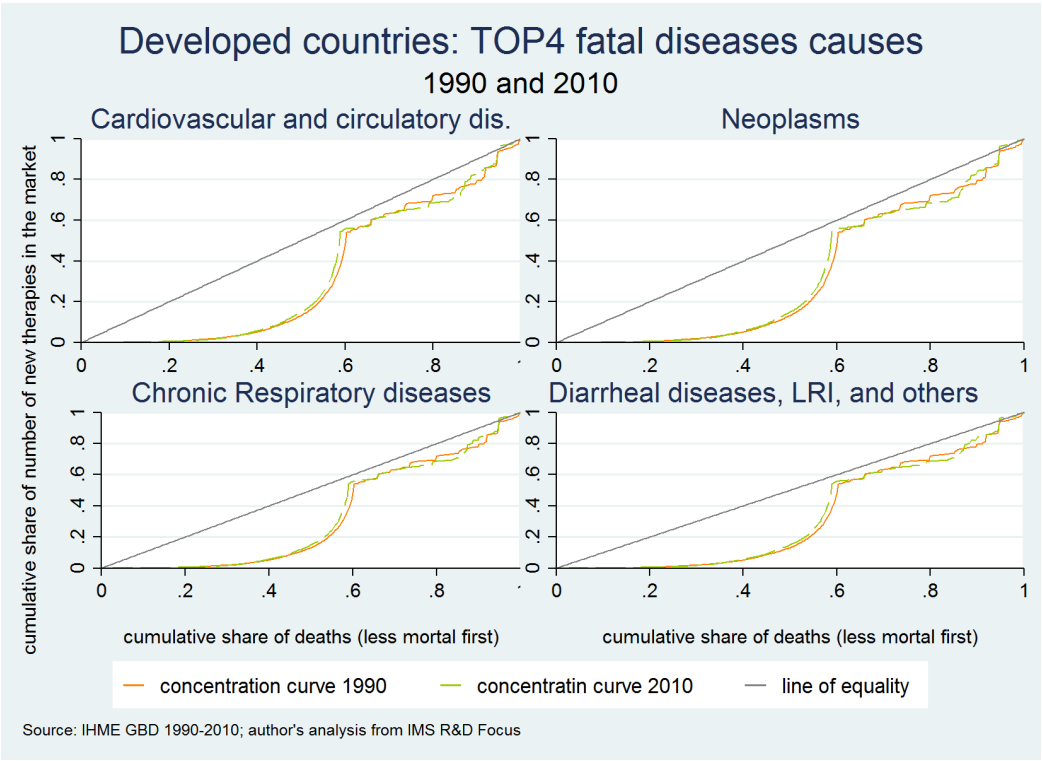


Figure 6.7. Developing and developed countries TOP4 causes of death and market launches

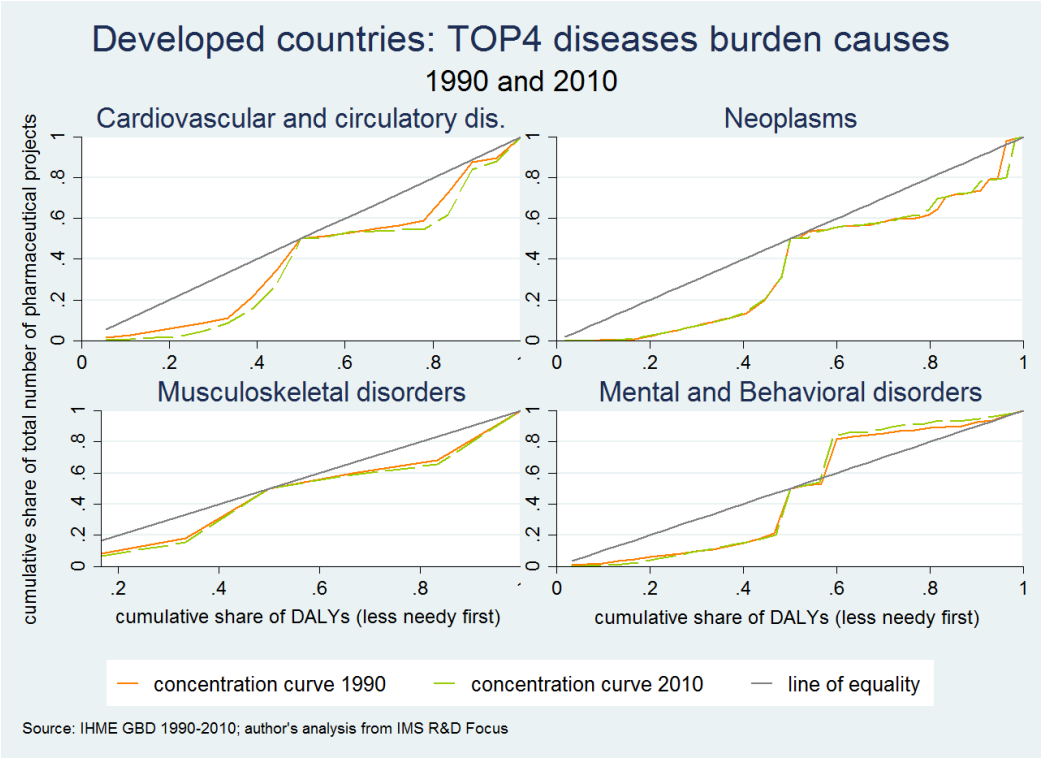


Figure 6.8. Developed countries TOP4 causes of disease burden and total R&D activity

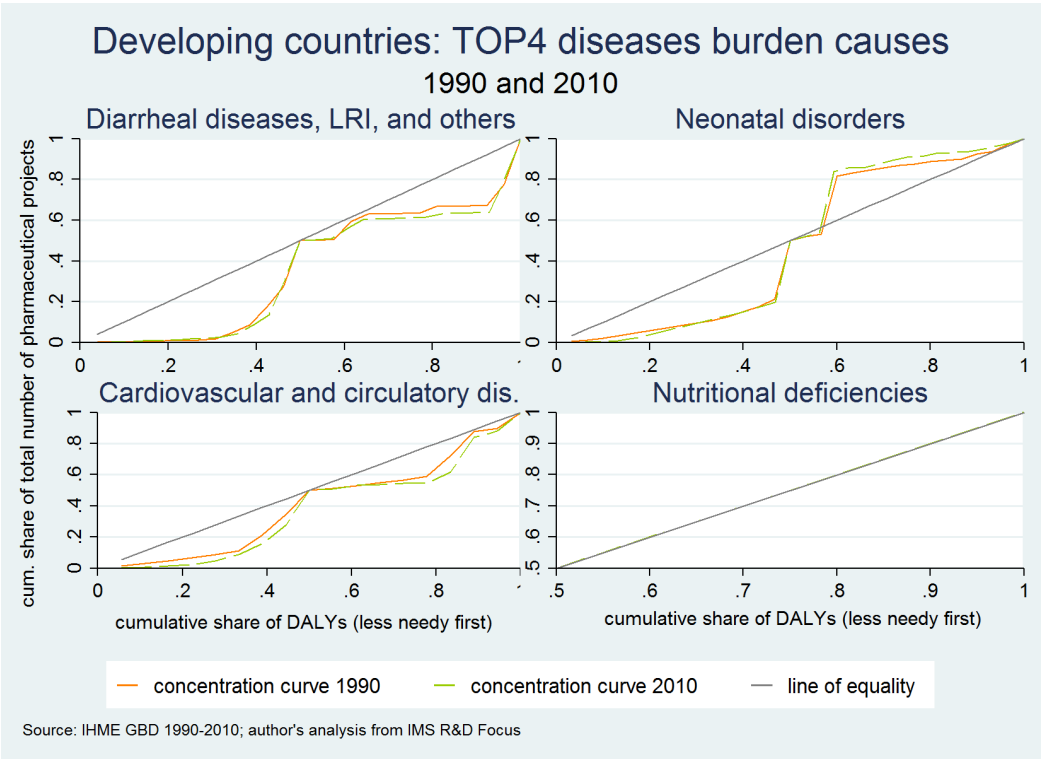


Figure 6.9. Developing countries TOP4 causes of disease burden and and total R&D activity

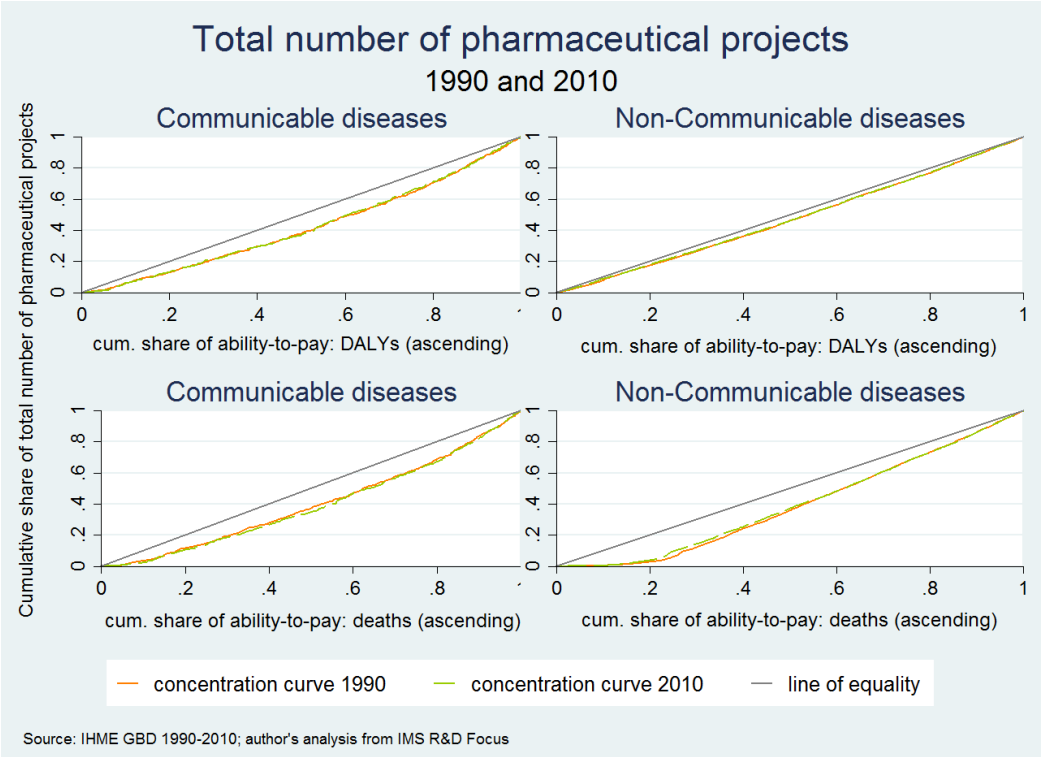


Figure 6.10. Total number of pharmaceutical projects and disease market attractiveness

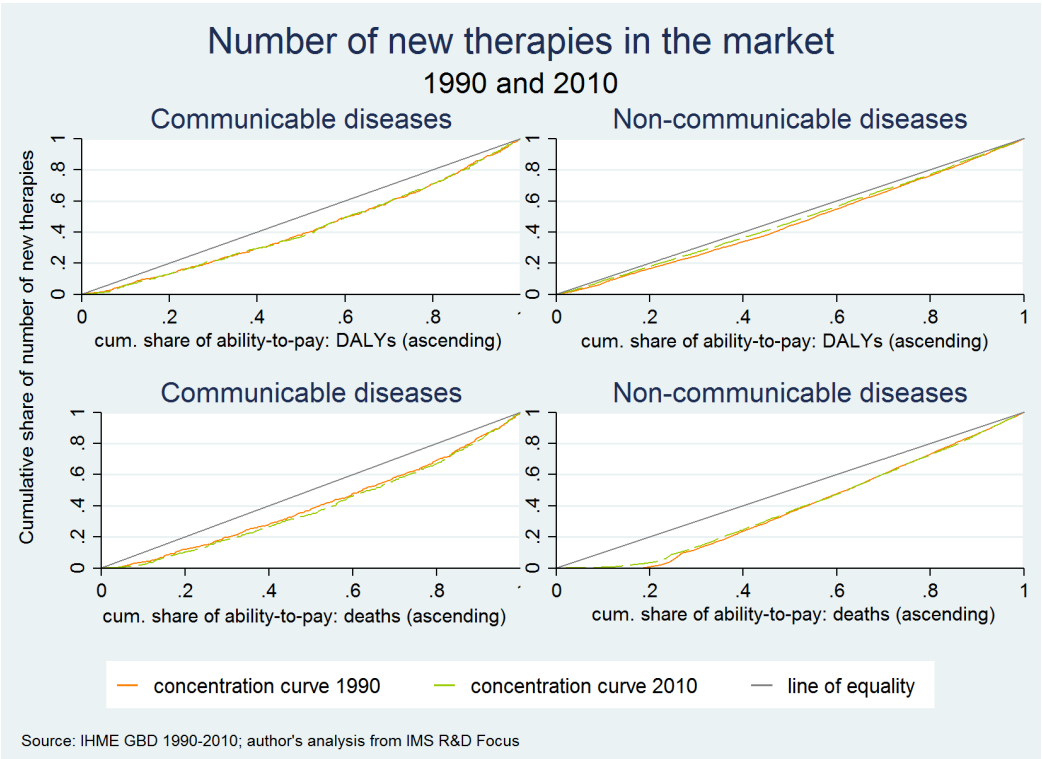


Figure 6.11. Number of new therapies and disease market attractiveness

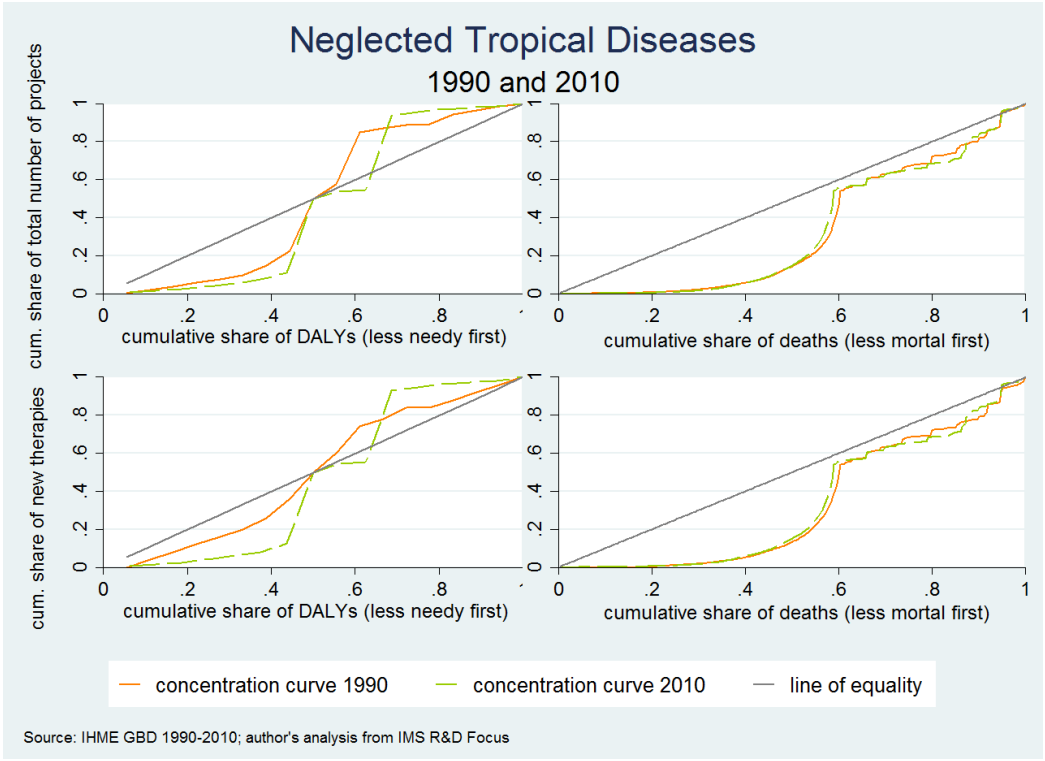


Figure 6.12. Neglected Tropical Diseases and health need

Chapter 7

Concluding remarks

In the last century, biomedical and pharmaceutical innovation have contributed considerably to improvements in population health through increased life expectancy and improved quality-of-life (Acemoglu and Johnson, 2006). These sustained effects have further contributed to improvements in living conditions and social welfare. Therefore, investment in drug innovation is crucial in shaping the distribution of future disease burden and promoting economic growth and wellbeing.

Despite these achievements, investment in R&D is still scant in many disease areas associated with high levels of morbidity and mortality (Kremer, 2002; Kremer and Glennerster, 2004; Pecoul et al., 1999; Trouiller et al., 2002). It is often said that part of the explanation for the lack of investment in these areas is the lack of appropriate incentive mechanisms aligning private sector interests with societal needs. Another explanation that has been put forward is the proliferation of policies across health systems that fail to reconcile short term cost containment with mechanisms that promote long term incentives for innovation (Morton and Kyle, 2012; OECD, 2014; Refoios Camejo et al., 2011). Health technology assessment is an example of an increasingly used approach to contain drug spending that potentially fails to provide the appropriate incentives for innovation to the pharmaceutical industry. If it fails to provide right incentives, health systems aiming to contain health care costs may unintentionally affect the investment in under-researched disease areas that could contribute to the reduction of global burden of disease.

At the same time, the pharmaceutical industry has argued that the increase in market competition, the complexity of scientific targets, the level of financial commitment required for sustaining a lengthy R&D process that potentially ends in failure, combined with market access regulation, makes it increasingly difficult to innovate in several disease areas (Dhankhar, 2012; PWC, 2011b).

In this context, it is therefore important to understand the determinants of pharmaceutical innovation and the value it brings to society.

This thesis contributed to the conceptual understanding and empirical analysis of the determinants of pharmaceutical innovation.

The conceptual contribution of this thesis is presented in Chapter 2, where we critically appraise the literature on the determinants of pharmaceutical innovation and constructed a framework overview of the factors associated with scientific discovery and development of new therapies. The review unveiled the gaps in the literature that need further research, and allowed the identification of factors to be considered in the empirical contributions in Chapters 4, 5 and 6.

The current literature is scarce in providing evidence at three levels that guided the analyses performed in Chapters 4, 5 and 6 respectively.

First, there is need to analyse the R&D process thoroughly when assessing the factors determining the innovation performance of the industry. Most of the studies do not consider the level of heterogeneity

of the R&D process or focus solely on the market launch stage of the innovation process. We gave further insight to this in Chapter 4 by modelling with robust methodologies the factors associated with project failure across the different stages of R&D.

Secondly, there is lack of a dynamic analysis of demand on firms' R&D decisions, taking into market competition and other supply-side conditions, and the organisation of healthcare systems and public policies. We developed in Chapter 5 a dynamic model that simulates the future R&D landscape by considering the impact of demand factors, industrial factors and firm-specific characteristics on R&D decisions under no further policy intervention. This can be further applied to evaluate the impact of policy changes on the future R&D landscape and availability of new therapies.

Thirdly, there is insufficient research on the welfare implications of current policies to the pharmaceutical innovation. In Chapter 6 we have assessed the inequalities of the access to new therapies and R&D activity, by measuring dispersion of global drug R&D activity and market launches in terms of global health need in the last two decades.

When outlining the results, and in particular in Chapter 4, we measured the association of competition and alliances between firms with the probability of failure in the different R&D phases as a function of the duration of R&D projects. This empirical contribution focuses on the analysis of the factors associated with project failure across the different stages of the R&D process, by estimating how R&D phase-specific failure rates correlate with competition and alliances between firms, taking into account the specificities of the product being developed and the relevant market. With the limitations driven by data constraints explored extensively in Chapters 3 and 4, there are three main results arising from this analysis. First, results suggest that the determinants of failure differ across the different stages of R&D. Second, competition is significantly associated with failure in discovery and Phase 2 in two ways. On one hand, the likelihood of failure declines in the number of potential competitors, and, on the other hand, it increases with fiercer competition from drugs established in the market for less than 5 years when projects are in Phase 2. Third, there is mixed association between alliances and the probability of failure. However, results are clear for the participation of at least one big firm in Phase 3 trials that is associated with a lower risk of failure of projects, and with the participation of academia in discovery that is associated with higher risk of failure.

In Chapter 5, we developed a dynamic micro-simulation model that forecasts the stock of new drug launches in the market in the next decades in the absence of any further policy intervention. This empirical contribution focuses on the prediction of drug availability in the market in the absence of further policy interventions and consists of the design of a policy evaluation tool that can be further explored in the analysis of other policy scenarios.

Using historical data to calibrate the microsimulation model and assuming a no-policy intervention scenario, there are three main results suggested by the analysis. First, that the regional disparities in terms of pharmaceutical innovation are going to persist. There is a relatively significant higher number of launches of NCEs in USA, event though the model specification and forecast show relatively similar R&D landscapes between USA and EMA given the non-significant differences between both regions in terms of number of new projects, duration of projects in pipeline, probabilities of market launch and failure. EMAP and Japan with relatively similar trends for the market launches for NCEs, similar average number of new projects being born, probabilities of market launch and duration of projects. Second, there are significant disparities at therapeutic level. For the market for NCEs, USA leads significantly the market launches in drugs targeting the nervous system and the musculoskeletal system. Moreover, EMAP is

predicted to have a large impact over the launch of antineoplastic and immunomodulating agents. Third, for the market for Biologics, USA, EMA and Japan lead the market launch for blood and blood forming organs, dermatological conditions, systematic hormonal preparations, respiratory systems and sensory organs, whereas EMAP and rest of the world show negligent innovation activity.

Finally in Chapter 6, we measured inequalities in global drug R&D activity in terms of population health needs. This chapter contributes to the first analysis on the equity implications of drug innovation in the last two decades. The results suggest that over the last twenty years, R&D activity has been concentrated in more commercially attractive disease areas associated with high mortality in the richest countries. The results show that the distribution of the R&D activity and disease burden have not changed significantly over the last two decades. This suggests that morbidity, as measured in terms of DALYs, seems not to influence firms' R&D decisions. Although this distribution of innovation effort may be considered efficient given the current income distribution, it is surely debatable from an equity point of view, and certainly socially undesirable. Moreover, analysis of Chapter 5 suggests that these inequalities will persist in the next decades if no policy intervention is implemented.

For the purpose of these analyses we have built a rich global panel dataset that uniquely links data on the lifecycle of R&D processes with global health data, and modelled the dynamic nature of R&D decisions using advanced statistical methods. The contribution of the construction of this dataset is explained in detail in Chapter 3.

Overall, this thesis focused on the analysis of the factors affecting the failure of pharmaceutical R&D activity, taking a global perspective on health needs and the decisions of firms, and the implications for the future R&D landscape and the access to new therapies for populations. Three key messages emerge from this thesis that are discussed in the following section.

I. Policy implications and value for decision makers

The first key message relates to the efficiency concerns of the current global pharmaceutical R&D model, and the potential consequences for the future distribution of health gains and socioeconomic benefits across populations (Munos, 2009; Paul et al., 2010; Morton and Kyle, 2012; Refoios Camejo et al., 2011). The vast majority of innovation in new therapies relies on market incentives to invest in long and uncertain projects that present a high likelihood of failure. The lack of dynamic incentive mechanisms to promote the optimal rate and level of innovation can lead to inefficiencies that might generate distortions in the potential health gains and welfare distribution (Vallerie Paris, 2013; OECD, 2014; Morton and Kyle, 2012). Furthermore, from a societal perspective, we might be paying a high price for innovations to compensate for the large number of failures that the industry faces. The availability of value-for-money therapies might be compromised and depends, to a large extent, on efficiency gains in the R&D process.

Therefore, understanding the causes of failure is crucial to designing optimal incentive mechanisms that result in efficiency gains, while promoting innovation in areas of need. This insight is supported by the results on the role of market and innovation competition, and collaboration between innovators presented in Chapter 4. Market competition may exert a harmful pressure that can intensify the level of failure in innovation if too many young drugs are competing in the market. This suggests that policies regulating market entry potentially affect the future incentives for firms to innovate. Evaluations of these policies should take the costs related with the potential future sub-optimal levels of innovation into consideration.

On the other hand, the negligible effect of competition by incumbents on failure questions the adequacy

of current reimbursement policies and health technology assessment tools in establishing the relevant drug comparators for the cost-effectiveness analysis of new drugs. This means that, not only clinical and efficacy criteria are important to consider in the design of these policies, but also the novelty and market exclusivity of the products that are designated as the benchmark for the cost-effectiveness analyses. On the other hand, there is scope to reduce failure levels by promoting more competition between innovators in the R&D pipeline. Indeed, our results suggest that within therapeutic areas that exhibit substantially higher levels of failure, the promotion of multiple investments may generate significant positive spillover effects across innovators.

The second main insight that emerges from this thesis regards the equity concerns posed by the current global pharmaceutical R&D model. In the absence of optimal incentive mechanisms targeted at promoting innovation in riskier or less profitable disease areas, the current model has failed to deliver innovations in therapeutic areas with potentially large economic impact (Kremer, 2002; Kremer and Glennerster, 2004; Morton and Kyle, 2012). Results from Chapter 6 suggest a mismatch between unmet global health needs and the pharmaceutical R&D strategies. Over the last twenty years, R&D activity has been concentrated in more commercially attractive disease areas associated with high mortality in the richest countries. The results show that the distribution of the R&D activity and disease burden have not changed significantly over the last two decades. This suggests that morbidity, as measured in terms of DALYs, seems not to influence firms' R&D decisions. Although this distribution of innovation effort may be considered efficient given the current income distribution, it is surely debatable from an equity point of view, and certainly socially undesirable.

Results from Chapter 5 suggest that such equity concerns are likely to persist unless new policy interventions to address inequalities in R&D and access to new therapies are introduced. Indeed, our results suggest that historical trends will persist with the USA, followed by the EMA region, leading the global market launches of new chemical and biologic therapies, and driving the global innovation in new therapies targeting the nervous system of both chemical and biological composition, and chemicals targeting musculoskeletal conditions.

The third key message relates to challenges posed by the expected increasing role of emerging markets on the innovation process and launch of new therapies in areas traditionally led by the established economies (IMS, 2010; Danzon et al., 2013). Results from Chapter 5 suggest that emerging economies are expected to launch a promising number of new anticancer drugs in the market, as well as other products such as vaccines. The increasing role of emerging markets can potentially affect the global R&D investment and the policy environment. The alleged "free riding" on innovation and access to drugs from less developed countries has often been of concern to the most innovative countries (Danzon, 2011). Differentials in national legislations at the level of IP rights, price regulation and safety and clinical surveillance have caused a number of arbitrage opportunities (Danzon, 2011; Phill et al., 2012; OECD, 2014). Because of these opportunities, firms and regulators in the US and Europe have questioned the sustainability of such global differences in legislation given the supposedly relatively higher innovation costs and drug prices imposed by stricter regulation policies in these countries.

Economic analysis has provided insight on some of these issues, but many fundamental questions remain. Further research is needed to assess the effects of these arbitrage opportunities on R&D decisions and innovation. Indeed, there is a need for further research on how to address parallel trade of drugs while maintaining the incentives to innovate. Furthermore, more research is required on how to design optimal IP rights that consider these arbitrage opportunities and promote innovation in drugs that target

the less developed world.

The methodological contribution of this thesis has been to show how duration models and dynamic micro-simulation models can inform policy design to promote innovation with great impact on population health and social welfare. The application of these methods should not be seen as an end in itself but a starting point that provides a rich domain for further research towards understanding the flaws of the current pharmaceutical R&D process and the opportunities for optimally designing incentive mechanisms for drug innovation that can promote population health and welfare benefits.

Our results also highlight some other potential areas for further research.

The first is to examine the impact of organisational models of the innovation process in the pharmaceutical industry, in terms of efficiency and equity.

The waste of resources that failure causes may, in fact, be caused by the way markets strategically operate. Failure of innovation may be due to a combination of regulatory pressures, scientific/clinical non-achievements, or strategic decisions of firms to withdraw projects. Firms' strategic decisions may have significant efficiency and equity implications. Indeed as shown in Chapter 4, as competition from peers in the R&D process is associated with lower levels of failure, firms might be, not only, benefiting from positive externalities, but also cooperating strategically. This strategic cooperation could be related with firms wanting differentiate themselves in order to focus on areas with reduced competition in the market. In future research, this analysis can be extended to explore to what extent firms adopt differentiation strategies at target level to reduce competition, and the impact of such strategies on availability of new therapies and social welfare.

Furthermore, it is increasingly pertinent to examine the impact of organisational models of R&D initiatives that have been targeting low-income country diseases. A number of R&D initiatives by not-for-profit organisations and the industry have been encouraging innovation in disease areas that are not economically attractive for the industry but for which there is large health un-met need. However, these types of initiatives tend to be financially dependent on public and philanthropic sponsorship, and are often criticised for, on one hand, not being sufficiently dynamic in interacting with the private sector, because of difficulties in sharing knowledge between the different stakeholders and, on the other, not being sufficiently efficient in delivering large scale technologies (Siegel et al., 2003; Perkmann and Walsh, 2007). A crucial step in promoting the availability of drugs in such disease areas involves measuring the performance of these organisational models within the R&D process. It is natural to extend the analysis from Chapter 6 to examine how the incentives of different organisational models determine the decision to invest in those areas and the performance throughout the innovation process. Another potential research avenue is the assessment of the performance of different organisational models when not-for-profit institutions are leading the R&D process. Policy making would indeed benefit from further insight in the optimality of organisational models for innovation that promote efficiency and equity in health systems.

The second future research avenue draws from the previous point and relates to the need of optimal policy design in disease areas that generate large public health costs and social welfare losses across the globe. This is urgent in many disease areas, where the market sub-optimally delivers innovation. The market has shown difficulties particularly in innovation and development of new vaccines and drugs to address drug resistant infections.

Vaccination is traditionally an area where firms decide not to invest given its low profitability (Trouiller et al., 2002; Acemoglu and Linn, 2004). Even when considering the prevention and the therapeutic use of vaccines, the sales volume is low in comparison to other medicines because people are vaccinated only

once, while, for example, patients with chronic conditions are frequently dependent on using of drugs for the rest of their lives. Also, profit margins are lower because governments, being larger buyers of prevention often exert their countervailing power in negotiating low prices in bulk acquisitions. This results in a lack of R&D investment in new vaccines by pharmaceutical companies (Kremer and Snyder, 2003).

Also, in the area of antimicrobials, the industry has shown difficulties in innovating by developing new drugs to address drug resistant infections. In the past, companies have relied on a standard *blockbluster* business model to generate mass sales volumes which has led to a large use and demise of existing antibiotics (Smith and Coast, 2002). Partly due to over-utilization of these treatments, antimicrobials are increasingly less effective leading to antimicrobial resistance (Smith and Coast, 2002; Rosamond and Allsop, 2000). However, to some extent, the incentives to invest in new antibiotics are relying on large scale repeated prescription of these drugs, while recent governmental initiatives have restricted the use of antibiotics. These restrictions thus pose a challenge to the profitability of investment in new antimicrobials (ECDC, 2013; DoH, 2012). Therefore, the design of new organisational models that accommodate the particular market conditions of these areas need further theoretical and empirical insight.

The third future area of research relates to the necessity of policy evaluations that consider the simultaneous interactions between the different agents that play a role in the drug innovation process. These agents include firms, payers and providers of healthcare, regulators, and patients. The current literature lacks policy evaluation at the pharmaceutical innovation level that considers the dynamic analysis of demand and supply of R&D decisions, and their interaction with third-parties intervening in the R&D process. The level of regulation in this industry provides an abundance of examples. However, given the frequent use of clinical guidelines and reimbursement policies in Europe, Australia and, increasingly in the United States (Vallerie Paris, 2013; Morton and Kyle, 2012), the analysis of the impact of these policies on drug R&D decisions is an area with a particular need for further research (Claxton, 2007; Danzon and Towse, 2003). The analysis in Chapter 5 could be extended in order to accommodate these objectives. This could be done by incorporating policy options and evaluation in the micro-simulation model, while considering the simultaneous interactions between demand and supply for drug innovation, and unpredictable shocks originated by changes in the industry and economy.

Bibliography

Abrantes-Metz, R., Adams, C., Metz, A., 2004. Pharmaceutical development phases: A duration analysis. Tech. rep., US Federal Trade Commission, Washington DC.

Acemoglu, D., Johnson, S., 2006. Disease and development: the effect of life expectancy on economic growth. Tech. rep., National Bureau of Economic Research.

Acemoglu, D., Linn, J., 2004. Market size in innovation: theory and evidence from the pharmaceutical industry. The Quarterly Journal of Economics 119 (3), 1049–1090.

Adams, C. P., Brantner, V. V., 2006. Estimating the cost of new drug development. Health Affairs 25 (2), 420–428.

Adams, C. P., Brantner, V. V., 2010. Spending on new drug development. Health economics 19 (2), 130–141.

Aghion, P., Bloom, N., Blundell, R., Griffith, R., Howitt, P., 2005. Competition and innovation: An inverted u relationship. Quarterly Journal of Economics 120 (2), 701–728.

Aghion, P., Howitt, P., 1992. A model of growth through creative destruction. Econometrica.

Aharonson, B. S., Baum, J. A. C., Feldman, M. P., 2007. Desperately seeking spillovers? increasing returns, industrial organization and the location of new entrants in geographic and technological space. Industrial and Corporate Change 16 (1), 89–130.

Akaike, H., 1974. A new look at the statistical model identification. Automatic Control, IEEE Transactions on 19 (6), 716–723.

Alexander, D., Flynn, J., Linkins, L., 1995. Innovation, r&d productivity, and global market share in the pharmaceutical industry. Review of Industrial Organization 10 (2), 197–207.

Allison, P. D., 1984. Event history analysis: Regression for longitudinal event data. SAGE Publications, Incorporated.

Arrowsmith, J., Miller, P., 2013. Trial watch: Phase ii and phase iii attrition rates 2011-2012. Nature Reviews Drug Discovery 12 (8), 569–569.

Arundel, A., Kabla, I., 1998. What percentage of innovations are patented? empirical estimates for european firms. Research Policy 27 (2), 127–141.

Backhaus, J. G., 2012. Claudia desogus: Competition and innovation in the eu regulation of pharmaceuticals, the case of parallel trade. European Journal of Law and Economics 34 (2), 423–423.

Baily, M. N., 1972. Research and development costs and returns: the us pharmaceutical industry. The Journal of Political Economy, 70–85.

Barker, C., Green, A., 1996. Opening the debate on dalys. Health Policy and Planning 11 (2), 179–183.

Barro, R. J., Sala-i Martin, X., 1992. Public finance in models of economic growth. The Review of Economic Studies 59 (4), 645–661.

Barzel, Y., 1968. Optimal timing of innovations. The Review of Economics and Statistics, 348–355.

BCG, 2008. The future of pharma: Adjusting the pharma r&d model. Tech. rep., Boston Consulting Group.

Berndt, E., Bui, L., Lucking-Reiley, D., Urban, G., 1996. The roles of marketing, product quality, and price competition in the growth and composition of the us antiulcer drug industry.

Berndt, E., Cockburn, I., Thiers, F., 2007. The globalization of clinical trials for new medicines into emerging economies: Where are they going and why, paper to be presented at the UNU-MERIT Conference.

Berndt, E. R., 2001. The us pharmaceutical industry: why major growth in times of cost containment? Health Affairs 20 (2), 100–114.

Berndt, E. R., 2002. Pharmaceuticals in us health care: determinants of quantity and price. Journal of Economic Perspectives, 45–66.

Bhattacharya, J., Packalen, M., 2011. Opportunities and benefits as determinants of the direction of scientific research. Journal of Health Economics 30 (4), 603–615.

Boasson, V., Boasson, E., MacPherson, A., Shin, H.-H., 2005. Firm value and geographic competitive advantage: evidence from the us pharmaceutical industry. Journal of Business 78 (6).

Bouvier, M., 2007. Will academic research rescue the drug discovery crisis? Medecine sciences: M/S 23 (3), 227–228.

Bresnahan, T. F., Reiss, P. C., 1990. Entry in monopoly market. The Review of Economic Studies 57 (4), 531–553.

Bresnahan, T. F., Reiss, P. C., 1991. Entry and competition in concentrated markets. Journal of Political Economy, 977–1009.

Bresnahan, T. F., Reiss, P. C., Willig, R., Stigler, G. J., 1987. Do entry conditions vary across markets? Brookings Papers on Economic Activity, 833–881.

Briggs, A. H., Claxton, K., Sculpher, M. J., 2006. Decision modelling for health economic evaluation. Oxford University Press.

Cameron, A., Trivedi, P., 2005. Microeconometrics: methods and applications. Cambridge Univ Press.

Carpenter, D., Turenne, M., 2001. Why do bigger firms receive faster drug approvals?, presentation at the Harvard-MIT Workshop on Positive Political Economy.

Carpenter, D. P., 2004. The political economy of fda drug review: processing, politics, and lessons for policy. Health Affairs 23 (1), 52–63.

Catalá-López, F., García-Altés, A., Álvarez-Martín, E., Gènova-Maleras, R., Morant-Ginestar, C., Parada, A., 2011. Burden of disease and economic evaluation of healthcare interventions: are we investigating what really matters? BMC health services research 11 (1), 75.

Cicciò, L., 2004. Inequities in medical research and development: The case of neglected diseases. Health Policy and Development 2 (3).

Civan, A., Maloney, M. T., 2006. The determinants of pharmaceutical research and development investments. Contributions in Economic Analysis & Policy 5 (1).

Clarke, P. M., Gerdtham, U.-G., Johannesson, M., Bingefors, K., Smith, L., 2002. On the measurement of relative and absolute income-related health inequality. Social Science & Medicine 55 (11), 1923–1928.

Claxton, K., 2007. Oft, vbp: Qed? Health Economics 16 (6), 545–558.

Claxton, K., Briggs, A., Buxton, M. J., Culyer, A. J., McCabe, C., Walker, S., Sculpher, M. J., 2008. Value based pricing for nhs drugs: an opportunity not to be missed? BMJ: British Medical Journal 336 (7638), 251.

Claxton, K., Culyer, A. J., 2007. Rights, responsibilities and nice: a rejoinder to harris. Journal of Medical Ethics 33 (8), 462–464.

Cleves, M., William, G., Gutierrez, R. G., Marchenko, Y. V., 2010. An Introduction to Survival Analysis using Stata, third edition Edition. StatCorp LP, Texas.

ClinicalTrials.gov, February 2015.
URL `https://clinicaltrials.gov/`

Cockburn, I., Henderson, R., 1994. Racing to invest? the dynamics of competition in ethical drug discovery. Journal of Economics & Management Strategy 3 (3), 481–519.

Cockburn, I. M., 2007. Is the pharmaceutical industry in a productivity crisis? In: Innovation Policy and the Economy, Volume 7. MIT Press, pp. 1–32.

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.

Cohen, J., Dibner, M. S., Wilson, A., 2010. Development of and access to products for neglected diseases. PLoS One 5 (5), e10610.

Cohen, W. M., Levinthal, D. A., 1989. Innovation and learning: the two faces of r & d. The economic journal, 569–596.

Comanor, W. S., Scherer, F. M., 2013. Mergers and innovation in the pharmaceutical industry. Journal of Health Economics 32 (1), 106–113.

Culyer, A. J., Wagstaff, A., 1993. Equity and equality in health and health care. Journal of Health Economics 12 (4), 431–457.

Danzon, P., 2011. The Economics of Biopharmaceutical Industry. Oxford University Press, Oxford, Ch. 22, pp. 520–554.

Danzon, P. M., Chao, L.-W., 2000. Does regulation drive out competition in pharmaceutical markets?*. The Journal of Law and Economics 43 (2), 311–358.

Danzon, P. M., Epstein, A., Nicholson, S., 2007. Mergers and acquisitions in the pharmaceutical and biotech industries. Managerial and Decision Economics 28 (4-5), 307–328.

Danzon, P. M., Epstein, A. J., 2012. Effects of regulation on drug launch and pricing in interdependent markets. Vol. 23. Emerald Group Publishing Limited.

Danzon, P. M., Mulcahy, A. W., Towse, A. K., 2013. Pharmaceutical pricing in emerging markets: Effects of income, competition, and procurement. Health Economics.

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

Danzon, P. M., Pereira, N. S., Tejwani, S. S., 2005b. Vaccine supply: a cross-national perspective. Health Affairs 24 (3), 706–717.

Danzon, P. M., Towse, A., 2003. Differential pricing for pharmaceuticals: reconciling access, r&d and patents. International Journal of Health Care Finance and Economics 3 (3), 183–205.

Danzon, P. M., Wang, Y. R., Wang, L., 2005c. The impact of price regulation on the launch delay of new drugsevidence from twenty-five major markets in the 1990s. Health Economics 14 (3), 269–292.

Dasgupta, P., Stiglitz, J., 1980. Uncertainty, industrial structure, and the speed of r&d. The Bell Journal of Economics, 1–28.

Dhankhar, Ahay; Matthias Evers, M. M., 2012. Escaping the sword of damocles: Towards a new future for pharmaceutical r&d. Tech. rep., McKinsey Co.

DiMasi, J., Hansen, R., Grabowski, H., 2003. The price of innovation: new estimates of drug development costs. Journal of Health Economics 22 (2), 151–185.

DiMasi, J. A., 2002. The value of improving the productivity of the drug development process: Faster times and better decisions. Pharmacoeconomics 20 (15), 1–10.

DiMasi, J. A., Feldman, L., Seckler, A., Wilson, A., 2010. Trends in risks associated with new drug development: success rates for investigational drugs. Clinical Pharmacology Therapeutics 87 (3), 272–277.

DiMasi, J. A., Grabowski, H. G., 2007. The cost of biopharmaceutical rd: is biotech different? Managerial and Decision Economics 28 (45), 469–479.

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.

Dingermann, T., Windisch, J., Cornes, P., 2012. Biosimilar monoclonal antibodies: the next frontier. a summary report from a symposium held at eahp 2012. European Journal of Hospital Pharmacy: Science and Practice 19 (3), 315–317.

Dixit, A. K., 1994. Investment under uncertainty. Princeton University Press.

Djankov, S., La Porta, R., Lopez-de Silanes, F., Shleifer, A., 2002. The regulation of entry. The Quarterly Journal of Economics 117 (1), 1–37.

DoH, 2012. Uk five year antimicrobial resistance strategy 2013 to 2018. Tech. rep., Department of Health, London.

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

Drummond, M. F., Mason, A. R., 2007. European perspective on the costs and cost-effectiveness of cancer therapies. Journal of Clinical Oncology 25 (2), 191–195.

Drummond, M. F., Sculpher, M., Torrance, G., OBrien, B., Stoddart, G., 2005. Methods for the economic evaluation of health care programmes. Oxford University Press, New York.

Drummond, M. F., Wilson, D. A., Kanavos, P., Ubel, P., Rovira, J., 2007. Assessing the economic challenges posed by orphan drugs. International Journal of Technology Assessment in Health Care 23 (01), 36–42.

Dubois, P., de Mouzon, O., Scott Morton, F., Seabright, P., 2011. Market size and pharmaceutical innovation. TSE Working Paper 11.

Dukes, M., 2002. Accountability of the pharmaceutical industry. The Lancet 360 (9346), 1682–1684.

ECDC, 2013. Antimicrobial resistance antimicrobial resistance surveillance in europe.

EMA, February 2015. Eu clinical trials register.
URL https://www.clinicaltrialsregister.eu/ctr-search/search;jsessionid=NfN_nL-fc6H1v1_ruOBYWlbrcS-oVFvmcfTZWAaZQ7UYEByeObs6!-2078813522

EphMRA, 2013. Ephmra anatomical classification.

Erreygers, G., 2009. Correcting the concentration index. Journal of Health Economics 28 (2), 504–515.

Erreygers, G., Van Ourti, T., 2011. Measuring socioeconomic inequality in health, health care and health financing by means of rank-dependent indices: a recipe for good practice. Journal of Health Economics 30 (4), 685–694.

Evans, D. B., Etienne, C., 2010. Health systems financing and the path to universal coverage. Bulletin of the World Health Organization 88 (6), 402–403.

EY, 2010. Progressions: Pharma 3.0.

FDA, February 2015a. Food and drug administration amendments act (fdaaa) of 2007.
URL http://www.fda.gov/regulatoryinformation/legislation/federalfooddrugandcosmeticactfdcact/significantamendmentstothefdcact/foodanddrugadministrationamendmentsactof2007/default.htm

FDA, February 2015b. Regulatory information.
URL http://www.fda.gov/RegulatoryInformation/Legislation/FederalFoodDrugandCosmeticActFDCAct/SignificantAmendmentstotheFDCAct/FDAMA/default.htm

Finkelstein, S., Temin, P., 2008. Reasonable Rx: solving the drug price crisis. FT Press.

Fisk, N. M., McKee, M., Atun, R., 2011. Relative and absolute addressability of global disease burden in maternal and perinatal health by investment in r&d. Tropical Medicine & International Health 16 (6), 662–668.

Fleurbaey, M., Schokkaert, E., 2009. Unfair inequalities in health and health care. Journal of Health Economics 28 (1), 73–90.

Fortune, April 2014. Fortune 500 2011.
URL http://fortune.com/fortune500/2011/

Fortune, July 2009. Fortune global 500, 2009 pharmaceutical industry. Fortune.

Furman, J. L., Kyle, M. K., Cockburn, I. M., Henderson, R., 2006. Public & private spillovers, location and the productivity of pharmaceutical research. Tech. rep., National Bureau of Economic Research.

Ganslandt, M., Maskus, K. E., 2004. Parallel imports and the pricing of pharmaceutical products: evidence from the european union. Journal of health economics 23 (5), 1035–1057.

Gardner, M. J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R. W., Carlton, J. M., Pain, A., Nelson, K. E., Bowman, S., 2002. Genome sequence of the human malaria parasite plasmodium falciparum. Nature 419 (6906), 498–511.

Giaccotto, C., Santerre, R. E., Vernon, J. A., 2005. Drug prices and research and development investment behavior in the pharmaceutical industry. Journal of Law and Economics 48 (1), 195–214.

Gold, M. R., Siegel, J., Russell, L., Weinstein, M., et al., 1996. Cost-effectiveness in health and medicine: report of the panel on cost-effectiveness in health and medicine. New York: Oxford Univ Pr.

Grabowski, H., 2008. Follow-on biologics: data exclusivity and the balance between innovation and competition. Nature Reviews Drug Discovery 7 (6), 479–488.

Grabowski, H., Vernon, J., 2000a. The determinants of pharmaceutical research and development expenditures. Journal of Evolutionary Economics 10 (1), 201–215.

Grabowski, H. G., Moe, J. L., 2008. Impact of economic, regulatory, and patent policies on innovation in cancer chemoprevention. Cancer Prev Res (Phila) 1 (2), 84–90.

Grabowski, H. G., Vernon, J., 2000b. The distribution of sales revenues from pharmaceutical innovation. Pharmacoeconomics 18 Suppl 1, 21–32.

Grabowski, H. G., Vernon, J. M., Thomas, L. G., 1978. Estimating the effects of regulation on innovation: an international comparative analysis of the pharmaceutical industry. Journal of Law and Economics, 133–163.

Gravelle, H., Morris, S., Sutton, M., 2006. Economic studies of equity in the consumption of health care. Edward Elgar Publishing.

Graves, S. B., Langowitz, N. S., 1993. Innovative productivity and returns to scale in the pharmaceutical industry. Strategic Management Journal 14 (8), 593–605.

Han, A., Hausman, J. A., 1990. Flexible parametric estimation of duration and competing risk models. Journal of Applied Econometrics 5 (1), 1–28.

Harris, J., 2005. Nice and not so nice. Journal of Medical Ethics 31 (12), 685–688.

Henderson, R., Cockburn, I., 1994. Measuring competence? exploring firm effects in pharmaceutical research. Strategic Management Journal 15 (S1), 63–84.

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.

Hirai, Y., Kinoshita, H., Kusama, M., Yasuda, K., Sugiyama, Y., Ono, S., 2010. Delays in new drug applications in japan and industrial r&d strategies. Clin Pharmacol Ther 87 (2), 212–8.

Hosmer, D. W. J., Lemeshow, S., May, S., 2011. Applied survival analysis: regression modeling of time to event data. Vol. 618. Wiley-Interscience.

Huskamp, H. A., Deverka, P. A., Epstein, A. M., Epstein, R. S., McGuigan, K. A., Frank, R. G., 2003. The effect of incentive-based formularies on prescription-drug utilization and spending. New England Journal of Medicine 349 (23), 2224–2232.

Huskamp, H. A., Rosenthal, M. B., Frank, R. G., Newhouse, J. P., 2000. The medicare prescription drug benefit: how will the game be played? Health Affairs 19 (2), 8–23.

Hwang, T. J., Carpenter, D., Kesselheim, A. S., 2014. Assessement of us pathway for approving medical devices for rare conditions. BMJ: British Medical Journal 348.

IHME, 2009. Global burden of disease study operations manual.

IHME, 2013. The global burden of disease: Generating evidence, guiding policy.

IHME, October 2014. Global burden of disease.
URL `http://www.healthdata.org/gbd`

IMAP, 2011. Pharmaceuticals biotech industry global report.

IMS, 2010. Pharmerging shake-up.

IMS, 2011. Ims lifecycle user guide 2011.

IMShealth, October 2012 2012. Imshealth: Lifecycle r&d focus.
URL `http://www.imshealth.com/portal/site/imshealth`

Jenkins, S. P., 2005. Survival analysis. Unpublished manuscript, Institute for Social and Economic Research, University of Essex, Colchester, UK.

Jensen, E. J., 1987. Research expenditures and the discovery of new drugs. The Journal of Industrial Economics, 83–95.

Joglekar, P., Paterson, M. L., 1986. A closer look at the returns and risks of pharmaceutical r&d. Journal of Health Economics 5 (2), 153–177.

Kanavos, P., Costa-Font, J., 2005. Pharmaceutical parallel trade in europe: stakeholder and competition effects. Economic Policy 20 (44), 751–798.

Kanavos, P., Vandoros, S., 2010. Competition in prescription drug markets: is parallel trade the answer? Managerial and Decision Economics 31 (5), 325–338.

Kelly, C. J., Mir, F. A., 2009. Economics of biological therapies. BMJ 339.

Kim, J., Lee, S. J., Marschke, G., 2009. Inventor productivity and firm size: Evidence from panel data on inventors. Pacific Economic Review 14 (4), 516–531.

Kola, I., Landis, J., 2004. Can the pharmaceutical industry reduce attrition rates? Nature Reviews Drug Discovery 3 (8), 711–716.

Kremer, M., 2002. Pharmaceuticals and the developing world. Journal of Economic Perspectives, 67–90.

Kremer, M., Glennerster, R., 2004. Strong medicine: creating incentives for pharmaceutical research on neglected diseases. Princeton University Press.

Kremer, M., Snyder, C. M., 2003. Why are drugs more profitable than vaccines?, nBER Working Paper.

Kyle, M., 2007. Pharmaceutical price controls and entry strategies. The Review of Economics and Statistics 89 (1), 88–99.

Kyle, M. K., 2006. The role of firm characteristics in pharmaceutical product launches. The RAND journal of economics 37 (3), 602–618.

Kyle, M. K., Allsbrook, J. S., Schulman, K. A., 2008. Does reimportation reduce price differences for prescription drugs? lessons from the european union. Health Services Research 43 (4), 1308–1324.

Kyle, M. K., McGahan, A. M., 2012. Investments in pharmaceuticals before and after trips. Review of Economics and Statistics 94 (4), 1157–1172.

Lambert, P., 2001. The distribution and redistribution of income. Manchester University Press.

Lancaster, T., 1992. The econometric analysis of transition data. Vol. 17. Cambridge Univ Pr.

Lancet, T., 2010. Striving for universal health coverage. The Lancet 376 (9755), 1799.

Lancet, T., 2012a. Neurological diseases remain neglected and ignored. The Lancet 379 (9813), 287.

Lancet, T., 2012b. Thyroid diseasemore research needed. The Lancet 379 (9821), 1076.

Lancet, T., 2012c. Tuberculosis control and elimination in 2012 and beyond. The Lancet 379 (9821), 1076.

Lanjouw, J. O., 2005. Patents, price controls and access to new drugs: how policy affects global market entry. Tech. rep., National Bureau of Economic Research.

Lanjouw, J. O., Cockburn, I. M., 2001. New pills for poor people? empirical evidence after gatt. World Development 29 (2), 265–289.

LeGrand, J., 1978. The distribution of public expenditure: the case of health care. Economica, 125–142.

Lichtenberg, F. R., 2005. Pharmaceutical innovation and the burden of disease in developing and developed countries 1. Journal of Medicine and Philosophy 30 (6), 663–690.

Lichtenberg, F. R., 2007. Importation and innovation. Economics of Innovation and New Technology 16 (6), 403–417.

Lopez, A. D., Murray, C., 1998. The global burden of disease. Nat Med 4 (11), 1241–3.

Mahlich, J., Roediger-Schluga, T., 2006. The determinants of pharmaceutical rd expenditures: evidence from japan. Review of Industrial Organization 28 (2), 145–164.

Manual, O., 2005. The measurement of scientific and technological activities. Proposed guidelines for collecting and interpreting innovation data.

Mathers, C., Fat, D. M., Boerma, J., 2008. The global burden of disease: 2004 update. World Health Organization.

Mathers, C. D., Lopez, A. D., Murray, C. J., et al., 2006. The burden of disease and mortality by condition: data, methods, and results for 2001. Global burden of disease and risk factors 45, 88.

Mestre-Ferrandiz, J., Sussex, J., Towse, A., 2012. The r&d cost of a new medicine. Tech. rep., Office of Healh Economics.

Mihaescu, O., Rudholm, N., 2013. Defining relevant product markets for pharmaceuticals.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., Group, P., et al., 2010. Preferred reporting items for systematic reviews and meta-analyses: the prisma statement. International Journal of Surgery 8 (5), 336–341.

Morton, F. S., Kyle, M., 2012. Markets for pharmaceutical products. Elsevier B.V.

Munos, B., 2009. Lessons from 60 years of pharmaceutical innovation. Nature Reviews Drug Discovery 8 (12), 959–968.

Murray, C. J., Acharya, A. K., 1997. Understanding dalys. Journal of Health Economics 16 (6), 703–730.

Murray, C. J., Vos, T., Lozano, R., Naghavi, M., Flaxman, A. D., Michaud, C., Ezzati, M., Shibuya, K., Salomon, J. A., Abdalla, S., et al., 2013. Disability-adjusted life years (dalys) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the global burden of disease study 2010. The Lancet 380 (9859), 2197–2223.

Murray, F., Stern, S., 2007. Do formal intellectual property rights hinder the free flow of scientific knowledge?: An empirical test of the anti-commons hypothesis. Journal of Economic Behavior & Organization 63 (4), 648–687.

Nelson, R. R., 1996. The sources of economic growth. Wiley Online Library.

Neumann, P. J., Rosen, A. B., Greenberg, D., Olchanski, N. V., Pande, R., Chapman, R. H., Stone, P. W., Ondategui-Parra, S., Nadai, J., Siegel, J. E., et al., 2005. Can we better prioritize resources for cost-utility research? Medical Decision Making 25 (4), 429–436.

Odling-Smee, L., 2007. Biomedical philanthropy: The giving machine. Nature 447 (7142), 254–254.

O'Donnell, O. A., Wagstaff, A., et al., 2008. Analyzing health equity using household survey data: a guide to techniques and their implementation. World Bank Publications.

OECD, 2011. Health at a Glance 2011: OECD Indicators. OECD.

OECD (Ed.), February 2014. Competition Issues in the Distribution of Pharmaceuticals. OECD Competition Committee.

Oughton, C., Landabaso, M., Morgan, K., 2002. The regional innovation paradox: innovation policy and industrial policy. The Journal of Technology Transfer 27 (1), 97–110.

Palmer, S., Smith, P. C., 2000. Incorporating option values into the economic evaluation of health care technologies. Journal of Health Economics 19 (5), 755–766.

Pammolli, F., Gambardella, A., Orsenigo, L., 2001. Global competitiveness in pharmaceuticals: a European perspective. Office for Official Publications of the European Communities.

Pammolli, F., Magazzini, L., Riccaboni, M., 2011. The productivity crisis in pharmaceutical r&d. Nature Reviews Drug Discovery 10 (6), 428–438.

Paul, S. M., Mytelka, D. S., Dunwiddie, C. T., Persinger, C. C., Munos, B. H., Lindborg, S. R., Schacht, A. L., 2010. How to improve r&d productivity: the pharmaceutical industry's grand challenge. Nature reviews Drug discovery 9 (3), 203–214.

Pecoul, B., Chirac, P., Trouiller, P., Pinel, J., 1999. Access to essential drugs in poor countries: a lost battle? Jama 281 (4), 361–367.

Pennings, E., Sereno, L., 2011. Evaluating pharmaceutical r&d under technical and economic uncertainty. European Journal of Operational Research 212 (2), 374–385.

Perkmann, M., Walsh, K., 2007. University–industry relationships and open innovation: Towards a research agenda. International Journal of Management Reviews 9 (4), 259–280.

Phill, O., Ruth, P.-P., Jorge, M.-F., Jon, S., 2012. International comparisons of medicine prices. Tech. rep., OHE.

Pisano, G. P., 1990. The r&d boundaries of the firm: an empirical analysis. Administrative Science Quarterly, 153–176.

PWC, 2011a. Pharma 2020: Supplying the future - which path will you take?

PWC, 2011b. Pharma 2020: Virtual rd which path will you take?

Rajapakse, A., Titchener-Hooker, N. J., Farid, S. S., 2005. Modelling of the biopharmaceutical drug development pathway and portfolio management. Computers & chemical engineering 29 (6), 1357–1368.

Refoios Camejo, R., McGrath, C., Herings, R., 2011. A dynamic perspective on pharmaceutical competition, drug development and cost effectiveness. Health policy 100 (1), 18–24.

Reinganum, J. F., 1981. Dynamic games of innovation. Journal of Economic Theory 25 (1), 21–41.

Reuters, T., 2009. Cmr international 2009 factbook. Tech. rep., Thompson Reuters.

Rhoten, D., Powell, W. W., 2007. The frontiers of intellectual property: Expanded protection versus new models of open science. Annu. Rev. Law Soc. Sci. 3, 345–373.

Ridley, D., Grabowski, H., Moe, J., 2006. Developing drugs for developing countries. Health Affairs 25 (2), 313.

Rogers, M. J., Gupta, A., Maranas, C. D., 2002. Real options based analysis of optimal pharmaceutical research and development portfolios. Industrial & Engineering Chemistry Research 41 (25), 6607–6620.

Romer, P. M., 1990. Endogenous technological change. Journal of Political Economy, S71–S102.

Rosamond, J., Allsop, A., 2000. Harnessing the power of the genome in the search for new antibiotics. Science 287 (5460), 1973–1976.

Royston, P., Lambert, P. C., 2011. Flexible parametric survival analysis using stata: beyond the cox model. Stata Press books.

Salomon, J. A., Mathers, C. D., Chatterji, S., Sadana, R., Ustun, T. B., Murray, C. J., 2003. Quantifying individual levels of health: definitions, concepts and measurement issues. Health Systems Performance Assessment: Debate, Methods, and Empiricism, 301–18.

Scannell, J. W., Blanckley, A., Boldon, H., Warrington, B., 2012. Diagnosing the decline in pharmaceutical r&d efficiency. Nature Reviews Drug discovery 11 (3), 191–200.

Scherer, F. M., 1967. Market structure and the employment of scientists and engineers. The American Economic Review, 524–531.

Scherer, F. M., Watal, J., 2002. Post-trips options for access to patented medicines in developing nations. Journal of International Economic Law 5 (4), 913–939.

Schumpeter, J., 1942. Creative destruction. Capitalism, socialism and democracy.

Scotchmer, S., 2004. Innovation and incentives. MIT press.

Scott Morton, F., 1999. Entry decisions in the generic pharmaceutical industry. The RAND Journal of Economics, 421–440.

ScripIntelligence, 2013. The scrip 100 ranking.

ScripIntelligence, April 2014. Pharma league tables 2011.
URL http://www.scripintelligence.com/leaguetables/

Siegel, D. S., Waldman, D. A., Atwater, L. E., Link, A. N., 2003. Commercial knowledge transfers from universities to firms: improving the effectiveness of university–industry collaboration. The Journal of High Technology Management Research 14 (1), 111–133.

Smith, R. D., Coast, J., 2002. Antimicrobial resistance: a global response. Bulletin of the World Health Organization 80 (2), 126–133.

Strauss, S., 2010. Pharma embraces open source models. Nature Biotechnology 28 (7), 631–634.

Subramanian, D., Pekny, J. F., Reklaitis, G. V., Blau, G. E., 2003. Simulation-optimization framework for stochastic optimization of r&d pipeline management. AIChE Journal 49 (1), 96–112.

Thiers, F. A., Sinskey, A. J., Berndt, E. R., 2008. Trends in the globalization of clinical trials. Nature Reviews Drug Discovery 7 (1), 13–14.

Thomas, L. G., 1994. Implicit industrial policy: the triumph of britain and the failure of france in global pharmaceuticals. Industrial and Corporate Change 3 (2), 451–489.

Trouiller, P., Olliaro, P., Torreele, E., Orbinski, J., Laing, R., Ford, N., 2002. Drug development for neglected diseases: a deficient market and a public-health policy failure. The Lancet 359 (9324), 2188–2194.

Ulku, H., 2007. Rd, innovation, and growth: Evidence from four manufacturing sectors in oecd countries. Oxford Economic Papers 59 (3), 513–535.

Valentin, F., Jensen, R. L., 2007. Effects on academia-industry collaboration of extending university property rights. The Journal of Technology Transfer 32 (3), 251–276.

Vallerie Paris, A. B., 2013. Value in pharmaceutical pricing.

Van den Berg, G., 2001. Duration models: specification, identification and multiple durations. Handbook of Econometrics 5, 3381–3460.

Van Doorslaer, E., Van Ourti, T., 2011. Measuring inequality and inequity in health and health care. Oxford University Press, Oxford, 837–869.

Vernon, J. A., 2004. Pharmaceutical r&d investment and cash flows: an instrumental variables approach to testing for capital market imperfections. Journal of Pharmaceutical Finance Economics and Policy 13 (4), 3.

Vernon, J. A., 2005. Examining the link between price regulation and pharmaceutical r&d investment. Health Economics 14 (1), 1–16.

Vernon, J. A., Golec, J. H., Dimasi, J. A., 2010. Drug development costs when financial risk is measured using the fama-french three-factor model. Health Economics 19 (8), 1002–1005.

Vernon, J. M., Gusen, P., 1974. Technical change and firm size: the pharmaceutical industry. The Review of Economics and Statistics, 294–302.

Viergever, R. F., Terry, R. F., Karam, G., 2013. Use of data from registered clinical trials to identify gaps in health research and development. Bulletin of the World Health Organization 91 (6), 416–425C.

Wadman, M., 2007. Biomedical philanthropy: state of the donation. Nature 447 (7142), 248–250.

Wagstaff, A., 2005. The bounds of the concentration index when the variable of interest is binary, with an application to immunization inequality. Health Economics 14 (4), 429–432.

Wagstaff, A., Paci, P., Van Doorslaer, E., 1991. On the measurement of inequalities in health. Social Science & Medicine 33 (5), 545–557.

Wagstaff, A., Van Doorslaer, E., 2000. Equity in health care finance and delivery. Handbook of Health Economics 1, 1803–1862.

Wagstaff, A., Van Doorslaer, E., Paci, P., 1989. Equity in the finance and delivery of health care: some tentative cross-country comparisons. Oxford Review of Economic Policy, 89–112.

WHO, October 2014. International statistical classification of diseases and related health problems 10th revision.
URL `http://apps.who.int/classifications/icd10/browse/2015/en`

Wiggins, S. N., 1981. Product quality regulation and new drug introductions: some new evidence from the 1970s. The Review of Economics and Statistics, 615–619.

Wiggins, S. N., 1983. The impact of regulation on pharmaceutical research expenditures: A dynamic approach. Economic Inquiry 21 (1), 115–128.

Willekens, F., 2009. Continuous-time microsimulation in longitudinal analysis. New Frontiers in Microsimulation Modelling, 353–376.

Williams, A., Cookson, R., 2000. Equity in health. Handbook of Health Economics 1, 1863–1910.

WorldBank, April 2014. Global economic prospects.
URL `http://www.worldbank.org/en/publication/global-economic-prospects`

Yin, W., 2008. Market incentives and pharmaceutical innovation. Journal of Health Economics 27 (4), 1060–1077.

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA EDUCAÇÃO E CIÊNCIA

POPH

UNIÃO EUROPEIA
Fundo Social Europeu

QREN QUADRO DE REFERÊNCIA ESTRATÉGICO NACIONAL
PORTUGAL 2007-2013