The determinants of failure in drug development: a duration analysis

Eliana Barrenho*
Imperial College London, Business School

October 2014

Abstract

Global research and development data on pharmaceutical drugs from 1980 to 2012 are used to estimate the risk of attrition of a project in each R&D stage as a function of the duration of each stage, and competition and alliances between firms. We find that markets with more new drugs and less R&D competitors experience higher attrition. These effects are particularly important in the discovery and phase 2 clinical trials. Competition from longer established drugs appears to not significantly affect attrition. The impact of alliances is not conclusive, bringing into question role of alliances on pharmaceutical innovation.

Keywords: pharmaceutical innovation, failure rates, competition, alliances.

JEL codes: I11, I19, L65, O32

*I am extremely indebted to Professor Peter C. Smith and Professor Marisa Miraldo for continual support and encouragement. I would like to thank Dr Simone Ghislandi, Dr Paolo Li Donni, Dr Jorge Mestre-Ferrandiz, Professor Deborah Ashby, Professor Rossana Tarricone, Professor Giovanni Fattore, and Dr Alessia Melegaro and attendants at internal seminars at Imperial College London, Bocconi CERGAS, 13*th* Portuguese Health Economics Conference 2013 in Braga, International Health Economics Association conference 2013 in Sydney, 2*nd* 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. I would also like to thank the Fundação para a Ciência e a Tecnologia for financial support. Contact: eliana.barrenho@imperial.ac.uk, Address: Imperial College Business School, Tanaka Building, South Kensington Campus, London SW7 2AZ. Please do not cite or circulate without permission.

1 Background

Product innovation in the pharmaceutical industry is costly, risky and time-consuming. With a decreased number of new chemical entities being discovered, and research and development (R&D) costs representing a high proportion of the industry revenues, the pharmaceutical industry is facing unprecedented challenges to its R&D model. 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 paper, we measure the association of the probability of failure in the different R&D phases with the potential factors that can affect innovation 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 (**?**), 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, 2000; 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 drugs 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 projects history and characteristics in the analysis of phase-specific success rates.

Danzon et al. (2005b) 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 firms 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 older drugs

seems to be greater than that of more recently introduced drugs. Also, markets with many potential competitors 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 originators 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 to some earlier contributions, we use two different semi-parametric proportional models to estimate the impact of market structure 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. (2005b) 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 drugs observable characteristics (i.e., therapy category,

route of administration) and companys size and location on the success of transition in clinical phases. Danzon et al. (2005b) 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.

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 on 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, 2000; 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:2000vo), 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 mar-

ket 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., 2005b) 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. (2005b) 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 paper is organised as follows. In section 2 we present the specification model and estimation strategy. In section 3 we describe the data. In section 4 we discuss the descriptive statistics and non-parametric analysis, and we present the main results from the semi-parametric analysis. Finally, in section 5 we provide a discussion of the results and conclusions.

2 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:

$$h(t) = P(\text{failure at time } t \mid \text{R\&D until time } t) = \frac{P(\text{failure at time } t)}{P(\text{R\&D until time } t)} \quad (1)$$

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

$$h(t) = \lim_{dt \to 0} \frac{P(t \leq T < t + dt) \mid T \geq t)}{dt} = \frac{P(t \leq T < t + dt) \mid T \geq t)}{P(T \geq t)} \quad (2)$$

If we represent the duration distribution function as $P(T < t) = F(t)$[1], where $t \geq 0$

[1]In the context of duration models,

$$F(t) = 1 - S(t) \quad (3)$$

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

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:

$$h(t) = \frac{F(t+dt) - F(t))}{(1 - F(t)} = \lim_{dt \to 0} \frac{F(t+\delta t) - F(t))}{dt} \cdot \frac{1}{1 - F(t)} = \\ = F\prime(t) \cdot \frac{1}{1 - F(t)} = \frac{f(t)}{1 - F(t)} \quad (6)$$

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

$$h(t, X) = h_0(t)\theta(X) \quad (7)$$

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.

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

it 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:

$$h(t) = -\frac{\frac{\delta S(t)}{dt}}{S(t)} \quad (5)$$

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

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:

$$h_i^j(t, X, \tau) = h_0(t)\theta(X)\tau = h_0(t)\theta(X) \tag{8}$$

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:

$$\theta(X) = \exp(X\prime\beta) \tag{9}$$

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 base-

line hazard function, which can be estimated non-parametrically (Lancaster, 1992). Further detail on this can be found in Appendix.

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

$$T = min\{T_c, T_j, T_k\} \tag{10}$$

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:

$$h(t) = P(\text{failure at time } t \mid t+ dt) = \\ = P(t \leq T_j < t + dt \mid \text{survival to t and all other } \{T_c, T_k\} \geq t + dt) \quad (11)$$

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:

$$\theta_i(t, X, \tau_j) = \theta_i^j + \theta_i^k \quad (12)$$

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:

$$\theta_i = \begin{cases} \lambda_j(t_j) \times \theta_{0,j}(X)\tau_j & \text{, if } j \text{ happens} \\ \lambda_k(t_k) \times \theta_{0,k}(X)\tau_k & \text{, if } k \text{ happens} \end{cases} \quad (13)$$

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 Appendix. 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 Akaikes 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 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) .

3 Data and variables

We have built a unique dataset by merging IMS Health R&D Focus of pharmaceutical projects with World Bank data on country level population and GDP 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; Intelligence, 2013).

The IMS Health RD Focus contains information on global pharmaceutical R&D activity from 1980 until 2012. The dataset contains all pharmaceutical 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

[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

$$\ln h(t, X, \beta) = \ln h_0(t) + X\prime\beta \tag{14}$$

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).

3, and market/registration (states). Therefore, the dataset contains information for each project for the time (in years) spent within each R&D phase (state duration), including the abandonment of projects by the firm (which we label *failure*), and the dates of transition between states. The dataset also records the timing of failure of a project. Using this dataset, we constructed time-to-event data that preserves transitions between different R&D stages, including the abandonment of a project, and the timing of all events.

Each observation is a project-year (actually each project might have several targets and we consider each target as a different project), reflecting the 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.

Each project is assigned to one of the 17 broad disease areas (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). Within this classification, drugs are classified into groups at five different therapeutic levels. We have used the third level of the ATC code, which indicates one from the total 262 therapeutic pharmacological subgroups to define the relevant market, similarly to other papers in the literature (see, for example, Kyle (2006)).

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, 2011) 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).

Finally, some projects are first observed in the database at an R&D stage different from basic discovery. This is because information on pre-clinical stage is more difficult to get hold for very good reasons, as some pre-clinical research is not project or indication-specific. This is one of the reasons for the limited literature on success probability for preclinical stages (Paul et al., 2010). We accommodate this issue by using appropriate modelling options in our semi-parametric models to account for the left-censored data.

Variables

A complete list of variables labels and description can be found in Table 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 $\{discovery, Phase1, Phase2, Phase3\}$, and the value for zero 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).

[Table 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.

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]. 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. 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. Also, further concerns with endogeneity would arise if we would have

[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

included them in our analysis given the important role that pharmaceuticals play in reducing disease prevalence. 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.

4 Results

4.1 Descriptive statistics and nonparametric analysis

Our sample consists of 18,252 projects, 4,230 of which have failed. Table 2 summarises the descriptive statistics across the years between 1980 and 2012 for failures and successes, duration of the projects, competition, alliances and market size proxies.

[Table 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 1). This may be partially due to more "births" of new projects: Figure 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 1 and Figure 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) anti-neoplastics (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: gGeneral anti-infectives) (Table 3).

[Table 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 3). Also, projects tend to fail less in the discovery stage when facing more competition within the R&D process (Figure 4). Results are less clear when looking at the other R&D stages.

[Figure 3 and Figure 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 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 6). Also, academic partners seem to linked with more successful projects in pre-clinical stages of development.

[Figure 5 and Figure 6 here]

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

4.2 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 reports estimation results of single risk hazard model for all R&D phases, whilst Table 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 both specifications of the baseline hazard. 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 and Table 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 GDP per capita 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.

5 Discussion and conclusions

This paper 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.

Secondly, we show competition to be significantly associated with failure in discov-

ery 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 incremental 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)). 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 signalled 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.

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 decreasing returns 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 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. 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 of patenting struture 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.

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

Finally, projects may clearly fail to complete a specific R&D phase due to several reasons. Failure may be a combination of rejection by the regulatory bodies, withdrawal by the firm, merger and/or acquisition by a competitor, or even scientific/clinical non-achievement. With our data cannot separately identify the reasons behind project abandonment.

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 their 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.

Appendix A

Table 1: Definition of variables

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

Table 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
#New 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
Nr of observations	54600			11109			9644			5789		
Nr of drug projects	10952	3277		3257	444		2446	361		1597	148	
Proportion of failures		29.90%			13.60%			14.80%			9%	
Nr therapeutic classes	78											
Years covered		1980-2012										

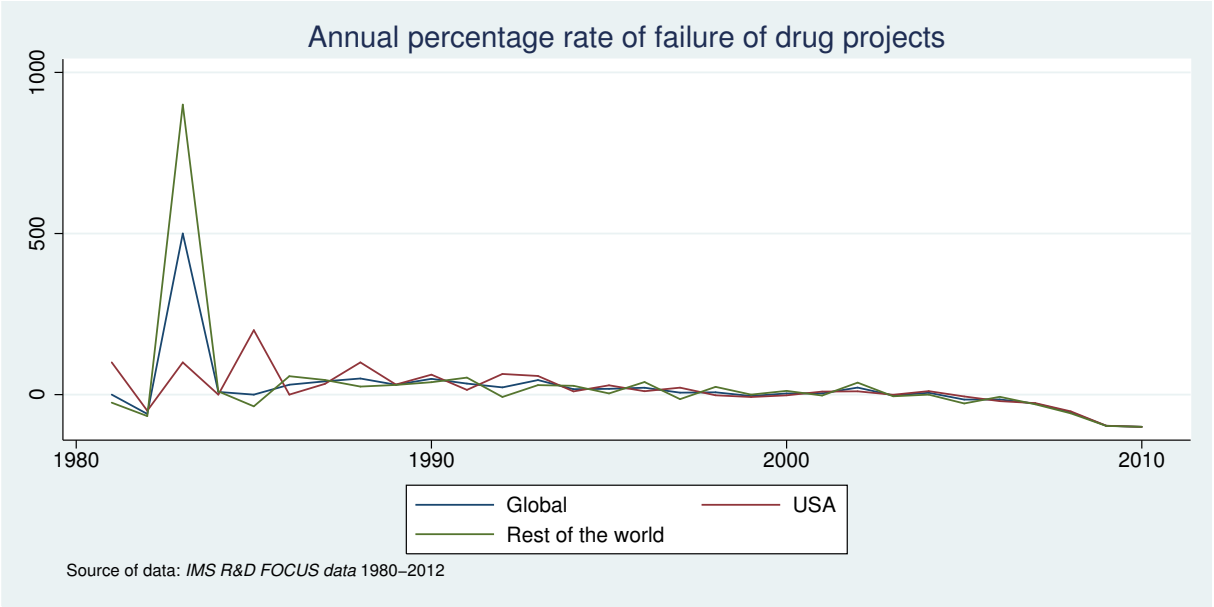

Figure 1: Annual percentage rate of failure of drug projects

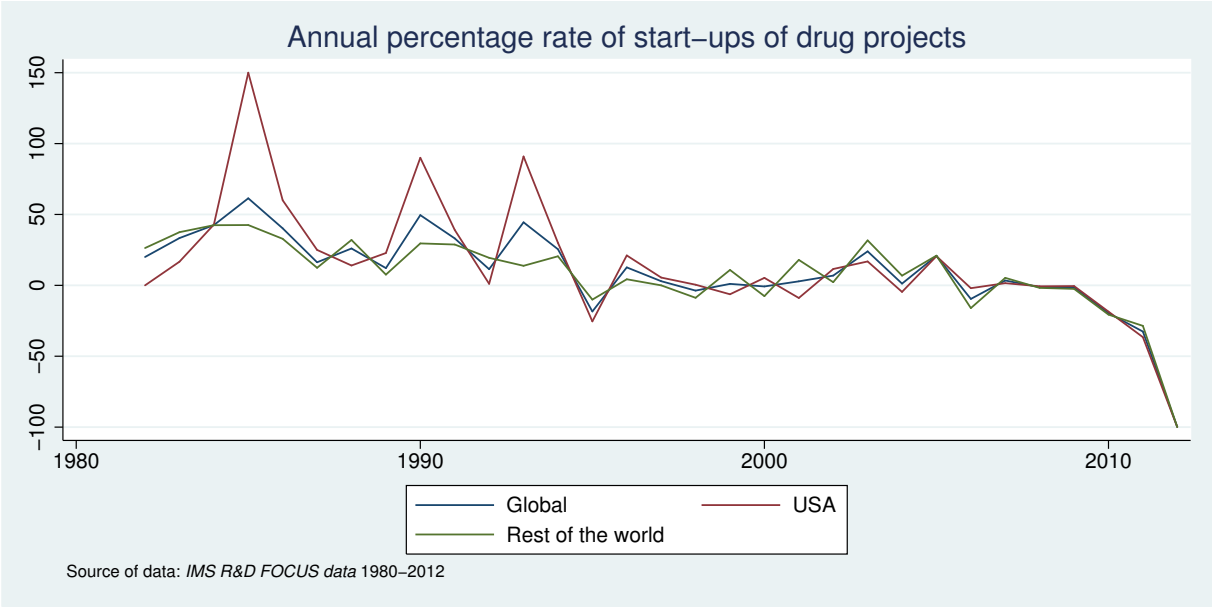

Figure 2: Annual percentage rate of start-ups of drug projects

Table 3: Distribution of R&D activity and failure by therapeutic category

	Total number of R&D projects	Proportion of failures as a proportion of total failures in each ATC	Proportion of failures as a proportion of total failures
Total number of projects: 18,252			
Failed projects: 4,230			
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

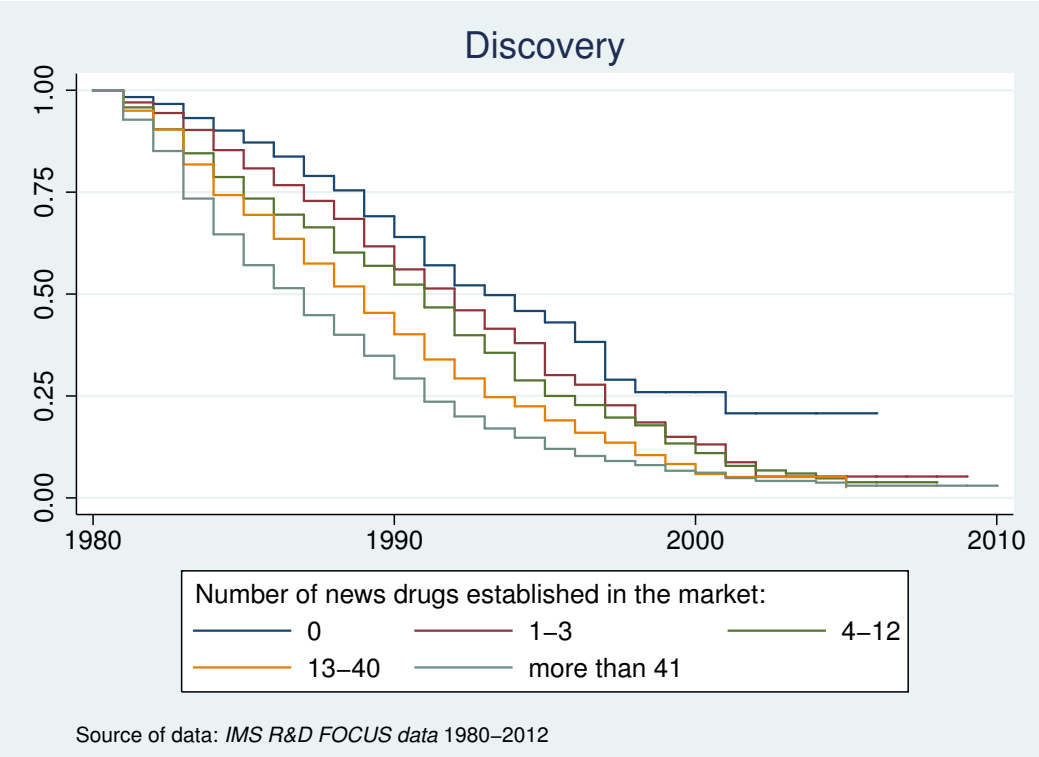

Figure 3: Kaplan-Meier survival function: competition in the final product market from young drugs

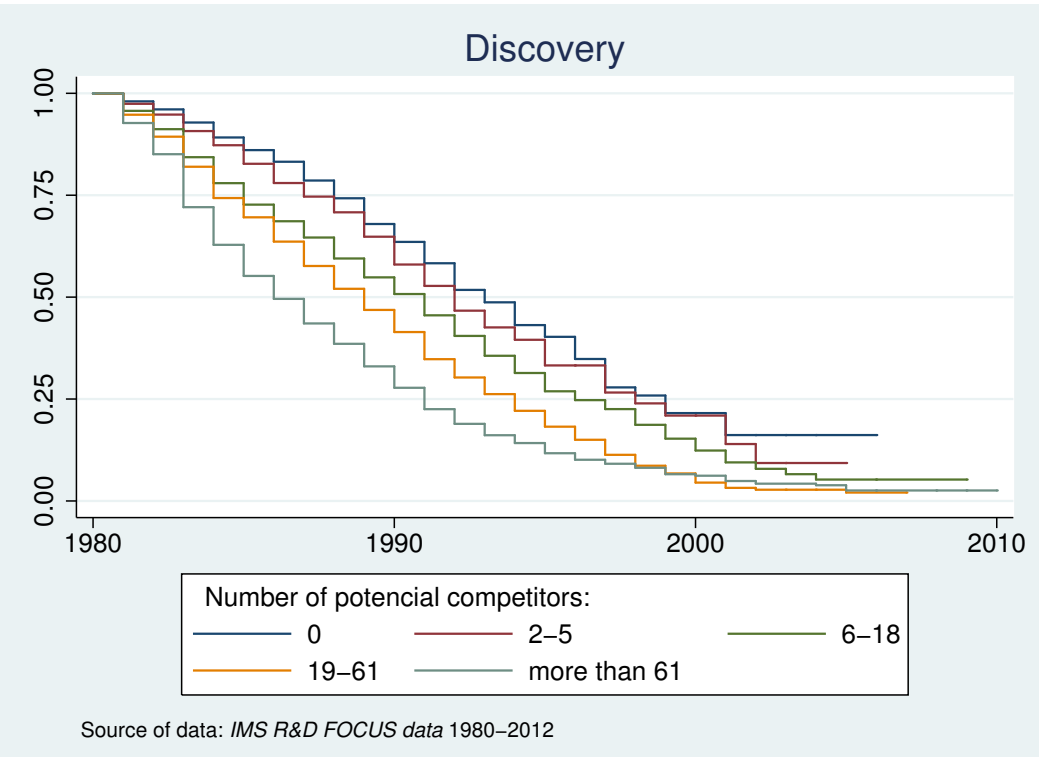

Figure 4: Kaplan-Meier survival function: competition in the final product market from potential competitors

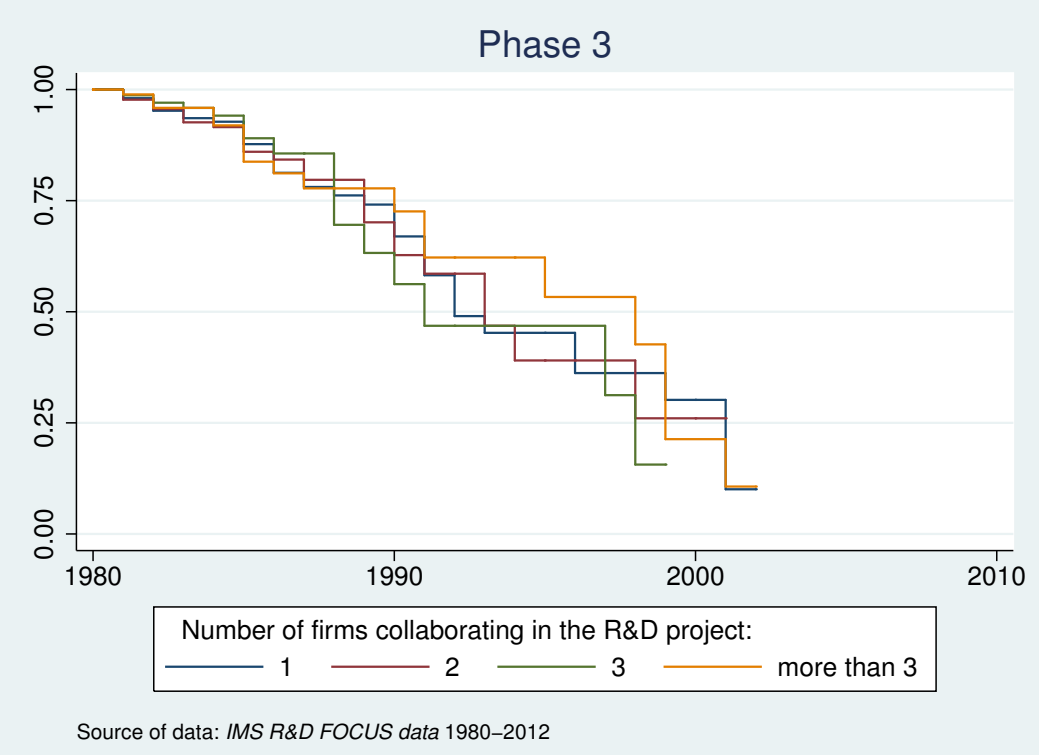

Figure 5: Kaplan-Meier survival function: intensity of alliances

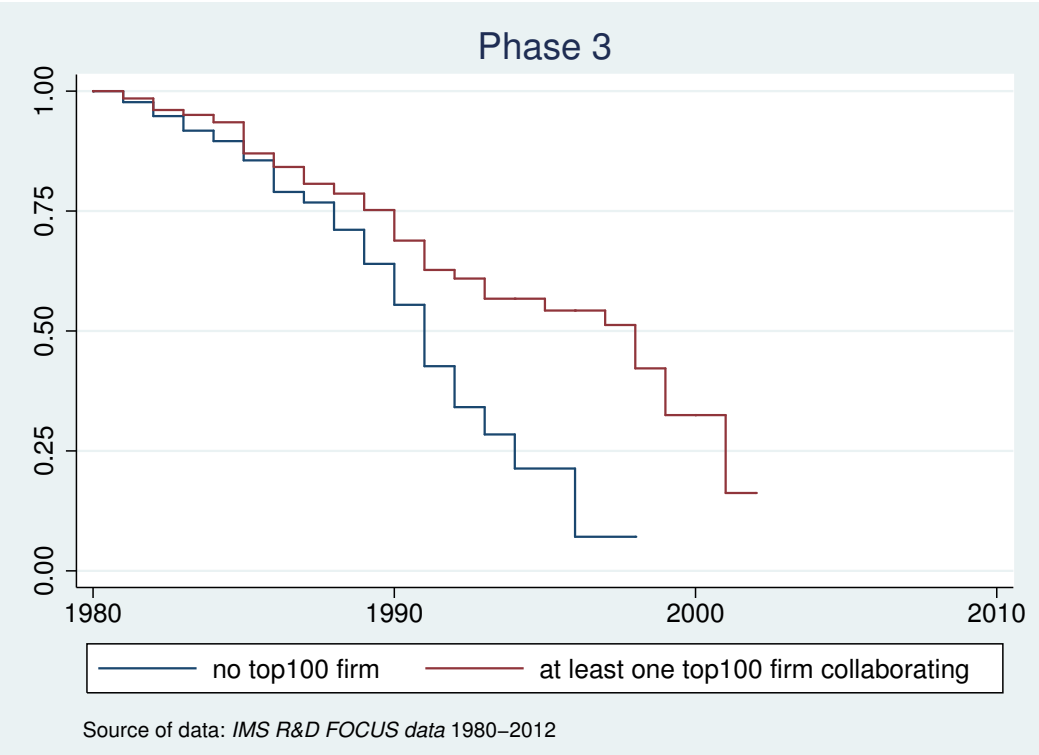

Figure 6: Kaplan-Meier survival function: participation of big firms

Table 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 pc	1	1	1	1
	(-2.22E-7)	(-6.54E-7)	(-4.07E-7)	(-7.26E-7)
GDP pc 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 participating	1.282***	1.107	0.871	0.457***
	(-0.0583)	(-0.147)	(-0.139)	(-0.131)
Academia participating	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 area	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 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 pc	1	1	1	1
	(-0.0000097)	(-0.0000173)	(-0.0000182)	(-0.000076)
GDP pc 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 participating	0.912***	0.948	0.995	0.610*
	(-0.0255)	(-0.0511)	(-0.0742)	(-0.17)
Academia participating	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.

Appendix B

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 (8):

$$h_i^j(t, X, \tau_j) = h_0(t) \exp(X\prime\beta) = h_0(t)\theta(X) \tag{15}$$

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:

$$PL = \prod_{i=1}^{S} \mathcal{L}_i \tag{16}$$

$\mathcal{L}_i = P$(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:

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

$$\begin{cases} A = P(\text{event experienced by } d{=}5 \text{ and not } d{=}6) = [h_5(13)S_5(13)\delta t][S_6(13)] \\ B = P(\text{event experienced by } d{=}6 \text{ and not } d{=}5) = [h_6(13)S_6(13)\delta t][S_5(13)] \end{cases} \tag{17}$$

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

Table 6: Illustrative dataset

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

is equal to

$$\mathcal{L}_5 = \frac{A}{A+B} = \frac{h_5(13)}{h_5(13)+h_6(13)} \tag{18}$$

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:

$$\mathcal{L}_5 = \frac{h_5(13)}{h_5(13)+h_6(13)} = \frac{h_5(13)\theta_5}{h_5(13)\theta_5+h_6(13)\theta_6)} = \frac{\theta_5}{\theta_5+\theta_6} \tag{19}$$

The baseline hazard contributions cancel out. Similarly,

$$\mathcal{L}_1 = \frac{\theta_1}{\theta_1+\theta_2+\ldots+\theta_6} \tag{20}$$

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 $s^t h$ 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 evaluated during 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 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

$$ln\mathcal{L} = \{\delta^j[ln\ \theta_i^j] + lnS_j\} + \{\delta^k[ln\ \theta_i^k] + ln\ S_k(T)\} \tag{21}$$

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

Assumption A. 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:

$$\begin{aligned} S(t) = \exp[-\int_0^t h(u)\,\mathrm{d}u] = \exp[-\int_0^t [\theta_i^j(u) + \theta_i^k(u)]\,\mathrm{d}u] = \\ = \exp[-\int_0^t [\theta_i^j(u)]\,\mathrm{d}u] \times \exp[-\int_0^t [\theta_i^k(u)]\,\mathrm{d}u] = S_j(t)S_k(t) \end{aligned} \tag{22}$$

Assumption B. 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:

$$\delta^j = \begin{cases} 1 & \text{, if transition to } j \\ 0 & \text{, otherwise (either exit to } k \text{ or censored)} \end{cases} \tag{23}$$

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

$$\begin{aligned} \mathcal{L} = (\mathcal{L})^{\delta^j}(\mathcal{L}^k)^{\delta^k}(\mathcal{L}^C)^{1-\delta^j-\delta^k} = \\ = [f_j(T)S_k(T)]^{\delta^j}[f_k(T)S_j(T)]^{\delta^k}[S_j(T)S_k(T)]^{\delta 1-\delta^j-\delta^k} = \\ = \left[\frac{f_j(T)}{S_j(T)}\right]^{\delta^j} S_j(T) \left[\frac{f_k(T)}{S_k(T}\right]^{\delta^k} S_k(T) = \\ = [h_j]^{\delta^j} S_j(T)[h_k]^{\delta^k} S_k(T) = \{[h_j]^{\delta^j} S_j(T)\}\{[h_k]^{\delta^k} S_k(T)\} \end{aligned} \tag{24}$$

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.

$$ln\ \mathcal{L} = \{\delta^j[ln\ \theta_i^j] + lnS_j\ \}\ +\ \{\delta^k[ln\ \theta_i^k]\ + ln\ S_k(T)\} \tag{25}$$

Akaike's criteria

Akaike's (AIC) criteria: for each model specification the value is computed as: $AIC = -2 \times (\text{log-likelihood}) + 2(p + 1 + s)$, 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 groups 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).

References

Abrantes-Metz, R., Adams, C., Metz, A., 2004. Pharmaceutical development phases: A duration analysis. Tech. rep., US Federal Trade Commission, Washington DC.

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.

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.

Cameron, A., Trivedi, P., 2005. Microeconometrics: methods and applications. Cambridge Univ Press.

Cleves, M., William, G., Gutierrez, R. G., Marchenko, Y. V., 2010. An Introduction to Survival Analysis using Stata, third edition Edition. StatCorp LP, Texas.

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.

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., 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., Wang, Y. R., Wang, L., 2005b. 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.

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., 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.

Dixit, A. K., 1994. Investment under uncertainty. Princeton University Press.

Dubois, P., de Mouzon, O., Scott Morton, F., Seabright, P., 2011. Market size and pharmaceutical innovation. TSE Working Paper 11.

EphMRA, 2013. Ephmra anatomical classification.

Fortune, July 2009. Fortune global 500, 2009 pharmaceutical industry. Fortune.

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.

Grabowski, H., Vernon, J., 2000. 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.

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.

Intelligence, S., 2013. The scrip 100 ranking.

Jenkins, S. P., 2005. Survival analysis. Unpublished manuscript, Institute for Social and Economic Research, University of Essex, Colchester, UK.

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.

Kyle, M., 2006. The role of firm characteristics in pharmaceutical product launches. The RAND Journal of Economics 37 (3), 602–618.

Lancaster, T., 1992. The econometric analysis of transition data. Vol. 17. Cambridge Univ Pr.

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.

Mahlich, J., Roediger-Schluga, T., 2006. The determinants of pharmaceutical rd expenditures: evidence from japan. Review of Industrial Organization 28 (2), 145–164.

Mestre-Ferrandiz, J., Sussex, J., Towse, A., 2012. The r&d cost of a new medicine. Tech. rep., Office of Healh Economics.

Munos, B., 2009. Lessons from 60 years of pharmaceutical innovation. Nature Reviews Drug Discovery 8 (12), 959–968.

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.

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., 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.

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.

PWC, 2011. Pharma 2020: Supplying the future - which path will you take?

Reinganum, J. F., 1981. Dynamic games of innovation. Journal of Economic Theory 25 (1), 21–41.

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.

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.

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.

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.

Van den Berg, G., 2001. Duration models: specification, identification and multiple durations. Handbook of Econometrics 5, 3381–3460.

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.