Reactive Outsourcing of Technological Innovations

Manuel Hermosilla
Johns Hopkins University
mh@jhu.edu

December 2018

Abstract

Large innovative firms routinely strengthen their new product pipelines by outsourcing developing technologies from other (typically startup) companies. This practice is portrayed by scholars and practitioners as a high-involvement, carefully planned and executed process. We contribute to their understanding by highlighting a novel concept and documenting an associated trade-off: technologies outsourced "reactively" (following development setbacks of other projects in their pipelines) are less likely to reach the market. Our argument is supported by evidence from the pharmaceutical industry, where outsourcing (drug candidate licensing) plays a central role for innovation, and where Phase 3 outcomes provide a rich source of quasi-experimental variation that we exploit for inference. The analysis takes advantage of recent methodological advances that leverage machine learning for causal inference. Compared to traditional methods, these techniques make more efficient use of data and allow for a more detailed description of the effects at play. Results suggest that, whereas reactive outsourcing behavior is not rare in the pharmaceutical industry, its incidence and consequences can be predicted to occur in specific environments.

Keywords: New product development, pharmaceuticals, licensing, causal forest.

*This paper supersedes the manuscript "Search-and-Match in a Rush: Investigating Reactive Licensing in the Pharmaceutical Industry." I thank Laura Kornish, Joshua Krieger, Tinglong Dai, Carlos Noton, editors and anonymous referees, and seminar participants at Johns Hopkins Carey Business School, Cambridge University Judge Business School, and 2018 Marketing Science Conference for their helpful comments and suggestions. I gratefully acknowledge a Johns Hopkins Carey Business School internal research grant. All errors and omissions are my own.

1 Introduction

Large innovative firms routinely strengthen their new product pipelines by acquiring rights to commercialize developing technologies invented by other (typically startup) companies. This practice—known as technological outsourcing—is the focus of this article. For the acquiring firm ("innovator"), outsourcing offers the main advantages of shortening development cycles and selectively accessing frontier new technologies in environments of rapid and distributed technical progress. Fueled by these advantages, the model has grown in popularity and even lead to the restructuring of a variety of high technology industries, ranging from aerospace to pharmaceuticals (Quinn, 2000).

The contribution of this paper is to highlight a novel concept and empirically document an associated trade-off: technologies outsourced "reactively" by innovators (following setbacks of other products in their pipelines) are less likely to reach the market. Our motivation comes from the standard scholarly and practitioner portrayals of outsourcing, which describe it as a high-involvement, "proactive" process, shaped towards the implementation of the innovator's strategic goals, and which requires several preparatory and execution activities. For example, Quinn (2000) emphasizes that the model's success is contingent on it being "carefully pursued." Moreover, a World Intellectual Property Organization white paper (WIPO, 2015) states that "before the other party has been approached, a party may spend many months defining business objectives, assessing leverage, researching the other party, deciding positions on key terms, preparing documentation, and protecting IP, among other tasks." We posit that innovators who outsource reactively may "cut corners" on some of these activities, compromising in turn technologies' chances to reach the market.

The empirical arena is the pharmaceutical industry, which we argue is a suitable but also important case study. Unable to keep up with the rapid and diverse advances in basic science, large pharmaceutical firms have come to significantly rely on investigational therapies invented by biotech startups (Cockburn, 2004; Pisano, 2006). Implemented through licensing contracts, these outsourcing deals often trade rights for the more innovative and potentially impactful types of developing therapies (Kneller, 2010). This is why many deals get to be valued at multiple billions of dollars, and why insights that help fine tune licensing (outsourcing) efforts could yield materially large rewards. Concurrently, drug innovation is notorious for its large attrition: most therapies that enter the development process fail to reach the market. This statistic is the key ingredient for the industry's large average development costs (DiMasi et al., 2003, 2016), as well as a common justification for the high prices of pharmaceuticals. To the extent that we can identify managerial behaviors that contribute to this large attrition, we could shed light on ways alleviate these problems.

Our interest on this industry is also based on some anecdotal evidence suggesting that large pharmaceutical firms may indeed engage in reactive licensing. In particular, follow-

ing the Phase 3 failure of Pfizer's cardiovascular compound Torcetrapib, the company's CEO vowed to "bring increased focus and emphasis to its business development and licensing efforts in order to identify new products and technologies that will supplement its pipeline" (Pfizer Inc., 2006). Five months later, the company announced an agreement to license a cardiovascular-targeted compound (Pfizer Inc., 2007).

Based on this episode, our empirical analysis consists on investigating the impacts of Phase 3 failures ("P3Fs") on the firm's subsequent licensing behavior, and the development performance (i.e., advancement versus termination at each stage) of subsequently licensed therapies. Partly because few therapies reach Phase 3, P3Fs constitute large shocks for developers—"When Phase III clinical trial failures happen it is a painful blow—to the drug manufacturer, to investors and to patients" (Merrill, 2016). Whereas economic considerations may prompt firms to terminate development at earlier stages, discontinuation decisions at Phase 3 are primarily driven by the quality of randomized testing results. This is because, this late in the development process, most uncertainty has been cleared and investments sunk, making continuation a dominant strategy. (This is also why therapies are said to "fail" rather than be "terminated" at Phase 3.) P3Fs thus provide a rich source of quasi-experimental variation that we take advantage of, while accounting for firms' varying propensity to experience a P3F at a given point in time.

In an ideal empirical setting, our analysis would correlate P3Fs with data tracking firms' preparatory and execution activities. But because these data are not systematically available, we focus on testable implications for our hypothesis' two main pillars. Using comprehensive 2001-2015 data for twenty of the largest pharmaceutical firms, we ask:

- **[Q1, engagement with reactive licensing]** Do P3Fs increase the near-future probability of licensing?
- **[Q2, consequences of reactive licensing]** Do pre-licensing P3Fs worsen post-licensing development performance (i.e., more likely terminations)?

Simple data descriptives support our hypothesis, by suggesting that both questions have yes for an answer. Nevertheless, there exists a sizable set of potential confounders X. We account for these adopting a data-driven causal inference approach, where P3F incidence plays the role of "treatment."

Our analysis first considers a matching estimator, which draws inference by comparing outcomes of treated and non-treated observations matched on X. While estimated average treatment effects are directionally consistent with those of our simpler analyses, there remains a doubt on whether the data could be used more efficiently. In particular, an important drawback of matching models is that they are not equipped to determine which of the variables in X are important to match on (and which are not), thus requiring that all variables be matched. This "skeptical" approach creates a curse of dimensionality problem, as finding good matches becomes difficult for relatively "wide" X matrices. Many

observations are often left unmatched, while remaining "matched variation" is not rich enough to characterize treatment effect heterogeneity. Moreover, discarding unmatched observations can introduce a "matching bias," which cannot be signed without knowledge of the true distribution of heterogenous treatment effects. These curse of dimensionality problems stand out for Q1 and Q2, as X contains a number of variables that is large relative to that of observations.

To address these limitations, we turn to the causal forest estimator of Athey et al. (2018), which also draws inference by comparing observations that are similar in terms of X. The main functional difference with the matching estimator is that observations do not need to coincide on all the variables in X to be considered similar, but only on the subset thereof that drive treatment effect differences. This crucial capability is articulated by incorporating the random forests methodology (Breiman, 2001) into the estimation routine, whose role here is to "learn" which differences in X actually matter for treatment-control comparisons. In other words, the causal forest estimator does not operate in an skeptic way, and is therefore much less vulnerable to curse of dimensionality issues (no observations are dropped). Another important advantage of this method is that it returns heterogeneous treatment effects, which estimated at each point X (along with a confidence interval). Furthermore, these (X-)conditional average treatment effects (CATEs) are estimated non-parametrically, in that their distribution does not need to be assumed. This estimated variability allow us to inspect the validity of our hypothesis across different regions of the sample.

Like matching estimates, causal forest results indicate reactive licensing does in fact occur in our sample (Q1). However, the behavior is not adopted across the board. For those firms that react to P3Fs, the estimated effect is quite large on average. P3Fs increase the near-future (one year) probability of licensing by about 0.08, which represents an almost 50% increase relative to the baseline. Variation in the propensity to engage in reactive licensing appears to be primarily driven by a size effect: reactions are more likely among larger firms, with more active portfolios and robust market positions to defend. Some evidence for the "matching bias" mentioned above comes from noting that the matching average treatment effect is smaller than the average causal forest estimate. At the same time, observations dropped by the matching procedure are associated to generally larger conditional average treatment effect values.

The analysis of Q2 also supports the hypothesis, indicating that pre-licensing P3Fs can be linked to worse post-licensing development performance. However, as before, not all firms appear to suffer these consequences. When the effect has statistical significance, it implies that P3Fs increase the probability of post-licensing termination by about 15%. Importantly, the "performance" CATEs obtained from this analysis are not independent from "reactiveness" CATEs obtained from the analysis of Q1. In particular, their correlation suggests those that firms that are more likely to react to a P3T are also those more likely to suffer the consequences. This finding highlights the empirical relevance of the

reactive licensing trade-off, even though its occurrence can be circumscribed to specific circumstances. We discuss a series of potential mechanisms that may underlie the results of Q1 and Q2.

Our research connects to previous Marketing work on the management of innovation. Closest to our study, Chandy et al. (2006) present evidence suggesting that pharmaceutical firms that overly emphasize development speed risk higher attrition rates. The same general point is also made by Bayus (1997) and Smith (1999), among others. In turn, Griffin (1997) describes how project and process characteristics affect development cycle times. Beyond these authors' focus on internal rather than outsourced innovation, our context differs in that haste is circumstantially imposed (by P3Fs) rather than willingly adopted or environmentally given. Moorman and Miner (1998) develop the concept of improvisation in new product development, which they define essentially as decision-making without time for planning. Consistent with our results, they find that improvisation is more likely to occur in environments of turbulence while, at the same time, that it may reduce the effectiveness of innovation efforts.

Beyond Marketing, our work relates to the literature on Markets for Technology (Arora et al., 2004) or Markets for Ideas (Gans and Stern, 2010), for which licensing plays a central role. Arora and Gambardella (2010) emphasize that this literature has maintained a strong "supply side" focus. Our work addresses this gap, suggesting that the demand for and performance of licensed technologies can depend on the short-term pipeline contingencies faced by the demanding firm. Lastly, our study adds to a rich literature that analyzes the determinants of development performance in pharmaceutical innovation (e.g., Cockburn and Henderson, 2001; Dranove and Meltzer, 1994). Like in Guedj and Scharfstein (2004), our results raise the possibility that development attrition could be impacted managerial agency problems (a discussion of this point is presented in subsection 6.1). To the best of our knowledge, ours is the first paper in this literature to relate performance outcomes to the timing of decisions.

2 Industry background

2.1 The drug development process

This highly structured process starts when a compound's (drug candidate's) formulation is identified, and then fine-tuned in the lab. At this point, a set of potential therapeutic applications (i.e., compound/targeted disease combinations, henceforth called "therapies") are identified. These early formulation activities are followed by experiments on animal subjects ("pre-clinical trials"), aimed at assessing their potential. Our data source collapses these two set of activities into a single "Discovery" stage. Therapies with good enough potential move on to clinical trials on human subjects. If the development process

carries forward, each therapy requires a largely independent set of development activities.

The clinical trial protocol is sequential, composed of three phases, mostly based on randomized experiments. Phase 1 trials are designed to assess safety. Phase 2 trials are primarily used as proof-of-concept experiments to investigate efficacy. Phase 3 trials aim to provide definitive evidence of superior therapeutic value (thus usually take longer and require larger patient pools). Successful therapies can be presented to the regulator in application for a commercialization permit. For conciseness, we refer to this last stage simply as "review." As we mentioned in the introduction, drug development process is further characterized by a notoriously distinctive feature: its large development termination rates (attrition), which unfold across all stages. In subsection 3.3 we provide a detailed account of termination statistics, both within our sample and overall.

2.2 The licensing market and the incentives to license

Until the 1980s, pharmaceutical innovation was characterized by a high degree of vertical integration, in the sense that most innovation activities (discovery, development and commercialization) were carried out "in-house" within large pharmaceutical firms. This industry feature was rooted on the fact that drug discovery was primarily performed through "combinatorial" approaches (thousands of chemical combinations synthesized in search of symptom reduction), for which access to chemical libraries (storing results of previous experiments) was a crucial input. Due to their proprietary nature, these libraries acted as entry barriers into discovery activities, and so sustained the vertically integrated structure (Pisano, 2006). A series of scientific breakthroughs in the 1970s and 1980s created an alternative route to drug discovery, which in turn led to an important degree of vertical disintegration. Specifically, these advances enabled scientists to "engineer" (rather than "discover") new compounds. Chemical libraries became a less relevant input (and entry barrier) for drug innovation, and a fringe of biotech startups focusing on early-stage innovation activities was established.

Backed by venture capital, biotech startups are typically established by academic scientists, with the goal of "translating" their research findings into drug technologies. These firms have three primary ways to monetize their innovations: (i) licensing, (ii) "trade sales" (biotech startup is acquired by a larger company), and (iii) self-commercialization. Licensing means that the biotech firm sells the commercialization rights for an specific set of its developing therapies to another firm. These transactions are enabled by negotiated licensing contracts, which typically deliver most of potential compensation on a contingent basis (milestone payments and royalties on market revenues) as means to spread risk, mitigate informational problems, and insure the biotech firm's continued involvement in development activities (Mason et al., 2008).

A primary incentive for biotech firms to choose licensing over self-commercialization is benefiting from the (in-licensing) partner's "complementary assets" (Teece, 1986; Gans

et al., 2002). Besides funding, these capabilities may include "know-how" (e.g., regulatory affairs, implementation of clinical trials; Powell, 1996), as well as assets that are important for massive commercialization (e.g., branded reputation, established sales-forces; Levine, 2009). Thus, relative to self-commercialization, licensing is better suited for the development of therapies that target "large market" conditions, or those which require complex and/or costly clinical trials. From the point of view of the biotech firm's financial backers, licensing has the benefit of outlining a safer profile of returns (some compensation is received even if licensed therapies do not reach the market). Relative to "trade sales," licensing allows founding scientists and investors to retain control of the firm, while at the same time, participate in the financial upside of potential blockbuster therapies.[1] From the point of view of in-licensing pharmaceutical firms, licensing constitutes the a lean and expedite route to integrating cutting-edge advances into their pipelines.

Fueled by these motivations, technological outsourcing via licensing now plays a central role for drug innovation (Cockburn, 2004; Kneller, 2010), as is reflected by the large and growing global volume of trade, estimated at $40B, $43B, and $57B respectively for 2007, 2010, and 2016 (Giovannetti and Spence, 2017).[2]

2.3 Contracting frictions and proactive licensing

The acquired importance of licensing for drug innovation has come through despite the presence of important contracting frictions. Here we review their main sources, as well as the extensive efforts that in-licensors incur on to overcome them. The summation of these efforts embodies what we refer to as the "proactive" approach to licensing.

The main contracting frictions in this market are rooted on the progressively deepening scientific basis from which biotech innovation draws upon (Powell, 1996; Pisano, 2006). This scientific environment infuses the supply of licensable therapies with novel, diverse, and rapidly evolving technological approaches. For potential in-licensors, the resulting supply is not only difficult to navigate as they search for the "right" candidate, but also a fertile ground for informational asymmetries to flourish on.

Alcacer et al. (2008) illustrate the magnitude of costs and types of search activities that pharmaceutical firms engage in, as they strive to narrow down the set of "right" therapies to license. Drawing on evidence from interviews with pharmaceutical executives, these authors report that the major stages of the in-licensing process include a worldwide search and screening for licensing opportunities, followed by an in-depth analysis of each identified candidate, and a final screening. After screening candidates from 80-100 firms,

[1] Danzon et al. (2007) provide evidence suggesting that biotech firms adopt a "trade sale" as an exit strategy after encountering financing problems.

[2] These figures only consider deals that include US and European firms, and are expressed in "potential value" or "biobucks," which is the traditional yardstick used to measure drug candidate licensing activity. The key aspect of this metric is that it includes includes value attached to contingent payments, which may not be materialized.

this process can deliver one or two licensing opportunities. At this point, further due diligence, valuation and contract negotiations remain, all of which may take the better part of a calendar year or even longer (Truex, 2018).

The importance of due diligence arises in part due to the risk of adverse selection given the large informational advantages of biotech firms (Pisano, 1997; Hermosilla, 2016). In-licensors need to not only verify the integrity of the underlying science, but also insure that adequate intellectual property is available or underway, as well as the extent to which it could be challenged. Accordingly, pharmaceutical in-licensors deploy multifunctional teams (including lawyers, executives, and field-expert scientists) at every stage of the process (Alcacer et al., 2008). Jones (2007) highlights that due diligence and valuation activities can present significant challenges for in-licensors.

Furthermore, due to its novelty, much of the scientific knowledge embedded in biotech compounds is tacit at the time agreements are in place (Pisano, 2006). For this reason, licensing agreements require the continued involvement of biotech scientists in post-licensing development. These circumstances set the stage for potential moral hazard problems. For example, once an agreement is in place, biotech researchers may use available funds and time to advance their own (alternative) projects, or attempt to publish results related to the developing therapy earlier than the in-licensing firm would like to (Lerner and Malmendier, 2010). Licensing contracts thus include a wide variety of non-financial clauses (e.g., control, publication, residual ownership rights, etc.), carefully designed to maintain engagement while minimizing moral hazard-like behavior (Lerner and Merges, 1998; Lerner and Malmendier, 2010). Accordingly, the negotiation and design of contracts require high involvement, and often becomes a "drawn-out affair" (Giordano-Coltart and Calkins, 2007).

As a whole, these factors suggest that drug candidate licensing is afflicted by a number of frictions, including issues stemming from costly search and matching, as well as potential adverse selection and moral hazard issues. In-licensors seeking to minimize their negative consequences adopt a proactive approach, which includes careful preparatory market research (WIPO, 2015), extensive search and due diligence, and the careful design of contracts.

3 Data

3.1 Source, sample, and structure

Our main data source is Clarivate Analytics' *Cortellis* Competitive Intelligence. Cortellis is heralded as the most comprehensive and up-to-date repository of pharmaceutical innovation data. Information is obtained from company records, conferences, and other public sources, curated and updated daily by over 500 expert analysts. According to company

documentation, as of the third quarter of 2016, Cortellis included over 65,000 drug development histories and 48,000 deal reports, both of which date back several decades. We accessed the data in mid 2018.

From this source, we assemble a dataset focusing on the licensing and development activities of twenty of the largest pharmaceutical firms in the World, for the fifteen year period of 2001-2015. Firms entered the sample based on two criteria: (i) they were top in-licensors in the sample period (according to parameters described below), and (ii) firm-level financial information at the quarter level was generally available from COMPUSTAT. All but two of the selected firms are members of the Association of Pharmaceutical Research and Manufacturers of America (PhRMA), known as the trade association for big pharma firms.[3]

Firms in the sample actively developed therapies spanning 20 therapeutic areas. We retain data from 17 of these for which key variables have enough in-sample variation.[4] Because the economics and science of drug development has a large area-specific component, in our analysis we assume that licensing and termination decisions are made at the level of firm/therapeutic area pairs. For ease, we refer to these decision making entities simply as "units." Appendix Table A.1 lists the 230 units that appear in our sample. A unit enters the sample in the first quarter that we observe it actively developing a therapy.

It is important to note that, from the perspective of our analysis, there are essentially two datasets available. The first of these includes licensing and development histories of all therapies licensed by units in the sample, during the sample period. These data are the main focus of our study and are described in subsection 3.2 and 3.3. The second dataset includes remaining licensing and development records available from Cortellis. These secondary data will be used extensively with two primary purposes: (i) to codify P3F incidence (subsection 3.4) and construct an auxiliary variable that we call the "P3F risk score" (subsection 4.2), and (ii) to construct pipeline-related controls (subsection 4.1).

3.2 Licensing activity

Here we cover the main characteristics of the set of licensed therapies that is the focus of our analysis. An important first antecedent is that most licensing records available from Cortellis do not reflect the type of technological outsourcing model that we study. We focus on the category of licensing that is labelled as "Drug - Development/Commercialization" by Cortellis, which provide a tight empirical correlate for the technological outsourcing

[3]The list of firms is: Abbvie, Amgen, Astellas, Astra Zeneca, Bayer, Biogen, Bristol-Myers Squibb, Celgene, Daiichi Sankyo, Eli Lilly, Gilead, GlaxoSmithKline, Johnson & Johnson, Merck & Co, Novartis, Pfizer, Roche, Sanofi, Shire, and Takeda. Gilead and Roche are not PhRMA members (as of 10/2018).

[4]The list of covered therapeutic areas is: Autoimmune/Inflammatory, Bone Disease, Cancer, Cardiovascular, Central Nervous System, Dermatological, Endocrinological & Metabolic, Gastrointestinal, Genitourinary/Gynecological, Hematologic, Infectious, Ophthalmic, Psychiatric, Renal, Respiratory, Transplantation, and Other/Unclassified.

model that we are interested in.[5]

Licensing agreements vary significantly in terms of the geographical coverage of traded commercialization rights. In some cases, these rights are for specific countries or regions, in others, they are worldwide. We only consider worldwide licensing deals in our sample. In addition, as we have indicated above, licensing deals may bundle rights for more than one therapy (in our sample, the average is about 2.5), which may span across different therapeutic areas. For tractability, our analysis treats the licensing of each therapy as a individual licensing event. Lastly, because financial compensation terms are rarely available from the data, they play no part in our analysis.

After imposing these filters, we are left with 3,495 therapies licensed by units in the sample. About 48% of these belong to the cancer area. Licensing of therapies in the endocrinological & metabolic, and autoimmune/inflammatory are the distant followers (with 9% and 8% of licensed therapies, respectively). Most licensing happens early in the process: 25% of therapies were licensed before clinical trials and 62% before reaching Phase 3. The average unit in-licenses an average of about 2 therapies each year that it appears in the sample. Among the units that license more than once (little over half of units), the average time between licensing is little about 10 months.

Figure 1 presents some aggregate trends that help place this licensing activity into context. The solid line shows the number of actively developing therapies maintained by sample units across the sample period (right axis). The short dash-dot line at the bottom shows the percentage that these therapies represent out of the industry's total (left axis). Together, these trends tell us that the industry underwent non-stationary growth over the covered period, and that our sample accounts for a significant and stable percentage (about 15%) of the total. The dashed line tracks the percentage of therapies added to the portfolios of units in the sample, which were obtained via licensing.[6] By this trend, licensing constituted an important and relatively stable source of all portfolio additions for sample units, about 30% in the first half the sample and 20% in the second half (the drop that coincides with the 2008 recession). At the same time, therapies licensed by units in the sample represented between 30% and 40% of those licensed in the industry (long dash-dot line). Thus, our sample accounts for a non-negligible share of overall licensing activity.

[5]Cortellis documentation describes this category of deals by indicating that the "Partner [in-licensing] firm acquires a license from Principal [out-licensing] firm to develop and commercialize (sell) drug(s)." (Excerpt taken from Cortellis documentation. Bracketed comments added.). Additional types of licensing deals include early-stage joint ventures, subcontracting of manufacturing and research services, licensing of drug-delivery technologies, among others.

[6]The total number of portfolio additions also includes therapies discovered in-house, as well as those incorporated via M&As and join ventures.

3.3 Post-licensing development performance

We assemble post-licensing development histories for 2,787 of the therapies in the sample described above.[7] For them, we observe a total of 3,829 therapy/development stage pairs. About 27% of these development stages were still ongoing at the time of data download (outcomes not observed). For an additional 25%, we were not able to assemble the full list of covariates used in the analysis (described in subsection 4.1). Our analysis utilizes data from the remaining 1,777 "resolved" stages in the data, for which we observe both an outcome and all covariates. Column (1) of Table 1 shows their distribution across stages. There is good coverage for all stages in the process.[8]

We measure development performance at the therapy/stage level, through a variable that we call ADVANCE. This variable equals one if a therapy advances past a given stage (i.e., enters the next), and zero otherwise (termination). Column (2) presents this variable's average across stages. As we discuss below, these figures are comparable to benchmark results from the literature.

We identify three forms of termination from the data. Even though this distinction is irrelevant for our main analysis (which relies solely on ADVANCE), it is useful for a secondary analysis and to give context. To understand the differences between termination types, first note that a licensed therapy's development termination may or may not imply that the underlying licensing agreement is terminated (parties may continue to cooperate on the development of related therapies). Sometimes, it is also the case that licensing agreements are terminated without it implying the therapy's development termination, for example, if parties enter conflict and their collaboration is "derailed."

Column (3) shows statistics for the main type of termination in our data (which also corresponds to the type of attrition emphasized by specialized literature). We label these simply as "development" terminations. In terms of our codification, the defining trait of a "development" termination is that a therapy's development is declared terminated before the underlying agreements are. Column (4) shows rates of "agreement" terminations. These are situations in which the underlying licensing agreement is terminated before development outcomes are observed. Even though therapies in question may continue to be developed (by other firms), they are no longer part of the in-licensing unit's pipeline (which warrants treating them as terminations). In our data, these are a distant second in terms of frequency. Lastly, in Column (5), we present a type of termination that we call "idle state" termination. This case reflects a common (though rarely acknowledged) feature of pharmaceutical innovation datasets. Namely, the case where an outcome is not

[7] Cortellis licensing and development data are kept in separate data repositories, and the website does not (as of mid 2018) allow the user to systematically bridge between the two. We performed the matching procedure manually. Therapies that we observe in the licensing sample but for which no match was found were left out of the post-licensing performance sample.

[8] The larger number of observations for Phase 2 than Phase 1 is largely explained by a nuance of the drug development process. This is, many therapies "skip" Phase 1, if safety has been already demonstrated for other therapies of the same compound.

observed, despite data telling us that a therapy has spent an abnormally long period of time being developed in a particular stage.[9] Fortunately, these terminations are quite rare.

To investigate the external validity of performance statistics in our sample, we searched for and reviewed articles that report rates of success in pharmaceutical development.[10] These articles have important differences in terms of the used samples (number and types of firms, time periods), as well as some discrepancies in the way variables defined. Most importantly, none of these articles considers a sample that is purely composed of in-licensed therapies. Thus, comparability is limited. With this caveat in mind, we found that, while Phase 2 and 3 rates were close to the lower bounds of the assembled distributions, the statistics of Column (2) in Table 1 are within range of reported results.[11] The relatively lower advancement rates for Phases 2 and 3 in our data could be in part reverted by accounting for the advancements registered within the outcomes that we label as "agreement" terminations.

3.4 Phase 3 Failures (P3Fs)

We identified a total of 418 P3Fs experienced by units in our sample, during the covered period. While some of these correspond to therapies included in our analysis sample (described above), the majority (about 80%) occur for other therapies, incorporated into pipelines through different means (in-house discovery, M&As, etc). The distribution of P3Fs across time and areas is shown by Table A.2. Consistent with the industry's growth over the covered period, their frequency is increasing over time. Even though the Cancer area is a distant first in terms of frequency, the distribution is quite spread out across other areas.

We also note that, in addition to the tight scrutiny of investors, patients, and specialized media, most P3 trials are monitored by independent "data monitoring committees" (Freidlin et al., 1999). These committees review incoming trial data in an approximate real-time fashion, and provide recommendations on how to continue the trial. At any time, these entities may recommend that the trial is stopped (as it happened for Pfizer's Torcetrapib). Thus, sponsors have in general little scope to engage in strategic P3F information release.

[9] We say that a therapy is terminated by entering an "idle state" at a given stage and date t, if: (i) we don't observe an outcome for that therapy at that stage, and (ii) at date t, the therapy has spent an abnormally long period of time (95^{th} percentile of time-to-stage completion in the broader Cortellis sample) at that stage.

[10] The list of surveyed papers is: DiMasi et al. (2003, 2010, 2016), Kola and Landis (2004), Abrantes-Metz et al. (2004), Pammolli et al. (2011), Hay et al. (2014), Arrowsmith (2011), and Waring et al. (2015).

[11] Summarizing the distribution of surveyed success rates through triplets (min,mean,max), we obtained (0.31,0.61,0.81) for Phase 1, (0.18,0.4,0.58) for Phase 2, (0.48,0.58,0.75) for Phase 3, and (0.77,0.87,1) for Review. There were no statistics for the Discovery stage.

4 Empirical Strategy

Both sets of analysis announced in the introduction refer to the causal impacts of P3Fs. Q1 (engagement with reactive licensing) posits that a P3F may increase a unit's near-future probability of in-licensing, whereas Q2 (consequences of reactive licensing), that pre-licensing P3Fs may translate into poorer post-licensing performance. We investigate these hypotheses using a causal inference framework, which views P3F incidence as the "treatment."

Because the adequacy of the implemented methodologies is best appreciated with an understanding of the available controls (or covariates) X, we begin by describing their construction. We proceed by introducing the "P3F risk score," and finalize with a methodological overview of the considered estimators.

4.1 Covariates X

Our hypotheses focus on two outcomes: licensing incidence (Q1) and development performance (Q2). Here we seek to specify a set of independent variables (or controls, or covariates) X that includes leading confounders for these outcomes. The list comes from insights of a broad literature.

A key reference in this regard is Chan et al. (2007), who theoretically analyze the interplay of termination and licensing decisions. These authors suggest that these decisions may depend on a sponsor's pipeline strength (i.e., expected number to reach the market in short/medium-run), as well as on their portfolio of marketed therapies. The importance of pipeline strength is also highlighted by Girotra et al. (2007), who find that stronger pipelines soften the negative stock price impact of P3Fs.

Following Henderson and Cockburn (1994) and Cockburn and Henderson (2001), Danzon et al. (2005) and Arora et al. (2009) empirically analyze the development performance of licensed therapies. They draw attention to the role of sponsors' development experience and diversification as determinants of innovative productivity. These constructs are captured by the so-called "scale" and "scope" variables. In line with a broader literature investigating the "demand pull" effect on pharmaceutical innovation (Acemoglu and Linn, 2004; Finkelstein, 2004; Dranove et al., 2014; Dubois et al., 2015), Arora et al. (2009) also emphasize the impact of market potential on termination decisions. More recently, Hermosilla and Wu (2018) extend this line of work by showing that larger market potential may also increase the likelihood of licensing.

A last set of papers focuses on technological outsourcing decisions, but in the form of M&As (rather than licensing). We incorporate insights from this line of research on the basis that similar effects could also apply to licensing. Higgins and Rodriguez (2006) show that (declining) recent productivity may lead firms to acquire other firms as means

to "revitalize" their pipelines. Cunningham et al. (2018) present evidence supporting a somewhat opposite idea, namely, that pharmaceutical incumbents may both acquire firms and "kill" their developing therapies to guard against threats to their market shares.

In summary, these antecedents suggest that licensing and performance may depend on a broad set of factors, including sponsors': (i) portfolio of commercialized therapies, (ii) pipeline strength, (iii) recent productivity, and (iv) market potential of developing therapies. Using Cortellis data we construct variables to specifically address (i)-(iii). We further develop "competitive" versions of some of these variables, as described below. From COMPUSTAT, we retrieve firm-level financials (sales, assets, R&D expenditures), which help us control for effects stemming from liquidity, capabilities, size of internal capital markets, overall technological leadership, among others. (Size-related effects are also highlighted by Arora et al. (2009).) To address (iv) (market potential), we rely on therapeutic area- and year-specific indicators. Development stage and area indicators are used to account for baseline differences in performance across each respective dimension.

We face two practical issues to adequately implement this plan. First, we require a scheme to consistently weight the importance of quantities associated to different development stages. For example, because a therapy that is being developed at Phase 3 is much more likely to reach the market than one being developed at the Phase 1 (see table Table 1), we should judge a pipeline composed of a single Phase 3 therapy as being stronger than one composed of a single Phase 1 therapy. Similarly, we should register a poorer productivity record when recent terminations have occurred at more advanced stages. A second concern is based on the patterns of Figure 1, which shows non-stationary industry growth over the covered period, while at the same time, a relatively stable importance of licensing. If covariates X are defined in absolute terms, we run the risk of estimating our models mainly on temporal variation.

The approach used to construct variables combines weighting and normalization steps to account for these issues. We exemplify the procedure focusing on a representative variable, PIPESTR, which tracks a units' pipeline strengths. We start by defining a chronological measurement window. For Q2, this window corresponds to the calendar year ending the day t for which a specific milestone (licensing, stage completion) is reported. Next, for each unit j, we count the number of unique therapies that were active on each stage ($s = 0, .., 4$) over the window. Denote these counts by $\{A_{jts}\}$.[12]

The weighting step is carried out using "reach-the-market" probabilities $\{p_{ast}\}$ as weights. These represent the likelihood that a therapy of area a, which is at stage s in time t, continues its development all the way into the market (as opposed to failing along the way). We compute them using the large number of development histories available from Cortellis, on a rolling basis (the reason for the t subindex), as described in Appendix

[12]In rare cases, a therapy is active on two different stages over the measurement window. We use the most advanced stage to codify these cases.

A. With these elements we compute:

$$A_{jt} = \sum_{j} A_{jts} \cdot p_{a(j)ts}, \tag{1}$$

were the variable is expressed in terms of a "approved therapies-equivalent" scale. In the normalization step, we compute PIPESTR as $A_{jt}/(A_{jt} + A_{-jt})$, where A_{-jt} represents the "rest-of-area" figure (including data from all units in the broad Cortellis sample). The resulting variable is bound to the unit interval, and equal to zero if and only if j had no therapies on active development over the measurement window. The normalization with respect to each area's current "size" makes the variable robust to the uneven development of areas across time.

Panel A of Table 2 presents a summary of the generated variables, including a formal definition, the level over which variability unfolds, and associated construct. Using the weighting procedure of (1), we constructed the "approved therapies-equivalent" variables T (terminations other than P3Fs) and L (licensing), both measured within the same window as A. With these, we assembled RTER_OWN and RLIC_OWN to track recent productivity and licensing activity, respectively. From the insights of Krieger (2017), who argues that P3Fs can impacts other sponsors' termination decisions, we also generate a variable that measures "rest-of-area" recent productivity, RTER_RA. RLIC_RA is specified analogously, to help us to control for competitively-motivated licensing effects, as well as the possible influence of "licensing waves" following the establishment of new technological approaches.[13] Lacking access to revenue data, we define MKTSHARE as each unit's share of approved therapies.[14] Notice that, because PIPESTR and MKTSHARE are normalized to the unit interval, defining their "competitive" counterparts would be redundant. SCALE and SCOPE variables are defined as in previous research, in a way that larger values reflect larger accumulated experience (SCALE) and higher degrees of specialization (SCOPE). Financial variables (retrieved from COMPUSTAT at the quarterly level) are log-transformed after CPI-adjusting them to year 2010. To impose a temporal stationarity, we normalize them by subtracting the in-sample yearly mean.

4.2 P3F Risk Score

One identification challenge is that different units are not equally likely to experience a P3F at a given time. One reason is that, all else constant, a unit's likelihood of experiencing a P3F increases with the number of therapies on active P3 development ("P3D"), a number that varies across units and time. A second reason is rooted on the microdynamics of P3 clinical trials: P3Fs are more likely to occur the longer the therapy has

[13] For example, genetically-targeted therapies following the 2003 unveiling of the Human Genome (Hermosilla and Lemus, 2018).

[14] The number of approved therapies is tracked by the variable M, which does not use probability weighting

been on active P3D. Here we introduce a variable that we call "P3F risk score," which naturally combines these two sources of variation. Both of the considered estimators utilize this variable to account for disparities in the likelihood of experiencing a P3F (i.e., selection into treatment).

We define the P3F risk score $R_j(t_1, t_2)$ as the probability that unit j experiences a P3F during the time interval $[t_1, t_2]$.[15] To formalize it, we denote a unit j's portfolio of therapies on active P3D at time t_1 by $\mathcal{A}_j(t_1)$. In addition, by $r_k(t_1, t_2)$ we denote the probability that a therapy k introduced to P3D at an earlier time $\underline{t}_k$ fails during $[t_1, t_2]$. With these elements, the risk score is defined as:

$$R_j(t_1, t_2) = \begin{cases} 1 - \prod_{k \in \mathcal{A}_j(t_1)} 1 - r_k(t_1, t_2) & \text{if } |\mathcal{A}_j(t_1)| > 0 \\ 0 & \text{otherwise (empty } \mathcal{A}_j(t_1)) \end{cases} \tag{2}$$

Thus, R corresponds to the probability complement for the event in which none of the therapies of $\mathcal{A}_j(t_1)$ fails during $[t_1, t_2]$. The key ingredient in this formulation is r_k. By its definition in the previous paragraph, r_k corresponds to a duration "hazard" or quantity. One difference with the standard "hazard" quantity of Cox (1972) is that r_k is not instantaneous. Rather, it corresponds to the integral of an instantaneous hazard quantity over the time interval of interest. Furthermore, r_k is not only defined by trial duration (i.e., the event "P3 trial ends"), but also by the reason behind the trial's completion (i.e., the more specific event "P3 trial ends due to the project's failure"). Thus, to compute r_k we require a cause-specific hazard estimate, which we can then integrate over the time interval of interest. The competing risks framework of Fine and Gray (1999) provides us with "sub-hazard" estimates that fit this description.

From competing risks estimates obtained from data on P3 trial durations and outcomes (details provided below), we assemble functions $F_k(t)$. These functions represent the probability that a therapy k that was introduced to P3 at time $\underline{t}_k$ fails at or before time t. Panel A of Figure 2 illustrates one such function. According to this curve, therapy k has probability $F_k(t_1)$ of failing at P3D at or before time t_1, whereas its probability of failing anytime after t_1 is $F_k(\infty) - F_k(t_1)$. Note that, because not all therapies fail at P3, these probabilities do not add up to one. Instead, they add up to the overall probability of failure in P3, $F_k(\infty)$. With this, r_k is computed as the fraction of all failure mass $F_k(\infty)$ that has not been realized by t_1, but is expected to unfold during $[t_1, t_2]$. That is,

$$r_k(t_1, t_2) = \frac{F_k(t_2) - F_k(t_1)}{F_k(\infty) - F_k(t_1)} \cdot F_k(\infty)$$

We fit the competing risks model to the over 9,000 P3 clinical trial observations (du-

[15] For simplicity we are using the continuous-time notation $[t_1, t_2]$ to reference the discrete-time interval $\{t_1, .., t_2\}$.

ration and outcomes) available from Cortellis, excluding therapies that appear in our analysis sample. The model specification allows for variability across areas (area-specific indicators) and time period (five year-period indicators, 2001-2005, 2006-2010, etc.). To illustrate these results, Panel B of Figure 2 presents F estimates for the 2006-2010 period, where variability captured by boxes comes from differences across therapeutic areas.[16] Lastly, Panel C of the same Figure presents the (cumulative) distributions of P3F risk scores for all therapies in our analysis sample. P3 portfolios $\mathcal{A}$ are measured over the one year window that ends the day each therapy was licensed. To provide a sense of validation, we split this distribution by whether a P3F was observed over the measurement window. As expected, risk scores are generally larger for the P3F than the no-P3F distribution. Whereas about 40% of no-P3F therapies are associated to null risk scores, the minimum value risk score at which a P3F is observed is about 0.1.

4.3 Estimators

The causal inference problem is cast in the terms of the potential outcomes framework. Let i index observations and consider a dependent variable (or outcome) $Y_i \in \mathcal{Y}$, a dichotomic treatment variable $W_i \in \{0, 1\}$, and a covariate vector $X_i \in \mathcal{X}$. The outcome is conceptualized as a free-form function of the treatment, $Y_i = Y_i^{(W_i)}$. Our goal is to estimate conditional average treatment effects (CATEs), $\tau(x) = E[Y_i^{(1)} - Y_i^{(0)} | X_i = x]$. The key difficulty is that, for each i, we only observe one of the two potential outcomes: depending on the value of W_i, we observe either $Y_i^{(1)}$ or $Y_i^{(0)}$. Both the matching and causal forest estimators approach this problem relying on a set of "close" or "similar" observations in the covariate space $\mathcal{X}$—a "neighborhood."

4.3.1 Matching Estimator

The matching estimator (MTE) approaches the causal inference problem by relying on a counterfactual outcome, $\hat{Y}_i^{(1-W_i)}$. This value is generated by first determining a neighborhood set $\mathcal{N}(x) = \{i' : X_{i'} \text{ and } x \text{ are similar enough}\}$ for each $X = x$, and then averaging Y within the subset of observations in this set that receive i's opposite treatment, $1 - W_i$. There exists a wide variety of MTEs, which primarily differ in their implementation of what it means for two covariate vectors to be "similar enough." Here we will discuss the *radius* matching estimator, which allows us to communicate key points in the most general and simple way. Our implementation will use the closely related *k-nearest neighbor* estimator.

For the radius MTE, two vectors $X_{i'}$ and $X_{i''}$ are similar enough if the vector distance $d = ||X_{i'} - X_{i''}||$ is smaller than some constant $\kappa > 0$. Figure 3 illustrates what the

[16] For illustrative purposes, this Figure only considers F values at the day that each year begins. However, our implementation of the P3F risk score takes advantage of day-by-day variation (i.e., t corresponds to days).

neighborhood would look like for an observation $X = x$ (marked with a cross). Potential outcomes are computed as:

$$\hat{Y}_i^{(W)} = \begin{cases} Y_i & \text{for } W = W_i \\ \frac{1}{|\{i' : W_{i'} = 1 - W_i, i' \in \mathcal{N}(X_i)\}|} \sum_{\substack{i' \in \mathcal{N}(X_i), \\ W_{i'} = 1 - W_i}} Y_{i'} & \text{for } W = 1 - W_i, \end{cases}$$

where the second case represents the counterfactual outcome. Naturally, counterfactual outcomes are undefined if $\mathcal{N}$ is empty or there are no opposite-treatment observations in it. Observations in that situation are dropped. If there is a large enough number of matched observations at a given $X = x$, $\tau(x)$ can be estimated as the average of $\hat{Y}_i^{(1)} - \hat{Y}_i^{(0)} | X_i = x$.

Remark MTE1 (Covariates X and the curse of dimensionality). Even though some of the variables in X may not impact outcomes, the distance function d penalizes deviations of all variables in X equally. In this sense, the matching estimator is "skeptical," as it assumes that all variables in X ("suspects") are a source of bias ("guilty"). The analyst is responsible for feeding the estimator the "right" set of covariates. This feature of the estimator is often responsible for a curse of dimensionality problem. Analysts seeking to avoid omitted variable bias may increase the number of variables in X. It then becomes increasingly difficult to find good enough matches. As a result, the number of unmatched observations increases quickly with the width of X. This problem is particularly burdensome in our context, where the number of variables in X is large relative to the number of observations.

Remark MTE2 (Homogeneous treatment effect and matching bias). By the curse of dimensionality problem, matching procedures often fail to produce a large enough number of matched observations to allow for the estimation of (conditional) treatment effects at each $X = x$. As a result, researchers are most times forced to neglect treatment effect heterogeneity and settle for an (unconditional) average treatment effect (ATE). Theoretically, the ATE can be formulated as a frequency-weighted aggregation of CATEs. This is, $\sum_{x \in \mathcal{X}} \tau(x) \cdot \Pr(X{=}x)$. However, the estimated ATE may depart from this formulation if there are unmatched observations. In particular, if unmatched observations are heavily drawn from low-$\tau(X)$ regions of $\mathcal{X}$, the ATE will be overestimated (and vice versa). Some evidence suggests that this "matching" bias is at play in our sample.

Remark MTE3 (Selection into treatment and P3F risk). The consistency of estimated treatment effects relies on the assumption of unconfoundedness, formally stated as $W \perp\!\!\!\perp \{Y^{(1)}, Y^{(0)}\} | X$ (e.g., Imbens and Rubin, 2015). In practice, this condition assumes that treatment assignment is random among observations close in $\mathcal{X}$. Unconfoundedness is commonly enforced using propensity scores (Rosenbaum and Rubin, 1983), which in practice correspond to (first stage) estimates for the probability or receiving the treatment. Propensity scores π provide a natural platform to incorporate P3F risk score

(R) variability, so they will play an important role in our analysis. We will estimate them as $\pi = \hat{\Pr}(W = 1|R, X^P, X^F)$, also allowing for variation across therapeutic areas and time. Thus, the resulting π values give the probability that a unit j experiences a P3F over a determinate time window, as a function of the characteristics of its P3 portfolio and P3D micro-dynamics, as well as of the environmental factors picked up by X^P and X^F.[17]

Propensity scores can enter the analysis (and often do) as a one-for-one replacement for the distance function d (e.g., Dehejia and Wahba, 1999). In this case, the implicit assumption is that covariates X impact selection into treatment but not outcomes. By our review of subsection 4.1, this assumption is inadequate in our context. Our approach is then to match each observation to its nearest neighbor based on X, while demanding that matched observations have the same propensity scores value after rounding it to half decimals (i.e., increments of 0.05), and belong to the same therapeutic area.[18]

4.3.2 Causal Forest

The causal forest estimator (CFE) advances a similar set of principles as the MTE by leveraging the random forests technique of Breiman (2001). As we mentioned in the introduction, a key advantage of the CFE is that it returns a set of conditional average treatment effects (CATEs), estimated at each point $X = x$ without requiring distributional assumptions. While the original version of the CFE was introduced by Wager and Athey (2018), here we present a more recent version, implemented as a special case of the Generalized Random Forests (GRF) framework of Athey et al. (2018). We choose this implementation for its generality, but also because it outperformed the original implementation in Monte-Carlo experiments (Athey et al., 2018). Closely related methods have been recently used in Marketing by Ascarza (2018) and Guo et al. (2017). Our exposition begins with a brief overview of the GRF framework, then moves on to details that are specific to the CFE.

GRF exploits local moment conditions $\psi(\cdot)$ to estimate heterogeneous parameters of interest $\tau(x)$, as well as optional nuisance parameters $\mu(x)$ (e.g., constants) at each point $X = x$. The method fits equations of the form $\mathbb{E}[\psi_{\tau(x),\mu(x)}(X_i, W_i, Y_i)|X_i = x] = 0$, $\forall x$. Each specific pair of X-conditional estimates $(\hat{\tau}(x), \hat{\mu}(x))$ is obtained as the solution to:

$$\min_{\tau(x),\mu(x)} \left\| \sum_{i=1}^{N} \alpha_i(x) \psi_{\tau(x),\mu(x)}(X_i, W_i, Y_i) \right\|_2, \tag{3}$$

[17] Experience-related covariates X^E are not suspected of impacting licensing decisions, so they are left out from these estimations.

[18] The "transformed outcomes" approach offers another alternative to incorporate propensity score variation into the analysis (Hirano et al., 2003). Here, propensity scores are used to compute the transformed outcome $Y_i^* = W_i \frac{Y_i}{\pi_i} - (1 - W_i) \frac{Y_i}{1-\pi_i}$, which is then used to compute treatment effects. Because propensity scores enter the denominators of Y^*, this estimator can be unstable. In our data, this approach delivered implausibly large ATE estimates.

where the summation is over all observations i in the sample. The fundamental role of random forests is to produce the set of weights α, which indicate how similar each observation i is to one at the test point $X = x$. We describe the construction of these weights in short. For now, note that if the treatment W was randomized, one could obtain CATEs by setting $\psi_{\tau(x),\mu(x)}(X_i, W_i, Y_i) = Y_i - \mu(x) - W_i \cdot \tau(x)$. As with the generalized method of moments, different models can be estimated by appropriately choosing ψ.

To see how weights α are computed, first note that a random forest uses the results of a series of classification/regression trees $b = 1, .., B$ (B usually in the thousands), which are grown by randomly choosing estimation ("training") subsamples and classification ("splitting") variables. Random forest results are then computed as averages across this ensemble of trees. Each tree estimates quantities with little bias but high variance, averaging quantities across many trees stabilizes these estimates.

Panel A of Figure 5 presents the dendogram representation for a sample tree classification. In this case, the covariate space has been split according to the values of two variables, X_1 and X_2. The resulting partition is composed of three terminal nodes, or "leaves" ($l_b = 1, 2, 3$). Panel B shows what this partition looks like in the covariate space. These results tell us that, according to the tree's optimally-determined partition, all observations within leaf $l_b = 3$ are similar to the test observation (marked with a cross), while observations on other leaves are not similar to it. Thus, at the tree level, similarity is dichotomic. The optimality criterion used to determine when a split is necessary purports to best explain the data (in validation samples) while, at the same time, maximizing the amount of detected heterogeneity in the parameter of interest.

The set of weights α used in (3) are built from tree-specific weights. These are computed as the relative participation that an observation i has on the leaf that contains the test point $X = x$. For the tree b, this is $\alpha_{ib}(x) = \frac{\mathbf{1}[X_i \in l_b(x)]}{|l_b(x)|}$. The definitive weights are continuous variables, which result from averaging tree-specific values, $\alpha_i(x) = \frac{1}{B} \sum_{b=1}^{B} \alpha_{ib}(x)$. As an illustration, Panel A of Figure 5 presents a series of individual trees. Across trees, an observation i may or many not belong to the same leaf of the test observation $X = x$. If it does, its weight for estimate $\tau(x)$ increases by $1/|l_b(x)|$. The Figure's Panel B represents final weights by the size of each marker.[19]

In this context, the CFE is obtained through appropriate choice of the local moment function ψ. Based on earlier work (summarized by Nie and Wager, 2017), Athey et al. (2018) use:

$$\psi_{\tau(x)}(X_i, W_i, Y_i) = \underbrace{\big(Y_i - m(X_i)\big)}_{\substack{\text{Locally centered} \\ \text{outcome}}} - \underbrace{\big(W_i - \pi(X_i)\big)}_{\substack{\text{Locally centered} \\ \text{treatment assignment}}} \cdot \tau(x),$$

where $m(X_i) = \mathbb{E}[Y|X_i]$ and $\pi(X_i)$ is the propensity score, both of which result from purely predictive tasks performed by standard prediction forests. According to this specification, CATE estimates $\tau(x)$ are obtained from a (α-)weighted regression of "locally

[19] A similar figure is presented by Athey et al. (2018).

centered" outcome and treatment assignment variables.

Remark CFE1 (Optimal splits and the curse of dimensionality). Because classification splits are introduced optimally in each tree, the CFE does not operate with skepticism. This property makes it much less vulnerable to curse of dimensionality issues, compared to the MTE. In our context, for example, it could be that advancement rate differences between treated and non-treated observations is about the same across all therapeutic areas. In this sense, therapeutic area information would not "matter," and the CFE would learn to ignore it. By the same principle, the CFE may learn that pipeline strength differences do not "matter" continuously, but only if a certain threshold is crossed (as in the example of Figure 5). In contrast, the MTE operate on the premise that all these differences "count" and should be matched away.

Remark CFE2 (Implicit neighborhoods). Weights α summarize the results of the exercise described in the previous remark, namely, determining what differences in X are meaningful with respect to differences in mean outcomes. It can thus be said that, for a test observation $X = x$, the program (3) forms an implicit neighborhood by heightening (minimizing) the relevance of those observations that are close (distant) to it, only based on differences that "matter."

5 Results

5.1 Do P3Fs increase the near-future probability of licensing? (Q1)

The first step to address this question is to format the data. Our approach is summarized by Panel A of Figure 6. This Figure illustrates the measurement windows used to construct a sample observation, for unit j at quarter q. We measure the incidence of a P3F (W), P3F risk score (R), and quarterly financials (X^F) only using data of quarter q. To construct pipeline-related covariates (X^P), this one-quarter window is too narrow, however. This is because the impact (or determinants) of recent terminations and licensing are likely to be felt over several quarters. Thus, we measure covariates X^P over the wider window $\{q-3, .., q\}$.[20] The dependent variable, DLICENSE, equals one if unit j in-licenses at least one therapy during quarters $\{q+1, .., q+4\}$, and zero otherwise. This approach results in 5,071 observations at the jq level, about 8% of which is associated to a P3F (W=1).[21] The hypothesis that pharmaceutical sponsors engage in reactive licensing would be supported if W's impact on DLICENSE is positive.

[20] Note that for covariates that rely on a count of unique active/marketed therapies (PIPESTR, MKTSHARE), a wider window makes little difference, as units do not drastically alter the composition of their pipelines from one quarter to the other.

[21] A control observation for quarter q may be "contaminated" if the conjectured impacts of a P3F span the period $\{q-3, .., q\}$, or if a P3F is observed during $\{q+1, .., q+4\}$. These "contaminated" control observations are dropped from the sample.

DLICENSE averages 0.32 among treated observations (W=1) and 0.19 among non-treated ones. This comparison supports the idea that firms engage in reactive licensing. Column (1) of Panel A in Table 3 presents the results of a formal test for the difference of these means. The difference of 0.13 is highly significant ($p < 0.01$). Taking the no-P3F licensing probability as a base, this difference suggests that P3Fs increase near-future licensing probabilities by about 68%. This estimate, however, is afflicted by a major source of (upward) bias. Namely, units that innovate more also experience higher project turnover, which increases their inherent propensity to both experience P3Fs and in-license therapies. In Column (2) we present an estimate that alleviates this problem to some extent. This estimate corresponds to W's coefficient in a linear probability regression, which also includes X^P and X^F variables as controls, along with year and therapeutic area fixed effects. The size of W's estimated effect on DLICENSE shrinks by about half, while retaining statistical significance.

In Columns (3) and (4) of the same Panel we present ATE estimates from two versions of our matching estimator (MTE). In contrast to estimates of Columns (1) and (2), these estimates account for non-random treatment assignment (i.e., P3F incidence) by incorporating propensity score variation. We implement two MTE versions to illustrate the role of the P3F risk score (R): the propensity score used in Column (3) does not include R as a predictor; that in Column (4) does. Both propensity scores are estimated via Logit.

We estimate these specifications on the subsample of observations such that $R \geq 0.1$, which is the minimum approximate value of R at which we start to observe P3Fs. By imposing this overlap restriction, we are eliminating control group observations that are too unlikely to produce a P3F, and should thus not enter the comparison (most of these are due to empty P3 portfolios). Given that there are about 990 observations in the $R \geq 0.1$ subsample, and that about 890 of these are used in Columns (3) and (4), we can conclude that the MTE's "skepticism" is responsible for the loss of about 10% of observations in this case.

The estimated coefficient of Column (3) suggests that P3Fs increase the near-future licensing probability by 0.1. In Column (4), where R is accounted for by the propensity score, this estimate falls to 0.06. Taking as a base the 0.19 average of DLICENSE in the W=0 subsample, this estimate suggests that a P3F increases the near-future probability of licensing by about 32%.

Causal Forest CATE estimates are obtained from the full $R \geq 0.1$ subsample (no observations are dropped). Moreover, by capturing treatment effect heterogeneity, these estimates provide a richer set of results. The additional nuance manifests itself both in terms of the value of estimates and their statistical significance. To simultaneously represent both these dimensions, we present results in the form of a scatter plot, in Panel A of Figure 7. CATE magnitudes are presented in the horizontal axis; their z-scores,

on the vertical one (estimates have a normal asymptotic distribution). The difference between marker types is explained below as it becomes relevant. Note that CATE values can be interpreted as the willingness to engage in reactive licensing (as a function of X), shall a P3F be experienced. We refer to this construct simply as "reactiveness."

The average value of CATE estimates is about 0.07. Even though this value is close to the benchmark estimate of Column (4), it hides the fact that only a fraction of CATE estimates are statistically significant. About half of them are significant with 90% confidence and a quarter with 95% (dotted lines mark significance thresholds). Thus, while this evidence suggests that reactive licensing is not a rare behavior, it also suggests that we should not expect every unit to engage in it following every P3F. In subsection 5.3 we investigate the drivers of this heterogeneity. Only considering the half of the sample for which CATE estimates have some statistical significance, these results imply that P3Ts increase the probability of near-future licensing by an average of 47%. This estimate is considerably larger than its matching counterpart of 32%. Part of this difference could be explained by the matching bias that we have alluded to before. To illustrate this, we have marked with a red cross the CATE estimates that correspond to observations left unmatched and therefore dropped in our matching analysis (Table 3, Column (4)). These are predominantly scattered among upper ranges of the axis, suggesting that the lower MTE ATE could follow from the missing influence of these high-valued CATEs.

To conclude this section, we investigate whether P3Fs may also have an impact on the number of therapies or their maturity (stage of development) at licensing. To this end we consider two additional outcome variables, NLICENSE and ELICENSE. The first corresponds to the number of therapies licensed (over the same measurement window used for DLICENSE). ELICENSE is also constructed based on the number of therapies licensed during this period, with the difference that the probability weighting scheme of (1) is used for aggregation. That is, all else constant, the value of ELICENSE is larger the more mature therapies were at licensing.

Causal forest estimates are computed separately for each outcome variable, only using the subset of observations in which at least one therapy was licensed. Resulting estimates are generally negative, for both outcome variables (results are shown by Appendix Figure A.4). Statistical significance has a similar pattern to that in our analysis for DLICENSE. For the NLICENSE outcome, the negative sign of estimates could suggest that P3Fs prompt units to engage in more targeted licensing. In the case of the ELICENSE outcome, it may indicate that units acting reactively are particularly vulnerable to the lack of thickness that characterizes this market (Gans and Stern, 2010). This is, even though units may ideally want to replace the recently failed P3 therapy with a mature (i.e., late-stage) therapy obtained from the licensing market, such therapy may not be available.

5.2 Do pre-licensing P3Fs worsen post-licensing development performance? (Q2)

Panel B of Figure 6 illustrates how we assemble a dataset to implement this analysis. Observations in this dataset correspond to licensed therapy/development stage pairs. Two such observations (i and i') are referenced in the Figure. These correspond to the two post-licensing stages that a hypothetical therapy r completed after licensing. (Only completed stages are considered in this analysis.) The outcome indicator ADVANCE (activated if development continues onto the next stage, and zero otherwise) is recorded for each of these observations. Covariates suspected of conditioning the likelihood of advancement (performance) are measured over the year prior to each milestone's reported completion date. These are used as the matching and "splitting" variables, respectively by the MTE and CFE. Variables related to the incidence of P3Fs are measured over the year prior to each therapy's licensing date. These are used to estimate propensity scores at the therapy level.

Panel B of Table 3 presents benchmark estimates. These are produced as before. However, because there are marked differences in the probability of advancement across stages (see Table 1 and related discussion), we demand that MTE also matches exactly on development stages. As a result, a relatively larger percentage of observations are left unmatched (relative to our previous analysis)—the MTE's "skepticism" implies the loss of 53% and 71% of the data, respectively for the analysis of Columns (3) and (4).[22]

The negative estimates across columns of Panel B suggest that the incidence of a pre-licensing P3F exerts a negative impact on post-licensing performance. In particular, estimates of Columns (1) and (2) suggest that a pre-licensing P3F lowers the probability of stage advancement in post-licensing development by 0.04 (linear probability model includes stage indicators). These coefficients are imprecisely estimated, however. The two versions of the MTE estimator (with and without R) give a similar result, by which P3Ts lower the probability of advancement by 0.11-0.12. Considering that the mean value of ADVANCE is 0.42 when there is no pre-licensing P3F ($W = 0$), this estimate suggests that pre-licensing P3Fs lower advancement probabilities by about 28%, overall.

As with our analysis of the previous section, Causal Forest CATE estimates (presented in Panel B of Figure 7) suggest that treatment effect heterogeneity tells an important part of the story. The average CATE is about -0.04. However, only about 30% and 5% are significant at the usual statistical confidence levels. Within this 35% of observations, the negative average impact of P3Fs on post-licensing performance is about 15%—about half

[22]The actual minimum value of R at which a P3F is observed differs from that in the context of the reactiveness analysis above (measurement windows are different). For the current analysis, this value is reflected in Panel C of Figure 2. Nevertheless, these threshold values are quite similar (even though their distributions are not). In absence of a theoretical framework to help us determine what is the "right" threshold for each case—and for simplicity—we have thus decided to use the relatively "clean" cut-off of 0.1 for both analyses.

that implied by matching estimates. The distribution of the large number of unmatched observations could again here explain part of the difference. To see this note that unmatched observations have an agglomeration near zero.[23] These relatively small estimates (in absolute value) are not accounted for by the matching procedure, which is why the resulting ATE may be larger.

5.3 What factors drive estimated treatment effect heterogeneity?

Our results above unveil a large amount of treatment effect heterogeneity across the covariate space $\mathcal{X}$. Our goal here is to shed some light on what specific variables in X can be regarded as the more important drivers of this heterogeneity.

Recall that reactiveness CATE estimates $\hat{\tau}^R$ can be viewed as a sort of latent willingness to engage in reactive licensing (as a function of the circumstances X). Analogously, performance estimates $\hat{\tau}^P$ can be interpreted as a latent willingness to terminate the development of a therapy licensed reactively. Also recall that CFE CATE estimates vary at the observation level, and are available for all observations in the samples used for each analysis. Thus, we can regress these on covariates X. To facilitate interpretation, we use a linear specification and a apply a normal-standard transformation on X. Obtained estimates can then be interpreted as the effect of increasing each covariate by one standard deviation. To account for the fact that the dependent variable is the result of an estimation procedure, we employ the generalized linear model correction of Hanushek (1974), which down-weights observations according to the standard error of the dependent variable.

Column (1) of Table 4 presents regressions results for the reactiveness estimates. Besides the listed variables, this model includes therapeutic area and year fixed effects. The positive coefficients for MKTSHARE, PIPESTR, LOGASSETS and LOGRD all align with the idea that larger firms are more likely to react to P3Fs. LOGASSETS exerts the stronger influence among these. The negative estimate coefficient on LOGSALES is a priori puzzling, as one would expect firms with more cash to be more likely to engage in reactive licensing. However, the variable's negative coefficient suggests the opposite. But in fact, licensing deals require relatively little cash (most potential compensation is contingent).

To analyze performance treatment effect heterogeneity (in Column (2)), we use the absolute value of CATE estimates as dependent variable. The strongest drivers of these are firms' scale and scope. Their negative coefficients coincides with previous literature, in suggesting that broader experience and specialization better equips firms to move

[23] The percentage of unmatched observations is three points higher for above-median CATE (i.e., less negative) observations.

development forward, despite a potential lack of fit of reactively-licensed therapies. The positive coefficient of PIPESTR is also intuitive, in that it suggests that for units with stronger pipelines, terminating a reactively-licensed therapy is likely to be less costly (more options to fall back on). The related "size effect" is also observed, based on the positive coefficient of LOGASSETS. To interpret the positive estimate of RLIC_RA, note that more intensive licensing by other units in the same area may signal that a new wave of technologies is starting to permeate from the biotech supply. In this context, sponsors may be less eager to continue developing a reactively-licensed therapy.

5.4 Do managers face a trade-off in practice?

Given the wide heterogeneity of estimated treatment effects, it could be possible that licensing circumstances X that make a unit more reactive (large $\hat{\tau}^R(X)$) are typically paired with milestone completion circumstances X' that make reactive licensing less consequential in terms of performance (small $|\hat{\tau}^P(X')|$). This empirical correlation would undermine the relevance of the reactive licensing trade-off, by implying that managers at high risk of engaging in reactive licensing usually are a low risk of suffering its consequences. On the other hand, if larger values of $\hat{\tau}^R(X)$ tend to be paired with larger values of $|\hat{\tau}^P(X')|$, then evidence would suggest that the reactive licensing trade-off is indeed experienced by managers in practice. Whether reality is closer to one extreme or the other depends on the empirical distribution of (X, X') pairs and the nature of the $\hat{\tau}$ functions. We can investigate the matter simply by examining the sample correlation of $\hat{\tau}^R(X)$ and $|\hat{\tau}^P(X')|$ values.

The key step to implement this analysis is merging the two sets of CATE estimates. We do so by first considering performance estimates, which are available for the therapy/completed stage pairs used in the causal forest estimation of subsection 5.2. Then, we identify the reactiveness estimate associated to each of these therapies', at the licensing quarter. This gives us a dataset with observations $(\hat{\tau}^R(x), |\hat{\tau}^P(x')|)$, where x corresponds to covariates ("circumstances") measured prior to licensing, while x' to those measured prior to the completion of each milestone. Using a generalized linear model (which accounts for both variables' estimated standard errors) to regress $\hat{\tau}^P$ on $\hat{\tau}^R$ (including therapeutic area, stage and licensing year fixed effects), we get a coefficient of 0.22 ($p < 0.01$) for $\hat{\tau}^R$. This positive correlation suggests that units more likely to react to P3Fs tend to also be more likely to terminate reactively-licensed therapies. This result supports the notion that managers do face the reactive licensing trade-off in practice.

6 Discussion of mechanisms

6.1 Why may firms react to setbacks?

What is the specific mechanism that prompts (some) firms to engage in reactive licensing? One possible explanation could be based on the use managerial incentives that are tied to the firm's stock market performance. Indeed, there exists ample evidence that the structure of executive compensation impacts innovation decisions and outcomes in high technology industries (e.g., Hoskisson et al., 1993; Balkin et al., 2000; Lerner and Wulf, 2007; Baranchuk et al., 2014). In our particular context, this type of explanation has the additional support of evidence showing that P3Fs have a significant negative impact on sponsoring firms' stock market valuation (Sharma and Lacey, 2004; Girotra et al., 2007): managers would engage in reactive licensing in an attempt to (at least partially) restore their firms' stock price, and with it, their personal compensation. On the other hand, this explanation has the problem that licensing deals are small relative to firms' overall operation, and hence unlikely to impact their stock market valuation. To resolve the matter, we conducted a financial event-study analysis. We found no evidence to support the idea that licensing (reactive or not) impacts the in-licensing firms' stock market valuation. This analysis is presented in Appendix B.

The corporate finance argument that managers may grow firms as means to expand the scope of their control (Shleifer and Vishny, 1988) maps into a second possible explanation. The bottom line of this argument is that managers may be willing to sacrifice some of their firm's market value if an investment offers them a sizable personal benefit (Morck et al., 1990). The correlate for our context would be that reactive licensing could be used to make up for the scope of control that is eroded by P3Fs. Donaldson and Lorsch (1983) indicate that value-decreasing acquisitions can be used to promote the firm's survival (even when liquidation may be better for shareholders), while Morck et al. (1990) find that acquisitions that follow a period of poor performance of the acquiring manager, as well as those whose target has experienced rapid growth have systematically lower returns for the acquiring firms' shareholders. Beyond the reactive flavor of these cases, this evidence is telling in the sense that acquisitions usually are much larger transactions than licensing deals. They should thus be less likely to be scrutinized and contested, and hence at higher risk of being used for the implementation of managers' personal objectives.

Lastly, a third possible explanation aligns with the "prolonged commitment" (Boulding et al., 1997) or "escalation of commitment" (Schmidt and Calantone, 1998) effect, a behavioral bias whereby managers delay "pulling the plug" on poor projects. The analogy for our context would be that managers may refuse to accept development setbacks by using outsourcing to quickly fill the gap they leave behind.

Whereas these explanations suggest that reactive licensing could be sub-optimal from the point of view of shareholders, we note that this would be difficult to verify empiri-

cally. The primary reason is the structural complexity of the problem—which has salient elements matching, negotiation and contract design in a dynamic environment with asymmetric information. To this inherent complexity, one should add the unavailability of systematic and reliable contract design and expected market profitability data. Nevertheless, given the great emphasis that is placed on reducing attrition rates in this industry (e.g., Kola and Landis, 2004; Paul et al., 2010), our results do raise the possibility that reactive licensing could be detrimental for shareholders. In that case, our analysis would serve as a counter example to the usual implication of R&D management models based on real options theory (e.g. Schwartz and Trigeorgis, 2004), namely, that managerial flexibility is valuable.[24]

6.2 Why do reactively licensed therapies underperform?

Our goal in this section is to shed some light on the potential mechanisms that underlie the negative consequences of reactive licensing that are illustrated by the results of subsection 5.2. We envision two potential mechanisms. The first regards the idea that reactively-licensed therapies may exhibit poorer fit with the in-licensor's capabilities:

- **Technical mismatch mechanism.** Reactively-licensed therapies may underperform if they are more likely to rely on technical approaches that are new to the in-licensor. In this case, the in-licensing unit may have been unable to perform adequate due diligence, or effectively dealt with the contingencies of clinical trial development.[25] As a result, the value of continuing with development may be lower. The importance of technical matching is highlighted by research documenting licensing sorting based on technological relatedness (e.g., Mowery et al., 1998; Diestre and Rajagopalan, 2012). More explicitly, Mittra (2007) states that "licensing-in products to fill a portfolio gap is a risky strategy if the company lacks the in-house expertise."

The second potential mechanism highlights the fact that, in practice, licensing deals require parties to engage in a lasting collaboration. Organizational frictions can arise if contracts are designed hastily:

- **Derailed collaboration mechanism.** Contractual agreements for reactively-licensed therapies may be less likely to properly specify contingent decision rules,

[24] Huchzermeier and Loch (2001) make a similar point but based on a different rationale.

[25] For example, suppose that a unit with experience in the development of bronchodilator asthma therapies reactively licenses another asthma therapy, but which uses a monoclonal antibody technology that is novel to the unit. Because monoclonal antibodies can increase the risk of cardiovascular events (Hansel et al., 2010), their clinical trials may require, among others, carefully crafted exclusion restrictions (e.g., less focus on older populations) and specific analyses of drug-drug interactions (e.g., cholesterol medications). A unit without experience on monoclonal antibodies may be less prepared to properly manage these tasks, making clinical trials more likely to produce adverse events.

making parties more likely to enter conflict and terminate their collaboration. These "missing" decisions rules could refer, for example, to the timing and scope of biotech scientists' publication rights, or to the obligations of biotech scientists in supporting development activities. These aspects are highlighted by studies focusing on the governance of and contractibility for research-based collaborations (e.g., Lerner and Merges, 1998; Lerner and Malmendier, 2010), as well as by anecdotal evidence by practitioners indicating that "failure to arrive at a common understanding of contractual terms from the start of an agreement is a major source of risk that can jeopardize outcomes" (Jones, 2007).

To address the technological mismatch mechanism, we take advantage of two technological categorizations available from our data. This information is generally available both for the therapies developed inside and outside our analysis sample. Thus, we use these data to construct indicators for each licensed therapy's technical matching quality.

For the first of these categorizations, levels correspond to broad technical approaches used in the formulation and delivery of developing medicines. Categories include, for example, small molecules (i.e., chemically synthesized), large molecules (i.e., living organisms), dermatological or intravenous delivery, etc. Cortellis labels these simply as "technologies." There are 327 such technologies in the overall data, and therapies are often associated to more than one. The second categorization corresponds to "targeted based actions" (TBAs), which describe the precise way in which a compound produces a pharmacological effect on the body.[26] There are about 7,500 TBA levels in the overall data. Therapies are associated to more than one only occasionally.

We utilize these categorizations one at the time, in the following way. For each licensed therapy, we code the matching quality variable MATCH as one if the entire set of technologies (or TBAs) associated to it have been previously (i.e., prior to licensing) been employed by the in-licensing unit. The variable is coded as zero otherwise. For consistency with our earlier analysis, we restrict the to subsample therapies used in the causal forest analysis ($R \geq 0.1$). The conjectured mechanism requires the average of MATCH to be lower among treated than among non-treated therapies (with treatment defined as in subsection 5.2). When we use the first categorization (technologies), MATCH averages 0.60 among treated therapies (licensed within a year of a P3T) and 0.61 among non-treated ones ($p = 0.57$). When the second categorization (TBAs) is used instead, these averages are respectively 0.36 and 0.35 ($p = 0.7$). Results do not qualitatively change if we drop the $R \geq 0.1$ requirement, or if we code MATCH considering the entire set of

[26]More precisely, TBAs correspond to mechanism of action/targeted cell pairs. A compound's mechanism of action refers to the way the compound produces a pharmacological effect on the body. Some molecules may, for example, act by stimulating specialized cells (like adrenaline); others, by replacing them (like insulin). Because TBA pairs also specify the targeted cell, they provide an even more specific categorization for the technicalities involved in each compound's development. For example, the "p38 MAP kinase inhibitor" TBA in our data refers to the implementation of an inhibition mechanism, targeting the mitogen-activated protein kinases P38.

therapies previously developed by the firm (instead of unit). Thus, this evidence does not support the technical mismatch mechanism.

Turning to the derailed collaboration mechanism, we first return to the statistics of Table 1, which describe the frequencies of different types of terminations. We approach these statistics highlighting the difference in the definition of "development" and "agreement" terminations (Columns (3) and (4), respectively). In particular, recall that whereas "development" terminations imply that therapies in question are not developed further, "agreement" terminations do not. Instead, "agreement" terminations solely indicate that collaboration has reached an end. As such, these terminations may signal the deterioration of parties' relationship. Using termination events only, our test relies on the indicator DERAIL, which is activated only for "agreement" terminations. (As before, the analysis is restricted to the $R \geq 0.1$ subsample.) DERAIL averages 0.12 among treated observations and 0.06 among non-treated ones ($p < 0.05$). When the $R \geq 0.1$ requirement is dropped, the difference shrinks somewhat but retains statistical significance ($p < 0.1$). This evidence suggests that the underperformance of reactively licensed therapies may be fueled by organizational frictions. We nevertheless note that, because agreement terminations are a relatively small share of all terminations in the sample, this mechanism is unlikely to the sole driver of the underperformance of reactively-licensed therapies.

7 Conclusion and limitations

This paper highlights the concept of "reactive" outsourcing of technological innovations, which diverges from standard literature portrayals of outsourcing as a carefully planned and executed, "proactive" process. Evidence from the pharmaceutical industry suggests that firms occasionally engage it. When they do so, outsourced technologies are less likely to reach the market.

Some of the limitations of our analysis stem directly from data availability. We do not have access to data tracking the set of preparatory and execution activities that underpin our hypothesis. Neither do we systematically observe contracts financial compensation terms, or the type of incentives that may shape managerial decisions. For these reasons, our analysis is build over tests for the hypothesis' empirical implications. These types of data may be particularly useful to shed light on the mechanisms at play, a topic that we cannot conclusively resolve with the data at hand.

A second set of limitations comes from the structural complexity of the decision environment. For example, outsourcing deals in this industry are heavily negotiated and can vary in terms of exclusivity and geographical scope of traded rights, among others. Here we have adopted an "apples to apples" comparison strategy, circumscribing the analysis to particular type of deals (exclusive worldwide rights). We suspect that broadening the scope of considered outsourcing activities (to include other types of deals as well as M&A)

could shed further valuable insight. We hope to address some of these aspects in future research.

References

Abrantes-Metz, R., C. Adams, and A. Metz (2004). Pharmaceutical development phases: a duration analysis.

Acemoglu, D. and J. Linn (2004). Market size in innovation: theory and evidence from the pharmaceutical industry. *The Quarterly journal of economics 119*(3), 1049–1090.

Alcacer, J., J. Cantwell, and M. Gittelman (2008). Are licensing markets local? an analysis of the geography of vertical licensing agreements in bio-pharmaceuticals. Technical report.

Arora, A., A. Fosfuri, and A. Gambardella (2004). *Markets for technology: The economics of innovation and corporate strategy*. MIT press.

Arora, A. and A. Gambardella (2010). The market for technology. In *Handbook of the Economics of Innovation*, Volume 1, pp. 641–678. Elsevier.

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

Arrowsmith, J. (2011). Trial watch: phase iii and submission failures: 2007-2010. *Nature reviews. Drug discovery 10*(2), 87.

Ascarza, E. (2018). Retention futility: Targeting high-risk customers might be ineffective. *Journal of Marketing Research 55*(1), 80–98.

Athey, S., J. Tibshirani, and S. Wager (2018). Generalized random forests. *Annals of Statistics (forthcoming)*.

Balkin, D. B., G. D. Markman, and L. R. Gomez-Mejia (2000). Is ceo pay in high-technology firms related to innovation? *Academy of management journal*, 1118–1129.

Baranchuk, N., R. Kieschnick, and R. Moussawi (2014). Motivating innovation in newly public firms. *Journal of Financial Economics 111*(3), 578–588.

Bayus, B. L. (1997). Speed-to-market and new product performance trade-offs. *Journal of Product Innovation Management 14*(6), 485–497.

Boulding, W., R. Morgan, and R. Staelin (1997). Pulling the plug to stop the new product drain. *Journal of Marketing research*, 164–176.

Breiman, L. (2001). Random forests. *Machine learning 45*(1), 5–32.

Brown, S. J. and J. B. Warner (1985). Using daily stock returns: The case of event studies. *Journal of financial economics 14*(1), 3–31.

Cao, Z. and A. Sorescu (2013). Wedded bliss or tainted love? stock market reactions to the introduction of cobranded products. *Marketing Science 32*(6), 939–959.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of finance 52*(1), 57–82.

Chan, T., J. A. Nickerson, and H. Owan (2007). Strategic management of r&d pipelines with cospecialized investments and technology markets. *Management Science 53*(4), 667–682.

Chandy, R., B. Hopstaken, O. Narasimhan, and J. Prabhu (2006). From invention to innovation: Conversion ability in product development. *Journal of Marketing Research 43*(3), 494–508.

Chaney, P. K., T. M. Devinney, and R. S. Winer (1991). The impact of new product introductions on the market value of firms. *Journal of Business*, 573–610.

Cockburn, I. (2004). The changing structure of the pharmaceutical industry. *Health Affairs 23*(1), 10–22.

Cockburn, I. M. and R. M. Henderson (2001). Scale and scope in drug development: unpacking the advantages of size in pharmaceutical research. *Journal of health economics 20*(6), 1033–1057.

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

Cunningham, C., F. Ederer, and S. Ma (2018). Killer acquisitions. *Available at SSRN 3241707*.

Danzon, P. M., A. Epstein, and S. Nicholson (2007). Mergers and acquisitions in the pharmaceutical and biotech industries. *Managerial and Decision Economics 28*(4-5), 307–328.

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

Dehejia, R. H. and S. Wahba (1999). Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American statistical Association 94*(448), 1053–1062.

Diestre, L. and N. Rajagopalan (2012). Are all 'sharks' dangerous? new biotechnology ventures and partner selection in r&d alliances. *Strategic Management Journal 33*(10), 1115–1134.

DiMasi, J. A., L. Feldman, A. Seckler, and A. Wilson (2010). Trends in risks associated with new drug development: success rates for investigational drugs. *Clinical Pharmacology & Therapeutics 87*(3), 272–277.

DiMasi, J. A., H. G. Grabowski, and R. W. Hansen (2016). Innovation in the pharmaceutical industry: new estimates of r&d costs. *Journal of health economics 47*, 20–33.

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

Donaldson, G. and J. W. Lorsch (1983). Decion making at the top: The shaping of strategic direction.

Dranove, D., C. Garthwaite, and M. Hermosilla (2014). Pharmaceutical profits and the social value of innovation. Technical report, National Bureau of Economic Research.

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

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

Fine, J. P. and R. J. Gray (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American statistical association 94*(446), 496–509.

Finkelstein, A. (2004). Static and dynamic effects of health policy: Evidence from the vaccine industry. *The Quarterly Journal of Economics 119*(2), 527–564.

Freidlin, B., E. L. Korn, and S. L. George (1999). Data monitoring committees and interim monitoring guidelines. *Controlled clinical trials 20*(5), 395–407.

Gans, J. S., D. H. Hsu, and S. Stern (2002). When does start-up innovation spur the gale of creative destruction? *RAND Journal of Economics 33*(4), 571–587.

Gans, J. S. and S. Stern (2010). Is there a market for ideas? *Industrial and Corporate Change 19*(3), 805–837.

Giordano-Coltart, J. and C. W. Calkins (2007). Best practices in patent license negotiations. *Nature biotechnology 25*(12), 1347.

Giovannetti, G. and P. Spence (2017). Beyond borders: The global biotechnology report.

Girotra, K., C. Terwiesch, and K. T. Ulrich (2007). Valuing r&d projects in a portfolio: Evidence from the pharmaceutical industry. *Management Science 53*(9), 1452–1466.

Griffin, A. (1997). The effect of project and process characteristics on product development cycle time. *Journal of Marketing Research*, 24–35.

Guedj, I. and D. Scharfstein (2004). Organizational scope and investment: Evidence from the drug development strategies of biopharmaceutical firms. *NBER working paper No. 10933*.

Guo, T., S. Sriram, and P. Manchanda (2017). The effect of information disclosure on industry payments to physicians.

Hansel, T. T., H. Kropshofer, T. Singer, J. A. Mitchell, and A. J. George (2010). The safety and side effects of monoclonal antibodies. *Nature reviews Drug discovery 9*(4), 325.

Hanushek, E. A. (1974). Efficient estimators for regressing regression coefficients. *The American Statistician 28*(2), 66–67.

Hay, M., D. W. Thomas, J. L. Craighead, C. Economides, and J. Rosenthal (2014). Clinical development success rates for investigational drugs. *Nature biotechnology 32*(1), 40.

Henderson, R. and I. Cockburn (1994). Measuring competence? exploring firm effects in pharmaceutical research. *Strategic management journal 15*(S1), 63–84.

Hendricks, K. B. and V. R. Singhal (1997). Delays in new product introductions and the market value of the firm: The consequences of being late to the market. *Management Science 43*(4), 422–436.

Hermosilla, M. (2016). Imperfect outsourcing of technological innovations. *Mimeo*.

Hermosilla, M. and J. Lemus (2018). Therapeutic translation of genomic science: Opportunities and limitations of gwas. *Forthcoming in Economic Dimensions of Personalized and Precision Medicine, Berndt, Goldman, and Rowe.*.

Hermosilla, M. and Y. Wu (2018). Market size and innovation: The intermediary role of technology licensing. *Research Policy 47*(5), 980–991.

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

Hirano, K., G. W. Imbens, and G. Ridder (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica 71*(4), 1161–1189.

Hoskisson, R. E., M. A. Hitt, and C. W. Hill (1993). Managerial incentives and investment in r&d in large multiproduct firms. *Organization Science 4*(2), 325–341.

Huchzermeier, A. and C. H. Loch (2001). Project management under risk: Using the real options approach to evaluate flexibility in r... d. *Management Science 47*(1), 85–101.

Imbens, G. W. and D. B. Rubin (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

Jones, A. (2007). Minimizing leakage of value from r&d alliances. *Nature Reviews Drug Discovery 6*(9), 711.

Kneller, R. (2010). The importance of new companies for drug discovery: origins of a decade of new drugs. *Nature Reviews Drug Discovery 9*(11), 867.

Kola, I. and J. Landis (2004). Can the pharmaceutical industry reduce attrition rates? *Nature reviews Drug discovery 3*(8), 711.

Krieger, J. L. (2017). *Trials and Terminations: Learning from Competitors' R&D Failures*. Harvard Business School.

Lerner, J. and U. Malmendier (2010). Contractibility and the design of research agreements. *American Economic Review 100*(1), 214–46.

Lerner, J. and R. P. Merges (1998). The control of technology alliances: An empirical analysis of the biotechnology industry. *The Journal of Industrial Economics 46*(2), 125–156.

Lerner, J. and J. Wulf (2007). Innovation and incentives: Evidence from corporate r&d. *the Review of Economics and Statistics 89*(4), 634–644.

Levine, A. (2009). Licensing and scale economies in the biotechnology pharmaceutical industry. *Working Paper, Washington University in St. Louis*.

Mason, R., N. Savva, and S. Scholtes (2008). The economics of licensing contracts. *Nature biotechnology 26*(8), 855–857.

Merrill, J. (2016). Surprise! it's a phase iii failure. *Scrip Pharma Intelligence*. Available at https://scrip.pharmaintelligence.informa.com/SC097113/Surprise-Its-A-Phase-III-Failure.

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

Moorman, C. and A. S. Miner (1998). The convergence of planning and execution: Improvisation in new product development. *the Journal of Marketing*, 1–20.

Morck, R., A. Shleifer, and R. W. Vishny (1990). Do managerial objectives drive bad acquisitions? *The Journal of Finance 45*(1), 31–48.

Mowery, D. C., J. E. Oxley, and B. S. Silverman (1998). Technological overlap and interfirm cooperation: implications for the resource-based view of the firm. *Research policy 27*(5), 507–523.

Nie, X. and S. Wager (2017). Learning objectives for treatment effect estimation. *arXiv preprint arXiv:1712.04912*.

Pammolli, F., L. Magazzini, and M. Riccaboni (2011). The productivity crisis in pharmaceutical r&d. *Nature reviews Drug discovery 10*(6), 428.

Paul, S. M., D. S. Mytelka, C. T. Dunwiddie, C. C. Persinger, B. H. Munos, S. R. Lindborg, and A. L. Schacht (2010). How to improve r&d productivity: the pharmaceutical industry's grand challenge. *Nature reviews Drug discovery 9*(3), 203.

Pfizer Inc. (April 26, 2007). Bristol-myers squibb and pfizer announce worldwide collaboration to develop and commercialize anticoagulant and metabolic compounds. Pfizer Press Release.

Pfizer Inc. (December 2, 2006). In interests of patient safety, pfizer stops all torcetrapib clinical trials. Pfizer Press Release.

Pisano, G. (1997). R&d performance, collaborative arrangements and the market for know-how: A test of the "lemons" hypothesis in biotechnology. *Mimeo*.

Pisano, G. P. (2006). *Science business: The promise, the reality, and the future of biotech*. Harvard Business Press.

Powell, W. W. (1996). Inter-organizational collaboration in the biotechnology industry. *Journal of Institutional and Theoretical Economics*, 197–215.

Quinn, J. B. (2000). Outsourcing innovation: the new engine of growth. *Sloan management review 41*(4), 13–13.

Robertson, T. S., J. Eliashberg, and T. Rymon (1995). New product announcement signals and incumbent reactions. *The Journal of Marketing*, 1–15.

Rosenbaum, P. R. and D. B. Rubin (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika 70*(1), 41–55.

Schmidt, J. B. and R. J. Calantone (1998). Are really new product development projects harder to shut down? *Journal of product innovation management 15*(2), 111–123.

Schwartz, E. S. and L. Trigeorgis (2004). *Real options and investment under uncertainty: classical readings and recent contributions*. MIT press.

Sharma, A. and N. Lacey (2004). Linking product development outcomes to market valuation of the firm: The case of the us pharmaceutical industry. *Journal of Product Innovation Management 21*(5), 297–308.

Shleifer, A. and R. W. Vishny (1988). Value maximization and the acquisition process. *Journal of Economic Perspectives 2*(1), 7–20.

Smith, P. G. (1999). From experience: Reaping benefit from speed to market. *Journal of Product Innovation Management 16*(3), 222–230.

Srinivasan, S. and D. M. Hanssens (2009). Marketing and firm value: Metrics, methods, findings, and future directions. *Journal of Marketing research 46*(3), 293–312.

Teece, D. (1986). Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy. *Research Policy 15*(6), 285–305.

Truex, S. (2018). Just a little patience: In-licensing from biopharma. *Life Sci VC*. Available at https://lifescivc.com/2018/03/just-a-little-patience-in-licensing-from-biopharma/.

Wager, S. and S. Athey (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 1–15.

Waring, M. J., J. Arrowsmith, A. R. Leach, P. D. Leeson, S. Mandrell, R. M. Owen, G. Pairaudeau, W. D. Pennie, S. D. Pickett, J. Wang, et al. (2015). An analysis of the attrition of drug candidates from four major pharmaceutical companies. *Nature reviews Drug discovery 14*(7), 475.

WIPO (2015). Successful technology licensing. *World Intellectual Property Organization*.

Figures and Tables

Table 1: Post-licensing development performance.

Stage	(1) N	(2) Advance to next stage	(3) Development termination	(4) Agreement termination	(5) Idle stage termination
Discovery	509	0.46	0.46	0.05	0.03
Phase 1	325	0.38	0.49	0.07	0.06
Phase 2	540	0.20	0.72	0.05	0.02
Phase 3	229	0.51	0.41	0.05	0.03
Review	174	0.86	0.11	0.02	0.01
Total	1,777	0.41	0.50	0.05	0.03

Statistics in this table do not consider ongoing stages. A definition for the variables considered by each column is provided in text.

Table 2: Specification summary for covariates X.

Variable	Definition	Variability	Construct
	A. Pipelines-related covariates X^P		
PIPESTR	$A_{jt}/A_j + A_{-jt}$	Unit/date	Strength of unit's pipeline
RTER_OWN	T_{jt}/A_j	Unit/date	Unit's recent productivity
RTER_RA	T_{-jt}/A_{-jt}	Unit/date	Rest-of-area's recent productivity
RLIC_OWN	L_{jt}/A_{jt}	Unit/date	Unit's recent in-licensing
RLIC_RA	L_{-jt}/A_{-jt}	Unit/date	Rest-of-area's recent in-licensing
MKTSHARE	$M_{jt}/M_j + M_{-jt}$	Unit/date	Strength of unit's commercialized portfolio
	B. Experience-related covariates X^E		
SCOPE	1-HHI (across areas)	Firm/date	Firm's development diversification
SCALE	Log number of therapies historically developed by sponsor	Firm/date	Firm's development experience
	C. Quarterly financial covariates X^F		
LOGSALES	Demeaned log sales	Firm/quarter	Size, liquidity, internal capital markets
LOGASSETS	Demeaned log assets	Firm/quarter	Size, capabilities, internal capital markets
LOGRD	Demeaned log R&D expenditures	Firm/quarter	Size, technological leadership
	D. Additional controls		
Year, area, and development stage indicators			

Table 3: Benchmark estimates.

	(1) Difference of means	(2) Linear probability	(3) MTE 1	(4) MTE 2
A. Reactiveness analysis (Q1), outcome DLICENSE				
Estimate	0.13***	0.06***	0.10***	0.06***
Std. error	(0.01)	(0.02)	(0.03)	(0.01)
p-value	[0.00]	[0.01]	[0.00]	[0.00]
Observations	5,071	5,071	897	889
B. Performance analysis (Q2), outcome ADVANCE				
Estimate	-0.04	-0.04	-0.11***	-0.12***
Std. error	(0.03)	(0.03)	(0.01)	(0.01)
p-value	[0.17]	[0.23]	[0.00]	[0.00]
Observations	1,777	1,777	546	333
Propensity score	N/A	N/A	Function of X	Function of X and R
Sample	All	All	Matched $R \geq 0.1$	Matched $R \geq 0.1$

Models in this Table aim to estimate the impact of pre-licensing P3Fs on the short term probability of license (Panel A) and post-licensing development performance (Panel B). Outcomes and independent variables are constructed according to Figure 6. In columns (3) and (4), estimates correspond to ATEs. Further specification details are provided in text.

Table 4: Exploring treatment effect heterogeneity.

	(1)	(2)
	Reactiveness	Performance
MKTSHARE	0.005***	0.001
	(0.000)	(0.001)
	[0.000]	[0.348]
PIPESTR	0.005***	0.006***
	(0.001)	(0.001)
	[0.000]	[0.000]
RLIC_OWN	-0.000	-0.000
	(0.000)	(0.000)
	[0.267]	[0.776]
RLIC_RA	-0.004***	0.001*
	(0.000)	(0.000)
	[0.000]	[0.010]
RTER_OWN	0.001***	0.001***
	(0.000)	(0.000)
	[0.000]	[0.001]
RTER_RA	0.001	-0.000
	(0.000)	(0.000)
	[0.143]	[0.244]
LOGASSETS	0.009***	0.003***
	(0.000)	(0.001)
	[0.000]	[0.000]
LOGSALES	-0.005***	-0.005***
	(0.001)	(0.001)
	[0.000]	[0.000]
LOGRD	0.004***	0.000
	(0.000)	(0.001)
	[0.000]	[0.939]
SCALE		-0.031***
		(0.001)
		[0.000]
SCOPE		-0.012***
		(0.001)
		[0.000]
Therapeutic area FE	✓	✓
Year FE	✓	✓
Stage FE		✓
Observations	992	1,161

Generalized linear models for the dependent variables $\hat{\tau}^R$ and $|\hat{\tau}^P|$, respectively. Estimated models account for the standard errors associated to each outcome (both are estimated variables). All shown covariates are standardized by their mean and standard deviation.

Figure 1: Trends in pharmaceutical licensing and innovation.

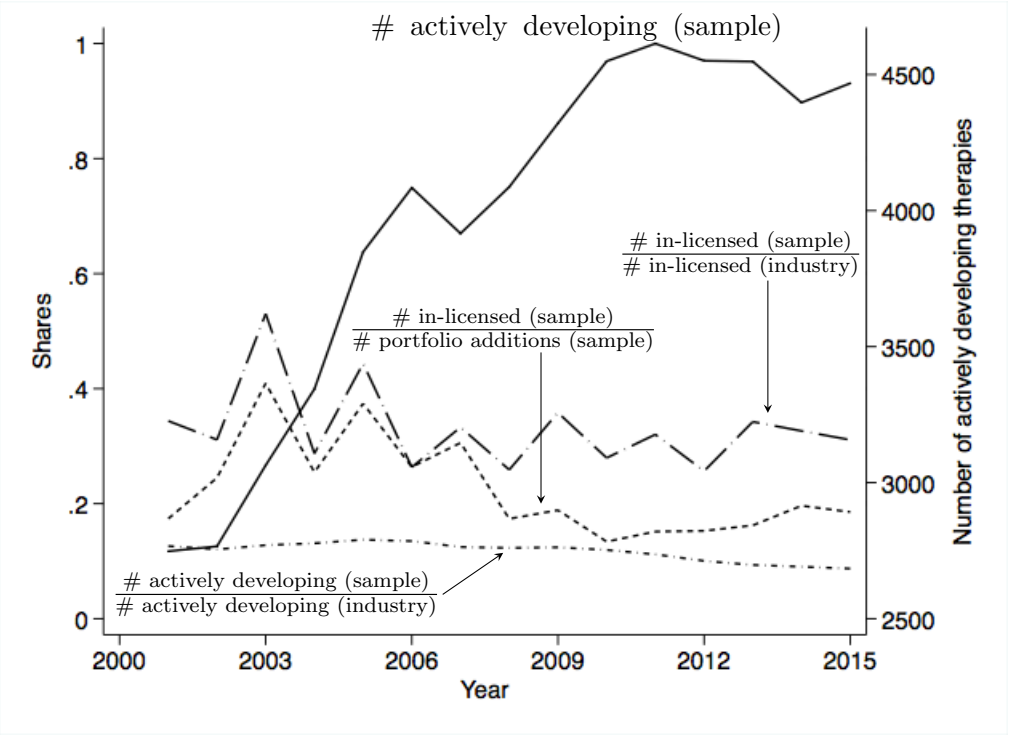
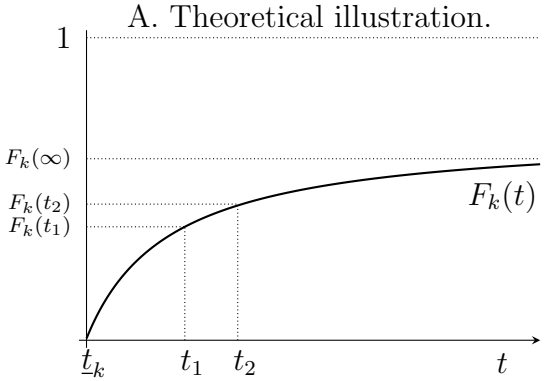
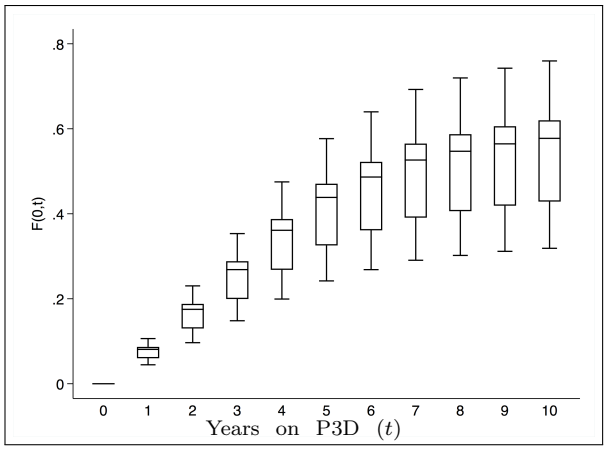


Figure 2: Construction elements and descriptives of the P3F risk score (R).

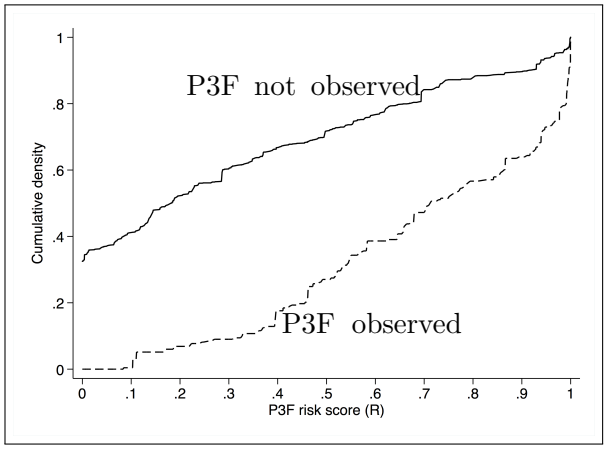
A. Theoretical illustration.



B. Estimated cumulative failure sub-hazards $F(0,t)$
(all therapeutic areas in period 2001-2005).



C. P3F risk scores (R) for licensed therapies
(measured over year ending at licensing date).



Panel A. The function $F_k(t)$ corresponds to the probability that a therapy k introduced to P3D at time $\underline{t}_k$ fails at or before time t. **Panel B.** Estimated cumulative P3D failure sub-hazards (empirical correlate of $F_k(t)$), estimated via competing risks (Fine and Gray, 1999) on a large sample of P3 development histories. Variability shown by boxes comes from differences across therapeutic areas. **Panel C.** Estimates of P3F risk scores for all licensed therapies, computed via expression (2), and where the measurement window $[t_1, t_2]$ corresponds to the calendar year ending the day each therapy is licensed.

Figure 3: Radius matching neighborhood for an observation at $X = x$.

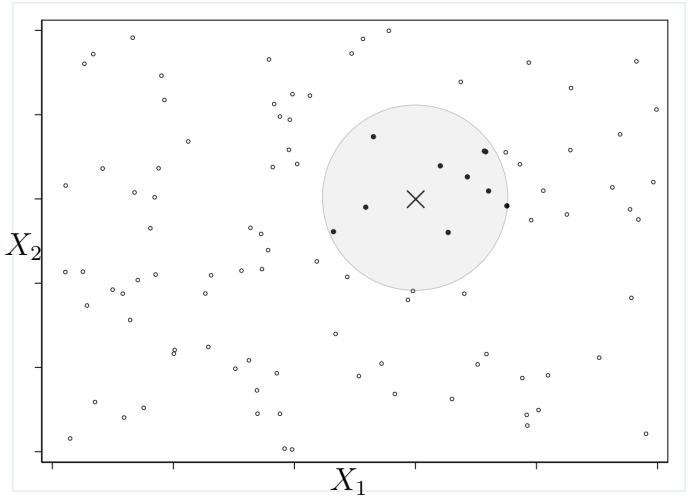
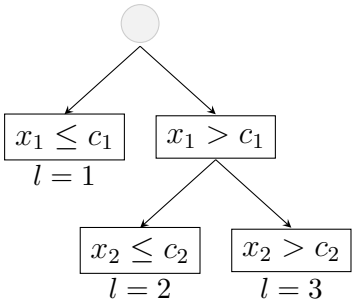


Figure 4: Classification and neighborhoods in tree-based methods.

A. Sample classification tree (dendogram representation).



B. Neighborhoods in tree-based classification.

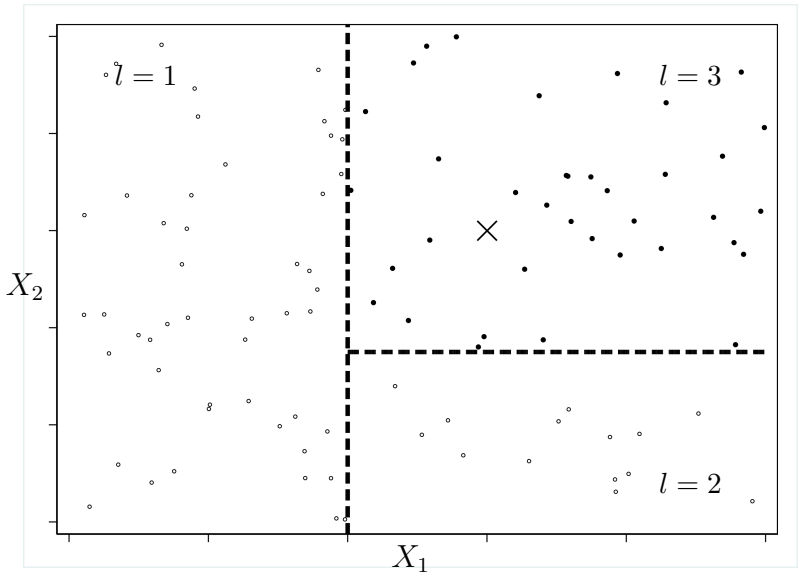
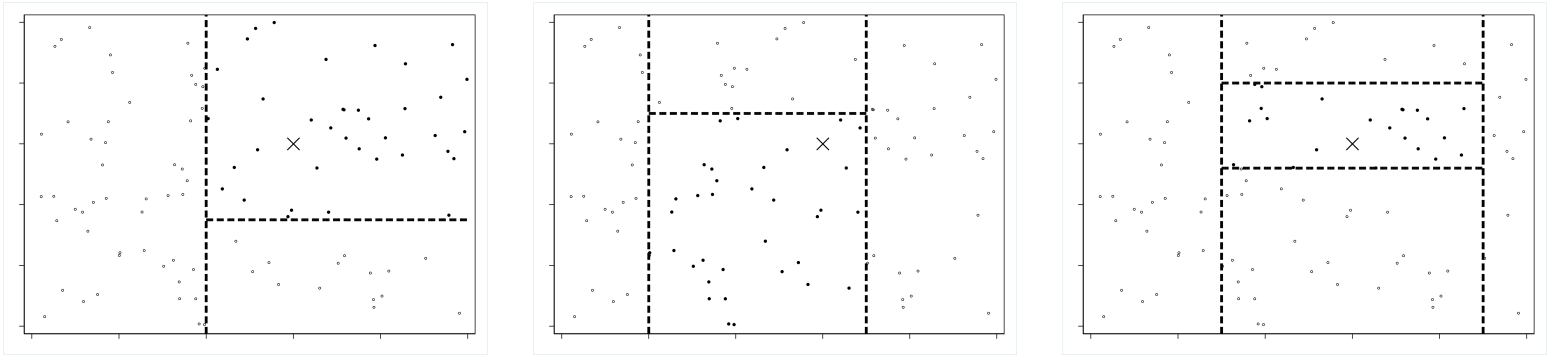


Figure 5: Classification and neighborhoods in tree-based methods.

A. Classification in three decision trees.



B. Resulting weights relative to a focal observation $X = x$, $\alpha(x)$ (given by each marker's size).

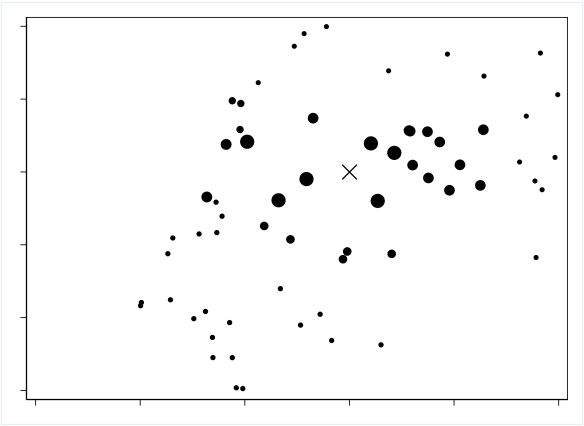
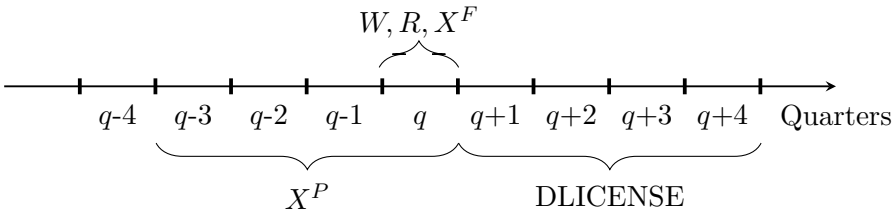


Figure 6: Measurement windows for variable construction.

A. Reactiveness analysis (Q1)



B. Performance analysis (Q2)

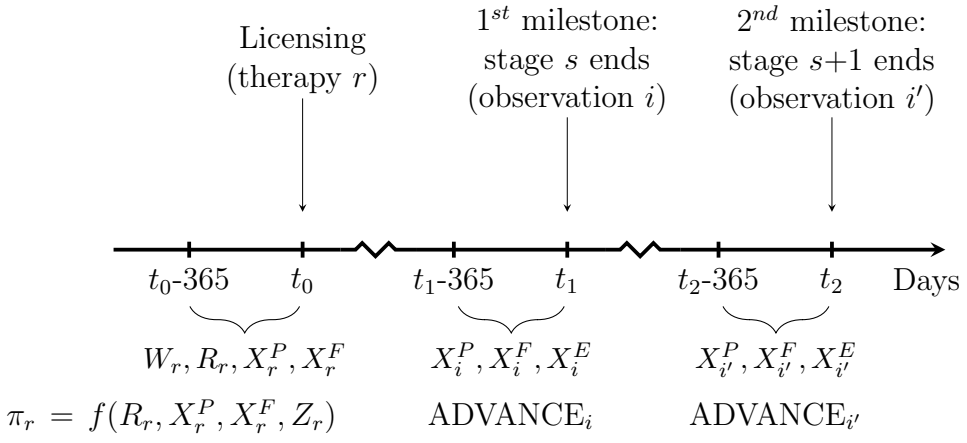
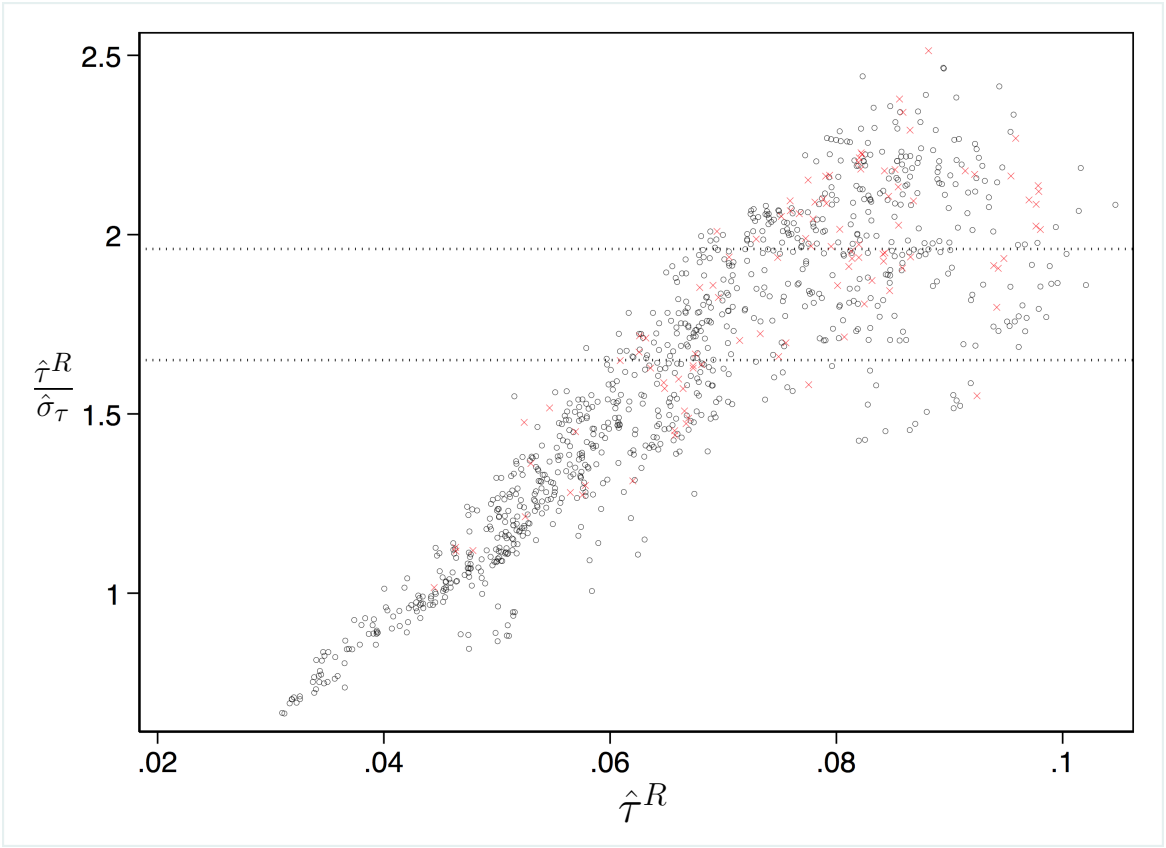
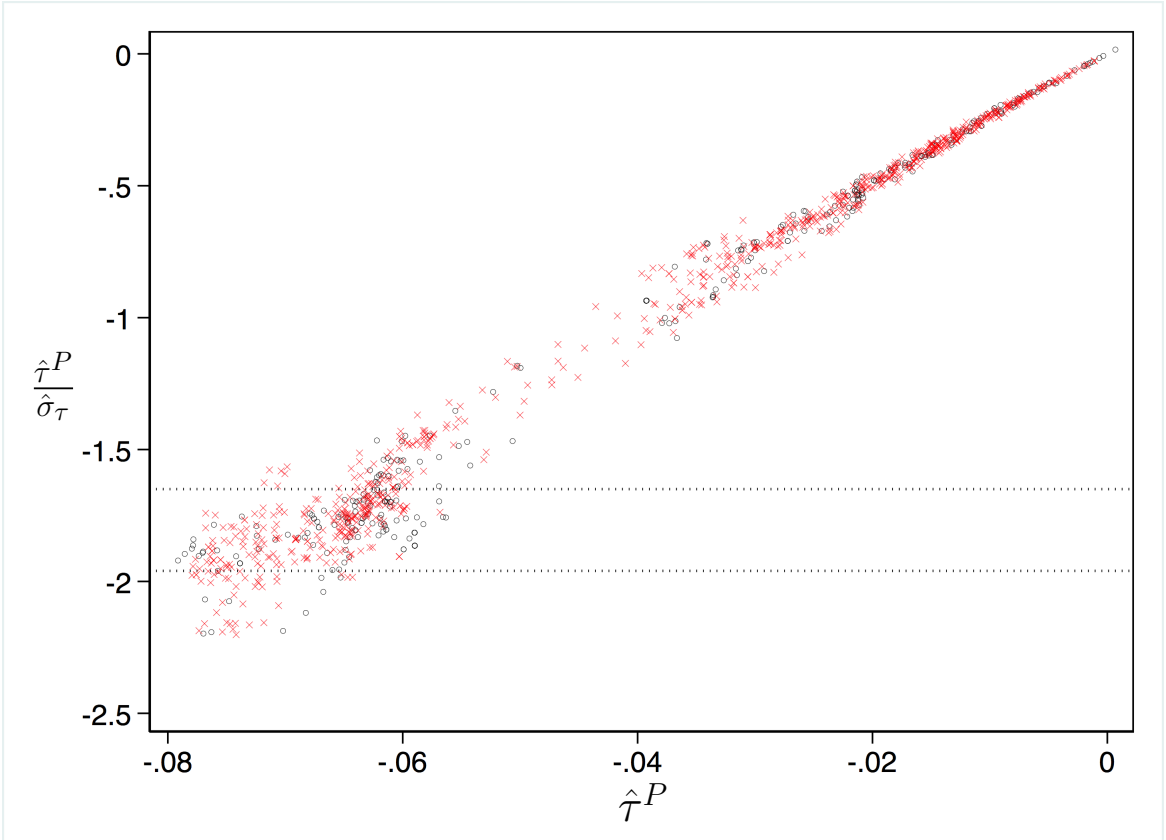


Figure 7: Causal Forest CATE estimates $\hat{\tau}$.

A. Reactiveness analysis (Q1)



B. Performance analysis (Q2)



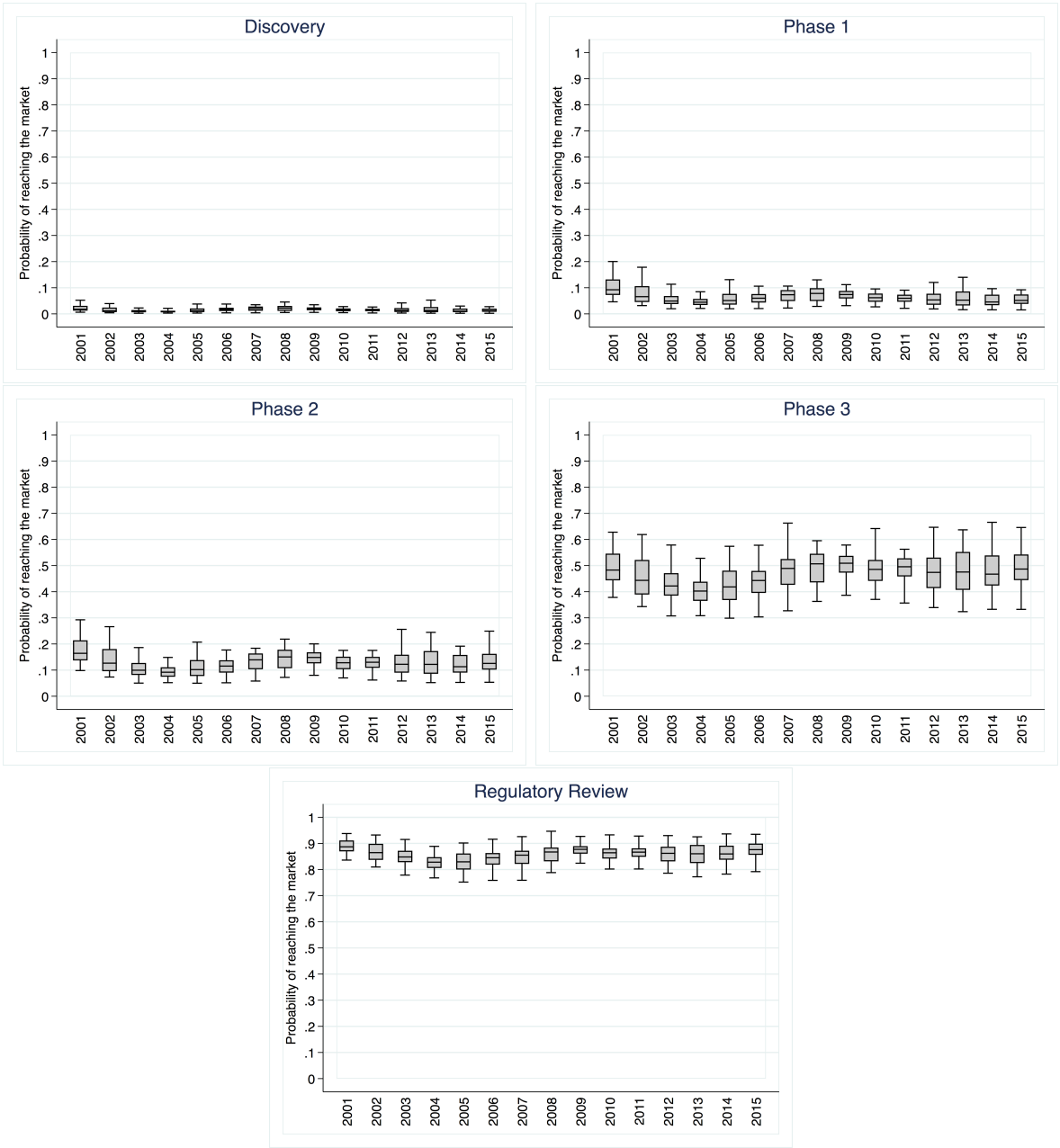
Appendices

A Stage-conditional reach-the-market probabilities

Here we describe the procedures used to construct the set of "stage-conditional reach-the-market" probabilities $\{p_{as}\}$ used for the weighted aggregation procedure of expression (1). We construct these probabilities on a rolling basis and at the therapeutic area level, using the about 80,000 development histories available from the broad Cortellis sample.

To see how we do this specifically, consider the problem of computing a unit's "marketed therapies-equivalent" number of active therapies over the one-year measurement window that ends on day t. In this case, we would consider all development outcomes that were observed between y-5 and y-1, for y =year(t). With these, we estimate a probit model for the probability of advancement for one stage at the time, using therapeutic area indicators as independent variables. The predictions of this model are stored, and the estimation is repeated for all other stages. Call the resulting probabilities q_{asy}. Stage-conditional probabilities of reaching the market are then computed as $p_{asy} = \prod_{s' \geq s} q_{as'y}$. Figure A.1 describes these estimates, for each year(t) in the sample. Boxes capture variability across therapeutic areas.

Figure A.1: Stage-conditional reach-the-market probabilities.



B Do licensing events impact stock market returns?

To address this question, we the event-study methodology that have been applied numerous times to study how innovation-related events impact firms' stock market returns.[27] Our results suggest that (i) licensing events do not impact the in-licensing firm's stock market valuation, and (ii) pre-licensing P3Fs do not appear to systematically mediate these impacts.

The analysis relies on an "abnormal returns" variable (AR), which computed at the daily level, based for the in-licensing firm's stock price returns. The "event day" corresponds to the day in a therapy j is licensed (as reported by Cortellis). We index this day by $t = 0$. Days following the event are successively indexed by $t = 1, 2, ..$; those that preced it, by $t = -1, -2, ...$ AR represents the difference between the stock's actual daily return and that predicted by an equilibrium pricing model.

Crucially, the event-study methodology postulate that ARs account for the impact of the information surfaced by the event under study on the firm's stock market returns. Inference is drawn by aggregating ARs into cumulative abnormal returns (CAR) over an event window $\{\underline{T}, .., \bar{T}\}$. To analyze the impact of a therapy j's licensing on its in-licensing firm's stock returns, CAR is defined as:

$$\text{CAR}_j = \sum_{t=\underline{T}}^{\bar{T}} \text{AR}_{ft},$$

where f indexes the in-licensing firm. Following other studies in this literature (e.g., Girotra et al., 2007), we set $\underline{T} = -1$ to account for the possibility that there may be "information leakage" prior to each deal's announcement. Keeping this threshold fixed, we consider the three windows given by $\bar{T} = 3, 5, 7$. Thus, CAR computes the sum of abnormal returns experienced by the firm around the deal's date. An estimated value of $\text{CAR} > 0$ would be consistent with the idea that licensing a positive impact on the firm's stock market valuation.

To compute ARs, we consider two equilibrium pricing models to predict expected returns $\mathbb{E}[\text{R}_{ft}]$. With these predictions, ARs are constructed as AR_{ft}=R_{ft}-$\mathbb{E}[\text{R}_{ft}]$. The employed pricing models are the Market Model (MM) of Brown and Warner (1985) and the Fama-French-Cahart 4-Factor Model (FFC) of Carhart (1997). These compute expected returns respectively as:

$$\begin{aligned} \mathbb{E}[R_{ft}] &= \alpha_j + \beta_j R_{mt} && \text{(MM)} \\ \mathbb{E}[R_{ft}] &= \alpha_j + \beta_j R_{mt} + \theta_j^1 \text{SMB}_t + \theta_j^2 \text{HML}_t + \theta_j^3 \text{UMD}_t && \text{(FFC)} \end{aligned}$$

Both models produce expected return predictions by invoking an equilibrium condition

[27] The methodologies employed here have been applied numerous times in the context of new product innovation. Sharma and Lacey (2004) use them to unveil the asymmetry of reactions to good and bad news in the pharmaceutical industry, while Girotra et al. (2007) to dissect the negative impacts of P3Fs based on the richness of the firm's developmental pipeline. Chaney et al. (1991) provide estimates of stock market reaction to developmental success across industries. Cao and Sorescu (2013) focus on the impact of co-branded products. Hendricks and Singhal (1997) study the impacts of product delays, and Robertson et al. (1995) how new product announcements impact the stock market valuation of competing firms. See Srinivasan and Hanssens (2009) for a thorough review of these methods and their application to marketing problems

that ties the return of the market portfolio (R_{mt}) to that of the firm's stock. The parameters governing this relationship are estimated independently for each licensing event through OLS regressions, using data from the pre-licensing window $\{-365, .., -11\}$. Models differ in terms of the number of additional factors that enter the specification as independent variables. Whereas MM only includes return of the market portfolio, FFC includes three other potential moderators. These are, (i) SMB ("small minus big") factor (return spread between small and large firms), (ii) HML ("high minus low" or "value premium") factor (differential return between stocks of companies with high and low book-to-market ratios), and (iii) UMD ("up minus down") factor (historical excess return of "winning stocks," compared to "loosing" ones). Factor data was obtained from Professor Kenneth French's website; stock market returns data, from the Center for Research in Security Prices (CRSP). We present results for the 1,775 licensed therapies for which we were able to compile all required data.

Estimated CAR values are small on average, suggesting that licensing does not have a systematic impact on the in-licensing firm's stock market value. When MM is used, these average 0.001, 0.002, and 0.004 for each window (narrowest to widest). Their FFC counterparts are about an order of magnitude smaller. Between 2% and 6% of these estimates are statistically significant with 95% confidence. (Standard errors are obtained from each regressions estimated error variance). These relatively few statistically significant CAR estimates are about evenly split between positive and negative values.

To further illustrate the absence of a systematically positive effect of licensing on the in-licensing firm's stock market valuation, Figure A.2 presents the (cumulative) distribution of estimates obtained in each case. To investigate the possibility that these effects may be present when licensing follows a recent P3F, we present separate distributions for each set of events (i.e., with and without a P3F having occurred within year ending on the licensing date). If the referenced impacts of licensing existed, these distributions should place relatively more mass on the positive domain of CAR. But this is not the case. Furthermore, the incidence of a pre-licensing P3F does appear to make a difference: dashed (P3F) and solid (no P3F) distributions are quite similar. The hypothesis that CAR estimates for licensing that follows a P3F are larger than their "no-P3F" counterparts is rejected within all sets of results (95% confidence).

As a last check, we evaluate the volume of stocks traded around the date of licensing $t = 0$. We base our analysis on the variable NVOL, defined as:

$$\text{NVOL}_{jt} = \frac{\text{VOL}_{jt}}{\bar{\text{VOL}}_j} - 1 \qquad \text{for } t = -10, .., 0, .., 10,$$

where VOL_{jt} represents the number of stocks traded on day t (of therapy j's in-licensing firm). The denominator $\bar{\text{VOL}}_j$ corresponds to the the average of VOL within the period $t = -365, .., -11$. Thus, NVOL gives a normalized measure of traded volume around licensing. If licensing has an impact on the in-licensing firm's equity, we should expect NVOL values to shift around licensing. The patterns described by Figure A.3 suggest that this is not the case. NVOL values are quite stable around zero throughout the considered window. In conclusion, this evidence does not support the idea that licensing impacts the in-licensing firm stock market valuation.

Figure A.2: Distributions of Cumulative Abnormal Returns following licensing events.

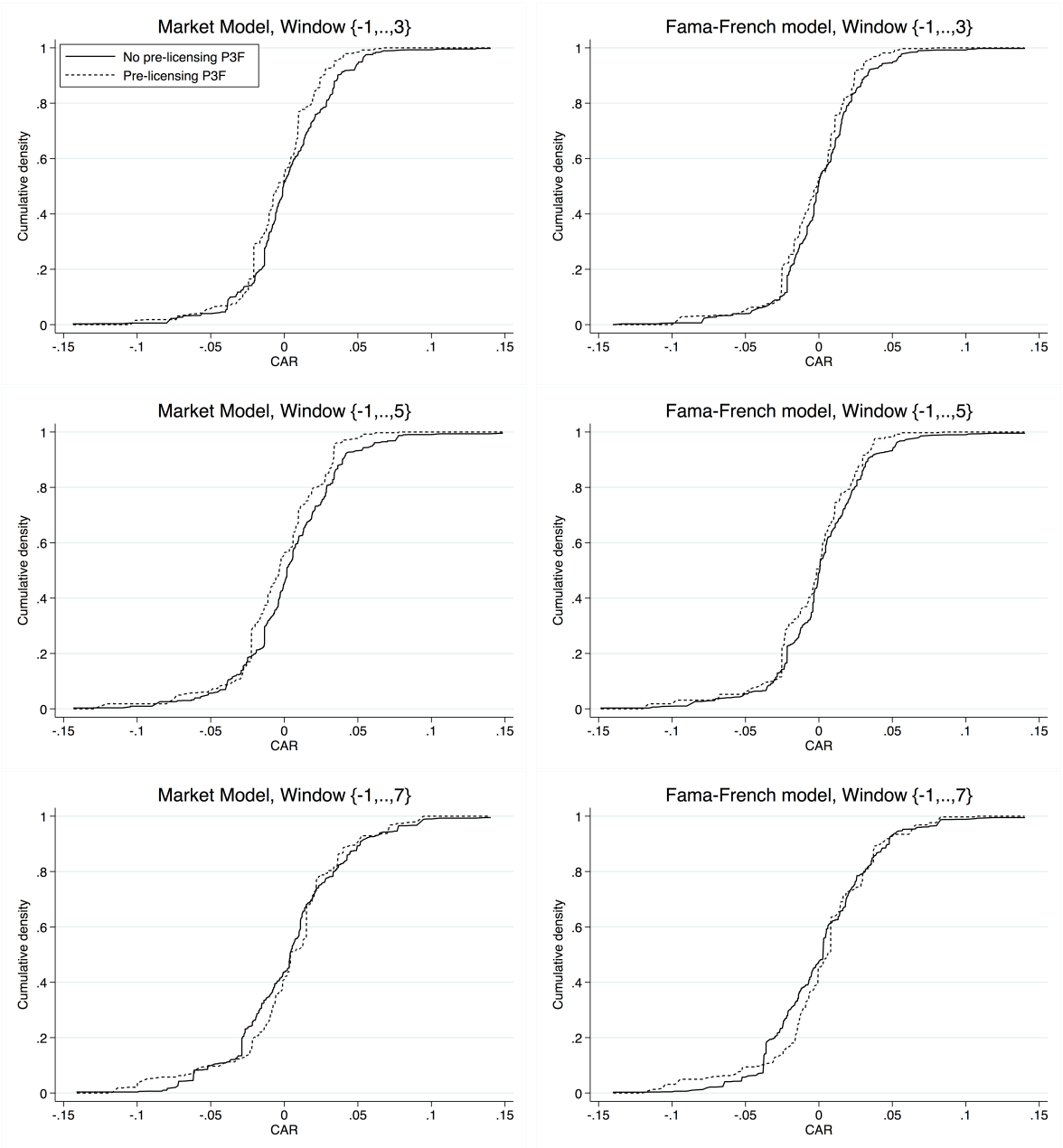
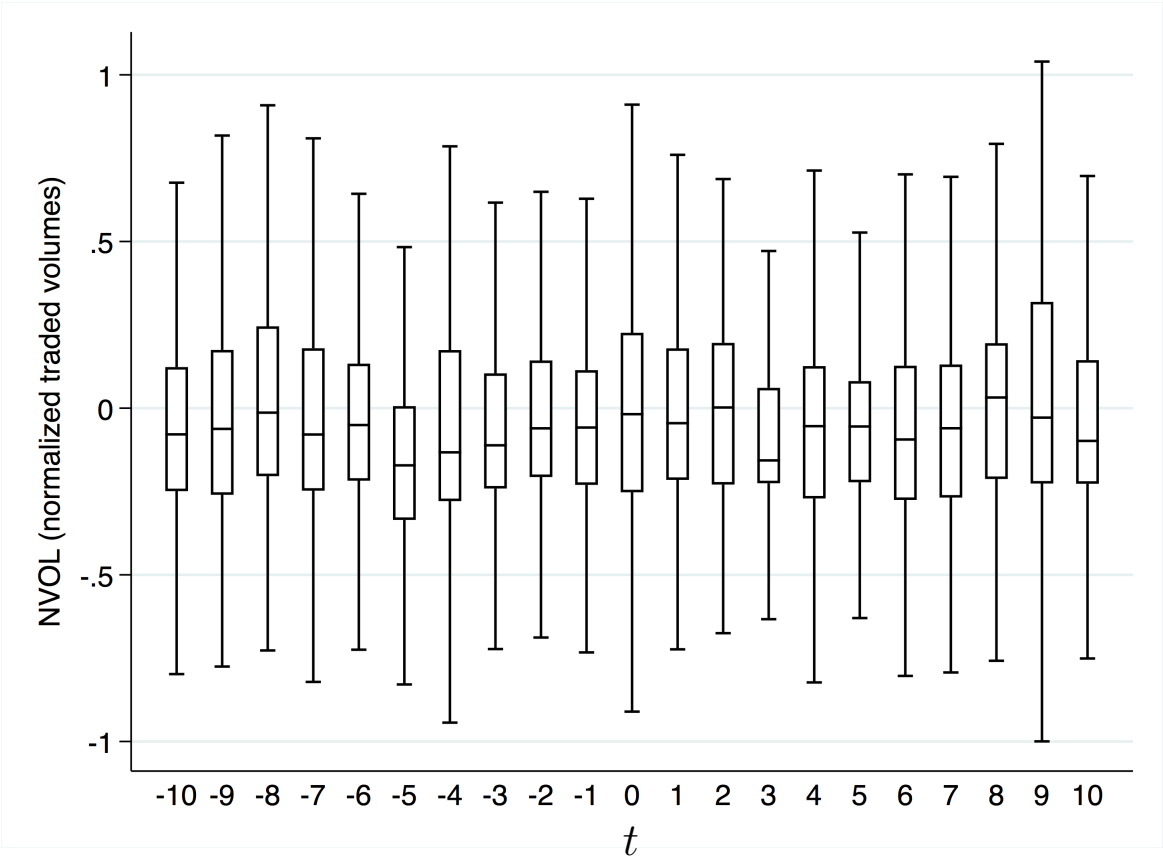


Figure A.3: Normalized traded volumes (NVOL) around the licensing date ($t = 0$).



C Additional Tables and Figures

Table A.1: Decision-making units in the sample.

Sponsor	Therapeutic area																
	Autoimmune/Inflammatory	Bone Disease	Cancer	Cardiovascular	Central Nervous System	Dermatologic	Endogrinological & Metabolic	Gastrointestinal	Genitourinary/Gynecologic	Hematologic	Infectious	Ophthalmic	Other/Unclassified	Psychiatric	Renal	Respiratory	Transplantation
ABBVIE	✓		✓	✓	✓	✓	✓	✓	✓		✓	✓			✓		
AMGEN	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓				✓		
ASTELLAS	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓
ASTRA ZENECA	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	
BAYER	✓		✓	✓	✓		✓		✓	✓	✓	✓	✓		✓	✓	
BIOGEN	✓		✓	✓	✓		✓			✓							
BRISTOL-MYERS SQUIBB	✓		✓	✓	✓	✓	✓	✓			✓		✓	✓	✓		✓
CELGENE	✓		✓			✓				✓				✓			
DAICHII SANKYO	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓		✓				
ELI LILLY	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓			
GILEAD			✓	✓			✓	✓			✓					✓	
GLAXO SMITH KLINE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓		✓	
JOHNSON & JOHNSON		✓	✓	✓	✓		✓	✓		✓	✓		✓	✓			
MERC & CO	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓		✓	
NOVARTIS	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓
PFIZER	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓
ROCHE	✓	✓	✓	✓	✓	✓	✓	✓		✓			✓	✓	✓		✓
SANOFI	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	
SHIRE			✓				✓	✓			✓	✓	✓	✓			
TAKEDA	✓	✓	✓	✓	✓		✓	✓		✓	✓		✓	✓			

Table A.2: Phase 3 Terminations across areas and time.

Therapeutic area	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	Total
Autoimmune/Inflammatory	0	0	1	0	0	0	0	2	3	4	3	2	5	2	2	24
Bone Disease	0	0	0	0	0	0	0	0	1	1	0	0	1	1	0	4
Cancer	4	2	5	4	8	3	2	5	12	13	11	10	18	7	9	113
Cardiovascular	3	0	1	1	3	3	7	5	7	7	3	1	4	1	6	52
Central Nervous System	3	1	1	4	1	0	2	1	4	3	3	3	3	3	0	32
Dermatologic	1	0	1	0	0	1	0	2	2	0	0	0	0	0	1	8
Endocrinological & Metabolic	2	1	3	0	2	4	3	7	3	5	5	0	5	4	2	46
Gastrointestinal	1	0	0	0	0	0	0	2	2	1	0	2	0	0	0	8
Genitourinary/Gynecologic	0	0	0	0	0	0	1	2	1	2	3	1	0	1	1	12
Hematologic				0	0	0	1	0	0	1	0	0	1	0	0	3
Infectious	1	1	5	1	1	0	0	3	2	4	5	2	2	3	3	33
Ophthalmic			0	0	0	0	0	0	0	0	1	0	0	0	1	2
Other/Unclassified	1	1	2	0	3	1	1	2	3	1	2	0	4	4	0	25
Psychiatric	0	1	3	1	1	2	2	5	3	3	3	3	2	4	2	35
Renal			0	0	1	0	1	0	2	0	2	0	0	0	0	6
Respiratory	1	0	0	0	0	1	1	1	3	1	1	1	0	0	1	11
Transplantation	0	0	0	1	0	0	0	0	1	0	0	1	1	0	0	4
Total	17	7	22	12	20	15	21	37	49	46	42	26	46	30	28	418

Figure A.4: Causal Forest CATE estimates: P3F impacts on NLICENSE and ELICENSE (conditional on DLICENSE=1).