Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Rushed Innovation: Evidence from Drug Licensing

Manuel Hermosilla

To cite this article:
Manuel Hermosilla (2020) Rushed Innovation: Evidence from Drug Licensing. Management Science

Published online in Articles in Advance 08 Apr 2020

. https://doi.org/10.1287/mnsc.2019.3530

http://pubsonline.informs.org/journal/mnsc

MANAGEMENT SCIENCE
Articles in Advance, pp. 1–22
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Rushed Innovation: Evidence from Drug Licensing

Manuel Hermosilla[a]

[a] Carey Business School, Johns Hopkins University, Baltimore, Maryland 21202
Contact: mh@jhu.edu, https://orcid.org/0000-0002-5851-2303 (MH)

Received: April 23, 2018
Revised: April 22, 2019; September 3, 2019
Accepted: October 18, 2019
Published Online in Articles in Advance: April 8, 2020

https://doi.org/10.1287/mnsc.2019.3530

Abstract. We study the drug licensing behavior (acquisition of rights for developing drugs) of large pharmaceutical firms in the aftermath of large negative shocks to their pipelines, phase 3 failures (P3Fs). We find that P3Fs lead to increased licensing within a year of the event. This result is significant, because one year is a short window given the usual timelines—licensing is a lengthy process that requires extensive planning and careful execution. Supported by a series of additional results, we interpret this finding as a reflection of *rushed* firm behavior. Correspondingly, our main finding is that drugs licensed in these circumstances (within a year of a P3F event) underperform in subsequent development: they are significantly less likely to reach the market compared with others licensed in normal conditions. Further analysis suggests that this underperformance may stem from the influence of rush on activities taking place in the "last mile" of the licensing process and could hinge on the quality of the agreements that firms converge to during contract negotiations.

History: Accepted by Juanjuan Zhang, marketing.
Funding: The author acknowledges the Johns Hopkins Carey Business School [Internal Research Grant].
Supplemental Material: The online appendix is available at https://doi.org/10.1287/mnsc.2019.3530.

Keywords: rush • new product development • new technology • innovation • pharmaceuticals • pharmaceutical productivity • drug licensing • causal inference

1. Introduction

On December 2, 2006, the pharmaceutical giant Pfizer announced "surprising and disappointing" news to its investors (Pfizer Inc. 2006). In its Phase 3 clinical trial, the cardiovascular-targeted experimental cholesterylester transfer protein (CETP) inhibitor Torcetrapib had produced a disproportionate amount of deaths. The trial was stopped, and the drug's development was discontinued immediately. The next trading day, Pfizer's stock price dropped by 11%. Responding to stakeholder pressure, the company's Chief Executive Officer 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, management made good on this promise by licensing the experimental anticoagulant Apaxiban (Pfizer Inc. 2007).

Taken at face value, this timeline is surprising. Drug licensing (the acquisition of development and commercialization rights for experimental drugs) is a highly involved and lengthy process that requires extensive planning and careful execution. Even after candidates are identified, extensive negotiations and due diligence take place, each of which could take the better part of a year. Did Pfizer relax its selection, planning, or execution standards to get this deal over the line quickly? Did the licensing of Apaxiban adhere strictly to Pfizer's strategic goals, or was it primarily shaped as a reaction to Torcetrapib's failure?

We engage with these questions through a systematic analysis of licensing and product development performance in large pharmaceutical firms. Our primary motivation is to illustrate the significance of a type of firm behavior that we think may be both common and consequential. This behavior is encapsulated by the layperson's notion of *rushing*—the proclivity to respond to pressure through immediate, extemporaneous action. The arena of drug licensing is unique in the sense that it allows us to measure both the incidence and consequences of rushed behavior in a context where stakes are large. Phase 3 failures (P3Fs) are large negative shocks that reportedly introduce a significant amount of stakeholder pressure on management. In turn, licensing deals are large transactions which, as noted, require a high level of premeditation. More importantly, the potential consequences of rush can be assessed through a rather incontrovertible performance measure: rates of post-licensing development success. Consistent with the presence of rush, we find that firms like Pfizer do tend to increase their licensing activity shortly after a P3F event. Correspondingly, our main result is that, compared with drugs licensed in normal conditions, those

licensed in these circumstances are significantly more likely to have their development subsequently terminated.

Our empirical strategy takes advantage of the substantive quasi-experimental component governing the incidence and timing of P3Fs. Models are estimated on a comprehensive dataset tracking development and licensing activities of 20 of the largest pharmaceutical firms in the world over a 15-year period (2001–2015). The licensing impact of P3Fs amounts to a 0.05 higher licensing probability (23% lift) within one year of the P3F event—a short window given usual licensing timelines. Further analyses reveal a series of signs that are consistent with rushed behavior. The first of these is a focus on near-future action It is illustrated by the finding that P3F-fueled licensing vanishes in slightly longer windows (two and three years). The second marker is lack of premeditation, and is suggested by three complementary results. Specifically, we find that the P3F-fueled licensing surge (i) does not seem to be motivated by the pursuit of stock market gains, (ii) does not seem to be grounded on pre-established policies linking pipeline events to licensing activity, and (iii) is concentrated in circumstances in which the amount of stakeholder pressure triggered by a P3F event is presumably larger. Based on these results, we take pre-licensing P3Fs as a signal for a rushed licensing process and say that drugs licensed shortly after such an event were *licensed in a rush*. Compared with others licensed in normal conditions, these have a 0.1 higher probability of termination in post-licensing development (16% lift). These results withstand falsification and a number of robustness checks. Analyzing the data through the lens of heterogeneous treatment effects, we find that firms' propensities to license drugs in a rush and later terminate their development may vary widely depending on their contexts (and may not always be significant). Importantly, this analysis also suggests that those firms that are more likely to engage in rushed licensing behavior are also more likely to suffer the negative performance consequences thereof.

We then turn our attention to mechanisms—what explains the underperformance of drugs licensed in a rush? Based on the characteristics of the licensing process, we outline three sets of activities that could be adversely impacted by rush: search, contracting, and due diligence. We implement high-level tests that directly address the former two channels and provide suggestive evidence of the effects at play. If drugs licensed in a rush were always identified through "fresh" search processes (i.e., processes initiated following the triggering P3F shock), we would expect them to exhibit a lower degree of firm/technology fit—rush would limit firms' ability to scout the "right" candidates. In our data, however, rushed licensing does not imply poorer technological fit. We neither find a rush-fueled difference in term of maturities (i.e., development stage at licensing). These results suggest that rushed licensing does not rely on fresh search, suggesting that drugs licensed in a rush likely correspond to "recycled old leads." Our suspicions for the motives of underperformance are thus turned to the possible acceleration of due diligence and/or contracting. Rushed contracting would manifest through agreements that are overly incomplete in the sense that they contain "lose ends," which increase the probability of interorganizational friction. Evidence is consistent with this conjecture: drugs licensed in a rush are twice as likely to be terminated in circumstances linked to an abnormal amount of such friction. Thus, the underperformance result seems to be rooted on activities taking place in the "last mile" of the licensing process and could hinge on the quality of the agreements that firms converge on.

As a whole, this evidence is consistent with the idea that licensing decisions of large pharmaceutical firms can be influenced by rush at the expense of nonnegligible development underperformance. This general result aligns with the views of some industry observers. For example, Mittra (2007) states that "licensing-in products to fill a portfolio gap is a risky strategy" (p. 294). Similarly, Sarnet and Lachman (2005) write that "companies stating that they are 'opportunistic' regarding licensing activities find themselves floundering with an inefficient and resource consuming process resulting in few successful licensing candidates." Licensed drugs now account for half of revenues and a third of the launches of large pharmaceutical firms (Goodman 2009, Kneller 2010). Moreover, the industry's historical struggles with attrition have continued to intensify (Paul et al. 2010, Pammolli et al. 2011)—according to the most optimistic recent estimate, one in over seven therapies tested in clinical trials reaches the market (Wong et al. 2019). Although we cannot formally qualify the (shareholder) optimality of rushed licensing behavior, these caveats suggest that few pharmaceutical executives would be indifferent to our findings. Our mechanistic results suggest that firms may have it in their control to soften the underperformance of rushed licensing, for example, by strengthening contracting protocols.

Our findings directly inform the new product development (NPD) literature which, despite paying significant attention to timing issues, has not yet addressed the concept of rush. One closely related framework showcases the tradeoff between development speed and NPD success (Cohen et al. 1996, Bayus 1997). Speed is conceptualized as a fully premeditated decision that is controlled over long horizons (full development cycle).[1] Although consistent with this framework, the notion of *rushed innovation*

that emerges from our analysis emphasizes distinctively different mechanics: rushed decisions lack premeditation and impact near-future actions. Moorman and Miner (1998) develop another related framework based on the notion of *NPD improvisation*. Defined as the convergence of planning and execution, improvisation can arise as a response to turbulence, and thus be potentially linked to positive outcomes through creative (although premeditated) adaptation. The detrimental impacts of rush that we illustrate are most relevant for contexts in which creative courses of action offer little gain.

Our research also adds to a rich literature focusing on pharmaceutical innovation, particularly to its strands emphasizing productivity (e.g., Cockburn and Henderson 2001), the role of alliances (e.g., Danzon et al. 2005), and the impact of product-related shocks on subsequent innovation (e.g., Ball et al. 2018). The structure of our result is similar to that of Guedj and Scharfstein (2004), who find that, compared with managers of larger biopharmaceutical startups, those of smaller, single-product ones make riskier early-stage development decisions that backfire later in the process. In addition, in a prominent study of drug development timing, Dranove and Meltzer (1994) find that clinically important drugs reach the market sooner. This result is interpreted as a reflection of firms' premeditated and consistent efforts to expedite the development of potential blockbusters.

Lastly, our research contributes to the study of innovation outsourcing. Higgins and Rodriguez (2006) find that pharmaceutical firms tend to engage in Mergers and Acquisitions (M&As) when their pipelines are "running dry." Contrasting with the intuition of our results on licensing, M&As seem to generally resolve the problem of weak pipelines. This substantial difference between the two sets of results likely stems from the relative size of transactions—M&As are much larger deals that require widespread stakeholder support. The process of gathering this support will likely weed out poor prospects and force premeditation. Given their smaller size, licensing deals are likely to receive much less organizational scrutiny, be more discretional, and thus more vulnerable to rush. In addition, by its focus on the demand side of the licensing market, our work also contributes to the literature on markets for technology (Arora et al. 2004), which has maintained a strong "supply-side" focus (Arora and Gambardella 2010).

2. Institutional Background

2.1. The Drug Development Process and Phase 3 Failures

2.1.1. The Process. This highly structured process starts when a molecule's formulation is identified and then fine tuned in the laboratory. At this point, a set of potential therapeutic applications (i.e., molecule/targeted disease combinations, henceforth called "therapies") is identified. These early formulation activities are followed by experiments on animal subjects ("preclinical" trials) aimed at assessing their potential. Our data source collapses these two sets of activities into a single "discovery" stage. Therapies with adequate-enough potential move on to clinical trials on humans. If the development process carries forward, each therapy requires a largely independent set of trials. The clinical trial protocol is sequential and composed of three phases of randomized experiments: Phases 1 (safety), 2 (efficacy), and 3 (safety and efficacy in larger populations and relative to alternatives). Therapies that are successful at all of these stages can be presented to the regulator in application for a commercialization permit ("review" stage).

2.1.2. P3Fs. Because of significant attrition at earlier stages, few therapies that enter the process reach Phase 3 (P3).[2] Therapies that succeed at P3 move on to the review stage, in which failure rates are the lowest. However, P3 success is not ensured—about 40% of therapies fail at this stage (Wong et al. 2019). Poor experimental results (lack of efficacy and adverse events) are by far the primary reason for failure (Hwang et al. 2016). P3 is therefore perceived as the last "big hurdle" to overcome to reach the market.

Because P3 trials aim at providing definitive proof of safety and efficacy, they are the longest and largest in the process. These factors build in significant randomness in terms of the amount of time that it takes for outcomes to be observed. Some of this randomness is introduced through patient enrollment—the "arrival" of patients is not fully controlled by the firm. This is an aspect that challenges sponsors given that patients need to meet several eligibility criteria in addition to being sick (i.e., actively suffering the targeted condition).[3] Another source of randomness comes from the nature of collected data. If these data provide strong-enough evidence against or in favor of the tested therapy, ethical imperatives require that the trial is stopped before its planned completion time. This was the case, for example, for the P3 trial of Torcetrapib that we cited in Section 1 (terminated early because of an abnormally large number of treatment-arm deaths).

Sponsors in general have little scope to engage in strategic information release in the wake of a P3F event. By the time that a therapy reaches P3 it has usually captured the attention (and hopes) of the firm's stockholders, patient organizations, analysts, and investors—"when Phase 3 clinical trial failures happen it is a painful blow—to the drug manufacturer, to investors and to patients" (Merrill 2016). Thus, the trial's evolution is tracked closely by several interest groups. More importantly, clinical trials

are conducted under tight regulations aimed at ensuring the scientific integrity of results and the ethical treatment of patients. The degree of oversight is accentuated for P3 trials, for example, as reflected by the widespread use of independent "data monitoring committees." These committees review incoming trial data on a regular basis and provide recommendations on whether/how to continue (Freidlin et al. 1999). At any time, these entities may recommend that the trial be stopped. Pfizer's announcement of Torcetrapib's discontinuation was made the same day (a Saturday) that the respective monitoring committee recommended the trial's termination (Pfizer Inc. 2006).

2.2. Drug Candidate Licensing

2.2.1. The Licensing Market. Until about 35 years ago, drug discovery was vertically integrated into large pharmaceutical firms along with development and commercialization activities. This configuration changed following a series of scientific breakthroughs beginning in the mid-1970s, which largely eliminated barriers of entry into drug discovery (Pisano 2006). A fringe of biotechnology startups focusing on early-stage innovation activities was established in the decades that followed. Backed by venture capital, these startups are typically led by academic scientists with the goal of "translating" their research findings into therapeutic technologies. Their innovations are monetized in three primary ways: (i) licensing, (ii) "trade sales" (firm is acquired by a larger company), and (iii) self-commercialization. The supply side of the licensing market is primarily populated by biotechnology firms choosing the first route.

Licensing means that the "inventing" firm sells the commercialization rights of a set of developing therapies to another firm. Selling firms are called "out-licensors" and buying firms "in-licensors." Because our focus here is on large pharmaceutical "buyers," we will simplify by just using "licensing" instead of "in-licensing" (unless differentiation is useful). These transactions are based on negotiated contracts, which typically deliver most of the potential compensation on a contingent basis (milestone payments and royalties on market revenues). The purposes of this contingent component are to spread risk, mitigate informational problems, and ensure the biotechnology firm's continued involvement in development activities (Mason et al. 2008). As reflected by the volume of U.S.- and European-based licensing deals—$40, $43, and $57 billion in 2007, 2010, and 2016, respectively (Giovannetti and Spence 2017)—this is a large and growing market.[4]

2.2.2. Incentives to License. From the point of view of in-licensing pharmaceutical firms, licensing constitutes the most lean and expedited route to integrate cutting edge advances into their pipelines (Cockburn 2004). For biotechnology firms, a primary incentive to choosing licensing over self-commercialization is to benefit from the (in-licensing) partner's "complementary assets" (Teece 1986, Gans et al. 2002). Other than funding, these capabilities may include "know-how" (e.g., regulatory affairs and implementation of clinical trials) (Powell 1996) as well as assets that are important for massive commercialization (e.g., branded reputation and established sales forces). Thus, relative to self-commercialization, licensing is better suited for the development of therapies that target large markets or those that require complex or costly clinical trials. In addition, from the point of view of the biotechnology firm's financial backers, licensing has the benefit of outlining a safer profile of returns (some compensation is received even if therapies do not reach the market). Relative to trade sales, licensing allows founding scientists and investors to retain control of the firm while participating in the financial upside of potential blockbuster therapies.[5]

2.2.3. Contracting Frictions and the Liabilities of Rush. Important contracting frictions are primarily rooted in the rapidly evolving and deepening scientific basis from which biotechnology innovation draws (Powell 1996, Pisano 2006). In this environment, the licensing supply is continuously infused with therapies that rely on novel, diverse, and rapidly evolving technological approaches. This dynamism makes it difficult for potential in-licensors to identify the "right" candidates. It also increases the need for thorough due diligence—scientific novelty and complexity create a fertile ground for informational asymmetries. Most large pharmaceutical firms have dedicated business development and licensing offices tasked with constantly scouting the market for licensing opportunities (Davies 2013).

Drawing on interview evidence, Alcacer et al. (2009) illustrate some of the orchestrated search activities performed by pharmaceutical in-licensors. Major stages of the process include a worldwide search and screening for licensing opportunities followed by an in-depth analysis of each identified candidate, and a final screening. For one of the (anonymous) interviewed firms, an initial screening of candidates from 80–100 firms typically delivered one or two licensing opportunities. The statistics of Davies (2013) suggest that these figures may pale in comparison with those of the very largest pharmaceutical firms (such as the ones in our sample). For example, Merck & Co. reviewed over 8,000 opportunities in 2011, of which less than 1% produced a licensing deal. Around the same time, Roche acted on less than 2% of reviewed opportunities. These low conversion rates suggest that good opportunities are not easy to find.

After licensing candidates have been selected, deep due diligence and valuation activities begin. These may take the better part of a calendar year, or longer (Truex 2018). The importance of due diligence arises in part from biotechnology firms' large informational advantages (Pisano 1997, Hermosilla 2016). In-licensors need to verify the integrity of the underlying science and intellectual property (breadth and vulnerabilities). Accordingly, pharmaceutical in-licensors deploy multifunctional teams (including lawyers, executives, and field expert scientists) at every stage of the process (Alcacer et al. 2009).

Lastly, it is important to note that much of the scientific knowledge embedded in licensed therapies is tacit at the time that agreements take place (Pisano 2006). For this reason, postlicensing development requires the continued involvement of the out-licensing firm's scientists. Accordingly, the negotiation and design of contracts are of particular importance, and as such, require a particularly high degree of involvement from the in-licensing firm. If parties fail to adequately specify boundaries, expectations, or contingent decision rules, development outcomes may be jeopardized. These contracting aspects will be revisited in Section 6, where we investigate the mechanisms underlying our main result.

3. Data

3.1. Source 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.[6] From this source, we assembled a dataset focusing on the licensing and development activities of 20 of the largest pharmaceutical firms globally for the 15-year period of 2001–2015.[7] During this period, selected firms had the highest licensing frequency within the full data (according to the parameters described below).

Firms in the sample actively developed therapies spanning 20 therapeutic areas. We retained data from the 17 areas in which key variables have enough variation. Because the economics and science of drug development have large area-specific components, we assume that licensing and termination decisions are made at the firm/therapeutic area-level. We thus call each firm/area pair a decision-making "unit." A unit enters the sample in the first quarter that it actively develops a therapy. The sample has a total of 230 units.

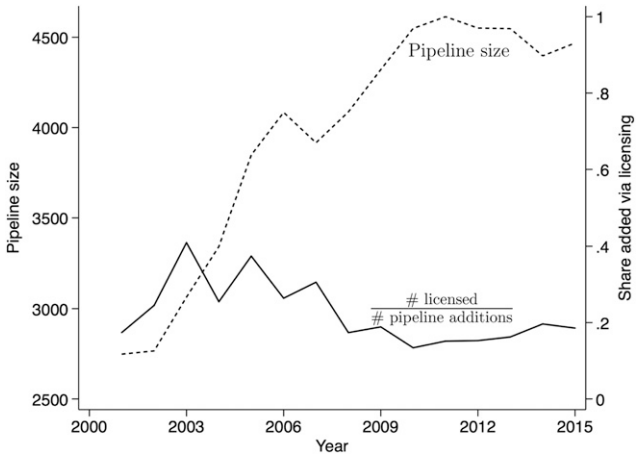
Variable construction and data processing are easier to follow if we consider two datasets. The first includes licensing and development histories of all therapies licensed by units in the sample over the covered period. We call this the "analysis sample." The second dataset includes the universe of all records available from Cortellis. These will be used to construct controls and secondary outcomes. We call this the "broad Cortellis sample."

3.2. Licensing

Here, we cover the main characteristics of the set of licensed therapies included in the analysis sample.[8] Licensing agreements vary in the geographical scope of traded rights—we focus on worldwide deals. Licensing deals may also bundle rights for more than 1 therapy (sample average of about 2.5 therapies), and bundled therapies may span across areas. We treat each licensed therapy as an individual licensing event.[9] Lastly, because financial compensation terms are rarely available from the data, they play no part in our analysis. The sample that results after imposing these filters includes 3,495 licensed therapies. About 48% of these are in the cancer area. Endocrinological and metabolic therapies and autoimmune/inflammatory therapies follow distantly (9% and 8%, respectively). Most licensed therapies are early stage: 25% are licensed before clinical trials, and 62% are licensed before phase 3. The average unit licenses an average of about two therapies each year, although the distribution is quite skewed.

Figure 1 presents some aggregate trends that help place this licensing activity into context. The dashed line of Figure 1 shows the number of actively developing therapies maintained by units in the analysis sample ("pipeline size" on the left axis). This trend suggests that pipelines experienced significant growth over the covered period.[10] The solid line plots the share of all therapies added to these pipelines

Figure 1. Main Innovation and Licensing Trends in the Analysis Sample



Notes. The analysis sample includes 20 large pharmaceutical firms. The dashed line corresponds to aggregate pipeline size, measured as the number of experimental therapies on active development. The solid line indicates the share of pipeline additions that are obtained via licensing.

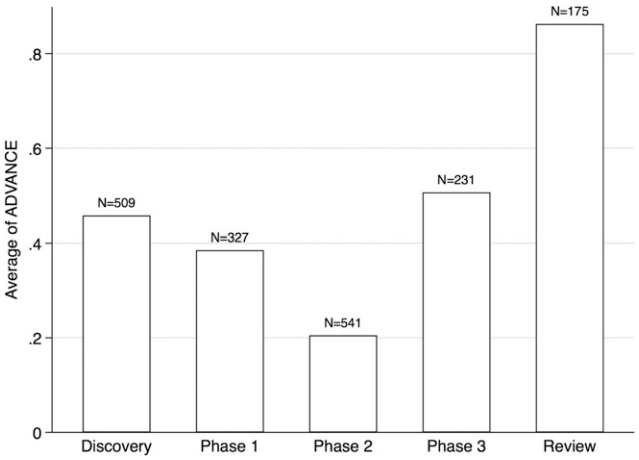
through licensing. This trend suggests that licensing was an important and relatively stable source of all pipeline additions during the sample period (20%–30%).

3.3. Post-licensing Performance

Post-licensing development performance is tracked at the therapy/stage level based on reported outcomes. The compiled sample includes 1,783 such outcomes, drawn from 1,375 therapies (from a total of 3,495 licensed in the sample). The primary reason for missing outcomes is the inability to link the identity of licensed therapies to their postlicensing development outcomes.[11] In addition, among therapies for which this link is available, data are missing if stage development was still ongoing at the time of data download.

We define the variable *ADVANCE* to measure performance. Following standard practice in the literature (e.g., Wong et al. 2019), *ADVANCE* is set to one if a therapy is observed advancing past a given stage (i.e., next-stage development is observed). If development is terminated at that stage, *ADVANCE* equals zero. Although the majority of terminations in the data are because of discontinued or indefinitely halted development (86%), others come from the dissolution of the underlying licensing agreement (9%).[12] The bars of Figure 2 describe averages of *ADVANCE* at different stages. P2 exhibits the lowest success rate (about 20%), and the review stage exhibits the highest (about 85%). For P3, the success rate is close to 50%. These success rates are all within range of previously reported figures, although generally closer to the lower bounds.[13] Also note that the sample has good coverage across stages and that the majority of observations is from stages prior to P3.[14] Thus, our analysis of post-licensing performance will primarily focus on cases where economic/strategic considerations may play a meaningful role.

Figure 2. Post-licensing Development Performance: Stage Advancement (Success) Rates



3.4. P3F Shocks

Given the units and period covered by the analysis sample, the P3 failures of 498 therapies are relevant for our analysis. The majority of these therapies (about 80%) is not part of the analysis sample. That is, most failed therapies were incorporated into pipelines through in-house discovery, M&As, or pre-sample licensing. About 24% of units experience one of these failures, 12% experience two, and 27% experience three or more. The remainder (37%) experience none (these will be useful as part of control groups). Consistent with the sustained growth of pipelines over the covered period, the frequency of these failures is increasing over time—about 70% of them occur in the second half of the sample. Even though the cancer area is a distant first in terms of frequency (31%; followed by cardiovascular and endocrinological/metabolic with about 11% each), the distribution is quite spread out across areas.[15]

To implement our research design, we consider "aggregate" P3F shocks. In practice, these shocks are defined by the following condition: did the unit in question experience the P3 failure of at least one of its therapies over a given time window? Aggregation windows vary for our two main analyses and consequently, also for the resulting number of aggregate P3F shocks that we can exploit in each case. (The resulting number is in the hundreds in both cases. Aggregation details and descriptives are provided in Section 4.4.) It is also worth noticing that, as a result of aggregation, some shocks are composed of the failure of more than one therapy. We characterize the frequency of these "non-standard" shocks and probe their influence on our results in a robustness analysis.

Another aspect of these aggregate P3F shocks is that they are not all fully independent events. This is rooted in a pervasive feature of drug development that we noted in Section 2.1: namely, the fact that molecules can be developed to treat more than one condition (i.e., a single molecular entity can be associated with more than one therapy). Specifically, in our case, the 498 therapies that failed in P3 come from 337 different molecules. About half of these therapies are "independent failures," in the sense that they represent a molecule's unique P3 failure in our data. The other half of failures are connected to another in the data through a common molecule. It is important to note that this last case does not always create interdependencies among aggregate P3F shocks. An interdependency is created only if a molecule's failed therapies span more than one therapeutic area (more than one unit is affected) or if the molecule's therapies fail over different aggregation windows.[16] These situations are, however, relatively infrequent (less than 16% of shocks are affected). In our analysis of non-standard shocks, we provide additional detail

and robustness. Evidence overall suggests that these issues do not drive our main conclusions.

4. Empirical Strategy

Our strategy exploits the incidence and timing of the aggregate P3F shocks described above. For simplicity, these are henceforth referred to simply as "shocks," "P3Fs," or "P3F events." Estimation methods are selected with the aim of holding constant the degree of "surprise" associated with each shock.

4.1. Estimation Methods

We first consider propensity score matching (Rosenbaum and Rubin 1983). Here, P3F incidence plays the role of treatment. It is coded as $W = 1$ for observations exposed to a shock (treated) and $W = 0$ otherwise (control). The propensity score corresponds to an estimate $\pi_i = \Pr(W = 1|Z_i)$, which represents the probability that an observation i receives the treatment given a set of predictors Z. Thus, estimated π values capture P3F surprise. Following standard practice, we estimate this probability using a logit model. Estimation holds P3F surprise constant through the computation of counterfactual outcomes. In particular, the counterfactual outcome $\hat{Y}_i^{(1-W_i)}$ is computed as the average outcome within a "neighborhood set" composed of observations with opposite treatment assignment but similarly valued propensity scores. The causal impact of P3F shocks is summarized by the average treatment effect (ATE), defined as $\tau = E[\hat{Y}_i^{(W=1)} - \hat{Y}_i^{(W=0)}]$, with $\hat{Y}_i^{(W_i)} = Y_i$. In our context, this ATE estimate is interpreted as the expected (sample-level) impact of a P3F on the outcome Y. For inference, we use the heteroskedasticity-consistent standard errors of Abadie and Imbens (2006). The set of predictors Z includes a broad set of contextual factors identified from related literature as well as a variable that proxies for a unit's risk of experiencing a P3F based on P3 microdynamics and portfolio composition. All of these variables are described below.

We also consider the causal forest estimator (Wager and Athey 2018, Athey et al. 2019), which also accounts for P3F surprise through a propensity score. The main advantage of this approach is that it delivers heterogeneous "conditional average treatment effect" (CATE) estimates, which are computed separately for each observation (along with standard errors) without requiring distributional assumptions. These heterogenous estimates will not only be useful for additional characterization and structural insight, but also to control for variation that cannot be accommodated by propensity score matching.

We implement this estimator using the generalized random forests framework of Athey et al. (2019).[17] Propensity scores are computed as before except that the flexible random forests technique (Breiman 2001) is used instead of the logit. The more important difference between methods lies in the way that treatment effects are computed. The CATE estimate for observation i, τ_i, is obtained by fitting the following weighted moment condition:

$$E[(Y_{i'} - m_{i'}) - (W_{i'} - \pi_{i'}) \cdot \tau_i] = 0, \tag{1}$$

where $m_{i'}$ corresponds to a random forests prediction of Y_i. Thus, τ_i is estimated from the deviations of outcomes and treatment assignments from their respective predictions. (Computation details are provided in Online Appendix A.) This expression is also helpful to illustrate one of the advantages over propensity score matching, namely, the ability to control for variation that is relevant for outcomes but not for propensity scores. In particular, propensity scores π will be computed using variables measured prior to licensing, whereas expected outcomes m will be computed using variables measured during the period that immediately precedes outcome completion. This aspect is of particular relevance for the analysis of post-licensing performance, because outcomes are usually observed several years after the treatment (i.e., pre-licensing P3F) is experienced.

4.2. P3F Risk Score

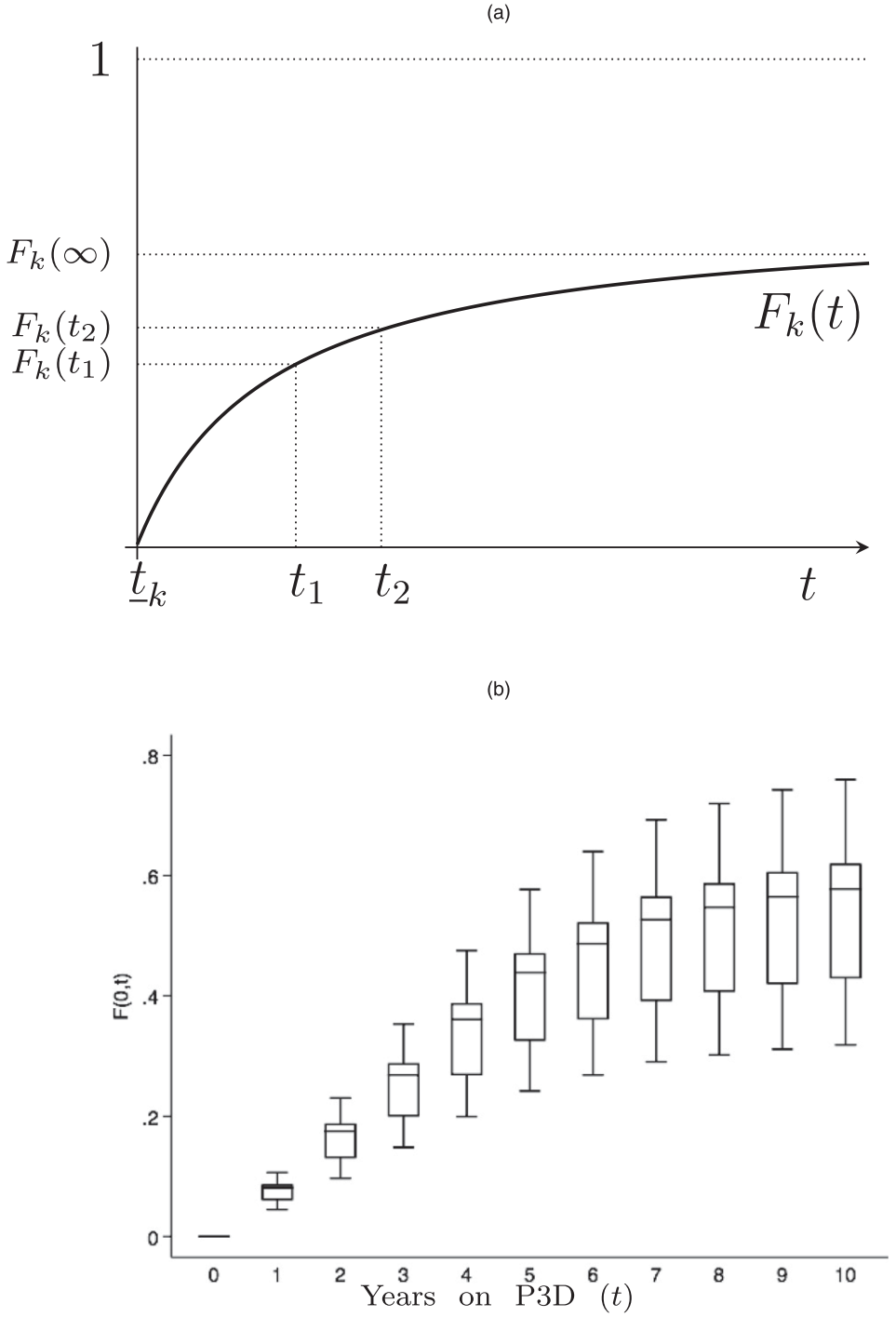
We introduce a "P3F risk score" to better distill the exogenous component of P3Fs. We first observe that the likelihood of experiencing a P3F should increase with the number of therapies that a unit has on active P3 development ("P3D"). This number varies across units and time. Even if this number is held constant, P3Fs may be more or less likely to occur depending on how long therapies have been on active P3D. The risk score that we introduce combines these two sources of variation in a natural way.

The score is formally defined as the probability that unit j experiences a P3F during the time interval $[t_1, t_2]$. To formalize it, denote unit j's active P3D portfolio at time t_1 as $\mathcal{A}_j(t_1)$. Also denote as $r_k(t_1, t_2)$ the probability that a therapy k introduced to P3D at an earlier time $\underline{t}_k$ fails during $[t_1, t_2]$—the gradient implied by r_k stems directly from the microdynamics of P3 trials. With these elements, the risk score is specified as

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

Thus, R reflects the probability complement for the event in which none of the therapies in $\mathcal{A}_j(t_1)$ fails during $[t_1, t_2]$. The key ingredient in this formulation is r_k. By its definition in the previous paragraph,

Figure 3. Elements for the Construction for the P3F Risk Score



Notes. (a) Theoretical illustration. 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. (b) Estimated cumulative P3 failure subhazards (empirical correlates of $F_k(t)$) estimated via competing risks (Fine and Gray 1999) on a large sample of P3 development histories (all therapeutic areas in period 2001–2005). Variability shown by boxes comes from differences across therapeutic areas.

r_k corresponds to a duration or "hazard" 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 $[t_1, t_2]$. Furthermore, r_k is defined not only 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 failure"). Thus, to compute r_k, we require a cause-specific hazard estimate that we can then integrate

over $[t_1, t_2]$. The competing risks framework of Fine and Gray (1999) provides us with "subhazard" estimates that fit this description.

From competing risks estimates obtained from data on P3 trial durations and outcomes, we assemble functions $F_k(t)$. These functions represent the probability that a therapy k that was introduced to P3 at time t_k fails at or before time t. Figure 3(a) illustrates one such function. According to this curve, therapy k has probability $F_k(t_1)$ of failing at or before time t_1. Its probability of failing any time after t_1 is $F_k(\infty) - F_k(t_1)$. Because not all therapies fail, these probabilities do not add up to one. Instead, they add up to the overall probability of P3 failure, $F_k(\infty)$. The value 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 be realized within $[t_1, t_2]$. 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 over 9,000 P3 observations (duration and outcomes) available from the broad Cortellis data, excluding therapies that appear in the analysis sample. The model is specified to allow for variability across areas (area-specific indicators) and time periods (five year-period indicators). To illustrate these results, Figure 3(b) presents F estimates for the 2006–2010 period (boxes capture variability across areas). The R values entering estimations are described below.

4.3. Contextual Factors

Licensing and advancement/termination outcomes may be influenced by the context faced by firms. Here we specify a set of variables that capture leading such contextual factors. These are compiled from the insights of related research and will play the role of controls (or predictors, or covariates) in our analysis.

4.3.1. Recent Productivity. The evidence of Higgins and Rodriguez (2006) suggests that declining productivity may prompt pharmaceutical firms to acquire other firms as a means to "revitalize" their pipelines. In the face of poor recent productivity, firms may also revert to the licensing market. We thus introduce the variable *RTER_OWN*, which tracks the number of a unit's recent terminations (excluding P3Fs).

We face two hurdles to adequately specify this variable. First, the importance of recent terminations depends on the development stage at which they unfold. For example, a therapy that is being developed at P2 is more likely to reach the market than one being developed at P1. We should thus register a poorer recent productivity record when the former is terminated. Accordingly, we require a scheme to consistently weight terminations that occur at different stages. Second, we also note that the relative importance of licensing is roughly stable despite the sustained growth of pipelines during the covered period (Figure 1). If *RTER_OWN* was defined in absolute terms (e.g., count of terminations), its influence on our estimates could primarily stem from its temporal (industry-wide) variation. We combine weighting and normalization steps to account for these issues.

The weighting step follows the approaches of Higgins and Rodriguez (2006) and Girotra et al. (2007). We begin by counting the number of terminations at each stage s for each unit j over a measurement window (window details provided below). We label each of these counts T_{js} and aggregate them using "reach-the-market" probabilities $\{p_{as}\}$ as weights. These weights represent the likelihood that an area a therapy at stage s continues its development all of the way into the market.[18] (We are thus assuming that the importance of a therapy's termination is proportional to its statistical proximity to the market.) With these elements, we compute unit j's weighted number of terminations as $\hat{T}_j = \sum_s T_{js} \cdot p_{as}$. In the normalization step we compute *RTER_OWN* as $100 \cdot (\hat{T}_j / \hat{A}_j)$, where $\hat{A}_j$ represents the unit's (importance-weighted) number of actively developing therapies (computed in the same fashion as $\hat{T}_j$). Therefore, *RTER_OWN* represents the importance of a unit's recent terminations relative to the strength of its pipeline during the measurement window.

4.3.2. Pipeline Strength and Precompetitive Environment. Girotra et al. (2007) find that stronger pipelines soften the negative impact of P3Fs on the sponsoring firm's stock market valuation and Chan et al. (2007) suggest that termination and licensing decisions may jointly depend on a sponsor's pipeline strength. We therefore introduce a pipeline strength variable *PIPESTR*. We construct it based on the (importance-weighted) number of actively developing therapies that we introduced above, $\hat{A}$. For a unit j, *PIPESTR* is formally defined as $100 \cdot (\hat{A}_j / (\hat{A}_j + \hat{A}_{-j}))$, where $\hat{A}_{-j}$ is the summation of $\hat{A}_{j'}$ values for all units $j' \neq j$ in the area.[19] Thus, *PIPESTR* measures the strength of j's pipeline relative to the area's overall. Also note that *PIPESTR* also reflects the intensity of premarket competition (100 − *PIPESTR*).

We also create a variable tracking the importance of recent licensing, *RLIC_OWN*. This variable helps us to account for a units' baseline licensing propensities and potentially confounding "licensing waves."[20] Using the weighting procedure, we compile the unit's recent licensing activity into the variable $\hat{L}_j$ (aggregates over therapies licensed at different stages), and then set $RLIC\_OWN_j = 100 \cdot (\hat{L}_j / \hat{A}_j)$. The potential

influence of competitively motivated licensing effects is captured by the analogous "rest-of-area" variable $RLIC\_RA_j = 100 \cdot (\hat{L}_{-j}/\hat{A}_{-j})$. Thus, both of these variables measure recent licensing activity relative to contemporaneous pipeline strength. Because the productivity of competitors may impact future market opportunities, we also track "rest-of-area" productivity through $RTER\_RA_j = 100 \cdot (\hat{T}_{-j}/\hat{A}_{-j})$.

4.3.3. Competitive Environment. Chan et al. (2007) also stress that licensing and development decisions may depend on a firm's portfolio of marketed therapies. Firms with broad and "young" portfolios may be under less pressure to supplement their pipelines (or relax advancement thresholds) than others with weak or "old" portfolios.[21] We construct *MKTSHARE* to control for this effect. The variable is defined as the percentage of all marketed therapies in j's area that has been brought to market by unit j.[22]

4.3.4. Experience. Previous research has highlighted the role of sponsors' development experience ("scale") and diversification thereof ("scope") as determinants of innovative productivity (Henderson and Cockburn 1996, Cockburn and Henderson 2001), including in the context of licensing (Danzon et al. 2005, Arora et al. 2009). Following the approach of these papers, we incorporate these constructs through the variables *SCALE* (each unit's historical number of therapies introduced to development; logged) and *SCOPE* (each unit's Herfindahl–Hirschman concentration index of therapies historically introduced to development; across targeted diseases). Larger values of *SCALE* point to larger amounts of accumulated experience; larger values of *SCOPE*, to less diversified experience.

4.3.5. Financials. Licensing and termination decisions may depend on the sponsoring firm's financial conditions, for example, through liquidity effects. We retrieved sales data from COMPUSTAT to account for this type of effects. Values are logged and recorded by *LOGSALES*. Possible technological leadership effects are accounted for similarly, through *LOGRD* (R&D expenditures). Both variables are at the quarter/firm level.

4.3.6. Secular Variation. The literature studying drug development performance unveils significant differences in success rates across stages and areas (e.g., Wong et al. 2019). Therapeutic area indicators will thus be included in all of our analyses. The analysis of performance will also include development stage indicators. In addition, Hermosilla (2016) argues that the licensing market is afflicted by an adverse selection problem that is less problematic at more advanced stages. Hence, to analyze performance, we will also use licensing -stage indicators. Lastly, an extensive literature highlights the role of market size as a determinant of pharmaceutical innovation (Acemoglu and Linn 2004, Dranove et al. 2014, Dubois et al. 2015), including the propensity to license (Hermosilla and Wu 2018). To control for market size variability, area indicators will be supplemented with year fixed effects.

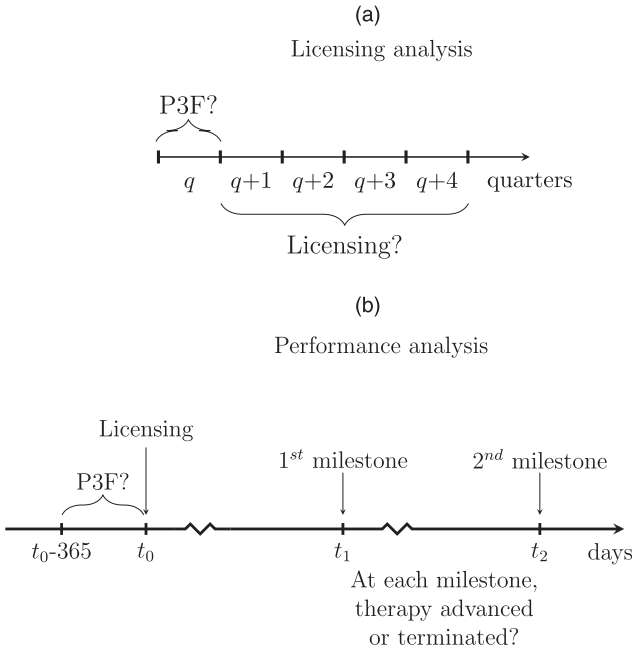
4.4. Measurement Windows and Descriptive Statistics

4.4.1. Licensing Analysis. Our goal in this first analysis will be to assess whether P3Fs increase licensing activity in the near future. Figure 4(a) illustrates our research design. Observations are defined at the unit/quarter level. The incidence and risk of P3F shocks are measured at each focal quarter q, yielding a total of 418 P3F shocks. The outcome *DLICENSE* equals one if the unit in question licenses at least one therapy in $\{q+1, .., q+4\}$ and zero otherwise. Recent terminations and licensing are measured over $\{q-3, .., q\}$. This last window is also used to measure *PIPESTR* and *MKTSHARE*. Experience variables are measured from the broad sample based on the unit's entire development history up to (and including) quarter $q-1$.

The resulting data set contains 6,479 observations, 6.5% of which are exposed to P3F shocks (treated).[23] Columns (1) and (2) of Table 1 present means and interquartile ranges (IQRs) for each variable. Most variables exhibit a right-skewed distribution. Given that licensing events are relatively infrequent, this skewness is particularly pronounced for a unit's recent licensing (*RLIC_OWN*). The solid and dashed lines of Figure 5 show the cumulative distributions of P3F risk scores entering this analysis. These range between 0 and 0.6. The dashed distribution (treated observations) first-order stochastically dominates the solid one (control observations). That is, treated observations tend to have higher risk scores than control ones. This feature is reassuring because it suggests that our formulation is effectively capturing P3F risk. (Recall that hazard estimates used to construct risk scores did not utilize analysis sample data.) Lastly, the minimum risk score among treated observations is approximately 0.01. About 5% of observations are below this threshold. For these, P3F shocks are arguably too unlikely. They are therefore dropped for matching and causal forest estimations.

4.4.2. Performance Analysis. The question here is whether pre-licensing P3Fs impact post-licensing development performance. Figure 4(b) describes the approach used to construct variables. P3F shocks, their risk, and contextual factors are measured within the year (365 days) that ends the day prior to each therapy's licensing. These variables will be used to

Figure 4. Empirical Design



estimate propensity scores.[24] The outcome variable *ADVANCE* is measured at all completed milestones. Because these outcomes may depend on the recent contexts faced by each unit, contextual factors are measured again within the year prior to milestones' completion. There is a total of 1,783 observations (287 treated), each of which corresponds to a licensed therapy/completed milestone pair.

Columns (3) and (4) of Table 1 show descriptive statistics for variables measured prior to licensing. Because the sample focuses exclusively on contexts leading up to licensing, some of the differences with respect to statistics of columns (1) and (2) are due to selection. Descriptives for factors measured prior to the completion of each milestone are presented in columns (5) and (6). The average time between licensing and the completion of the first milestone is about 3.5 years, and about 2.5 between the second and third (when available). The relatively minor differences between columns (3) and (5) are explained by the changes in the contexts faced by units over these periods. The dotted lines of Figure 5 present the distributions of risk scores measured over the same window. Because of the wider measurement windows, risk scores are generally larger than in the licensing analysis. The minimum score at which a P3F is observed is about 0.1.

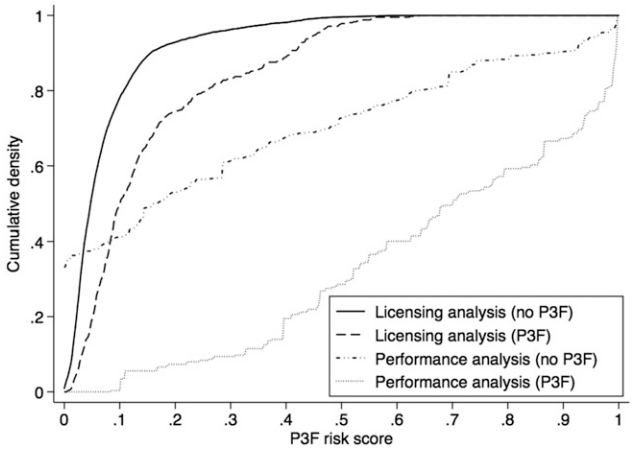
4.5. How Random Are P3F Shocks?

Recall that our inference relies on variability coming from the incidence and timing of P3Fs, and that our methods use propensity scores to isolate the random component in it. Here, we characterize this random component by asking whether contextual factors predict P3F incidence. We use the dataset constructed for the licensing analysis, in which observations are specified at the unit/quarter level.[25] Results are composed of two sets of logit estimates, which are presented in Table 2. In addition to all shown contextual factors, both specifications include year and area fixed effects. The only difference between specifications is the inclusion of the P3F risk score, which is omitted from the specification of column (1). The estimated coefficients for *PIPESTR* and *SCALE* in this

Table 1. Descriptive Statistics for Contextual Factors

		(1)	(2)	(3)	(4)	(5)	(6)
		Licensing analysis		Performance analysis			
		At focal quarter		Prior to licensing		Prior to milestone completion	
Variable	Construct	Mean	IQR	Mean	IQR	Mean	IQR
RTER_OWN	Recent productivity (unit)	5.52	8.81	8.26	11.18	8.05	10.43
PIPESTR	Pipeline strength (unit)	1.17	0.97	1.15	1.02	1.12	0.81
RLIC_OWN	Recent licensing (unit)	2.56	0.00	5.96	4.74	2.70	2.01
RLIC_RA	Recent licensing (rest of area)	1.26	1.10	1.77	1.32	1.12	0.93
RTER_RA	Recent productivity (rest of area)	9.66	3.62	8.89	3.23	9.21	3.84
MKTSHARE	Market share (unit)	1.94	2.01	2.04	2.18	1.96	1.72
SCALE	Development experience (unit)	3.59	1.54	4.83	1.47	4.39	1.36
SCOPE	Development diversification (unit)	19.13	13.59	13.14	11.87	15.33	12.51
LOGSALES	Log sales (firm)	0.67	1.23	0.61	1.09	0.59	1.11
LOGRD	Log R&D expenditures (firm)	0.81	1.09	0.82	0.88	0.82	0.86
Observations		6,479		1,783		1,783	

Figure 5. Estimated P3F Risk Score Distributions



column suggest that P3Fs are more likely to arise from units with more robust pipelines. The main result is that the statistical significance of contextual factors vanishes when the risk score is included in column (2). That is, none of the contextual factors can be taken as a systematic predictor of P3F incidence—all regularities are captured by the risk score.[26] These results thus suggest that, after controlling for a unit's P3F risk, all remaining P3F incidence variation can be viewed as random.

5. Main Results

5.1. Licensing

A positive impact of the P3F treatment on *DLICENSE* would support the idea that P3Fs fuel licensing activity. Panel A of Table 3 presents a series of estimates that speak to this effect. Column (1) presents a difference-of-means estimate computed as the average of *DLICENSE* for treated observations minus that for control ones. Because the latter is 0.22, the estimated difference of 0.1 suggests that P3Fs lift the near-future licensing probability by about a 45%. The estimate of column (2) is obtained from a linear regression of *DLICENSE* on the treatment indicator and contextual factors (including area and year fixed effects). The estimated coefficient of 0.33 also supports the hypothesis. Column (3) of presents the propensity score matching ATE. Although smaller, the estimated impact aligns with previous estimates. It indicates that P3Fs increase the licensing probability by 0.05 (23% lift).[27]

A series of checks is performed to see if we can trust these results. First, we repeat the analysis but considering a falsified (randomly assigned) treatment. Results are shown in panel B of Table 3. Estimates shrink, and statistical significance goes away. Second, we test whether our results could be influenced by "competitive" P3Fs (i.e., P3Fs experienced by other units in the same therapeutic area). We enrich the linear probability specification of column (2) with an indicator for the incidence competitive P3Fs. Whereas the estimated coefficient for the baseline ("own") P3F indicator remains about constant (estimate of 0.33, $p < 0.05$), that for competitive P3F incidence is small and insignificant (estimate of −0.01, $p = 0.184$).[26] Third, we consider the issue of non-standard P3F shocks outlined in Section 3.4. Recall that non-standard P3F shocks arise due to aggregation (more than one failed therapy) or interrelatedness (a molecule's multiple therapies failing at different times or areas). We reestimate the linear probability specification, successively removing the sets of treated observations affected by each of these issues. Estimated coefficients remain positive, and statistical significance is preserved in the majority of cases. An additional specification that controls for non-standard treatments through indicators produces the same qualitative result. A characterization of each issue and estimation results are presented in Online Appendix D. We conclude that we can trust the estimates of panel A of Table 3 as the causal effect of P3Fs on near-future licensing.

We now investigate the extent to which this P3F-fueled licensing activity exhibits signs of rushed behavior. We consider two markers of rush: (i) a focus

Table 2. Logit Estimates for P3F Incidence

	(1)	(2)
P3F risk score		6.072***
		(0.741)
PIPESTR	0.227**	0.099
	(0.104)	(0.107)
MKTSHARE	−0.048	−0.003
	(0.052)	(0.051)
RLIC_OWN	−0.018**	−0.014
	(0.008)	(0.008)
RLIC_RA	−0.019	0.017
	(0.096)	(0.096)
RTER_OWN	0.001	0.010
	(0.008)	(0.008)
RTER_RA	0.046	0.047
	(0.036)	(0.037)
SCALE	0.415***	0.137
	(0.124)	(0.131)
SCOPE	0.003	0.000
	(0.009)	(0.008)
R&D	−0.147	−0.247
	(0.149)	(0.150)
SALES	0.143	0.176
	(0.126)	(0.131)
Observations	6,479	6,479

Notes. Observations are at the unit/quarter level. The dependent variable equals one if a P3F event is observed and zero otherwise. Both models include year and therapeutic area fixed effects. Robust standard errors are presented in parentheses.

**$p < 0.05$; ***$p < 0.01$.

Table 3. Main Results for the Licensing Analysis

	(1)	(2)	(3)
	Difference of means	Linear probability	Propensity score matching
Panel A: Outcome *DLICENSE*			
Estimate	0.100***	0.033***	0.050*
Standard error	(0.021)	(0.014)	(0.029)
Observations	6,479	6,479	5,955
Panel B: Falsified treatment			
Estimate	0.003	−0.005	−0.005
Standard error	(0.021)	(0.018)	(0.025)
Observations	6,479	6,328	6,159
Panel C: Outcome *DLICENSE2*			
Estimate	0.105***	0.018	0.039
Standard error	(0.025)	(0.016)	(0.038)
Observations	6,479	6,479	5,955
Panel D: Outcome *DLICENSE3*			
Estimate	0.094***	−0.010	0.024
Standard error	(0.025)	(0.014)	(0.035)
Observations	6,479	6,479	5,955
Panel E: P3S treatment			
Estimate	0.069***	0.015	0.036
Standard error	(0.021)	(0.030)	(0.036)
Observations	6,328	6,328	5,880

Notes. Column (1) shows the average outcome difference between treated and control observations. Column (2) shows the coefficient estimated for the treatment indicator in a linear probability specification (specification described in the text; robust standard errors reported). Column (3) shows the ATE estimate obtained from propensity score matching (Abadie–Imbens heteroskedasticity-consistent robust standard errors). The smaller number of observations in this column follows from the selected sample (common propensity score support, matched observations, $R \geq 0.01$).

*$p < 0.1$; ***$p < 0.01$.

on near-future actions and (ii) lack of premeditation. To evaluate the first of these we define two additional outcomes, *DLICENSE2* and *DLICENSE3*. The first is an indicator for any licensing activity occurring in $\{q+1,..,q+8\}$; the second, one for $\{q+1,..,q+12\}$. Panels C and D of Table 3 present estimates obtained using these outcomes. Relative to the baseline (*DLICENSE*), linear probability and propensity score estimates shrink and lose significance. This suggests that the licensing impacts of P3Fs do not span beyond the immediate near-future (i.e., one year).[29]

To evaluate the level of premeditation behind the licensing surge, we first ask whether firms could be motivated by a stock market payoff. This analysis is prompted by previous research showing that P3Fs adversely impact the sponsoring firm's stock market valuation (Sharma and Lacey 2004, Girotra et al. 2007). Following a P3F, it would thus be conceivable for firms to deliberately increase the amount of licensing as means to restore their stock market valuation. We investigate this idea through a standard event study analysis (Srinivasan and Hansens, 2009). According to this methodology, the conjectured motivation would be supported if we found that there are positive cumulative abnormal (stock price) returns around licensing events. These effects are undetectable from the data, however, suggesting that this motivation may be weak if present at all. The analysis and results are presented in Online Appendix E.

We next evaluate the licensing impacts of P3 success events (P3Ss). These are positive pipeline shocks, roughly symmetrical to P3Fs. Our rationale is that symmetric licensing effects would be consistent with pre-established policies linking pipeline events to licensing, and thus illustrative of premeditation. We identified 479 P3Ss in the sample and constructed a "P3S risk score."[30] The analysis was then reproduced, maintaining all previous data assembly and estimation protocols. Panel E of Table 3 shows the results. Linear probability and propensity score estimates are small and statistically insignificant. Thus, the referenced policies do not seem to be in place. The asymmetry

of the licensing responses to P3Fs and P3Ss is further consistent with the contrasting nature of the problems triggered by each type shock:whereas P3Fs create pipeline gaps, P3Ss create pipeline surpluses. Because the latter problem can be dealt with at full discretion and without uncertainty, P3Ss are not expected to introduce stakeholder pressure on management. Our last analysis probes this rationale further by testing whether the licensing surge is stronger in instances where the amount of stakeholder pressure triggered by P3Fs is presumably larger.

To implement the test, we borrow the notion of "desperation" from Higgins and Rodriguez (2006), which they use to describe pharmaceutical firms in weak innovative positions (i.e., pipelines or portfolios that are "running dry"). We identify these cases through the indicator *DESPERATE*, which is activated when a unit j's *RTER_OWN* value belongs to the top quartile of the variable's distribution. Thus, *DESPERATE* identifies units that have recently experienced the termination of an unusually large portion of their pipelines. Using this measure, we estimate the following equation:

$$DLICENSE_{jq} = \beta_0 + \beta_1 \cdot W_{jq} \cdot (1 - DESPERATE_{jq}) + \beta_2 \cdot W_{jq} \cdot DESPERATE_{jq} + \Theta X_{jq} + \epsilon_{jq},$$

where W is the P3F indicator and X contains all controls used in previous linear probability specifications (including *RTER_OWN* and fixed effects). We obtain $\hat{\beta}_1 = 0.017$ ($p = 0.28$) and $\hat{\beta}_2 = 0.073$ ($p < 0.01$). These estimates indicate that the P3F-fueled licensing surge primarily resides on "desperate" units.[31]

To summarize, we observe three signs that are consistent with lack of premeditation: (i) licensing does not have a perceptible stock market return, (ii) the P3F-fueled licensing surge does not seem to follow from pre-established policies linking pipeline events to licensing, and (iii) the surge is concentrated on instances where the amount stakeholder pressure introduced by P3Fs would be presumably larger.

5.2. Performance

Results above suggest that P3Fs increase near-future licensing activity in a way that is consistent with rushed behavior. Pre-licensing P3Fs can, therefore, be taken as a marker for a rushed licensing process. We thus test whether pre-licensing P3Fs have a detrimental impact on post-licensing development performance. Results are presented in panel A of Table 4. The means of *ADVANCE* are 0.42 among control therapies and 0.38 among treated ones. This associates pre-licensing P3Fs to a 0.05 higher probability of termination (column (1)). The estimate of column (2) corresponds to the coefficient for the treatment indicator in a linear probability model. As controls, this model includes contextual factors measured prior to milestone completion as well as fixed effects for area, year, and development and licensing stages. That is, this model controls for the context at milestone completion but not for that at licensing. The −0.036 estimate aligns

Table 4. Main Results for the Performance Analysis

	(1)	(2)	(3)
	Difference of means	Linear probability	Propensity score matching
Panel A: Outcome *ADVANCE*			
Estimate	−0.044	−0.036*	−0.092**
Standard error	(0.031)	(0.020)	(0.047)
Observations	1,783	1,783	1,092
Panel B: Falsified treatment			
Estimate	−0.039	−0.030	−0.049
Standard error	(0.031)	(0.028)	(0.040)
Observations	1,783	1,783	1,148
Panel C: P3S treatment			
Estimate	−0.003	−0.001	−0.046
Standard error	(0.030)	(0.032)	(0.067)
Observations	1,783	1,783	996

Notes. Column (1) shows the average outcome difference between treated and control observations. Column (2) shows the coefficient estimated for the treatment indicator in a linear probability specification (specification described in the text; robust standard errors reported). Column (3) shows the ATE estimate obtained from propensity score matching (Abadie–Imbens heteroskedasticity-consistent robust standard errors). The smaller number of observations in this column follows from the selected sample (common propensity score support, matched observations, $R \geq 0.1$).

*$p < 0.1$; **$p < 0.05$.

with the difference-of-means result. Column (3) presents the propensity score matching ATE. The estimated effect of almost −0.1 is considerably larger than the two previous estimates. It implies that prelicensing P3Fs lower post-licensing advancement rates by about 16%. Thus, therapies linked to a rushed licensing process underperform in post-licensing development relative to their counterparts licensed under normal conditions.

In panel B of Table 4, we present results obtained when the treatment is falsified through randomization, and in panel C those obtained when the P3Fs are replaced with P3Ss. As expected, these estimates are small and lacking in statistical significance. Results are also robust to the issues associated with nonstandard P3F shocks addressed before (see Online Appendix D for the analysis and results). We now consider whether the underperformance of therapies licensed in a rush may be driven by potential "capture and kill" behavior. Under this rationale, licensing following a P3F event would not aim at filling a pipeline gap but instead aim at hindering future competition—following a P3F, therapies would first be "captured" (licensed) and then "killed" (terminated). Our test relies on the idea that firms engaging in this type of behavior should have no incentives to continue to fund costly clinical trials for therapies that they plan to kill. We should thus observe that therapies licensed after a P3F event encounter termination sooner than those licensed under normal conditions. This difference is not present in our data however. (Testing details and results are presented in Online Appendix F.) We thus conclude that "capture and kill" behavior is unlikely to drive the underperformance of therapies licensed in a rush.

Lastly, we note that our inference could be vulnerable to possible anticipatory effects. These would occur if the teams in charge of running P3 trials received information signaling adverse results prior to the unfolding of P3F events.[32] If licensing activity was impacted by such information, the implied timelines would be relaxed. This possibility suggests that our estimates may include an attenuation bias. That is, the impacts of P3Fs that are fully nonanticipated should be larger than those described here.

5.3. Treatment Effect Heterogeneity

5.3.1. Licensing. Causal forest CATE estimates for the licensing analysis ($\hat{\tau}^L$) are presented in Figure 6(a). These represent the estimated impact of a P3F on *DLICENSE*, evaluated at each observation's vector of contextual factor values. Consistent with the results of Table 3, about three-quarters of estimated values are positive. However, a good share of estimates is estimated imprecisely and does not meet statistical significance standards. Nevertheless, all statistically significant estimates lie in the positive domain. Overall, these results indicate that units' inclinations to license in a rush vary significantly and are not strong enough in every context.

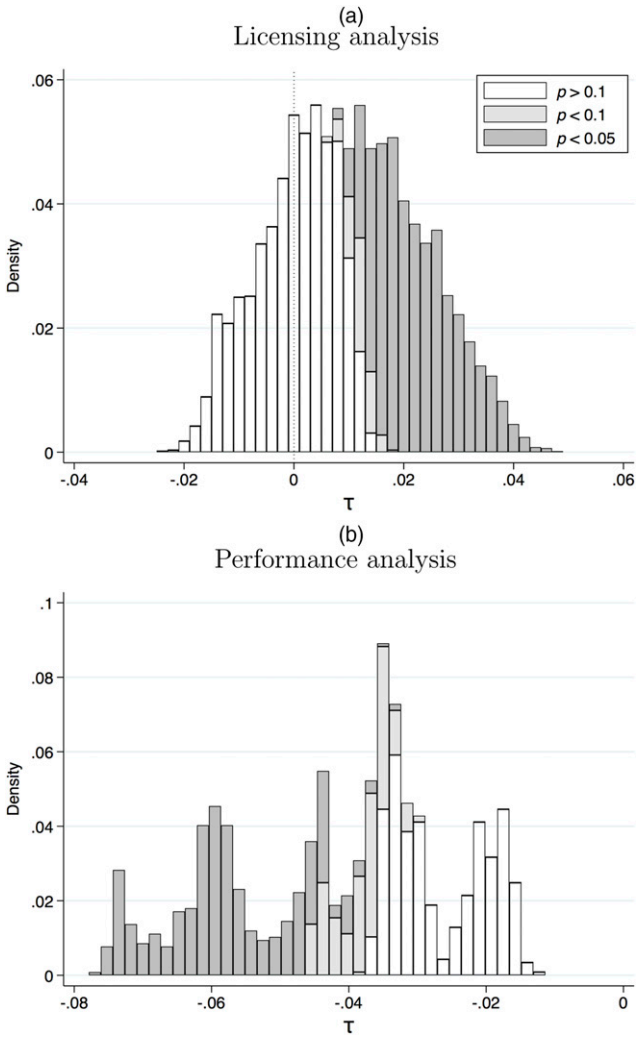
5.3.2. Performance. Recall from Equation (1) that CATE estimates are computed from deviations of random forest predictions for the outcome and treatment assignment (m and π, respectively). Compared with our previous performance analyses, this framework allows us to simultaneously account for sources of contextual variability at licensing (through π) and milestone completion (through m). Figure 6(b) presents the distribution of resulting performance CATE estimates ($\hat{\tau}^P$). Consistent with the underperformance result, all estimates are negative, although statistical significance is mixed. Estimates also point to a significant amount of treatment effect heterogeneity.

In Online Appendix G, we characterize the importance of the different contextual factors as drivers of treatment effect heterogeneity (licensing and performance). Among others, we find that the magnitude of performance treatment effects decreases at more advanced stages. That is, units are less inclined to terminate therapies licensed in a rush when they are (statistically) closer to the market. This result provides a measure of external validity, as it coincides with the idea that the value of continued development is larger when the market is closer.

5.3.3. Is There a Trade-off in Practice?. Motivated by the wide heterogeneity of CATE estimates, we ask whether the inclination to license in a rush tends to coincide with that to terminate a therapy that was licensed in a rush. To better grasp our question, note that it could be possible that the contexts that make a unit more inclined to engage in one behavior make it less inclined to engage in the other (i.e., positively correlated $\hat{\tau}^L$ and $\hat{\tau}^P$ values). This scenario would undermine the relevance of the trade-off implied by rushed behavior—units at higher risk of engaging in rushed licensing would usually be at lower risk of suffering its consequences. Our results support the opposite view, namely, that the trade-off is likely experienced in practice.

The key step to implement this analysis is in merging $\hat{\tau}^L$ and $\hat{\tau}^P$ estimates. The problem exists because the former is available at the unit/quarter level, whereas the latter is available at the completed milestone level. Thus, the merging step requires some form of aggregation. We first note that $\hat{\tau}^L$ values exhibit a high degree of temporal persistence within units. This is evidenced by the first-order autoregressive specification $\hat{\tau}^L_{jq} = \alpha + \rho\hat{\tau}^L_{jq-1} + \epsilon_{jq}$, where j indexes units and

Figure 6. Heterogeneous P3F Impacts: Causal Forest CATE Estimates



q quarters, and ϵ is an error. The large temporal persistence is illustrated by the high value of $\hat{\rho}$, 0.84 ($p < 0.01$). This result suggests that it is reasonable to aggregate $\hat{\tau}^L$ estimates at the unit level. We do so through the unit-specific fixed effects α_j obtained from $\hat{\tau}^L_{jq} = \alpha_j + \rho\hat{\tau}^L_{jq-1} + \lambda_{\text{year}(q)} + \epsilon_{jq}$, where λ corresponds to year fixed effects. Figure 7 presents the scatterplot of (unit-normalized) $\hat{\alpha}$ and $\hat{\tau}^P$ estimates. The key result is illustrated by the negative slope (best linear fit) of the dashed line of Figure 7: stronger inclinations to engage in rushed licensing tend to be paired with stronger inclinations to terminate therapies licensed in a rush. The overall correlation between the two variables nears −0.4 ($p < 0.01$). Observations shown by diamonds in Figure 7 reflect $\hat{\tau}^L$ estimates with p-values below usual significance thresholds. The correlation between these and $\hat{\alpha}$ estimates is more pronounced than overall.

6. Why Do Therapies Licensed in a Rush Underperform?

Our goal here is to shed some light on the possible mechanisms behind the underperformance result of the previous section. Based on our review in Section 2.2, we divide the licensing process into three main sets of activities—search, contracting, and due diligence—and argue that post-licensing performance could be impacted through each. We proceed by outlining these impacts, implementing tests, and casting an interpretation.

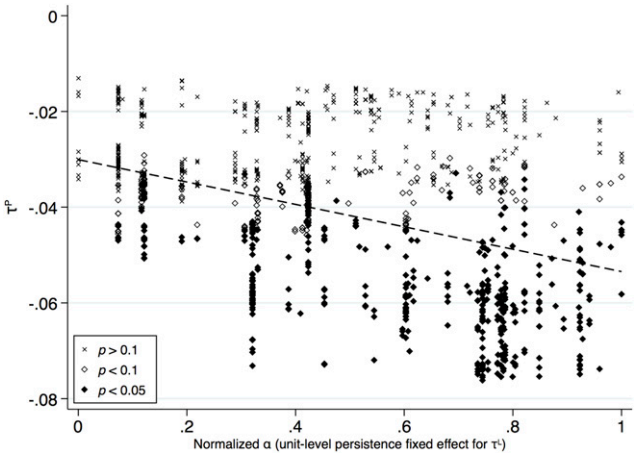
6.1. Potential Mechanisms

6.1.1. Search. As noted in Section 2.2, large pharmaceutical firms deploy comprehensive search efforts oriented at identifying the "right" candidates for licensing (i.e., therapies that match the firm's technical capabilities and strategic needs). It is easy to see how rush could hinder these efforts: firms may not have enough time to fully screen the landscape of opportunities, evaluate all possible leads, etc. An impact on post-licensing performance could then be expected based on the insights of previous research, indicating that the lack of proper firm/technology fit can jeopardize the developmental success of new technologies.[33]

6.1.2. Contracting. Because the full set of activities required to support a therapy's development are often ill-defined at the onset of collaborations, licensing contracts cannot rely on specific deliverables.[34] This aspect introduces significant difficulties into the contracting process, because firms must resolve the inherent contract incompleteness through a collection of (monetary and nonmonetary) incentives and the allocation of (control and residual property) rights (Lerner and Merges 1998, Mason et al. 2008, Lerner and Malmendier 2010). We posit that, if the contracting process is carried out in a rush, the set of terms that parties agree to may contain "loose ends" that increase the probability of organizational friction. This friction may, in turn, increase the likelihood of project termination. The potential relevance of this mechanism has been insinuated, for example, by Jones (2007), who states 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 (p. 716)." Rhodes et al. (2003) further indicate that common pitfalls associated with failed collaborations include poorly defined responsibilities and inadequate structures to manage and resolve conflicts.

6.1.3. Due Diligence. As mentioned earlier, due diligence activities are primarily aimed at unearthing

Figure 7. Correlating CATE Estimates obtained from Licensing and Performance Analyses



Notes. Horizontal values represent each decision-making unit's overall inclination to engage in rushed licensing. They are computed as unit-level fixed effects in a first-order autoregressive model for licensing CATE estimates.

weaknesses in two key determinants of post-licensing performance: the underlying science and intellectual property protection of licensed therapies. Rush could adversely impact the quality of due diligence if teams are not given enough time to do their research. The importance of due diligence for post-licensing success has been widely stressed by observers (Rhodes et al. 2003, Mason et al. 2008) and academic research (Palermo et al. 2019) and brought to the fore by the case of Erbitux. This was a prominent experimental anticancer antibody licensed by Bristol-Meyers Squibb (BMS) from ImClone. Following the Food and Drug Administration's 2001 decision to reject the therapy for commercialization, BMS investors sued the company on the basis that it had performed inadequate due diligence (Prudhome 2013). Incidentally, BMS reportedly licensed Erbitux at a time that its cancer pipeline was "running dry" (Prudhome 2013).

6.2. Tests

In his defense to the above-mentioned lawsuit, the head of BMS oncology stated that his team had reviewed "every piece of information: lab data, X-ray scans, the toxicity-every single thing" (Prudhome 2013). This quote illustrates the difficulty of directly testing for the due diligence channel. That is, by definition, the quality of due diligence activities reflects on variables that are unobservable to the analyst (laboratory data, x-rays, etc.). Our testing is thus circumscribed to the search and contracting channels.

6.2.1. Search Channel. If this mechanism was at play, therapies licensed in a rush should less often represent the "right" candidate for the licensing firm (compared to those licensed under normal conditions). To implement our test, we thus construct a technical matching quality indicator (*MATCH*), leveraging two fine-grained technological categorizations available from the Cortellis data. In the first of these, categories correspond to broad technologies used in the formulation and delivery of drugs. These include, for example, small molecules (i.e., chemically synthesized), large molecules (i.e., living organisms), topical or intravenous delivery, etc. There is a total of almost 330 unique such technologies in the broad sample, and therapies are often associated with more than one. For each licensed therapy, *MATCH* equals one if its entire set of associated technologies associated had been previously used by the licensing unit and zero otherwise. Thus, *MATCH* = 1 indicates that, at the time that the deal was struck, the licensing unit had expertise developing the specific set of technologies deployed by the licensed therapy. A differences-of-means estimate suggests that rushed licensing does not entail a matching quality penalty: *MATCH* averages 0.61 among therapies licensed in a rush and 0.60 among the rest ($p = 0.96$). We obtained the same qualitative result when considering the second and even finer technological categorization, which is based on "target-based actions" (i.e., precise descriptions of the way that therapies produce a pharmacological effect in the human body).[35]

6.2.2. Contracting Channel. We analyze the incidence of a specific type of development termination that arguably represents a signal of organizational friction and can, therefore, be taken as a proxy for the relevance of this channel. Recall from Section 3.3 that most terminations in our data (86%) correspond to cases in which a therapy's development is declared terminated before the associated licensing agreement is. This scenario aligns with standard industry views for what causes of attrition—terminations primarily stem from the nature of testing results, which span beyond organizational remits. We therefore refer to these as "standard" terminations. For a small fraction of terminations (9%), we instead observe that the licensing agreement is dissolved prior to the termination of development.[36] Because the organizational element takes precedence in this case, we assert that these "agreement" terminations can indicate organizational friction and as such, signal rushed contracting. A comparison of the frequencies of agreement terminations supports this view: they correspond to 12% of terminations of therapies licensed in a rush but only 6% of terminations of therapies licensed under normal conditions ($p < 0.05$).

The low overall frequency of agreement terminations challenges the generalizability of this result. To address this weakness, we perform an additional

analysis utilizing performance CATE estimates ($\hat{\tau}^P$). Recall that these estimates reflect units' inclinations to forsake the development of therapies licensed in a rush. Also recall that they are obtained without utilizing termination-type information and that their heterogeneity is pegged to contextual factors measured at milestone completion. We evaluate whether $\hat{\tau}^P$ estimates predict termination type. Finding that smaller (i.e., more negative) $\hat{\tau}^P$ values are associated with higher probabilities of agreement termination (relative to standard termination) would support the generalizability of the above result. In particular, this would suggest that the contexts that give rise to stronger inclinations to terminate therapies licensed in a rush also warrant higher likelihoods of agreement termination. Considering termination outcomes only, we regress an agreement termination indicator on $\hat{\tau}^P$ values. Contextual factors are added as controls along with year, area, development, and licensing stage fixed effects. Results support generalizability: a one-standard deviation smaller (i.e., more negative) $\hat{\tau}^P$ value is associated with a 0.17 larger probability of agreement termination ($p < 0.05$).

6.3. Interpretation and Caveats

Testing results for the search channel suggest that, based on observable characteristics, rush does not hinder firms' ability to license a technologically fitting therapy. An additional analysis reinforces this conclusion: there is no evidence of a rush-fueled difference in terms of maturities (stages at licensing).[37] These results would be difficult to rationalize if therapies licensed in a rush were identified through "fresh" search (i.e., search initiated following the triggering P3F). Instead, these findings may reflect that therapies licensed in a rush are generally drawn from a pool of candidates identified prior to the triggering P3F event. That is, therapies licensed in a rush may correspond to old leads that are "recycled," or to ongoing licensing processes that are accelerated. We favor this interpretation on different grounds. Foremost, the timing of the licensing surge makes it implausible for a fresh search process to take place—search would have to be completed within months. This seems unlikely given characteristics of the process (see Section 2.2) and the notorious lack of thickness of technology licensing markets (Gans and Stern 2010). In addition, as noted previously, large pharmaceutical firms engage in continuous scouting. This suggests that, at the time that a P3F shock hits, firms likely possess a number of advanced leads to pursue further. As such, our suspicions are turned away from the search mechanism and to the possible acceleration of the remaining set of activities, due diligence, and contracting.

Given the importance of contracting aspects for development performance, our results in this regard are conspicuous. We interpret them as a strong signal that the underperformance result could hinge on the quality of the agreements that firms converge to during negotiation. However, being unable to directly test for the due diligence mechanism, we cannot rule out that part of the underperformance effect may stem from a heightened vulnerability to "unobservable" (scientific/intellectual property) weaknesses. Despite this limitation, our results seem to converge on the idea that the underperformance result is rooted on activities taking place in the "last mile" of the licensing process. As such, firms may have it in their control to alleviate the potential consequences of licensing under pressure, for example, through the strengthening of protocols aimed at minimizing contracting "lose ends" or by revamping the amount personnel and resources devoted to due diligence.

We conclude with a couple of caveats. First, if biotechnology firms on the supply were aware of the post-licensing underperformance of therapies licensed in a rush, they may be reluctant to transact under these conditions. This could imply a form of selection on unobservables that would further obfuscate our interpretation. In this case, the scope of actions aimed at limiting the detrimental consequences of rushed licensing would be reduced. Second, if research teams are able to anticipate P3Fs (as noted in Section 5.2), the implied timeframes would be relaxed, and fresh search could play a role. Based on these caveats and the high-level formulation of our tests, we treat this evidence as primarily suggestive. A more powerful analysis would rely on records tracking the kinds and amount of activities performed in preparation for a licensing deal.

7. Conclusions and Limitations

Our findings suggest that large pharmaceutical firms impacted by large negative shocks to their pipelines (P3Fs) may use drug licensing to "fill" the resulting "pipeline gaps." The effect's timing is what makes it significant: it unfolds over a window that is quite short given the usual licensing timelines. We thus interpret this finding as a reflection of rushed innovation behavior. Consistently, we find that drugs licensed shortly after a P3F event underperform in subsequent development: they are significantly less likely to reach the market compared with others licensed under normal conditions. The significance of this result, in turn, stems from the importance of development attrition: the bulk of pharmaceutical R&D budgets is spent on experimental therapies that never reach the market. For this reason, the development underperformance of therapies licensed in a rush stands out as a heavy penalty for any benefits that could be derived from agile decision-making.

Because P3 failures are not rare events in the industry, rushed decisions may be more than anecdotal. Other negative shocks such as Food and Drug Administration Public Health Advisories (Krieger et al. 2018) and product recalls (Ball et al. 2018) could induce the same type of behavior and consequences.

There is, however, no simple way to conclusively determine whether rushed licensing is suboptimal from the point of view of shareholders. The primary difficulty comes from the complexity of the typical financial compensation structure. Licensing contracts rely heavily on contingent payments, which means that development terminations also imply a reduction in the stream of payments made by the in-licensing firm. Moreover, counterfactual cost calculations would require the researcher to a take stand on the specific mechanics of contract negotiations (and whether rush affects them). This exercise would be further challenged by the fact that it is difficult to compile systematic contract design and clinical trial cost data.

The primary limitations of this study stem directly from data availability. As we noted earlier, data tracking the set of preparatory and execution activities for the licensing process would be of great help to better ascertain the mechanisms underlying the underperformance result. Because we lack this type of data, we have only been able to provide suggestive mechanistic evidence. A second set of limitations stems from structural complexity. For example, licensing deals vary in terms of exclusivity and geographical scope. Here, we have opted for a clean, although inefficient, solution, which is to circumscribe the analysis to a particular type of deals (i.e., exclusive worldwide rights). Another challenging feature of drug licensing corresponds to the bundling of multiple therapies within a single licensing deal. Again in this case, we have opted for tractability, treating each licensed therapy as a separate licensing event. We hope that future research is able to overcome these limitations.

Acknowledgments

The author thanks Laura Kornish, Joshua Krieger, Ariel Dora Stern, Matt Higgins, Mahdu Wiswanathan, Lex Zhao, Tinglong Dai, Carlos Noton, editors, and anonymous referees as well as seminar participants at Johns Hopkins and Cambridge Universities, the 2018 Marketing Science conference, the 2019 AEA meetings, the 2019 UT Dallas FORMS conference, the 2019 DC Management Research Consortium, and the 2019 USPTO/Northwestern Searle Conference on Innovation Economics for their helpful comments and suggestions. The author acknowledges the research assistantships of Kun Bao and Saanika Pillai. All errors and omissions are the author's own.

Endnotes

[1] Chandy et al. (2006) illustrate this tradeoff by examining firm-level pharmaceutical conversion rates (patenting to launch). Like these authors, we are unable to investigate impacts on revenues owing to the small number of products that reach the market.

[2] Based on the comprehensive estimates of Wong et al. (2019) (who use a large sample that covers the same period as ours), the probability of reaching P3 conditional on entering P1 is about 0.3.

[3] The problem's magnitude is reflected by the statistic that about 80% of U.S. trials fail to meet enrollment timelines (see https://www.drugdevelopment-technology.com/features/featureclinical-trial-patient-recruitment/).

[4] These figures are expressed in the industry's "biobucks" yardstick (i.e., contingent payments are accounted for in nominal value).

[5] The evidence of Danzon et al. (2007) suggests that trade sales are used as an exit strategy after encountering financing problems.

[6] Information is obtained from company records, conferences, and other public sources; curated and updated daily by over 500 expert analysts. As of 2016 Q3, Cortellis included over 65,000 drug development histories and 48,000 deal reports, both of which date back several decades. Data were last accessed in mid-2018.

[7] 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. As of October 2018, all but two of these firms (Gilead and Roche) were members of the Association of Pharmaceutical Research and Manufacturers of America, which is known as the trade association for big pharma.

[8] We focus on Cortellis' "Drug-Development/Commercialization" licensing deals, described as "Partner [in-licensing] firm acquires a license from Principal [out-licensing] firm to develop and commercialize (sell) drug(s)" (from Cortellis documentation). Other types of licensing deals in the broad sample reflect alternative business models.

[9] This approach is adopted for tractability. Some support is offered by the finding that bundles of licensed therapies are elastic to the in-licensing firm's goals (Hermosilla and Wu 2018).

[10] By this metric, the analysis sample represents significant and stable percentage (about 15%) of the broad Cortellis sample.

[11] 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.

[12] The complementary 5% corresponds to a third category that we call "idle state" termination. This arises when abnormally long stage development times coincide with unreported outcomes.

[13] We attribute this to the large participation of cancer therapies in our sample, which are typically associated with the lowest success rates (Wong et al. 2019). Success rates in our sample were compared with those of DiMasi et al. (2003, 2010, 2016), Abrantes-Metz et al. (2004), Kola and Landis (2004), Arrowsmith (2011), Pammolli et al. (2011), Hay et al. (2014), Waring et al. (2015), Smietana et al. (2016), and Wong et al. (2019).

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

[15] The Herfindahl-Hirschman Index of concentration is about 15/100.

[16] A molecule's therapies may fail at different times if therapies are tested through independent P3 trials. This often occurs if there are large-enough differences among targeted conditions, populations, or dosages required to test each therapy. If the molecule's therapies have similar-enough properties, they may be tested through a unique trial (as for Pfizer's Torcetrapib).

[17] Causal forests were first proposed by Wager and Athey (2018). The generalized random forest implementation that we use was proposed

by the same authors in a later paper (Athey et al. 2019). We prefer this approach given its demonstrated superior performance recovering "true" CATEs (see Athey et al. 2019).

[18] These probability weights also vary across time. They are estimated using tens of thousands of development histories available from the broad Cortellis sample without including those in the analysis sample. Online Appendix B outlines the procedure and provides descriptives.

[19] $\hat{A}_{-j}$ is computed from industry-wide data (all units in the broad Cortellis sample).

[20] Licensing waves could arise when basic science breakthroughs infuse the supply of licensable candidates with new types of candidates. Hermosilla and Lemus (2019) illustrate such a case following the 2003 unveiling of the Human Genome.

[21] The empirical relevance of this effect is nevertheless downplayed by the findings of Higgins and Rodriguez (2006). They find that, whereas pipeline strength systematically predicts a pharmaceutical firm acquiring other to revitalize its position, portfolio strength is "consistently not significant" (Higgins and Rodriguez 2006, p. 370).

[22] To construct this variable, one would ideally leverage time series sales and market exclusivity data for each marketed therapy. This approach is unfeasible in our case: the portfolio of marketed therapies is empty for many units in our sample (between 40% and 50% of observations depending on the analysis—for these, there is no product lifecycle information). For units with nonempty portfolios, the omission of this information will only be problematic if there are systematic treated/control differences in the strength and "youth" of portfolios. We are unable to directly test this. Because we are including a host of related factors (including *MKTSHARE* and firm-level sales), we nevertheless presume that, if present, these imbalances will be small. Moreover, the results of Higgins and Rodriguez (2006) (see footnote 21) suggest that, even if present, these imbalances are unlikely to introduce bias.

[23] This number results after dropping about 286 "contaminated" control observations. Contamination occurs when a P3F shock is observed within $\{q-3,..,q\}$ or $\{q+1,..,q+4\}$.

[24] Note that propensity scores here reflect the probability that a licensing event has been preceded by a P3F.

[25] Recall that in this dataset there are 418 P3Fs and almost 6,500 observations. "Contaminated" observations (see footnote 22) are not problematic in this case so they are kept.

[26] According to the estimate, a 0.1 larger risk score yields a 0.044 higher P3F probability.

[27] Balancing statistics suggest that the matching procedure does a good job reducing treated/control covariate imbalances in the raw sample, and that resulting differences are acceptable. These statistics are presented in Online Appendix C.

[28] This result do not qualitatively change when we replace the indicator for a count of competitive P3Fs.

[29] The differences-of-means estimate of column (1) of Table 3 remains about constant. This result could follow from the confounding effects of contextual differences (in particular, by the fact that units with stronger pipelines license more often).

[30] P3Ss are identified through the definitive marker of successful P3 completion—the submission of the new drug application package (i.e., initiation of review stage). P3S risk scores were constructed using analogous procedures to those used for the P3F score. We note that, because a therapy's P3S requires the completion of scheduled trials (without having experienced a P3F), P3Ss should be less surprising that P3Fs and therefore entail shocks of smaller magnitude. However, the rate of P3 success for cancer therapies is well below 0.5 (0.36 according to the best available estimate of Wong et al. 2019). This means that, for cancer therapies, successful events should be intrinsically more surprising than negative ones. Because cancer therapies dominate our sample, P3Ss may carry relatively more importance in our data set than in the population. These caveats suggest that P3Fs and P3Ss should not be viewed as strictly symmetric shocks. We move forward assuming rough symmetry.

[31] For robustness, we also estimate a specification in which the terms $\beta_1 \cdot W_{jq} \cdot (1 - DESPERATE_{jq}) + \beta_2 \cdot W_{jq} \cdot DESPERATE_{jq}$ are replaced for $\beta_1 \cdot W_{jq} + \beta_2 \cdot W_{jq} \cdot NORM\_RTER\_OWN_{jq}$, where *NORM_RTER_OWN* is a standard normalization of *RTER_OWN*. We obtain $\hat{\beta}_1 = 0.028$ ($p = 0.048$) and $\hat{\beta}_2 = 0.040$ ($p < 0.003$). These estimates suggest that one standard deviation of *RTER_OWN* increases the impact of P3F on the near-future licensing probability by 0.04.

[32] These signals could, for example, be based on the dropout rates of patients enrolled in the trial (Chan and Hamilton 2006).

[33] For example, in a seminal article, Teece (1986) argues that technological innovators may fail to profit from their innovations if these are not paired with the "right" (broadly defined) complementary assets. Gans and Stern (2010) find consistent evidence in the context of commercialization strategies for startup innovation. Mowery et al. (1998) and Diestre and Rajagopalan (2012) illustrate the importance of matching on a narrower set of capabilities (technical expertise) by documenting equilibrium assortative matching between (buying) firms' technical capabilities and the type of technologies that they license.

[34] Aghion and Tirole (1994) highlight this as a general observation for R&D collaborations.

[35] Formally, target-based actions (TBAs) correspond to mechanism of action/targeted cell pairs. A compound's mechanism of action refers to the way that the compound produces a pharmacological effect on the body. Some molecules may, for example, act by stimulating specialized cells (like adrenaline); others may act 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. There are about 7,500 TBA levels in the overall data. Therapies are associated with more than one only occasionally. Defining *MATCH* based on these data, we find that it averages 0.26 among therapies licensed in a rush and 0.22 among the rest. The difference is not statistically significant ($p = 0.24$).

[36] Depending on the nature of the original agreement, this scenario can be followed by either or both firms developing the therapy individually or with other partners, or by the therapy's definitive scrapping.

[37] We tested for a difference in the distribution of stages at licensing using a chi-squared test. The p-value for the null hypothesis of no difference is 0.242.

References

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

Abrantes-Metz R, Adams C, Metz A (2004) Pharmaceutical development phases: A duration analysis. Federal Trade Commission, Bureau of Economics Working Paper No. 274, Federal Trade Commission, Washington, DC.

Acemoglu D, Linn J (2004) Market size in innovation: Theory and evidence from the pharmaceutical industry. *Quart. J. Econom.* 119(3):1049–1090.

Aghion P, Tirole J (1994) The management of innovation. *Quart. J. Econom.* 109(4):1185–1209.

Alcacer J, Cantwell J, Gittelman M (2009) Are licensing markets local? An analysis of the geography of vertical licensing agreements in bio-pharmaceuticals. Working paper, Harvard Business School, Boston.

Arora A, Gambardella A (2010) The market for technology. Hall BH, Rosenberg, eds. *Handbook of the Economics of Innovation*, vol. 1 (Elsevier, Amsterdam), 641–678.

Arora A, Fosfuri A, Gambardella A (2004) *Markets for Technology: The Economics of Innovation and Corporate Strategy* (MIT Press, Cambridge, MA).

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

Arrowsmith J (2011) Trial watch: Phase III and submission failures: 2007-2010. *Nature Rev. Drug Discovery* 10(2):87.

Athey S, Tibshirani J, Wager S (2019) Generalized random forests. *Ann. Statist.* 47(2):1148–1178.

Ball GP, Macher JT, Stern AD (2018) Recalls, innovation, and competitor response: Evidence from medical device firms. Harvard Business School Working Paper No. 19-028, Harvard Business School, Boston.

Bayus BL (1997) Speed-to-market and new product performance trade-offs. *J. Product Innovation Management* 14(6):485–497.

Breiman L (2001) Random forests. *Machine Learn.* 45(1):5–32.

Chan T, Nickerson JA, Owan H (2007) Strategic management of R&D pipelines with cospecialized investments and technology markets. *Management Sci.* 53(4):667–682.

Chan TY, Hamilton BH (2006) Learning, private information, and the economic evaluation of randomized experiments. *J. Political Econom.* 114(6):997–1040.

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

Chaney PK, Devinney TM, Winer RS (1991) The impact of new product introductions on the market value of firms. *J. Bus.* 64(4):573–610.

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

Cockburn IM, Henderson RM (2001) Scale and scope in drug development: Unpacking the advantages of size in pharmaceutical research. *J. Health Econom.* 20(6):1033–1057.

Cohen J (1988) *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. (Lawrence Erlbaum Associates, Hillsdale, NJ).

Cohen MA, Eliasberg J, Ho T-H (1996) New product development: The performance and time-to-market tradeoff. *Management Sci.* 42(2):173–186.

Cox D (1972) Regression models and life-tables. *J. Royal Statist. Soc. Series B: Methodological* 34(2):187–220.

Danzon PM, Epstein A, Nicholson S (2007) Mergers and acquisitions in the pharmaceutical and biotech industries. *Managerial Decision Econom.* 28(4–5):307–328.

Danzon PM, Nicholson S, Pereira NS (2005) Productivity in pharmaceutical–biotechnology R&D: The role of experience and alliances. *J. Health Econom.* 24(2):317–339.

Davies R (2013) The relevance and importance of business development and licensing in the biopharmaceutical industry. *J. Commercial Biotechnology* 19(3), https://doi.org/10.5912/jcb592.

Diestre L, Rajagopalan N (2012) Are all 'sharks' dangerous? New biotechnology ventures and partner selection in R&D alliances. *Strategic Management J.* 33(10):1115–1134.

DiMasi JA, Grabowski HG, Hansen RW (2016) Innovation in the pharmaceutical industry: New estimates of R&D costs. *J. Health Econom.* 47(2016):20–33.

DiMasi JA, Hansen RW, Grabowski HG (2003) The price of innovation: New estimates of drug development costs. *J. Health Econom.* 22(2):151–185.

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

Dranove D, Meltzer D (1994) Do important drugs reach the market sooner? *RAND J. Econom.* 25(3):402–423.

Dranove D, Garthwaite C, Hermosilla M (2014) Pharmaceutical profits and the social value of innovation. NBER Working Paper No. 20212, National Bureau of Economic Research, Cambridge, MA.

Dubois P, De Mouzon O, Scott-Morton F, Seabright P (2015) Market size and pharmaceutical innovation. *RAND J. Econom.* 46(4):844–871.

Fine JP, Gray RJ (1999) A proportional hazards model for the subdistribution of a competing risk. *J. Amer. Statist. Assoc.* 94(446): 496–509.

Freidlin B, Korn EL, George SL (1999) Data monitoring committees and interim monitoring guidelines. *Controlled Clinical Trials* 20(5):395–407.

Gans JS, Stern S (2010) Is there a market for ideas? *Indust. Corporate Change* 19(3):805–837.

Gans JS, Hsu DH, Stern S (2002) When does start-up innovation spur the gale of creative destruction? *RAND J. Econom.* 33(4):571–587.

Giovannetti G, Spence P (2017) Beyond borders: The global biotechnology report. Accessed November 19, 2019, https://www.ey.com/Publication/vwLUAssets/ey-biotechnology-report-2017-beyond-borders-staying-the-course/$FILE/ey-biotechnology-report-2017-beyond-borders-staying-the-course.pdf.

Girotra K, Terwiesch C, Ulrich KT (2007) Valuing R&D projects in a portfolio: Evidence from the pharmaceutical industry. *Management Sci.* 53(9):1452–1466.

Goodman M (2009) Market watch: Sales of biologics to show robust growth through to 2013. *Nat. Rev. Drug Discovery* 8(11):837.

Guedj I, Scharfstein D (2004) Organizational scope and investment: Evidence from the drug development strategies of biopharmaceutical firms. NBER Working Paper No. 10933, National Bureau of Economic Research, Cambridge, MA.

Hay M, Thomas DW, Craighead JL, Economides C, Rosenthal J (2014) Clinical development success rates for investigational drugs. *Nature Biotechnology* 32(1):40–51.

Henderson R, Cockburn I (1996) Scale, scope, and spillovers: The determinants of research productivity in drug discovery. *RAND J. Econom.* 27(1):32–59.

Hendricks KB, Singhal VR (1997) Delays in new product introductions and the market value of the firm: The consequences of being late to the market. *Management Sci.* 43(4):422–436.

Hermosilla M (2016) Imperfect outsourcing of technological innovations. Working paper, Johns Hopkins Carey Business School, Baltimore.

Hermosilla M, Lemus J (2019) Therapeutic translation of genomic science: Opportunities and limitations of GWAS. Berndt ER, Goldman DP, Rowe JW, eds. *Economic Dimensions of Personalized and Precision Medicine* (University of Chicago Press, Chicago), 21–52.

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

Higgins MJ, Rodriguez D (2006) The outsourcing of R&D through acquisitions in the pharmaceutical industry. *J. Financial Econom.* 80(2):351–383.

Hwang TJ, Carpenter D, Lauffenburger JC, Wang B, Franklin JM, Kesselheim AS (2016) Failure of investigational drugs in late-stage clinical development and publication of trial results. *JAMA Internal Medicine* 176(12):1826–1833.

Jones A (2007) Minimizing leakage of value from R&D alliances. *Nature Rev. Drug Discovery* 6(9):711–719.

Kimes M (2008) The case of Pfizer. *Fortune Magazine* (December 1).

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

Kola I, Landis J (2004) Can the pharmaceutical industry reduce attrition rates? *Nature Rev. Drug Discovery* 3(8):711.

Krieger J, Li X, Thakor RT (2018) Find and replace: R&D investment following the erosion of existing products. Preprint, submitted September 17, http://dx.doi.org/10.2139/ssrn.3240344.

Lerner J, Malmendier U (2010) Contractibility and the design of research agreements. *Amer. Econom. Rev.* 100(1):214–246.

Lerner J, Merges RP (1998) The control of technology alliances: An empirical analysis of the biotechnology industry. *J. Indust. Econom.* 46(2):125–156.

Mason R, Savva N, Scholtes S (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* (August 12), 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. *Tech. Anal. Strategic Management* 19(3):279–301.

Moorman C, Miner AS (1998) The convergence of planning and execution: Improvisation in new product development. *J. Marketing* 62(3):1–20.

Mowery DC, Oxley JE, Silverman BS (1998) Technological overlap and interfirm cooperation: Implications for the resource-based view of the firm. *Res. Policy* 27(5):507–523.

Palermo V, Higgins MJ, Ceccagnoli M (2019) How reliable is the market for technology? *Rev. Econom. Statist.* 101(1):107–120.

Pammolli F, Magazzini L, Riccaboni M (2011) The productivity crisis in pharmaceutical R&D. *Nature Rev. Drug Discovery* 10(6):428.

Paul SM, Mytelka DS, Dunwiddie CT, Persinger CC, Munos BH, Lindborg SR, Schacht AL (2010) How to improve R&D productivity: The pharmaceutical industry's grand challenge. *Nature Rev. Drug Discovery* 9(3):203–214.

Pfizer Inc. (2006) In interests of patient safety, Pfizer stops all torcetrapib clinical trials. Press release, Pfizer, New York. Accessed November 19, 2019, https://www.businesswire.com/news/home/20070226005663/en/Interests-Patient-Safety-Pfizer-Stops-Torcetrapib-Clinical.

Pfizer Inc. (2007) Bristol-Myers Squibb and Pfizer announce worldwide collaboration to develop and commercialize anticoagulant and metabolic compounds. Press release, Pfizer, New York. Accessed November 19, 2019, https://www.pfizer.com/news/press-release/press-release-detail/bristol_myers_squibb_and_pfizer_announce_worldwide_collaboration_to_develop_and_commercialize_anticoagulant_and_metabolic_compounds.

Pisano G (1997) R&D performance, collaborative arrangements and the market for know-how: A test of the "lemons" hypothesis in biotechnology. Working paper, Harvard Business School, Boston.

Pisano GP (2006) *Science Business: The Promise, the Reality, and the Future of Biotech* (Harvard Business Press, Cambridge, MA).

Powell WW (1996) Interorganizational collaboration in the biotechnology industry. *J. Institutional Theoret. Econom.* 152(1):197–215.

Prudhome A (2013) Investigating imclone: Why did the FDA give a thumbs down to the $2 billion breakthrough? *Vanity Fair* (December 3), https://www.pfizer.com/news/press-release/press-release-detail/bristol_myers_squibb_and_pfizer_announce_worldwide_collaboration_to_develop_and_commercialize_anticoagulant_and_metabolic_compounds.

Rhodes I, Nelson C, Berman G (2003) The key to successful collaborations: Rigorous and independent due diligence. *J. Commercial Biotechnology* 9(4):297–304.

Robertson TS, Eliashberg J, Rymon T (1995) New product announcement signals and incumbent reactions. *J. Marketing* 59(3):1–15.

Rosenbaum PR, Rubin DB (1983) The central role of the propensity score in observational studies for causal effects. *Biometrika* 70(1):41–55.

Rosenbaum PR, Rubin DB (1985) Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *Amer. Statist.* 39(1):33–38.

Sarnet M, Lachman R (2005) The art and process of successful in-licensing. *Biopharmaceuticals Internat.* 18(6).

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

Smietana K, Siatkowski M, Moller M (2016) Trends in clinical success rates. *Nature Rev. Drug Discovery* 15(6):379–381.

Srinivasan S, Hanssens DM (2009) Marketing and firm value: Metrics, methods, findings, and future directions. *J. Marketing Res.* 46(3):293–312.

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

Truex S (2018) Just a little patience: In-licensing from biopharma. *Life Sci. VC* (March 20), https://lifescivc.com/2018/03/just-a-little-patience-in-licensing-from-biopharma/.

Wager S, Athey S (2018) Estimation and inference of heterogeneous treatment effects using random forests. *J. Amer. Statist. Assoc.* 113(523):1228–1242.

Waring MJ, Arrowsmith J, Leach AR, Leeson PD, Mandrell S, Owen RM, Pairaudeau G, et al. (2015) An analysis of the attrition of drug candidates from four major pharmaceutical companies. *Nature Rev. Drug Discovery* 14(7):475–486.

Wong CH, Siah KW, Lo AW (2019) Estimation of clinical trial success rates and related parameters. *Biostatistics* 20(2):273–286.