NBER WORKING PAPER SERIES

# EVALUATION AND LEARNING IN R&D INVESTMENT

Alexander P. Frankel
Joshua L. Krieger
Danielle Li
Dimitris Papanikolaou

Working Paper 31290
http://www.nber.org/papers/w31290

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2023

We have no relevant interests to disclose. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Evaluation and Learning in R&D Investment
Alexander P. Frankel, Joshua L. Krieger, Danielle Li, and Dimitris Papanikolaou
NBER Working Paper No. 31290
May 2023
JEL No. G11,L65,O31,O32,O34


## ABSTRACT

We examine the role of spillover learning in shaping the value of exploratory versus incremental R&D. Using data from drug development, we show that novel drug candidates generate more knowledge spillovers than incremental ones. Despite being less likely to reach regulatory approval, they are more likely to inspire subsequent successful drugs. We introduce a model where firms are better able to evaluate the viability of incremental drugs, but where investing in novel drugs helps firms learn about future projects. Firms appear to put more value on evaluation versus learning, and those patterns are in-part driven by the appropriability of spillovers.



Alexander P. Frankel
Stanford Graduate School of Business
518 Memorial Way
Stanford, CA 94305
afrankel@chicagobooth.edu

Joshua L. Krieger
Harvard Business School
Rock 213
Boston, MA 02163
jkrieger@hbs.edu

Danielle Li
MIT Sloan School of Management
100 Main St, E62-484
Cambridge, MA 02142
and NBER
d_li@mit.edu

Dimitris Papanikolaou
Kellogg School of Management
Northwestern University
2211 Campus Drive, Office 4319
Evanston, IL 60208
and NBER
d-papanikolaou@kellogg.northwestern.edu

Scientific breakthroughs often build on earlier research efforts, including those that initially end in failure.[1] Yet despite the importance of exploration and learning for innovation (Romer, 1990; Aghion and Howitt, 1992; Scotchmer, 1991; Furman and Stern, 2011), research and development projects are often evaluated only on their potential for direct success. For instance, traditional net present value (NPV) calculations multiply a project's probability of success times its profits if successful, assuming zero benefits if the project does not succeed. In this paper, we examine how firms prioritize learning spillovers in their present R&D decisions.[2]

Our empirical work focuses on pharmaceutical R&D, a setting where breakthrough innovations can lead to enormous gains in welfare, where cumulative learning is important, and where failure is common. We first develop a new measure of cross-product knowledge spillovers. Using this measure, we show that novel drug candidates generate greater spillovers, and that much spillover value comes from drug candidates that themselves fail. Next, we use these facts to motivate a model of R&D investments. In our model, a firm must decide whether to invest in developing a drug candidate, which can either be incremental or novel. Firms are better able to evaluate the success likelihoods of incremental drugs because they are related to ideas that have previously been investigated, enabling them to screen out weak candidates. Firms know less about novel drugs, but by investigating a new area, they generate knowledge spillovers and learn about the success of future related drugs. Using this model as a guide, we provide empirical evidence that firms prioritize evaluation over learning, i.e. they are more reluctant to invest in novel relative to incremental drugs.

Following Krieger, Li, and Papanikolaou (2021), we measure a drug candidate's novelty by comparing its molecular structure with that of previously developed candidates. In this paper, we complement these backward-looking molecular linkages by defining new measures

[1] For example, in 2013, a large clinical trial for one of the most promising HIV vaccine candidates to date was halted due to lack of efficacy. Known as DNA/rAd5, the proposed vaccine sought to prime the immune system by injecting DNA plasmids that code for the production of protein structures also present in HIV. For this approach to be effective, the injected DNA must successfully enter a cell's nucleus so that its instructions can be carried out. Yet, in postmortems, scientists worried that DNA plasmids were too easily destroyed before reaching the nucleus. These concerns suggested two avenues for follow on research: vaccines based on mRNA (which need only enter a cell's cytoplasm to be effective) and delivery mechanisms that protect genetic material (Harris, 2021). Both these innovations are present in Moderna and Pfizer-BioNTech's vaccines for SARS-CoV-2, which deliver mRNA wrapped in lipid nanoparticles.

[2] Risk-adjusted NPV (rNPV), a modern tool widely adopted by pharmaceutical firms, refines the traditional NPV calculation by adjusting for a drug candidate's likelihood of progressing through each stage of R&D. Both the traditional the risk-adjusted approaches would assign zero value to the failed DNA/rAd5 vaccine of footnote 1, though, ignoring its potential to inform future vaccine development efforts.

of a drug candidate's spillover value based on forward-looking molecular linkages. Specifically, a drug candidate's "successors" are future drug candidates that are molecularly derivative of this focal drug, but not of any earlier drugs. A drug's "successor revenue" is the total revenue of its successors. This methodology allows us to calculate spillover value for all drug candidates, including those that never reach regulatory approval. Our results indicate that drug candidates regularly inspire the development of successor drugs, and that the majority of successor revenues accrue to failed focal drugs. This is, to our knowledge, the first direct estimate of the revenue generated by failed projects in any industry.[3]

We provide evidence that the benefits of investing in more novel drugs are backloaded relative to the benefits of investing in more incremental ones: novel drug candidates generate greater successor revenues than incremental drug candidates, despite being significantly less likely to reach the market. As such, ignoring spillover value, as in traditional NPV calculations, disproportionately reduces the value of novel drugs.

In order to understand the tradeoffs firms face, we introduce a model of R&D investment that incorporates the value of knowledge spillovers from exploratory research. Firms have an information advantage in evaluating incremental drugs, but investing in novel drugs generates spillovers that enable learning about the viability of future drugs.

In the first period of the model, a firm is presented with a novel or an incremental drug candidate. The firm is initially uncertain both about whether the drug will reach regulatory approval, and what its revenues would be if approved. We assume that the science around incremental drugs is better understood: if a drug is incremental, then the firm observes an initial signal of its probability of success. Based on this information, the firm is able to screen out some incremental projects that are likely to fail at the outset. For candidates that pass this initial screen, the firm then learns about the drug's expected revenues if successful, and decides whether to begin a sequence of costly investments to develop the drug. After the last stage of development, the firm receives revenue if this drug is a success.

[3]Prior work in the management literature has sought to examine how organizations learn from failures (for a review of this literature, see Desai, Maslach, and Madsen (2020)). For example, studies have asked how airlines, railroad companies, and NASA have altered their performance and safety records following accidents and wrecks (Haunschild and Sullivan, 2002; Baum and Dahlin, 2007; Madsen and Desai, 2010). However, empirical challenges usually prevent researchers from reliably attributing learning benefits to specific failures (Bennett and Snyder, 2017). Closer to our setting, Magazzini, Pammolli, and Riccaboni (2012) and Chiou, Magazzini, Pammolli, and Riccaboni (2016) find that patents associated with successful drugs are cited more often than those associated with failed drugs. Maslach (2016) studies medical device development, finding that firms are more likely to persist in developing products after adverse events involving incremental, rather than novel, innovations.

In the second period, either the focal firm or another firm may have the opportunity to invest in drug candidates that are related to the focal drug from the first period. If the drug considered in the first period was incremental, or if it was novel but the firm chose not to invest in development, then no additional knowledge was gained and the second period is essentially identical to the first. However, if a novel drug was developed in the first period, then the second period firm obtains an additional signal about the viability of related drugs. Put differently, after investing in a novel project, firms learn information about later drugs that essentially renders these projects incremental. This information is generated regardless of whether the original idea was successful, and can be observed by other firms in the market.

The model delivers an empirical diagnostic for assessing the relative value that firms place on better evaluation of incremental drugs versus future learning from novel drugs. In particular, when the private value of learning is high, firms are willing to develop novel drugs even when their direct revenues are expected to be lower than those of incremental drugs. If this is the case, then novel drugs should be more likely to enter development than incremental drugs, and have lower revenue if approved (reflecting a lower revenue threshold for developing these drugs). By contrast, when evaluation is relatively more important, firms are less likely to develop novel drugs, and their revenues on approval will be higher. Crucially, this diagnostic allows us to infer the value of spillover learning using only quantities that are observable in our data: the development rate of novel and incremental drugs and their revenues conditional on approval.

We provide evidence consistent with the view that firms prioritize evaluation over learning. Among pre-clinical candidates that are assessed in the same year for the same disease condition, novel drugs are substantially less likely to be brought into the next stage of development, i.e., clinical trials. At the same time, among the set of drugs that receive regulatory approval, novel drugs generate substantially more direct revenue than incremental drugs. Firms therefore appear to be more selective about developing novel drugs, favoring incremental drugs instead.

In the last part of the paper, we explore one reason firms may be reluctant to develop novel drugs: concerns about appropriability. While firms receive all direct revenues from the drug candidates they bring to market, they cannot be assured of receiving revenues from successor drugs. And in more crowded research areas, it is more likely that successor drugs will be developed by competing firms. Firms facing more competition will therefore have a lower private value of learning and be even more selective when developing novel drugs.

To bring this prediction to the data, we create a drug-firm level measure of research competition and then examine how firms' investments vary with this measure. We find evidence that competition leads firms to become even more selective in developing novel relative to incremental drugs. In more competitive research areas, we show that firms bring fewer novel drugs into development but that, conditional on approval, novel drugs generate even more revenues, relative to incremental drugs. Taken together, these results are consistent with the idea that competition reduces the value that firms place on learning, leading them to favor easier to evaluate incremental drugs.

Our analysis highlights a key insight, which is that the observed risk and returns associated with novel and incremental drugs are shaped by—and therefore potentially diagnostic of—firms' R&D priorities. Rather than viewing novel drug candidates as fundamentally "high risk, high reward," we show how this notion can emerge endogenously from how firms trade off being able to evaluate projects versus learn from them. Perhaps counterintuitively, our model shows that *higher* direct revenues for novel drugs is actually evidence that firms place *less* value on learning—otherwise firms would have been willing to invest in novel drugs with lower expected revenues. This result echoes arguments from the labor discrimination literature: if minorities are subject to increased scrutiny in the hiring process then, among those who are hired, minorities should outperform.

Our paper highlights, both theoretically and empirically, a dynamic channel through which the choice to "explore or exploit" today impacts the knowledge available to all firms in the future. We provide a unified explanation for three key facts: a) novel drugs generate more successor revenue; b) firms are less likely to bring novel drugs into development but these drugs generate more revenue conditional on approval; and c) this tendency is exacerbated by competition. While we recognize that these findings can be explained by a combination of other factors, our goal is to highlight the explanatory power the simple tension between evaluation in the present and learning the future.

In doing so, we connect several literatures. First, we extend a rich literature on cumulative innovation. Existing work in this area has developed new ways of tracing knowledge flows across academic and private sector while focusing on how disclosure mechanisms, intellectual property, and funding shape the rate and direction of follow-on innovation (Furman and Stern, 2011; Murray and Stern, 2007; Williams, 2013; Murray, Aghion, Dewatripont, Kolev, and Stern, 2016; Sampat and Williams, 2019; Azoulay, Graff Zivin, Li, and Sampat, 2018). Our paper complements this literature by exploring a new set of questions linking the anticipated

value of follow-on innovation to the initial decision of whether to engage in exploration. Methodologically, we contribute a new measure of knowledge spillovers that is based on a product's inherent physical properties, rather than on socially contingent patent citations.

Second, our work expands on the factors governing the decision to explore or exploit. In a bandit model, agents must choose between actions with uncertain payoff distributions and alternatives with known outcome parameters (Bergemann and Valimaki, 2006). Incentive schemes, corporate governance, financing constraints and business cycle effects may all impact the relative returns to exploration, but a typical assumption in these models is that the payoffs of these technologies are independent (Manso, 2011; Azoulay, Graff Zivin, and Manso, 2011a; Ferreira, Manso, and Silva, 2014; Balsmeier, Fleming, and Manso, 2017). In drug development, however, this assumption does not apply: investing in a new drug reveals information about other similar drugs. This makes our setting closer to the Brownian policy model of Callander (2011) and related papers such as Garfagnini and Strulovici (2016), Callander and Matouschek (2019), and Callander, Lambert, and Matouschek (2022).

Our work also the literature on competition and innovation. In canonical models such as Dixit and Stiglitz (1977); Aghion, Bloom, Blundell, Griffith, and Howitt (2005), firms' incentives to invest in novel R&D are shaped by product market competition for the focal innovation, which reduces direct revenues. Our model considers a complementary dynamic mechanism: competition reduces firms' ability to appropriate the value of future learning from the most novel R&D.

Finally, our work contributes to a growing literature highlighting the value of tolerating (or even embracing) failure in order to achieve innovative outcomes (Aghion and Tirole, 1994; Azoulay, Graff Zivin, and Manso, 2011b; Tian and Wang, 2011; Hvide and Panos, 2014; Krieger, 2021). Using a novel measure of follow-on innovation, we provide concrete empirical evidence that failed projects generate substantial spillover knowledge.

# 1 Setting and Data

## 1.1 Institutional Background

The drug development process is typically divided into three stages: discovery, pre-clinical research, and human clinical trials. In the discovery stage, firms consider potential compounds, many of which may only exist as a concept. Firms may then create computer models of

how a particular compound is predicted to behave, or they may synthesize the compound and examine whether it has any effect on the biological target of interest. At the end of the discovery stage, firms apply for patents on promising candidates. In the pre-clinical stage, researchers focus on understanding how the drug impacts the body (its pharmacodynamics) and, in turn, how the body impacts the drug (its pharmacokinetics). These tests are conducted in test tube cell cultures and in animal models. Finally, if a drug performs well in pre-clinical testing, firms may choose to develop the drug and file an application to begin human clinical trials. Clinical trials have three phases. Phase 1 clinical trials focus primarily on establishing a drug's safety, usually in a healthy population; Phase 2 trials provide preliminary information on a drug's efficacy among patients, and Phase 3 trials are large trials that become the basis of a regulator's decision as to whether or not to approve the drug.

Successful drug development can be quite lucrative. Recent estimates from Aryal, Ciliberto, Farmer, and Khmelnitskaya (2023) report a mean expected value for approved drugs of $1.63 billion. However, investments in drug development are expensive and risky: DiMasi, Grabowski, and Hansen (2016) estimate that the direct cost of developing a single approved drug is over $1.4 billion.[4] This cost is spread unevenly across the stages of drug development, with clinical (that is, Phase 1 and beyond) trials accounting for the bulk of development expenses. In addition, failure is common, with over 90% of drugs entering clinical trials never making it to market (DiMasi et al., 2016).

When firms develop a drug, they learn about a variety of issues—efficacy against disease, toxicity at different levels of dosing, unintended and "off target" benefits and side effects, interactions with other drugs, and differences in drug metabolism across patient groups—that are also informative about how related drugs will function. And because pharmaceutical firms disclose their research via scientific articles, patents, and other mandated filings, much of this information is accessible to other firms. Indeed, in a cross-industry study, Qiu and Wan (2015) show that pharmaceutical firms rank at the top in terms of generating and benefiting from knowledge spillovers.

Anecdotally, pharmaceutical firms routinely build on insights obtained from *failed* projects. For example, the first cholesterol-reducing statin drug tested in animals was compactin, developed by the Japanese firm Sankyo in the late 1970s (Endo, 2010). While compactin was found to reduce cholesterol in animals, its development was discontinued because of adverse

[4]The authors in fact estimate combined direct and indirect costs of $2.6 billion, but others have argued that these numbers are too high. See, for instance, http://www.nytimes.com/2014/11/19/upshot/calculating-the-real-costs-of-developing-a-new-drug.html.

effects. Merck, however, remained inspired by the drug's potential and worked to develop its own chemical analog to compactin. That compound, lovastatin (Mevacor), went on to be the first approved statin in 1987. Lovastatin then paved the way for a series of chemically similar statins, including Merck's simvastatin (Zocor) and Pfizer's atorvastatin (Lipitor).

## 1.2 Data and Sample

Our sample of drug candidates comes from Clarivate Analytics' Cortellis Investigational Drugs database (Cortellis), a business-intelligence database that focuses on tracking the progression of candidates from pre-clinical investigation, to clinical development, to approval.

Drugs enter Cortellis when they appear in public documents such as patent filings or shareholder reports. Because patents are typically taken out at the end of the discovery stage, we observe most drugs in pre-clinical investigation, but not in the earlier discovery stage.[5] We think of these pre-clinical candidates as a sample of "potential projects," which we observe regardless of whether a firm ultimately develops them further.

We obtain information on revenues for approved drugs from Evaluate Pharma, a commercial provider of drug sales data. We use a combination of exact matching, fuzzy matching, and manual confirmation to match Evaluate data to Cortellis, relying on drug name and company sponsor. Our data allow us to observe year-by-year sales associated with marketed drugs. This revenue data is necessarily censored: we observe only sales that have occurred, not the full stream of lifetime revenues. Because drugs that have been marketed for a longer period of time will naturally have greater total sales, we focus on a drug's *average annual revenue* for the years in which it appears in our sample.

In summary, we are able to track individual drug candidates from preclinical development onward. For each candidate, we observe whether it is developed (e.g., enters clinical trials), whether it is approved, and its revenues conditional on approval.

Finally, we make two types of sample restrictions throughout our analysis. Because our measures of novelty and follow-on activity are based on analyses of chemical similarity that only work for so-called "small molecule" drugs, we restrict our analysis to the 80% of drugs in

[5]Drugs in the discovery stage of research are often too nascent to be observable by Cortellis. In many cases, they have not yet been named, do not yet have publicly disclosed patent applications, and may not even be generally known within the firm that is developing them (Hughes, Rees, Kalindjian, and Philpott, 2011). We note that Cortellis's coverage for drugs in pre-clinical trials may be incomplete, especially for drugs in the earliest stages of investigation: the subset that we observe should be thought of as relatively serious contenders in pre-clinical development.

our data that fit this criterion. Our analysis therefore excludes biologic drug candidates such as monoclonal antibodies and vaccines. We also restrict our sample to drugs investigated in the United States, where our data coverage is more complete.

## 1.3 Variable Construction

### Drug novelty

We focus on a firm's decision to engage in incremental versus exploratory innovation. To measure this, we follow Krieger et al. (2021) and define an individual drug candidate's novelty in terms of its molecular distance from previously developed drugs. Our measure is based on a notion of molecular similarity known as a "Tanimoto" score. Tanimoto scores are widely used in pharmaceutical chemistry to identify relatedness among molecular compounds. A score of 0 indicates that two molecules do not share any common chemical substructures while a score of 1 indicates that the two molecules are identical in their atoms and bonding, up to stereosymmetry.[6]

We compute pairwise similarity scores between an initial drug candidate $i$ and all other drug candidates $j$ that entered human clinical trials prior to the focal candidate $i$'s earliest development date.[7] We define candidate $i$'s *novelty* as one minus its maximum pairwise Tanimoto score with prior drugs $j$, $T_{ij}$:

$$\text{Novelty}_i \equiv 1 - \max_{j \text{ prior to } i} T_{ij}. \tag{1}$$

By this measure, drug candidates are novel if they are molecularly different from previously developed drugs. Krieger et al. (2021) provides additional details and validation tests.

[6]See Wawer, Li, Gustafsdottir, Ljosa, Bodycombe, Marton, Sokolnicki, Bray, Kemp, Winchester, Taylor, Grant, Hon, Duvall, Wilson, Bittker, Dančík, Narayan, Subramanian, Winckler, Golub, Carpenter, Shamji, Schreiber, and Clemons (2014), Bickerton, Paolini, Besnard, Muresan, and Hopkins (2012), and Maggiora, Vogt, Stumpfe, and Bajorath (2014) for a discussion of Tanimoto scores in chemistry. In economics, Krieger et al. (2021), validate this measure by showing that drugs that target the same patient populations or share the same biological mechanism of action (MOA) have significantly higher average similarity scores than pairs unrelated through disease or MOAs.

[7]Because clinical trial reporting is mandated and public, this restriction ensures that the structures of these prior drugs were publicly known at the time the initial drug was developed, reducing the possibly that we mistakenly credit a drug as being derivative when it was in fact simultaneously and independently developed. For additional discussion and validation, see Krieger et al. (2021).

### Successors and Successor Revenues

To measure the spillover value of drug investments, we link drug candidates to subsequent product variants. Specifically, we define a drug $j$ as a *successor* to a focal drug $i$ if the following conditions are met: 1) the two drugs are molecularly similar, defined as having a Tanimoto score greater than 0.75; 2) the successor drug $j$ entered pre-clinical development after the focal drug $i$ entered Phase 1 clinical trials; and 3) drug $i$ is the earliest molecularly similar drug to drug $j$.[8]

Figure 1 presents an example of a focal and successor drug. Telapristone acetate (Proellex) entered development for the treatment of uterine fibroids in 2004 but its clinical trial was put on hold due to safety concerns. Despite never reaching the market, Proellex generated five successor candidates. One of these candidates, ulipristal acetate (Ella), ultimately reached the market and is currently is on the World Health Organization's List of Essential Medicines. Ella is pictured alongside Proellex in Figure 1: the two drugs have a Tanimoto score of 0.81.

Having defined the notion of a successor drug, we define *successor revenue* to a drug $i$ as total revenues across all of its successor drugs $j$:

$$\text{Successor Revenue}_i \equiv \sum_{j \text{ successor to } i} \text{Revenue}_j \tag{2}$$

In Equation (2), $\text{Revenue}_j$ is the average annual revenue associated with successor drug $j$. Unsuccessful successor drugs are included as having zero revenues.

Before continuing, we highlight several limitations. First, our data are truncated: more recently developed drugs are less likely to be linked to successor drugs simply because they had less time to generate successors. In addition, our measure of average annual revenues does not account for differences in earnings over the lifecycle of a drug. More recently approved drugs will have annual sales measured during a period the drug is on patent whereas older drugs will have some of their revenue years come after the drug faces generic competition. As a result, our main analyses will always control for development-year-quarter fixed effects, so

[8] A Tanimoto score of greater than 0.75 indicates a level of similarity that is apparent even to a layperson's eye. Figure 1, for example, shows two drugs that have an overlap of 0.81. A threshold of 0.75 captures drug candidates that are clearly related in their structure, which, in turn, is highly informative of similarity in function (Maggiora et al., 2014; Krieger et al., 2021; Krieger, 2021). Additionally, we focus on a drug's earliest molecular antecedent because it is often the case that one drug inspires many follow-on drugs which reach development at various times, rather than each drug inspiring the next in sequence. This approach also ensures that the existence and associated revenue of any given drug is credited to at most one predecessor drug.

that we are comparing drugs within the same cohort of development. Finally, we note that while we can measure revenue, we cannot measure profitability because we do not observe information on costs.[9]

Despite these limitations, our approach to identifying successor drugs has several advantages relative to patent-based measures of relatedness, e.g., based on citations or text. Foremost, our measure is based on a concrete product characteristic, molecular structure, which uniquely determines a drug. Pharmaceutical patents, by contrast, represent a wide range of intellectual protections (active ingredients, methods of delivery or manufacture) that cannot necessarily be cleanly associated with a specific drug.[10] Second, our similarity measure is based on objective chemical properties and does not rely on a patent examiner's discretion in determining which citations are relevant. Lei and Wright (2017), for instance, questions the reliability of patent citations as measures of intellectual relatedness by identifying many cases in which citations appear to be unrelated to the original patent.[11] Last, recent studies have also used patent-text based methods to measure similarity (Kelly, Papanikolaou, Seru, and Taddy, 2021; Kuhn and Thompson, 2019). Text-based measures may miss differences in design that are only be represented in schematics/diagrams or technical formulas, an important consideration in the context of patents for chemical molecules.

# 2 Motivating Facts

We next document a set of stylized facts regarding the spillovers from novel and incremental drug candidates.

[9]The majority of drug development costs are determined by the cost of clinical trials. Trial costs can vary substantially based on the number of patients in a trial, how long the trial is run, and the difficulty of enrolling these patients. These factors can vary greatly across disease types—trials for non-lethal cancers, for instance, require larger sample sizes and longer trials to achieve statistical significance relative to trials for very lethal cancers—but tend to be similar within disease type (our analysis will include disease level fixed effects).

[10]Even in the cases where it is possible to identify a drug's active ingredient patent, these patents are often cover molecular classes that include tens of thousands of specific molecules, only one of which may represent the ingredient in question.

[11]Consider patent 6,368,227 for "Method of swinging on a swing", issued to Steven Olson (aged 5) on April 2002. The patent has 20 patent citations as of 2022; it is cited, among others, by patent 8,420,782 for "Modular DNA-binding domains and methods of use" and patent 8,586,526 for "DNA-binding proteins and uses thereof." Many of these citations were added by the patent examiner.

## 2.1 Quantifying Spillovers

Panel A of Figure 2 plots the average number of successor candidates associated with focal drug candidates, by the highest phase of development they reached. While successful drugs (those that have been approved by the FDA) generate substantially more successors on average, failed drugs (those that have not been approved) still generate follow-on activity.[12] For example, the average drug that does not make it past Phase 1 trials is linked to 0.1 successor drugs.

Panels B and C consider successor revenues. In Panel B, we find that the successors associated with approved focal drugs generate a substantially higher amount of revenue, on average, compared to failed focal drugs (almost $30 million per year versus about $6 million). Yet, because there are so many more failed than successful drugs, the total value of spillovers generated by failed drugs is still large. This point is illustrated in Panel C, where we present total successor revenues associated with successful and failed focal drug candidates, aggregated across our sample. Total successor revenues attributed to successful focal drugs is approximately $10 billion per year, whereas this figure is over $16 billion among failed drugs. Due to their sheer number, failed drug candidates make up the primary source of ideas that are linked to commercially successful follow-on innovation.

We note that a comparison of the raw differences in number of successors or successor revenues among approved and non approved drugs is likely to understate spillovers to failed drugs. This is because our definition of "failed" drugs includes both drugs that were abandoned during development and those that are simply to developed too recently to have reached approval. These recent drugs (which we may incorrectly classify as failed), will also have had less time to accrue successors or successor revenues. In our analysis going forward, we will include drug development year-quarter fixed effects to control for cohort-level differences in outcomes.

[12] Our use of the term "failed" is somewhat imprecise. Because our data do not have termination dates, we cannot definitively identify whether a drug has failed, so this set of drugs should be thought of as those that have either failed in development or ones whose ultimate outcome has not yet been realized. For brevity, we will use the terms "not approved" and "failed" interchangeably, although the former is more accurate.

## 2.2 Novelty and Spillovers

Next, we consider how spillover value varies by drug novelty using the following specification:

$$\text{Outcome}_i = a_0 + a_1 \text{Novelty}_i + \delta_t + \delta_d + \varepsilon_i \tag{3}$$

Equation (3) is estimated at the drug level, and includes all candidates that enter human clinical trials in the United States. The main explanatory variable of interest is $\text{Novelty}_i$ which represents a drug candidate's molecular novelty (0 being molecularly identical to a prior drug candidate and 1 being completely novel). In our primary specifications, we include controls for cohort fixed effects (development-year-quarter) $\delta_t$ as well as fixed effects $\delta_d$ for a drug candidate's lead disease indication (ICD-9).

Overall, we find that novel drug candidates are associated with greater successor revenues. Figure 3 plots the results and Table 1 presents the accompanying regression results. Panel A of Figure 3 shows that molecularly novel drug candidates generate more successor attempts. Column 2 of Table 1 provides the accompanying magnitude: among drugs developed at the same time for the same condition, a one standard deviation (0.23) increase in a drug's Tanimoto novelty score is associated with $0.43 \times 0.23 = 0.10$ more successor drug candidates, or a 29% increase from a baseline mean of 0.35. Panel B of Figure 3 also shows that novel drugs generate more successor revenue. In terms of magnitudes, Column 4 of Table 1 indicates that a one standard deviation increase in novelty correlates with increased successor revenues of $\$9.7 \times 0.23 = \$2.2$ million annually, a 32% increase from a mean of \$6.8 million.

Panels C and D provide additional context for these results. In Panel C, we plot the relationship between a drug candidate's novelty and its expected direct revenues (unsuccessful drugs are included as having zero revenues). Here, we find that novel drugs tend to generate a similar amount of direct revenue (the slope is positive but our estimates in Columns 5 and 6 of Table 1 lose significance in some specifications). Panel D finally combines insights from Panels B and C to show that, among novel drugs, successor revenue makes up a greater share of total attributed revenues (direct plus successor).[13] This result can be explained by the fact that novel drug candidates are substantially more likely to fail in the development process, thereby generating no direct revenues. Given this, successor revenues are a relatively more important part of the total value of novel drug candidates.

[13]Because share successor revenue is often undefined at the individual drug level (many drugs have no direct or successor revenues), the analysis in Panel D is done by aggregating total direct and successor revenues across all drugs in a given decile of novelty.

One potential concern is that these findings follow mechanically from our definition of a successor drug. Recall that we attribute a successor drug to the first molecular antecedent that has a pairwise Tanimoto score of 0.75 or greater. Definitionally, then, a focal drug that is itself highly incremental may be less likely to be credited with successor innovation because its own molecular predecessor claims this credit. There are three points worth noting here. First, pairwise similarity is not generally transitive so highly incremental drugs can still be linked to successors—put differently, a drug's immediate predecessor and successor need not be pairwise similar to each other. Second, our results are not driven by this measurement choice because we find a roughly linear and positive relation between novelty scores and successor revenue. If our results had been driven by special behavior around our 0.75 Tanimoto similarity cutoff, then we would expect the number of successor drugs and revenues to rise discontinuously for drugs with novelty less than 0.25 (i.e., $1 - 0.75$). In Appendix Table A.1, we show that these descriptive patterns continue to hold when we use alternative measures of a focal drug's novelty that do not rely on measures of molecular similarity.[14] Last, we view this feature as partly reflective of what it means to inspire new research: in many cases, we observe a single novel drug candidate that inspires a series of successors that arrive at a similar time. These successors should be thought of all being inspired by the original focal drug, rather than serially by each other.

## 2.3 Discussion

Our empirical evidence so far suggests that exploratory projects contribute disproportionately more value through spillovers ignored in traditional valuation. Novel projects generate more successor revenue than incremental projects, and successor revenues comprise a greater share of the total revenues associated with novel projects. Taken together, this suggests a potential trade off: firms may be less able to evaluate the prospects of novel drugs, but the process of developing such drugs generates greater opportunities for future learning.

In the next section, we formalize this tradeoff in a model. We show that firms may set a higher or lower threshold for developing novel drug candidates, depending on how much

[14] In particular, we define a focal drug as novel if it either focuses on a new "target" (e.g. it attempts to impact a new biological object such as a protein), or if it focuses on a new target with a new mechanism of action (e.g. if it is the first to seek to impact that a target in a new way, e.g. if it is the first to inhibit the expression of that protein). Across both definitions, we continue to find that novel drugs generate more successors and successor revenue, and that a greater share of the total revenue we associated with them come from spillover channels.

they benefit from learning spillovers. The model then generates several empirical predictions which we bring to the data.

# 3 A model of drug development

We present a two-period model with one *firm* and the *rest of the market*. In the first period, the firm is given the opportunity to develop a drug, which can either be *novel* or *incremental*. Incremental drugs are easier to screen: the firm is better able to predict which drugs will or will not succeed. Firms know less about novel drugs, but developing them generates information about the success of related drug candidates that this firm or the rest of the market may consider developing in the second period. Essentially, we think of incremental drugs as ones that are related to—and thus informed by—previously-developed novel drugs.

## 3.1 The model

There are two periods. Period 2 revenues and costs are discounted at factor $\beta \in (0, 1]$ relative to period 1.

**Period 1.** In period 1, the firm considers whether to develop an initial drug. This drug has *novelty* $N \in \{0, 1\}$, with $N = 0$ indicating an incremental drug and $N = 1$ indicating a novel drug; *success* if developed of $S \in \{0, 1\}$; and expected *revenue* if successful of $R$. We assume that the probability of success if developed is $Pr(S = 1) = \pi \in (0, 1)$ and that revenue $R$ is drawn according to a distribution $F_R$ on $\mathbb{R}_+$, with the same distributions of success and revenue regardless of drug novelty.[15] (Unless otherwise specified, all random variables are taken to be independent of one another.) The timeline of period 1 is then as follows.

**Stage (i): Discovery.**

**(a) Awareness.** The firm is made aware of the initial drug and observes whether the drug is novel or incremental.

[15] We interpret $R$ as the firm's expected revenue, rather than the drug's realized revenue, if the drug were to be successfully developed. The role of this expectation is to account for the noise that we will observe when we bring this model to data. In particular, in the data, there may be drugs that are successfully developed and yet yield negligible revenue. In the model, these would be drugs whose realized revenue underperformed relative to expectations.

**(b) Screening incremental drugs.** If the drug is incremental, the firm observes a signal $\sigma^0$ of whether the drug would be successful if developed. Specifically, drugs that would be successful ($S = 1$) always generate a good signal $\sigma^0 = g$. Drugs that would fail ($S = 0$) generate a bad signal $\sigma^0 = b$ with probability $q^0 \in (0, 1)$ and a good signal $\sigma^0 = g$ with probability $1 - q^0$. That is, if an incremental drug is low quality, the firm learns this fact with probability $q^0$.

If the drug is incremental and $\sigma^0 = b$, then the firm screens the drug out ($P = 0$) and the period is over. Otherwise—if $\sigma^0 = g$, or if the drug is novel—then the firm proceeds to *pre-clinical testing* ($P = 1$), Stage (ii).[16]

**Stage (ii): Pre-clinical testing.** The firm observes the drug's revenue $R$. The firm then decides whether to pay a cost $C^1 > 0$ to bring the drug into *development* ($D^1 = 1$), Stage (iii). If not ($D^1 = 0$), the period is over.

**Stage (iii): Development.** Development proceeds in $M \geq 1$ phases, corresponding to phases of human clinical trials.

In each development phase $t \in \{1, ..., M - 1\}$: The firm observes a signal $\sigma^t$ of success $S$. If $S = 1$, the firm observes a good signal $\sigma^t = g$ with certainty. If $S = 0$, the firm observes a bad signal $\sigma^t = b$ with probability $q^t \in (0, 1)$ and a good signal $\sigma^t = g$ with probability $1 - q^t$. Then the firm decides whether to pay a cost $C^{t+1} > 0$ to bring the drug into phase $t+1$ of development ($D^{t+1} = 1$). If not ($D^{t+1} = 0$), the period is over.[17]

In the final development phase $t = M$: Success $S$ is revealed. If $S = 1$, the firm brings the drug to *market*, Stage (iv). Otherwise the period is over.

**Stage (iv): Market.** The drug realizes revenue for the firm, equal in expectation to $R$.

**Period 2.** If the initial drug is incremental, then there is no second period decision. In this case we normalize the firm's second-period payoff to 0.

If the initial drug is novel, some number of related drugs may arrive. Each related drug $j$ has success if developed of $S_j \in \{0, 1\}$ and revenue if successful of $R_j$. Assume that $Pr(S_j = 1) = \pi_2 \in (0, 1)$ and $R_j \sim F_{2R}$. If the initial drug was brought to development, i.e.,

[16] An incremental drug with $\sigma^0 = b$ does not enter pre-clinical testing because it will fail with certainty. Implicitly, one can think of there as being a small cost of proceeding to the next stage.

[17] If the drug does not make it to development phase $t$, then we take $D^{t+1} = 0$: the drug is also not brought to development phase $t + 1$.

if $D^1 = 1$, then these related drugs are essentially rendered incremental, and we call them *successor drugs*. For each related drug that arrives, the second period proceeds similarly to the first, where there will now be additional information about successor drugs.

**Stage (i): Discovery.**

**(a) Awareness.** A non-negative number $Q$ of related drugs arrives, with $Q$ drawn from $F_Q$ and having mean $\mu_Q > 0$. For each related drug $j \in \{1, ..., Q\}$, either the focal firm ($A_j = 1$) or the rest of the market ($A_j = 0$) is made aware of the drug, meaning that this player has the exclusive opportunity to investigate and develop it. The probability that the focal firm is made aware of any given related drug is $Pr(A_j = 1) = \alpha \in (0, 1]$.

**(b) Screening successor drugs.** If the initial drug was developed ($D^1 = 1$), then the player who is made aware of successor drug $j$ observes a signal $\sigma_j^0$ of $S_j$. Specifically, drugs that would succeed ($S_j = 1$) generate a good signal $\sigma_j^0 = g$ with certainty, while drugs that would fail ($S_j = 0$) generate a bad signal $\sigma_j^0 = b$ with probability $q^0 \in (0, 1)$ and a good signal $\sigma_j^0 = g$ with probability $1 - q^0$.

If drug $j$ is a successor drug and $\sigma_j^0 = b$, then the relevant player screens the drug out. Otherwise—if $\sigma_j^0 = g$, or if the initial drug was not developed—then the drug proceeds to pre-clinical testing, Stage (ii).

**Stages (ii) – (iv).** These stages proceed exactly as in period 1, separately for each drug that enters pre-clinical testing, for the player who is made aware of that drug.[18]

We assume that all parameters and distributions that were not specified to be drawn from a distribution are commonly known at the start of the game: success probabilities $\pi$ and $\pi_2$, revenue distributions $F_R$ and $F_{2R}$, costs $C^t$, signal precisions $q^t$, the expected number of successor drugs $\mu_Q$, and the probability $\alpha$ that the firm has the opportunity to develop a given successor drug.

Putting all the payoffs together, the firm's realized profit over the two periods can be written out fully as

$$\underbrace{SD^M R - \left(\sum_{t=1}^{M} D^t C^t\right)}_{\text{Period-1 profit from initial drug}} + \underbrace{N\beta \sum_{j \in \{1,...Q\}} A_j \left(S_j D_j^M R_j - \left(\sum_{t=1}^{M} D_j^t C^t\right)\right)}_{\text{If novel: Period-2 profit from related drugs}}. \tag{4}$$

[18] For related drug $j$, we now replace $R$, $S$, $D^t$, and $\sigma^t$ with $R_j$, $S_j$, $D_j^t$, and $\sigma_j^t$.

The expression is the direct term benefit of investing in the drug, which is the benefit captured by a standard NPV calculation. If the firm is forward-looking ($\beta > 0$) it should also take into account the potential indirect benefits of its investment decisions.[19]

If the initial drug is novel, the rest of the market also realizes a payoff from developing related drugs (those with $A_j = 0$), given by

$$\underbrace{N\beta \sum_{j\in\{1,\dots Q\}} (1 - A_j) \left( S_j D_j^M R_j - \left( \sum_{t=1}^{M} D_j^t C^t \right) \right)}_{\text{If novel: Period-2 profit from related drugs}}. \tag{5}$$

As discussed below, after being made aware of a drug, the firm will develop that drug if its revenue ($R$ or $R_j$) is large enough. To guarantee that revenue can in fact be large enough so that a drug is developed, assume that the support of the revenue distribution $F_R$ extends above $\sum_{t=1}^{M} C^t/\pi$, and likewise the support of $F_{2R}$ extends above $\sum_{t=1}^{M} C^t/\pi_2$.[20] To guarantee that first-period revenue can be low enough so that direct profits alone are not enough to justify development, assume that the support of the revenue distribution $F_R$ extends below $C^1$.

## 3.2 Preliminary Analysis

We are primarily interested in the firm's first-period decision to invest in developing the initial drug, i.e., the choice of $D^1$. Recall that this decision is taken after the firm has already screened out some share of incremental drugs that would not have been successful (those with $\sigma^0 = b$), and after the firm learns the expected revenue $R$ for the drug.

To better understand this initial development decision, let us start by decomposing the firm's expected profit as a function of the choice of $D^1$ as

$$D^1 \cdot (V_1^N(R) - C^1) + \beta \left( W_2^N + N D^1 \mu_Q \alpha \cdot \Delta_2^{N=1} \right) \tag{6}$$

[19] Properly speaking, the direct revenue from the drug accrues over time and so the firm's value of this will also depend on discounting.

[20] In period 1, if a novel drug has $R > \sum_{t=1}^{M} C^t/\pi$, then committing to pay the cost of all development phases in advance would be more profitable in the first period than choosing not to develop the drug; an incremental drug with a positive signal $\sigma^0 = g$ would yield even higher profit. So for these drugs, the firm would choose $D^1 = 1$. Similarly in the second period if $R_j > \sum_{t=1}^{M} C^t/\pi_2$, both for drugs following a developed and an undeveloped first-period novel drug.

for some terms $V_1^N(R)$, $W_2^N$, and $\Delta_2^{N=1}$. (See Appendix A.1 for the full expansion of each term, along with a discussion of their properties.)[21] The first term $V_1^N(R)$ is the expected direct first-period payoff of developing the drug (excluding the initial cost $C^1$), which depends both on novelty and revenue. The second term $W_2^N$ is the portion of the firm's second-period profit that doesn't depend on the first-period development decision. The third term $\Delta_2^{N=1}$ is then the amount that second-period profit increases if a novel first-period drug is developed, for each related drug that the firm sees.

The key tradeoff in the model is captured by the two observations.

First, the following inequality holds:

$$V_1^{N=0}(R) - V_1^{N=1}(R) \geq 0 \text{ for every } R. \tag{7}$$

Equation (7) states that the firm has higher first-period profits from incremental drugs, for any revenue level.[22] This occurs because incremental drugs are easier to evaluate: the additional signal the firm receives allows it to screen out a subset of weak candidates. This "evaluation benefit" accrues entirely to the firm that makes the investment.

Second, we have:

$$\Delta_2^{N=1} > 0. \tag{8}$$

Equation (8) states that developing a novel drug gives a payoff bonus in the second-period. This holds because the development of novel drugs today improves the evaluation of successor drugs in the future. This benefits the firm in two distinct ways. First, by screening out successors that generate bad initial signals, the firm avoids making costly investments in drugs that are likely to fail. Second, the remaining successors that the firm develops will be positively selected and bring higher profits on average. However, unlike the direct evaluation benefit of developing incremental drugs, this "learning benefit" of developing novel drugs is shared by the firm and the rest of the market, both of whom may have the opportunity to develop successors.

[21] Importantly, none of the three terms depends on $D^1$; the parameters $\beta$, $\mu_Q$, and $\alpha$ do not appear in the decision-relevant terms $V_1^N(R)$ or $\Delta_2^{N=1}$; and expected revenue $R$ does not appear in $\Delta_2^{N=1}$. That is, all of the decision-relevant dependence of expression (6) on $D^1$, $\beta$, $\mu_Q$, $\alpha$, and $R$ is made explicit.

[22] The inequality is strict when there is a positive value of developing incremental drugs, i.e., when $V_1^{N=0}(R) > 0$.

Before proceeding, we add the following maintained assumption, that the firm's private informational benefit of developing a novel drug is less than the first phase development cost.

**Assumption 1.** The expected discounted increase in second-period profit from developing a novel drug is less than the cost of bringing such a drug to development: $\beta\mu_Q\alpha\Delta_2^{N=1} < C^1$.

If this assumption were violated, the firm would want to bring every novel drug into development, even if the drug had low enough revenue that the firm planned on "pulling the plug" in phase 2 (setting $D^2 = 0$) with certainty. Imposing this assumption, the firm will only develop a novel drug that it plans on taking to market if it continues to receive positive signals about the drug's success.

## 3.3 Model implications

We can now describe the firm's strategy in the first period of the game. For each type of drug $N$, incremental or novel, the firm will choose a revenue threshold $\overline{R}^N$. Drugs of type $N$ with expected revenue $R$ above the threshold $\overline{R}^N$ will be developed ($D^1 = 1$), and drugs with revenue below the threshold will not be. Conditional on arriving at development phase $t$, the firm proceeds to development phase $t+1$ ($D^{t+1} = 1$) if and only if it receives a positive signal ($\sigma^t = g$). See Appendix A.2 for a formalization.

Our model makes an unambiguous prediction that, conditional on having begun development, incremental drugs are more likely to progress at every phase. (All proofs are in Appendix A.3.)

**Proposition 1.** *The probability that a drug passes any given phase of development is higher for incremental drugs. That is, for $t \in \{1, ..., M-1\}$, it holds that $Pr(D^{t+1} = 1|D^t = 1, N = 0) > Pr(D^{t+1} = 1|D^t = 1, N = 1)$. In addition, incremental drugs have a higher probability of success conditional on reaching the final phase of development: $Pr(S = 1|D^M = 1, N = 0) > Pr(S = 1|D^M = 1, N = 1)$.*

This result arises from the fact that incremental drugs begin with an informational advantage: the firm screens out a subset of incremental drugs that generate a bad signal $\sigma^0 = b$ at the discovery stage. The remaining set of incremental drugs is positively selected to have a higher probability of success. Hence, conditional on arriving at any phase $t$ of development, incremental drugs are more likely to receive a positive signal $\sigma^t = g$, with the firm progressing to the next phase exactly when it sees such a signal.

Proposition 1 establishes that incremental drugs are more likely to progress at each stage of development. However, it is theoretically ambiguous whether incremental drugs are more likely to enter development in the first place. This occurs if the firm chooses to set a lower revenue threshold for developing incremental drugs than for novel drugs ($\overline{R}^0 < \overline{R}^1$). $\overline{R}^0$ may be lower because the firm is able to screen out a subset of weak incremental candidates so that the remaining candidates are safer bets. Alternatively, $R^1$ may be lower because developing novel drugs generates knowledge spillovers in addition to direct revenue.

Our model yields a diagnostic test for determining the relative thresholds for development. In particular, we connect a firms' thresholds for developing novel and incremental drugs to observable quantities: the likelihood of development conditional on entering pre-clinical testing, and revenue conditional on development.

**Proposition 2.** *One of the following two cases obtains:*[23]

1. High evaluation benefit. *If* $\overline{R}^{N=1} \geq \overline{R}^{N=0}$*, then novel drugs are less likely to be developed conditional on entering pre-clinical testing but have higher average revenues conditional on being successfully developed. That is,* $Pr(D^1 = 1|P = 1, N = 1) \leq Pr(D^1 = 1|P = 1, N = 0)$ *and* $\mathbb{E}[R|S \cdot D^M = 1, N = 1] \geq \mathbb{E}[R|S \cdot D^M = 1, N = 0]$.

2. High learning benefit. *If* $\overline{R}^{N=1} \leq \overline{R}^{N=0}$*, then novel drugs are more likely to be developed conditional on entering pre-clinical testing but have lower average revenues conditional on being successfully developed. That is,* $Pr(D^1 = 1|P = 1, N = 1) \geq Pr(D^1 = 1|P = 1, N = 0)$ *and* $\mathbb{E}[R|S \cdot D^M = 1, N = 1] \leq \mathbb{E}[R|S \cdot D^M = 1, N = 0]$.

Proposition 2 provides a diagnostic to help reveal firms' priorities. It states that one of two cases is possible. In the first, firms place a high value on evaluation today: because incremental projects are easier to screen, firms are more confident that the incremental drugs they bring into development will reach FDA approval. The lower risk of failure allows them to profitably invest in an incremental drug idea even when its projected revenue on approval is relatively low. This case implies that incremental projects will be more likely to enter development and will have lower revenues conditional on success. Alternatively, firms may place a high value on learning that will improve their ability to evaluate tomorrow: because firms value the knowledge spillovers when they explore new areas, they may be willing to invest in novel drugs even when their direct revenues are likely to be low. Here, firms should

[23] In the knife-edge case that $\overline{R}^{N=1} = \overline{R}^{N=0}$, both cases obtain.

be more willing to develop novel drug candidates and, conditional on approval, novel drugs will have lower revenues.

We also consider how the firm's development strategy varies with market characteristics.

**Proposition 3.** *Fix all parameters of the model other than $\beta$, $\mu_Q$, and $\alpha$.*

1. Incremental drugs: *The revenue threshold for development, $\overline{R}^{N=0}$, is constant in $\beta$, $\mu_Q$, and $\alpha$. Thus, the probability of development conditional on entering pre-clinical testing ($Pr(D^1 = 1|P = 1, N = 0)$) and the expected revenue conditional on successful development ($\mathbb{E}[R|S \cdot D^M = 1, N = 0]$) are constant in these parameters.*

2. Novel drugs: *The revenue threshold for development, $\overline{R}^{N=1}$, strictly decreases in the summary statistic $\beta\mu_Q\alpha$. So the probability of development conditional on entering pre-clinical testing ($Pr(D^1 = 1|P = 1, N = 1)$) weakly increases in $\beta\mu_Q\alpha$ and the expected revenue conditional on successful development ($\mathbb{E}[R|S \cdot D^M = 1, N = 0]$) weakly decreases in $\beta\mu_Q\alpha$.*

Intuitively, factors that reduce the spillover value of novel drugs lead firms to be more selective in the novel drugs that they pursue, resulting in the development of fewer novel relative to incremental drugs. Focusing on the appropriability parameter $\alpha$, we note that reducing $\alpha$ lowers a firm's private value of investing in novel drugs, but does not affect the social value of such investments. As such, reducing appropriability leads to inefficiently low levels of investment in novel drugs.

## 3.4 Model Discussion

### Empirical Content

Each Proposition generates an empirical prediction that is testable in our data. We note that while novelty in our model is binary, our empirical measure of novelty, defined in Equation (1), is a continuous measure ranging from 0 (no molecular overlap with previously developed candidates) to 1 (full overlap).

Proposition 1 predicts that, among drugs that enter human clinical trials, drugs that are less novel are more likely to progress at each stage: from Phase 1 development to Phase 2,

Phase 2 to Phase 3, and Phase 3 to approval. These quantities are observable in the Cortellis data.[24]

Proposition 2 uses observable quantities to infer the value that firms place on learning relative to evaluation. This is important because the value of information about a drug's success (the evaluation benefit in the present, and the learning benefit in the future) is difficult to measure directly. We can think of two distinct sources of value: spurring additional development of promising drugs, and shutting down research that would fail. While we can proxy for the former (a drug's direct revenue as well as successor counts and revenue), we have no way of directly measuring the value of avoiding bad projects, which never show up in our data. Proposition 2 allows us to infer the net value from both sources using information in our data: development decisions (which we define as progressing past pre-clinical testing to enter human clinical trials) and revenues conditional on success. In particular, it first makes the testable prediction that if novel drugs are more likely to be developed than incremental drugs, then they will also have lower average revenue conditional on success, and vice versa. It then tells us that whichever type of drug has a higher development probability (and lower average revenue) is the one favored by the firm.

Finally, Proposition 3 predicts that the value of learning falls when appropriability is lower. In areas where more firms are actively researching, the discoveries of one firm are more likely to benefit their competitors. That is, more competition corresponds to less appropriability, which reduces the focal firm's incentives to invest in developing novel drugs. We define a drug's "research area" as all treatments for a given disease using the same biological mechanism of action, and we measure competition as the number of new drugs introduced in a research area by other firms.

### Model Assumptions

Our model makes a number of simplifying assumptions. For instance, we assume that the arrival of different types of drugs is exogenous and undirected, i.e., the firm does not specifically seek out novel or incremental drugs. In addition, the knowledge generated by developing a novel drug accrues to all firms in the market, so that the developing firm does not have any specific advantage. These simplifications allow us to reduce the number of

[24]Using a similar dataset, Krieger et al. (2021) show that novel drugs are less likely to be approved, conditional on entering Phase 1.

parameters and assumptions in our model, so that we can focus on its central tradeoff of evaluation versus learning.

Beyond simplifying assumptions, we highlight a key substantive assumption, that the underlying distributions of success likelihoods and revenues are the same for novel and incremental drugs.[25] We begin with the idea that these drugs are distinguished only by a researcher's information. The same physical molecule, with the same underlying biological action, is novel if no similar molecule has been developed in the past, but would be incremental otherwise. Because any molecule could be novel or incremental depending on the state of prior research, we view both types of drug candidates as being drawn from the same distribution of project fundamentals.

Assuming that the underlying distribution of returns is the same for novel and incremental projects further allows us to highlight the role that selection plays in shaping the *observed* traits of drugs in development. In our model, differences in the risk and returns associated with novel and incremental drugs in development emerge endogenously from differences in the criteria that firms apply when deciding which projects to pursue. Our goal is to highlight how this simple mechanism can generate many of the empirical patterns we observe. In Section 4.2 we provide a more detailed discussion of alternative explanations for these patterns.

# 4 Empirical Results

## 4.1 Main Findings

Proposition 1 predicts that incremental drugs are more likely to progress through each stage of clinical trials and to be approved. Figure 4 confirms this prediction in our data, with corresponding regressions reported in Table 2. Panel A considers the relationship between a focal drug's novelty (one minus its maximum pairwise Tanimoto similarity score) and its likelihood of progressing from the first stage of clinical trials (Phase 1) into the second (Phase 2). Our specification compares drug candidates that are developed at the same time, for the same disease, but which differ in their novelty. We find statistically significant associations between a drug's novelty and its likelihood of progression through each stage of clinical

[25] Unfortunately this assumption is not directly testable because the initial distribution of quality for nascent ideas is fundamentally unobserved: beyond the fact that datasets such as Cortellis do not capture all drugs in discovery stages, many ideas occur to a researcher only for a moment and are dismissed before she even communicates them to anyone else or even writes them down for herself.

trials. Column 2 of Table 2 provides the analogous regression specification. The coefficient on novelty of -0.089 indicates that a one standard deviation increase in novelty (0.26), is associated with a $0.26 \times -0.089 = 0.023$ or 2.3 percentage point decrease in the likelihood of entering Phase 2 trials, conditional on entering Phase 1. This translates into a relatively modest 3.2 percent decrease from the mean. In Panels B and C (corresponding to Columns 4 and 6 of Table 2), we see larger negative relations between novelty and progression: one standard deviation increase in novelty is associated with a 13.7% decrease in progression from Phase to Phase 3 and a 9.2% decrease in progression from Phase 3 to launch. These results are consistent with our model's prediction that firms have an information advantage when developing incremental drugs: because they are able to quickly discard projects that are highly likely to fail, the remaining incremental drugs we observe are positively selected. The fact that we find the largest correlations for progression from Phase 2 to 3 and 3 to launch suggests that the information advantage that firms have about incremental drugs relates to the their efficacy rather than their toxicity in humans.[26]

We next consider how much firms value this information advantage. Proposition 2 states that if firms care more about evaluation, they will set a lower threshold for developing incremental drugs. In that case, incremental drugs will be more likely to be developed and their revenues will be lower on approval. In contrast, if firms care more about learning, we would find the opposite pattern.

Figure 5 provides evidence that firms care more about evaluation: that is, we are in Case 1 of Proposition 2. In Panel A, we show that, among drugs that enter pre-clinical investigation at the same time, for the same disease, novel drug candidates are substantially less likely to enter clinical development. The accompanying magnitudes, reported in Column 2 of Table 3, indicate that a one standard deviation increase in novelty (0.21 in the pre-clinical sample), is associated with a $0.21 \times 0.27 = 5.7$ percentage point decrease in the likelihood of entering development, or an approximately 15% decrease from our baseline development rate. At the same time, our findings in Panel B indicate that novel drugs generate more revenue conditional on approval. As reported in Column 4 of Table 3, a one standard deviation increase in novelty is associated with an increased annual revenue of just over 0.21*\$515=\$108 million, or an over 25% increase from the overall mean among launched drugs. This pattern is consistent with firms using a traditional NPV calculation (as if $\beta \approx 0$ in our model): if

[26]Drugs that enter clinical trials at all have passed some safety tests in animal models. Progression from Phase 1 to 2 hinges on drugs not having side effects in a population of healthy patients and over 70% of drugs in our sample pass this bar.

firms view novel drugs as riskier to develop, then they will only invest if they anticipate higher revenues on approval.

Before continuing, we note that in Panel C of Figure 3, we showed that novel drugs have higher expected direct revenues when compared to more incremental drugs. Without the context of a model, one might conclude that novel drugs are unambiguously better: they generate as much direct revenue, and they also provide knowledge spillovers for the future. Our model, however, highlights the role of selection. Firms apply a higher revenue threshold for developing novel drugs and, as such, the set of novel projects we observe in development will appear to be drawn from a better distribution.

## 4.2 Alternative explanations

A key assumption in our model is that novel and incremental drugs are initially drawn from the same distribution of success probabilities and revenues This assumption allows us to attribute differences among novel and incremental drugs observed in development to differences in firms' selection. In practice, however, differences in development rates or revenues between novel and incremental drugs could arise for reasons unrelated to those our model focuses on. For example, the successful development of one drug could mean that subsequent successors, addressing the same disease, face a more crowded product market with lower revenue potential. In this case, the successor drugs may be drawn from a distribution with lower revenue potential to begin with. Relatedly, if success is positively correlated across similar drugs, successors of already approved drugs may themselves be more likely to succeed. In this case, we would expect more incremental drugs to enter development, even if firms were applying the same development thresholds for novel and incremental drugs.

Even though other factors may lead novel and incremental drugs to be drawn from different initial distributions, we think that the channel highlighted by the model—additional endogenous differences driven by selection—is still relevant. For example, the alternative explanations provided above only explain why incremental drugs that are related to *successful* prior drugs may be more likely to be approved or have lower revenues conditional on success. As we discuss below, we find that our patterns are robust to comparing novel drugs only with incremental drugs based on *failed* predecessors. Such a pattern can be explained by our model (because firms learn about successors regardless of whether a focal drug succeeds or

fails), but not by these other stories. Thus we view the selection story highlighted in our model as a parsimonious explanation for the patterns we find in our data.

In Appendix Tables A.2 and A.3, we test the predictions of Propositions 1 and 2 after excluding follow-on drugs linked to predecessor drugs that are FDA approved. That is, we compare molecularly novel drugs to molecularly incremental drugs that are derivative of failed predecessors. Consistent with Proposition 1, our results in Appendix Table A.2 show that incremental drugs based on failed predecessors are still more likely to progress through development. In Appendix Table A.3, we look once again at firms' thresholds for developing novel versus incremental drugs. In Columns 1 and 2, we show that novel drugs are still less likely to enter development and, in Columns 3 and 4, that they continue to have higher revenues on approval. In all these regressions, our estimated magnitudes are similar. Taken together, this suggests that the process of developing a drug generates valuable information for future drugs, even if the drug itself fails. Were that not the case, we would not expect firms to favor the development of incremental drugs based on failed predecessors.

## 4.3 Competition and Appropriability

Our results so far suggest that firms do not place a high value on the potential knowledge spillovers generated by investing in novel drugs. Firms, however, make decisions based on private rather than social benefits. If the spillovers from developing in novel drugs cannot be fully appropriated, then this reduces the benefit of making such investments (Arrow, 1962).

In our model, we assumed that the opportunity to develop successor drugs can present itself to all firms in a market. Figure 6 provides empirical evidence for this phenomenon by examining the amount of successor revenue that accrues to the same firm that developed the focal drug versus the amount that accrues to the rest of the market. We find that, at any stage of development, half of a drug's future successor revenues accrue to rival firms.

Motivated by these facts, we next investigate the relationship between competition and investments in developing novel drugs. In Proposition 3 of our model, the value of future learning declines as competition increases ($\alpha$ decreases). Empirically, this implies that firms will set a higher revenue threshold for developing novel drugs when competition is high.

To test this hypothesis, we link focal drugs to their primary "research area," which we define as the nexus of their disease target (ICD-9) and mechanism of action (MOA) (for example, statins are designed to treat heart disease by inhibiting the HMG-CoA reductase

enzyme). We define a notion of research area competitiveness by examining the number of pre-clinical drug candidates in this research area that have been developed by other firms over the previous five years. Firms working in research areas with more active competitors may be less able to appropriate spillovers from their own investments.

In Figure 7, we test the predictions of Proposition 3. The regressions accompanying these figures are presented in Table 4. We caution, however, that the standard errors in these regressions are relatively large, so that our estimated coefficients by competition are not statistically different despite the magnitude of their differences being economically meaningful.

Panel A plots the coefficient on the predictive relation between a drug's novelty score and development likelihood, interacted with whether the focal drug's research area has above or below median competition. To understand these results, we first begin by noting that both plotted coefficients are negative. This is consistent with Figure 5 which shows a negative relation between a drug's molecular novelty and its likelihood of entering development. The results in Panel A split up this main effect and show that, when there are more active competitors in the same research area, firms become differentially less willing to invest in bringing novel drugs into development, relative to incremental drugs. In more competitive areas, a one standard deviation increase in novelty decreases a drug's chances of entering development by about 30 percent more than it does in less competitive areas

In Panel B, we examine a second prediction of Proposition 3, which states that approved novel drugs should have disproportionately higher revenues (relative to approved incremental drugs) in more competitive research areas. Specifically, we plot the coefficient on the relation between a drug's novelty score and revenues conditional on approval, interacted by competition. This time, both coefficient values are positive, consistent with Panel B of 5, which shows an overall positive relation between a drug's novelty and its revenues on development. Here, we show that this relation becomes more stark with competition: in more competitive research areas, approved novel drugs generate substantially more direct revenues than approved incremental drugs. The coefficient estimates presented in Column 2 indicate that a one standard deviation increase in novelty increases revenues on approval by more than twice as much as the same increase in novelty in less competitive areas.

A potential concern with interpreting these results is that research areas that see more entry from competitors may simply be more crowded, which may separately impact firms' decisions to invest in incremental versus novel innovation. To account for a research area's overall level of innovative potential, our analysis in both panels of Figure 5 includes a variety

of controls for the total amount of research being conducted in an area, defined several ways: the total number of drug candidates (irrespective of the originating firm) that have ever entered pre-clinical development in the focal drug's ICD-9-MOA to date, as well as the total entry for the entire disease area or mechanism of action. Our specifications also include development time fixed effects. As such, our results should be interpreted as the correlation between competitor research activity and various outcomes, holding constant total activity.

Taken together, these two sets of results provide suggestive evidence that, as competition increases, firms become more reluctant to invest in novel drugs. Relative to incremental projects, novel drugs are less likely to be approved, but generate greater revenues when they are approved. These findings are consistent with firms placing less value on learning when this learning provides more benefit to rival firms.

# 5 Conclusion

If past research informs current work, then firms face a choice. Is it more valuable to learn from exploring new areas, or to exploit the information revealed by past work?

In this paper, we show that despite novel drugs generating more learning, firms prefer to invest in incremental drugs, which are easier to evaluate. This presents a dynamic tension: while firms value the ability to discard incremental drugs that are unlikely to succeed, they are reluctant to make the types of exploratory R&D investments that improve future screening decisions.

In demonstrating these results, our paper makes several contributions that may inform future research.

First, we develop a concrete measure of knowledge spillovers that can be applied to projects regardless of whether they are successful. It is widely acknowledged that researchers learn from failure but, in many contexts, these spillovers are difficult to capture with existing measures. Patent citations, for instance, may understate the value of failed projects either because such projects are never patented or because the patents are under-cited. Our spillover measures rely on physical properties rather than discretionary citations and allow us to generate a conservative measure of the value of failed R&D in an important setting, pharmaceuticals.[27] More broadly, our approach of linking products over time using their physical evolution could

[27] Our measures of successors and successor revenue do not capture the additional value that failed projects can generate by teaching firms to abandon an unfruitful research area.

be applied in other settings where researchers observe prototypes (for example, using product design images embedded in patent documents).

Second, our model provides a new way of understanding observed differences between novel and incremental projects. In workhorse models of innovation, e.g., Manso (2011) or Akcigit and Kerr (2018), novel and incremental projects are assumed to differ in terms of their fundamental risk or returns. Our model shows that these assumed differences can instead emerge endogenously as a result of selection. Perhaps counterintuitively, the fact that novel drugs have higher expected returns can arise as a direct consequence of firms' reluctance to invest in exploration.

Finally, our results suggest that policymakers could provide stronger incentives for firms to develop novel drugs by making it easier for them to appropriate indirect revenues from potential successors. These revenues, which can exist even when the focal project fails, disproportionately increase the overall returns to investing in newer, more uncertain research areas. Such policies may include allowing drug makers to make more "reach through claims" in patents covering novel drug compounds, or modifying clinical trail disclosure rules such that novel drug developers can delay the flow of information across firm boundaries. Our work provides motivation to better understand the potential costs and benefits of such policies.

# References


Aghion, P., N. Bloom, R. Blundell, R. Griffith, and P. Howitt (2005, 05). Competition and Innovation: an Inverted-U Relationship*. *The Quarterly Journal of Economics 120*(2), 701–728.

Aghion, P. and P. Howitt (1992). A model of growth through creative destruction. *Econometrica 60*(2), 323–351.

Aghion, P. and J. Tirole (1994, 11). The Management of Innovation. *The Quarterly Journal of Economics 109*(4), 1185–1209.

Akcigit, U. and W. R. Kerr (2018). Growth through heterogeneous innovations. *Journal of Political Economy 126*(4), 1374–1443.

Arrow, K. (1962). Economic Welfare and the Allocation of Resources for Invention. In *The Rate and Direction of Inventive Activity: Economic and Social Factors*, NBER Chapters, pp. 609–626. National Bureau of Economic Research, Inc.

Aryal, G., F. Ciliberto, L. E. Farmer, and E. Khmelnitskaya (2023). Valuing pharmaceutical drug innovations.

Azoulay, P., J. S. Graff Zivin, D. Li, and B. N. Sampat (2018). Public r&d investments and private-sector patenting: Evidence from nih funding rules. *The Review of Economic Studies*, rdy034.

Azoulay, P., J. S. Graff Zivin, and G. Manso (2011a). Incentives and creativity: evidence from the academic life sciences. *The RAND Journal of Economics 42*(3), 527–554.

Azoulay, P., J. S. Graff Zivin, and G. Manso (2011b). Incentives and creativity: evidence from the academic life sciences. *The RAND Journal of Economics 42*(3), 527–554.

Balsmeier, B., L. Fleming, and G. Manso (2017). Independent boards and innovation. *Journal of Financial Economics 123*(3), 536–557.

Baum, J. A. and K. B. Dahlin (2007). Aspiration performance and railroads' patterns of learning from train wrecks and crashes. *Organization Science 18*(3), 368–385.

Bennett, V. M. and J. Snyder (2017). The empirics of learning from failure. *Strategy Science 2*(1), 1–12.

Bergemann, D. and J. Valimaki (2006). Bandit problems. Cowles Foundation Discussion Papers 1551, Cowles Foundation for Research in Economics, Yale University.

Bickerton, G. R., G. V. Paolini, J. Besnard, S. Muresan, and A. L. Hopkins (2012, 02). Quantifying the chemical beauty of drugs. *Nat Chem 4*(2), 90–98.

Callander, S. (2011, October). Searching and learning by trial and error. *American Economic Review 101*(6), 2277–2308.

Callander, S., N. Lambert, and N. Matouschek (2022). Innovation and competition on a rugged technological landscape.

Callander, S. and N. Matouschek (2019, February). The risk of failure: Trial and error learning and long-run performance. *American Economic Journal: Microeconomics 11*(1), 44–78.

Chiou, J.-Y., L. Magazzini, F. Pammolli, and M. Riccaboni (2016). Learning from successes and failures in pharmaceutical r&d. *Journal of Evolutionary Economics 26*, 271–290.

Desai, V. M., D. Maslach, and P. Madsen (2020, 01). 109Organizational Learning From Failure: Present Theory and Future Inquiries. In *The Oxford Handbook of Group and Organizational Learning*. Oxford University Press.

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

Dixit, A. K. and J. E. Stiglitz (1977). Monopolistic competition and optimum product diversity. *The American Economic Review 67*(3), 297–308.

Endo, A. (2010). A historical perspective on the discovery of statins. *Proceedings of the Japan Academy, Series B 86*(5), 484–493.

Ferreira, D., G. Manso, and A. Silva (2014). Incentives to innovate and the decision to go public or private. *Review of Financial Studies 27*(1), 256–300.

Furman, J. L. and S. Stern (2011, August). Climbing atop the shoulders of giants: The impact of institutions on cumulative research. *American Economic Review 101*(5), 1933–63.

Garfagnini, U. and B. Strulovici (2016). Social Experimentation with Interdependent and Expanding Technologies. *Review of Economic Studies 83*(4), 1579–1613.

Harris, J. E. (2021, March). The repeated setbacks of hiv vaccine development laid the groundwork for sars-cov-2 vaccines. Working Paper 28587, National Bureau of Economic Research.

Haunschild, P. R. and B. N. Sullivan (2002). Learning from complexity: Effects of prior accidents and incidents on airlines' learning. *Administrative science quarterly 47*(4), 609–643.

Hughes, J. P., S. Rees, S. B. Kalindjian, and K. L. Philpott (2011). Principles of early drug discovery. *British journal of pharmacology 162*(6), 1239–1249.

Hvide, H. K. and G. A. Panos (2014). Risk tolerance and entrepreneurship. *Journal of Financial Economics 111*(1), 200–223.

Kelly, B., D. Papanikolaou, A. Seru, and M. Taddy (2021, September). Measuring technological innovation over the long run. *American Economic Review: Insights 3*(3), 303–20.

Krieger, J., D. Li, and D. Papanikolaou (2021, 03). Missing Novelty in Drug Development*. *The Review of Financial Studies*. hhab024.

Krieger, J. L. (2021). Trials and terminations: Learning from competitors' r&d failures. *Management Science*.

Kuhn, J. M. and N. C. Thompson (2019). How to measure and draw causal inferences with patent scope. *International Journal of the Economics of Business 26*(1), 5–38.

Lei, Z. and B. D. Wright (2017). Why weak patents? testing the examiner ignorance hypothesis. *Journal of Public Economics 148*, 43 – 56.

Madsen, P. M. and V. Desai (2010). Failing to learn? the effects of failure and success on organizational learning in the global orbital launch vehicle industry. *Academy of management journal 53*(3), 451–476.

Magazzini, L., F. Pammolli, and M. Riccaboni (2012). Learning from failures or failing to learn? lessons from pharmaceutical r&d. *European Management Review 9*(1), 45–58.

Maggiora, G., M. Vogt, D. Stumpfe, and J. Bajorath (2014, 04). Molecular similarity in medicinal chemistry. *Journal of Medicinal Chemistry 57*(8), 3186–3204.

Manso, G. (2011). Motivating innovation. *The Journal of Finance 66*(5), 1823–1860.

Maslach, D. (2016). Change and persistence with failed technological innovation. *Strategic Management Journal 37*(4), 714–723.

Murray, F., P. Aghion, M. Dewatripont, J. Kolev, and S. Stern (2016, February). Of mice and academics: Examining the effect of openness on innovation. *American Economic Journal: Economic Policy 8*(1), 212–52.

Murray, F. and S. Stern (2007). Do formal intellectual property rights hinder the free flow of scientific knowledge?: An empirical test of the anti-commons hypothesis. *Journal of Economic Behavior and Organization 63*(4), 648–687. Academic Science and Entrepreneurship: Dual engines of growth.

Qiu, J. and C. Wan (2015). Technology spillovers and corporate cash holdings. *Journal of Financial Economics 115*(3), 558–573.

Romer, P. M. (1990). Endogenous technological change. *Journal of Political Economy 98*(5, Part 2), S71–S102.

Sampat, B. and H. L. Williams (2019, January). How do patents affect follow-on innovation? evidence from the human genome. *American Economic Review 109*(1), 203–36.

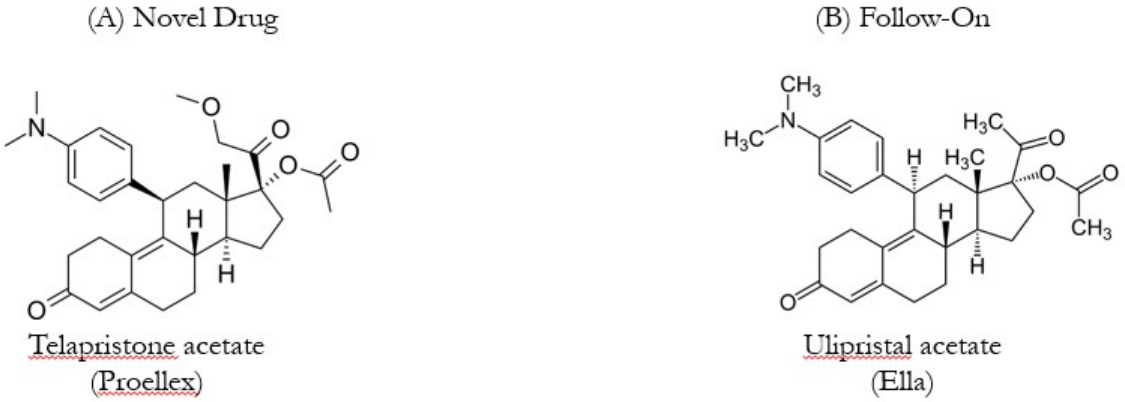
Scotchmer, S. (1991, March). Standing on the shoulders of giants: Cumulative research and the patent law. *Journal of Economic Perspectives 5*(1), 29–41.

Tian, X. and T. Y. Wang (2011, 12). Tolerance for Failure and Corporate Innovation. *The Review of Financial Studies 27*(1), 211–255.

Wawer, M. J., K. Li, S. M. Gustafsdottir, V. Ljosa, N. E. Bodycombe, M. A. Marton, K. L. Sokolnicki, M.-A. Bray, M. M. Kemp, E. Winchester, B. Taylor, G. B. Grant, C. S.-Y. Hon, J. R. Duvall, J. A. Wilson, J. A. Bittker, V. Dančík, R. Narayan, A. Subramanian, W. Winckler, T. R. Golub, A. E. Carpenter, A. F. Shamji, S. L. Schreiber, and P. A. Clemons (2014, July). Toward performance-diverse small-molecule libraries for cell-based phenotypic screening using multiplexed high-dimensional profiling. *Proceedings of the National Academy of Sciences 111*(30), 10911–10916.

Williams, H. L. (2013). Intellectual property rights and innovation: Evidence from the human genome. *Journal of Political Economy 121*(1), 1–27.

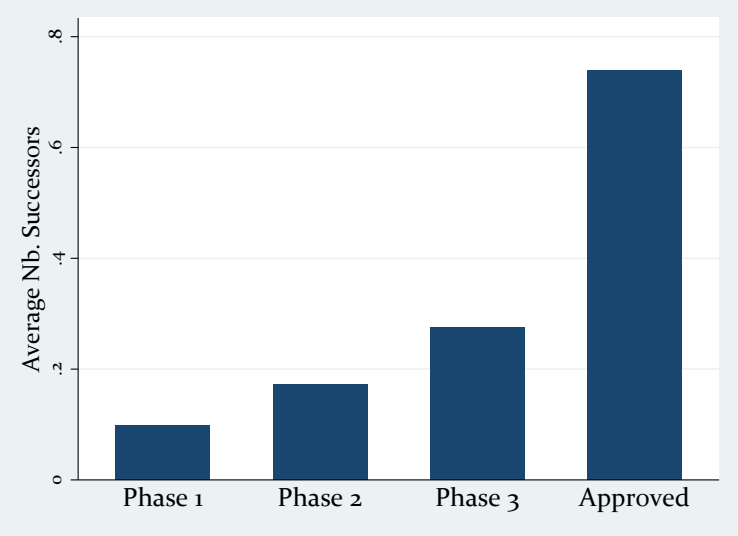
FIGURE 1: EXAMPLE OF NOVEL DRUG AND FOLLOW-ON



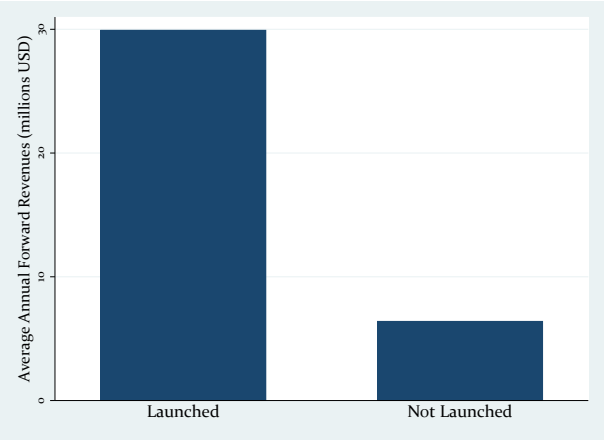

NOTES: This figure displays the molecular structure of two drugs that are linked in the data. The drug on the left (Panel A), telapristone acetate (Proellex), is a molecularly novel compound that entered clinical trials for the treatment of uterine fibriods in 2004, sponsored by the biotech company Repros Therapeutics. Several years later, during Phase 3 trials, development was halted due to patients experiencing liver toxicity issues. Despite this, Proellex inspired the development of 5 successor drugs. One of those drugs, ulipristal acetate (Ella) is pictured in Panel B, and has a Tanimoto similarity of 0.81 to Proellex. Ella was developed by a different drug company, HRA Pharma and was approved in 2010. It is currently on the World Health Organization's list of essential medicines.

FIGURE 2: SUCCESSOR REVENUE, BY DEVELOPMENT OUTCOMES

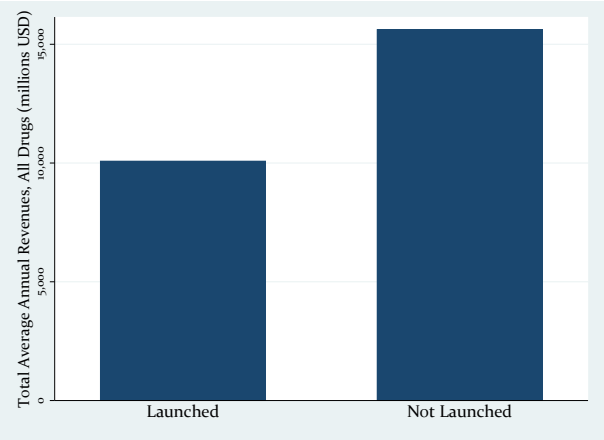
A. # Successors, by Focal Drug's Highest Development Phase




B. Average Successor Revenues

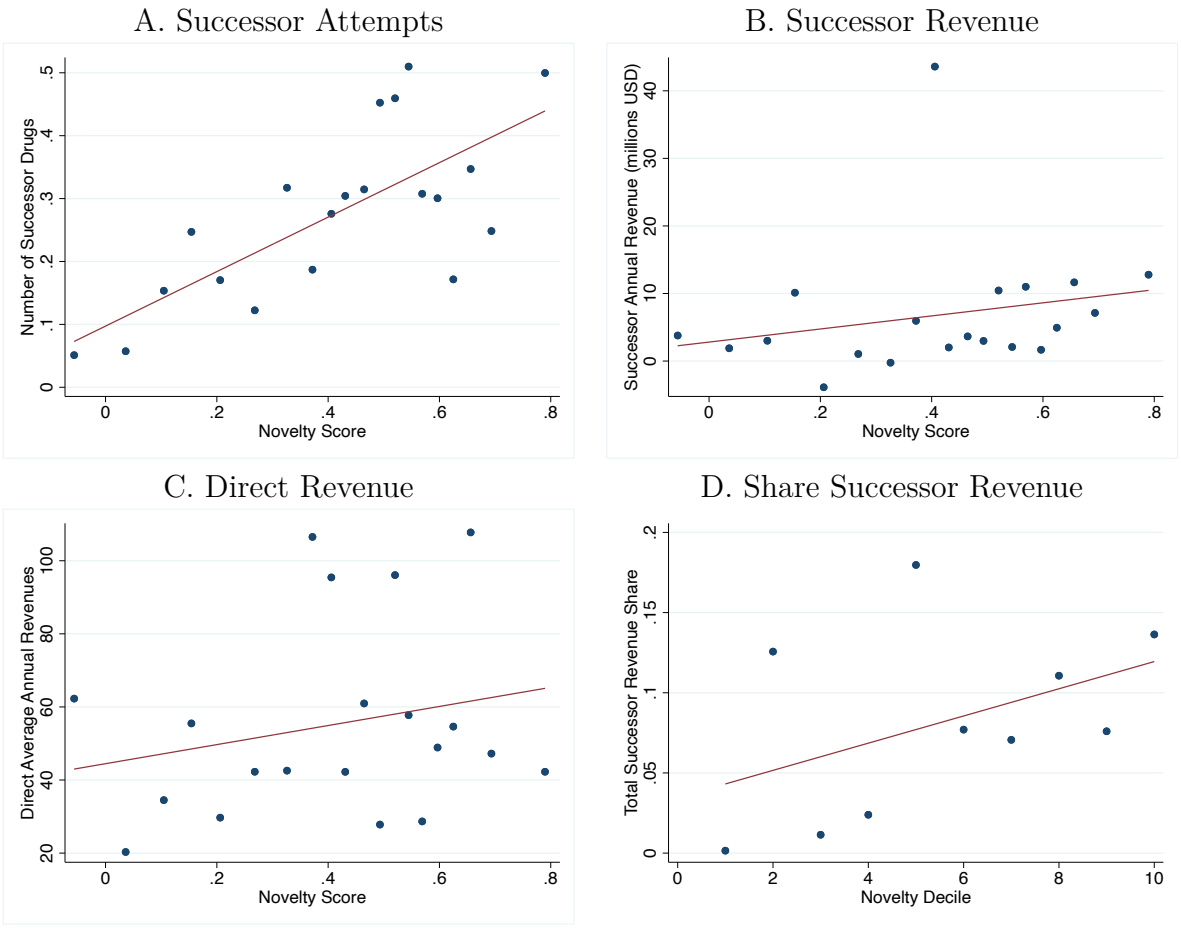



C. Aggregate Successor Revenues




NOTES: Figure 2 plots information on successor drugs and successor revenues for drug candidates in our data. Panel A shows the average number of successor drug attempts for drugs by their highest stage of development reached (e.g., phase 1 drugs are those that never graduated to phase 2 and beyond). Panel B plots average successor revenues per drug. Panel C provides the same information, except aggregated over all drugs in our sample. Successor revenues are defined as the sum of average annual revenues across all successor drugs to a given focal drug. Panels B and C present raw drug and revenue counts, and have not been adjusted for cohort differences. The sample includes drugs that enter clinical development in the United States.

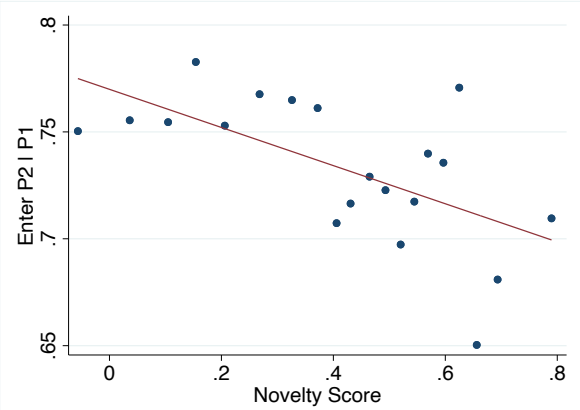
FIGURE 3: SUCCESSOR REVENUES, BY NOVELTY



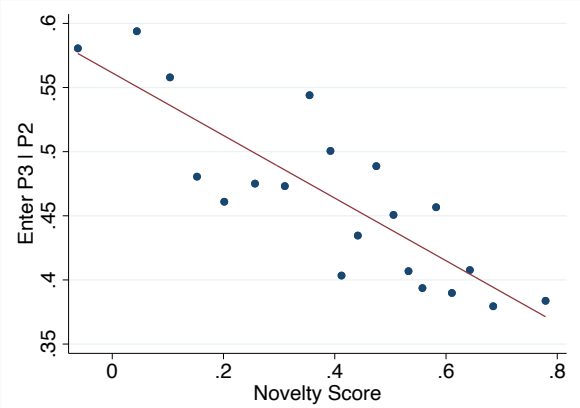

NOTES: Panels A-C of Figure 3 presents binned scatterplots of the relationship between a drug candidate's novelty and measures of successor activity and revenues, at the individual drug level. All specifications include controls for a focal drug's quarter of development and lead disease indication. The sample includes drugs that enter clinical development in the United States. To account for the fact that many drugs have no direct or successor revenues (so that their successor revenue share would be undefined), Panel D plots successor revenue shares aggregated by deciles of novelty. To compute this, we first residualize novelty by fixed effects for the drug's quarter of development and lead disease indication and then calculate the ratio of successor to successor plus direct revenues across all drugs that fall in each novelty decile. Table 1 presents accompanying regressions for Panels A-C, as well as for drug-level shares of successor revenues.

FIGURE 4: PROGRESSION THROUGH DEVELOPMENT, BY NOVELTY

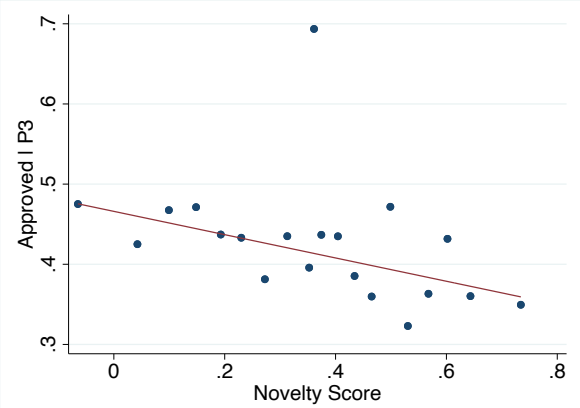
A. Phase 2, conditional on Phase 1




B. Phase 3, conditional on Phase 2



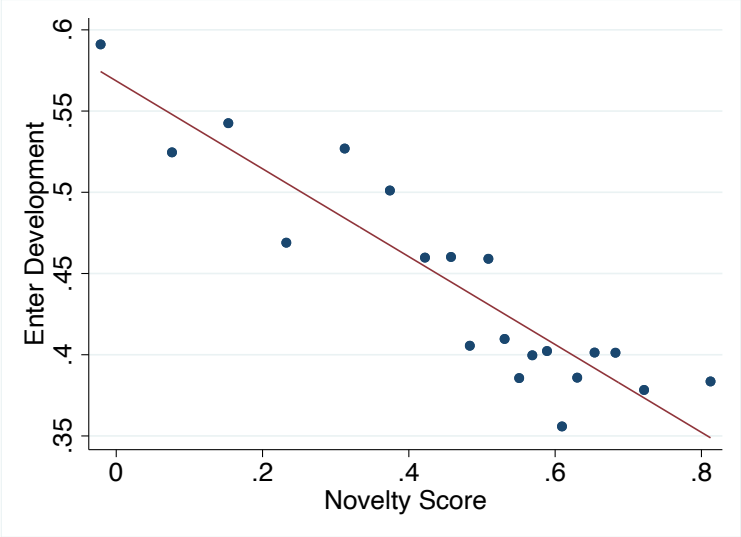

C. Approval, conditional on Phase 3



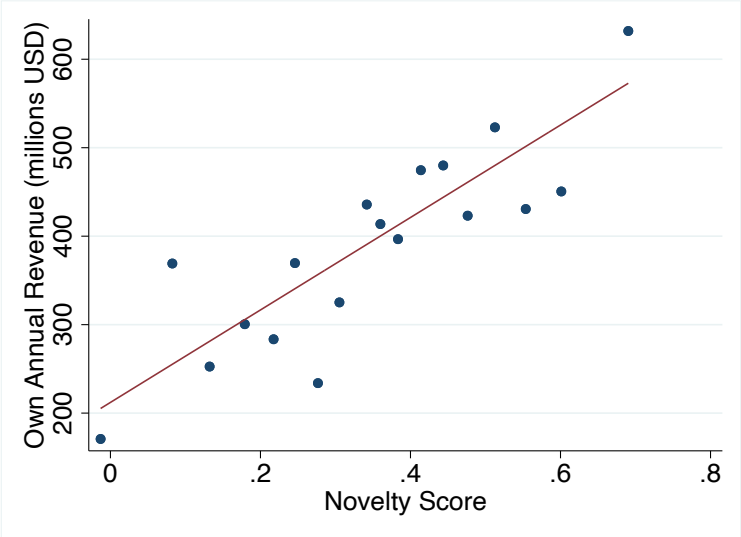

NOTES: Figure 4 present a test of Proposition 1. Each panel presents a binned scatter plot of the relationship between a drug candidate's molecular novelty and measures of its progression in clinical trials. All specifications include controls for quarter of development and disease indication. The sample includes drugs that enter clinical development in the United States. Accompanying regression estimates are presented in Table 2.

FIGURE 5: ENTRY INTO DEVELOPMENT AND REVENUES ON SUCCESS, BY PROJECT NOVELTY

A. Development Probability (Likelihood of Clinical Trial Entry)



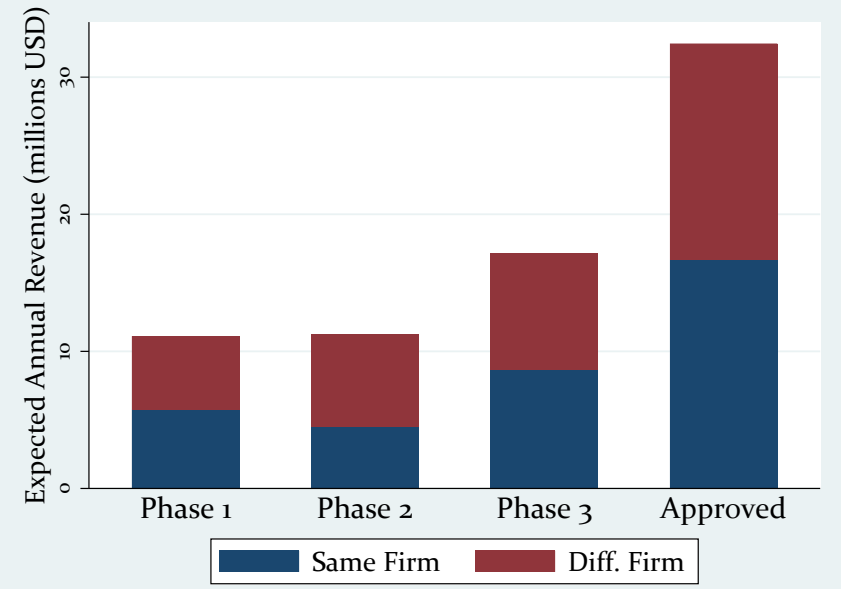

B. Revenue Conditional on Approval




NOTES: Figure 5 provides an empirical analysis of Proposition 2. Panel A presents binned scatterplots of the relationship between a focal drug's novelty and a firm's decision to invest in clinical drug development. Here, the sample is all drugs that are observed in Cortellis data for US pre-clinical development. In Panel B, we present binned scatterplots of the relation between novelty and average annual direct revenues, for the set of drugs that are approved in the US. All specifications include controls for quarter of development as well as for disease area FEs. The corresponding regression estimates are presented in Columns 2 and 4 of Table 3.

FIGURE 6: SUCCESSOR REVENUE, BY FIRM

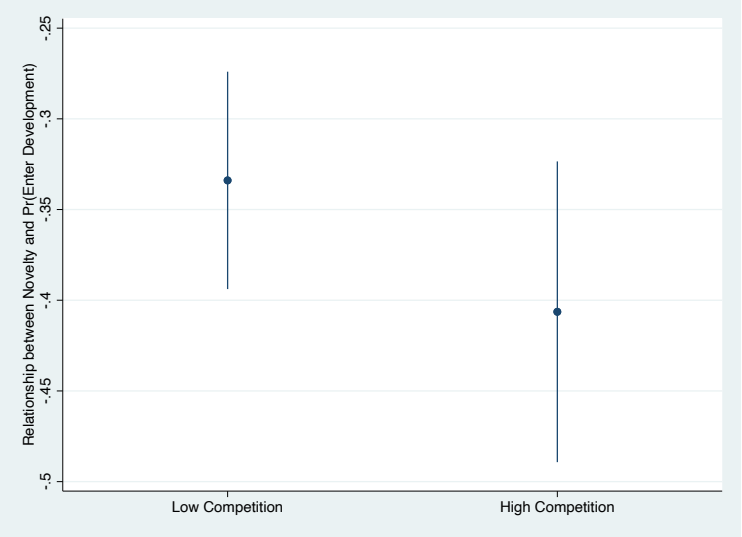
A. Average Successor Revenues, by Phase



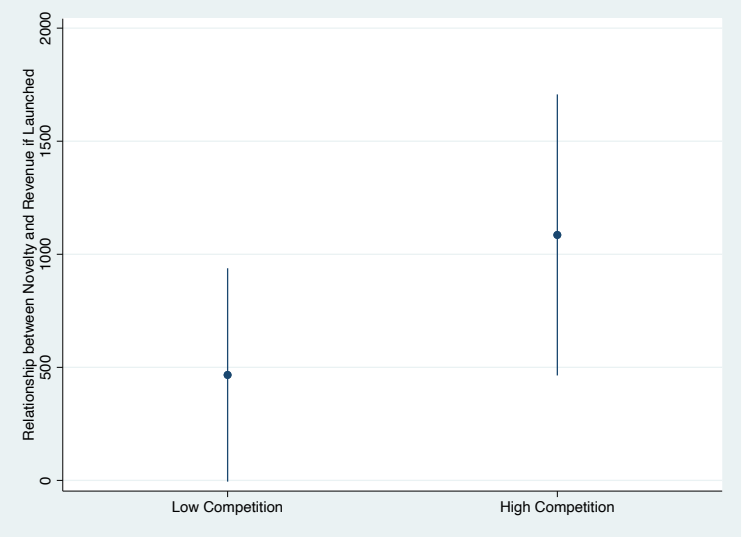

NOTES: Figure 6 plots the expected annual successor revenues, by their highest stage of development reached. The blue portion of each bar represents the average annual successor revenues for drugs developed by the same firm as the focal drug, while the red portion reflects successor drugs developed by different firms. Successor revenues are defined as the sum of average annual revenues across all successor drugs to a given focal drug. These graphs present raw drug and revenue counts, and have not been adjusted for cohort differences. The sample includes drugs that enter clinical development in the United States.

FIGURE 7: EXPECTED COMPETITION AND RELATIONSHIP BETWEEN NOVELTY AND INVESTMENT THRESHOLDS

A. Drug Novelty and Development Decision, By Competition




B. Drug Novelty and Revenue Conditional on Development, By Competition




NOTES: Panel A plots the estimated relationship between a focal drug's novelty and a firm's decision to invest in clinical drug development, interacted with indicators for whether research competition is high or low in the focal drug's area. Panel B plots the analogous relationship between novelty and revenue for successfully developed drugs. Research area competition is measured by the number of new drugs developed by competitor firms over the previous 5 years. To ensure that we are not identifying differences driven by the overall amount of research in an area, we control for various measures of total research activity in a research area. All specification include controls for quarter of development and disease area FEs. The corresponding regression estimates are presented in Columns 2 and 4 of Table 4.

Table 1: Drug Novelty: Successors (Number and Annual Revenue)

| VARIABLES | (1) # Successors | (2) # Successors | (3) Successor Revenue | (4) Successor Revenue | (5) Direct Revenue. | (6) Direct Revenue. | (7) Successor Rev. Share | (8) Successor Rev. Share |
|---|---|---|---|---|---|---|---|---|
| Novelty Score | 0.360*** | 0.433*** | 9.938* | 9.683* | 32.54** | 26.10 | 0.106*** | 0.128*** |
| | (0.0721) | (0.0955) | (5.948) | (5.734) | (15.38) | (17.09) | (0.0287) | (0.0375) |
| Observations | 3,913 | 3,844 | 3,913 | 3,844 | 3,913 | 3,844 | 843 | 781 |
| R-squared | 0.077 | 0.117 | 0.037 | 0.056 | 0.069 | 0.143 | 0.210 | 0.379 |
| Drug Cohort Year-Qtr FE | YES | YES | YES | YES | YES | YES | YES | YES |
| Lead Indication FE | | YES | | YES | | YES | | YES |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: Table 1 presents the relationship between a drug candidate's novelty and the number of follow-on successors, successor revenue, direct revenue, and share of direct revenue. All revenue figures, both direct and successor, are defined as average annual revenues for the time that the approved drug appears in Evaluate Pharma. The sample consists of all drug candidates that enter Phase 1 development in the United States.

Table 2: Drug Novelty: Progression by Phase

| VARIABLES | (1) Phase 1–2 | (2) Phase 1–2 | (3) Phase 2–3 | (4) Phase 2–3 | (5) P3–Launched | (6) P3–Launched |
|---|---|---|---|---|---|---|
| Novelty Score | -0.137*** | -0.0892*** | -0.302*** | -0.244*** | -0.178*** | -0.145** |
| | (0.0271) | (0.0296) | (0.0360) | (0.0391) | (0.0549) | (0.0623) |
| Observations | 3,913 | 3,844 | 2,871 | 2,802 | 1,329 | 1,266 |
| R-squared | 0.111 | 0.181 | 0.120 | 0.226 | 0.161 | 0.331 |
| Drug Cohort Year-Qtr FE | YES | YES | YES | YES | YES | YES |
| Lead Indication FE | | YES | | YES | | YES |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: Table 2 presents the relationship between a drug candidate's novelty and its probability of progressing through clinical trials (Phase 1 to Phase 2; Phase 2 to Phase 3; Phase 3 to approval). The sample consists of all drug candidates that enter Phase 1 development in the United States. All regressions include drug year-quarter of development fixed effects. Columns 2, 4, and 6 additionally include fixed effects for the first indication (ICD-9) for which the drug entered development.

Table 3: Development Thresholds, by Novelty

| VARIABLES | (1) Entered Phase 1 | (2) Entered Phase 1 | (3) Revenue if Approved | (4) Revenue if Approved |
|---|---|---|---|---|
| Novelty Score | -0.363*** | -0.270*** | 692.7*** | 515.2*** |
| | (0.0208) | (0.0221) | (117.3) | (149.5) |
| Observations | 9,451 | 9,384 | 551 | 490 |
| R-squared | 0.103 | 0.186 | 0.195 | 0.379 |
| Drug Cohort Year-Qtr FE | YES | YES | YES | YES |
| Lead Indication FE | | YES | | YES |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: Columns 1 and 2 show the relationship between drug novelty and the firm's development decision, as defined by the drug entering Phase 1 clinical trials. The sample in Columns 1 and 2 consists of all drugs that are observed in Cortellis data for US pre-clinical development. In Panel B, we present binned scatterplots of the relation between novelty and average annual direct revenues, for the subset of drugs that are approved in the US. All specifications include drug development cohort fixed effects, and Columns 2 and 4 additionally include fixed effects for the lead indication (first disease ICD-9 for which the drug was developed).

Table 4: Development Thresholds, by Competition

| VARIABLES | (1) Enter Phase 1 | (2) Enter Phase 1 | (3) Revenue if Approved | (4) Revenue if Approved |
|---|---|---|---|---|
| Novelty x Low Research Comp. | -0.334*** | -0.307*** | 466.4 | 453.3 |
| | (0.0364) | (0.0412) | (285.6) | (318.7) |
| Novelty x High Research Comp. | -0.406*** | -0.368*** | 1,086*** | 1,044** |
| | (0.0504) | (0.0552) | (376.2) | (405.4) |
| Total Research Area Entry | -0.00142*** | -0.00152** | 2.422 | 2.267 |
| | (0.000515) | (0.000602) | (3.618) | (4.784) |
| Observations | 6,702 | 6,702 | 348 | 348 |
| R-squared | 0.200 | 0.213 | 0.445 | 0.458 |
| Drug Cohort Year-Qtr FE | YES | YES | YES | YES |
| Lead Indication FE | YES | YES | YES | YES |
| Addl Research Area Controls | | YES | | YES |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: Table 4 is structured analogously to Table 3, but with additional interactions related to a drug's "research area." We define a drug's research area based on its lead disease indication (ICD-9) as well as its method of molecular action (MOA). Research competition is defined at the drug-company level. Low competition research areas are ones in which the number of new preclinical drugs introduced in that area by other firms is below the median in our sample. High competition areas are defined analogously. Total amount of research activity in an area is measured as the number of new drugs that have entered preclinical development to date in that research area. Additional research area controls include total entry into a drug's disease area and method of action area, defined separately.

# APPENDIX

Appendix Table A.1: Novelty and Successor Value, Alternative Definitions of Novelty

| VARIABLES | (1) # Successors | (2) Successor Revenue | (3) Successor Rev. Share | (4) # Successors | (5) Successor Revenue | (6) Successor Rev. Share |
|---|---|---|---|---|---|---|
| Pioneer Drug in Target | 0.901*** | 92.12** | 0.0943** | | | |
| | (0.256) | (41.94) | (0.0422) | | | |
| Pioneer Drug in Target-Action | | | | 0.388*** | 31.57** | 0.0577** |
| | | | | (0.109) | (16.01) | (0.0267) |
| Observations | 3,321 | 3,321 | 781 | 3,321 | 3,321 | 781 |
| R-squared | 0.166 | 0.132 | 0.381 | 0.160 | 0.127 | 0.379 |
| Drug Cohort Year-Qtr FE | YES | YES | YES | YES | YES | YES |
| Phase I Year-Qtr FE | YES | YES | YES | YES | YES | YES |
| Lead Indication FE | YES | YES | YES | YES | YES | YES |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: Appendix Table A.1 presents the relationship between a drug candidate's novelty (defined in two alternative ways) and the number of follow-on successors, successor revenue, and share of direct revenue. All revenue figures, both direct and successor, are defined as average annual revenues for the time that the approved drug appears in Evaluate Pharma. The sample consists of all drug candidates that enter Phase 1 development in the United States. Pioneer Drug in Target is an indicator for whether a drug attempts to impact a new biological object (e.g. a protein) while Pioneer Drug in Target-Action is an indicator for whether a drug is the first to attempt to impact a biological object in a particular way (e.g. inhibit the expression of a protein).

Appendix Table A.2: Drug Novelty: Progression by Phase—excluding successors of launched drugs

| VARIABLES | (1) Phase 1–2 | (2) Phase 1–2 | (3) Phase 2–3 | (4) Phase 2–3 | (5) P3–Launched | (6) P3–Launched |
|---|---|---|---|---|---|---|
| Novelty Score | -0.116*** | -0.0679 | -0.154*** | -0.120** | -0.0749 | -0.0869 |
| | (0.0412) | (0.0449) | (0.0541) | (0.0584) | (0.0870) | (0.0970) |
| Observations | 3,094 | 3,028 | 2,219 | 2,150 | 929 | 866 |
| R-squared | 0.117 | 0.190 | 0.119 | 0.229 | 0.201 | 0.385 |
| Drug Cohort Year-Qtr FE | YES | YES | YES | YES | YES | YES |
| Lead Indication FE | | YES | | YES | | YES |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: Table A.2 presents the relationship between a drug candidate's novelty and its probability of progressing through clinical trials (Phase 1 to Phase 2; Phase 2 to Phase 3; Phase 3 to approval). The sample consists of all drug candidates that enter Phase 1 development in the United States, excluding incremental drugs that are associated with approved prior drugs. All regressions include drug year-quarter of development fixed effects. Columns 2, 4, and 6 additionally include fixed effects for the first indication (ICD-9) for which the drug entered development.

Appendix Table A.3: Development Thresholds by Novelty—excluding successors of launched drugs

| VARIABLES | (1) Entered Phase 1 | (2) Entered Phase 1 | (3) Revenue if Approved | (4) Revenue if Approved |
|---|---|---|---|---|
| Novelty Score | -0.335*** | -0.237*** | 709.8*** | 231.9 |
| | (0.0304) | (0.0315) | (245.4) | (375.8) |
| Observations | 8,070 | 7,998 | 360 | 293 |
| R-squared | 0.084 | 0.172 | 0.199 | 0.469 |
| Drug Cohort Year-Qtr FE | YES | YES | YES | YES |
| Lead Indication FE | | YES | | YES |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: Appendix Table A.3 is analogous to Table 3, but excluding outcomes for incremental drugs associated with successfully marketed predecessors. Columns 1 and 2 show the relationship between drug novelty and the firm's development decision, as defined by the drug entering Phase 1 clinical trials. In Panel B, we present binned scatterplots of the relation between novelty and average annual direct revenues, for the subset of drugs that are approved in the US. All specifications include drug development cohort fixed effects, and Columns 2 and 4 additionally include fixed effects for the lead indication (first disease ICD-9 for which the drug was developed).

# A Additional details on model

## A.1 Decomposition of firm's expected profits in Expression (6)

Consider a drug (in the first or second period) that has been revealed to have revenue $\tilde{R}$ ($R$ or $R_j$), and for which the firm believes that the probability of success ($S$ or $S_j$) is $\tilde{\pi}$. Let the direct payoff within that period from making the initial development decision $\tilde{D}^1$ ($D^1$ or $D_j^1$), gross of the development cost $C^1$, be given by $V(\tilde{R}, \tilde{\pi})$.

If there are $M = 1$ development phases, then the firm makes no further decisions about whether to develop the drug before the success or failure is revealed. In this case, $V(\tilde{R}, \tilde{\pi}) = \tilde{\pi}\tilde{R}$.

If there are $M \geq 2$ development phases, then the firm makes additional decisions of whether to proceed. In particular, the firm optimally proceeds with the next round of development, choosing $\tilde{D}^2$ ($D^2$ or $D_j^2$) equal to 1 after observing signal $\tilde{\sigma}^1$ ($\sigma^1$ or $\sigma_j^1$) equal to $g$, if and only if it plans on proceeding with every other development phase conditional on continued positive signals $g$. So there is essentially only a single decision to make, over $\tilde{D}^2$ conditional on $\tilde{\sigma}^1 = g$. We can therefore write $V$ (for $M = 1$ or $M \geq 2$) as

$$V(\tilde{R}, \tilde{\pi}) \equiv \max_{\tilde{D}^2 \in \{0,1\}} \tilde{D}^2 \cdot \left( \tilde{\pi}(\tilde{R} - \sum_{t=1}^{M-1} C^{t+1}) - (1 - \tilde{\pi}) \sum_{t=1}^{M-1} \left( \prod_{s=1}^{t} (1 - q^t) \right) C^{t+1} \right). \tag{9}$$

**Lemma 1.** *$V$ satisfies the following properties.*

1. *For any $\tilde{\pi} \in (0,1)$ and $\tilde{R} \geq 0$, it holds that $V(\tilde{R}, \tilde{\pi}) \leq \tilde{R}$.*
2. *For any $\tilde{\pi} \in (0,1)$ and $\tilde{R} \geq 0$, it holds that $V(\tilde{R}, \tilde{\pi}) > C^1$ if $\tilde{R} > \sum_{t=1}^{M} C^t / \tilde{\pi}$.*
3. *For any $\tilde{\pi} \in (0,1)$, it holds that $V(\tilde{R}, \tilde{\pi})$ is weakly increasing in $\tilde{R}$, with $V(\tilde{R}, \tilde{\pi})$ strictly increasing in $\tilde{R}$ if $V(\tilde{R}, \tilde{\pi}) > 0$.*
4. *For any $\tilde{R} \in (0,1)$, it holds that $V(\tilde{R}, \tilde{\pi})$ is convex and weakly increasing in $\tilde{\pi}$, with $V(\tilde{R}, \tilde{\pi})$ strictly increasing in $\tilde{\pi}$ if $V(\tilde{R}, \tilde{\pi}) > 0$.*

**Proof of Lemma 1.** When $M = 1$ and $V_1^N = \pi^N R$, all of these properties are immediate. So, suppose that $M \geq 2$, in which $V_1^N$ is given by (9). We now show the desired properties.

1. Immediate from (9).

2. $V_1^N$ is bounded below by the expression in (9) with the max over $\tilde{D}^2$ replaced by $\tilde{D}^2 = 1$. The expression with $\tilde{D}^2 = 1$ is in turn bounded below by $\tilde{\pi}R - \sum_{t=1}^{M-1} C^{t+1}$.

3. The maximand of (9) is weakly increasing in $\tilde{R}$ for each $\tilde{D}^2$, and is strictly increasing if $\tilde{D}^2 = 1$. Moreover, if $V(\tilde{R}, \tilde{\pi}) > 0$, it must be that the expression is maximized by $D^2 = 1$.

4. The maximand of (9) is constant in $\tilde{\pi}$ if $\tilde{D}^2 = 0$. If this expression is maximized by $\tilde{D}^2 = 1$, then it must be that $\tilde{R} - \sum_{t=1}^{M-1} C^{t+1} > 0$, in which case the expression is strictly increasing, and linear, in $\tilde{\pi}$. The maximized expression is convex because it is a maximum over a constant function and a linear function.

■

We next observe that $V(\tilde{R}, \tilde{\pi})$ can be used to derive the terms in decomposition (6) of $V_1^N(R)$, $W_2^N$, and $\Delta_2^{N=1}$.

For the decision in period 1, let $\pi_1^N$ be the belief on success at the time the firm faces the initial development choice $D^1$. We have that $\pi_1^{N=1} = \pi$ and $\pi_1^{N=0} = \pi/(\pi + (1-\pi)(1-q^0))$, with $\pi_1^{N=0} > \pi_1^{N=1}$. With this notation, we can now write

$$V_1^N(R) = V(R, \pi_1^N). \tag{10}$$

Note that part 1 of Lemma 1 implies that, for each $N$, there are values of $R$ in the support of $F_R$ for which $V_1^N < C^1$. Part 2 implies that, for each $N$, there are values of $R$ in the support of $F_R$ for which $V_1^N > C^1$. Part 4 confirms the claim in the body of the paper that $V_1^N(R)$ is higher for $N = 0$ (recalling that $\pi_1^{N=0} > \pi_1^{N=1}$).

Continuing on, we have $W_2^{N=0} = 0$, as the second-period payoff conditional on $N = 0$ was normalized to 0. The term $W_2^{N=1}$ then gives the expected discounted second-period payoff for $N = 1$ when $D^1 = 0$:

$$W_2^{N=1} = \mu_Q \alpha \mathbb{E}_{\tilde{R} \sim F_{2R}} \max\{V(\tilde{R}, \pi_2) - C^1, 0\} \tag{11}$$

Finally, to determine the payoff for $N = 1$ when $D^1 = 1$, observe that for each successor drug that arrives, the initial signal takes the probability of success from $\pi_2$ to either $\tilde{\pi} = 0$ with probability $(1 - \pi_2)q^0$, or $\tilde{\pi} = \pi_2/(\pi_2 + (1-\pi_2)(1-q^0)) > \pi_2$ with probability $(\pi_2 + (1-\pi_2)(1-q^0))$. In the former case, the payoff for that successor drug will be 0; in the

latter case, it may be positive. So the expected payoff at the second period when $N = 1$ and $D^1 = 1$ is

$$\beta\mu_Q\alpha(\pi_2 + (1-\pi_2)(1-q^0))\mathbb{E}_{\tilde{R}\sim F_{2R}} \max\{V(\tilde{R}, \pi_2/(\pi_2 + (1-\pi_2)(1-q^0))) - C^1, 0\}. \quad (12)$$

Hence, $\Delta_2^{N=1}$ is equal to expression (12) minus $W_2^{N=1}$, all divided by coefficient $\beta\mu_Q\alpha$. We have that $\Delta_2^{N=1}$ is positive because better information – a mean-preserving spread of the belief $\tilde{\pi}$ – improves decisionmaking; this can be seen in the convexity of $V(\tilde{R}, \tilde{\pi})$ in $\pi$ from Lemma 1 part 4. The payoff is strictly positive because $F_{2R}$ has large enough support that $V(\tilde{R}, \pi_2/(\pi_2 + (1-\pi_2)(1-q^0)))$ is above $C^1$ with positive probability, and therefore the expectation of the maximand in (12) is positive; and because, for each $\tilde{R}$, it holds that $V(\tilde{R}, 0) = 0 < C^1$. Those facts imply that $\mathbb{E}_{\tilde{R}\sim F_{2R}} \max\{V(\tilde{R}, \tilde{\pi}) - C^1, 0\}$ is convex and is not linear over $\tilde{\pi} \in [0, \pi_2/(\pi_2 + (1-\pi_2)(1-q^0))]$, so the mean-preserving spread of beliefs from an interior point $\pi_2$ to these edges has strictly positive value.

## A.2 The firm's strategy in period 1

The following result confirms that – thanks to Assumption 1 – the firm's optimal strategy in the first period is as described in the body of the paper: on the equilibrium path, once development has begun, the firm proceeds with development when it sees a positive signal. By the equilibrium path, we mean at histories that the firm reaches with positive probability (conditioning on $N$ and $R$) given its strategy. Note that this lemma is only relevant if $M > 1$, since there are no development choices at $t \geq 1$ if $M = 1$.

**Lemma 2.** *Fix some revenue $R$ and novelty $N$. If the firm enters development phase $t \in \{1, ..., M-1\}$ on path during the first period, then the firm optimally chooses $D^{t+1} = 1$ if and only if $\sigma^t = g$.*

Given this lemma, we can summarize the firm's strategy in the first period through the single number of a revenue cutoff, $\overline{R}^N$, at each novelty value $N$. When revenue $R$ is weakly above the cutoff, the firm develops the drug conditional on pre-clinical testing ($D^1 = 1$); when $R$ is below the cutoff, the firm does not.

**Proof of Lemma 2.** Fix $N$ and $R$, and some $t \in \{1, ..., M-1\}$ that the firm may reach on path.

If $\sigma^t = b$, it is clear that the firm optimally chooses $D^{t+1} = 0$: there is a positive cost $C^{t+1}$ of proceeding with the drug development, and no possible benefit.

Let us now show that the firm does proceed when $\sigma^t = g$. First note that the firm has no benefit from mixing; there will always be a pure strategy that is optimal. So, we restrict attention to pure strategies. We now show the result by contradiction. Suppose that, after arriving to this history on the equilibrium path, the firm chooses $D^{t+1} = 0$ even if $\sigma^t = g$. That means that the firm never successfully develops a drug with this realization of $N$ and $R$, as the firm also chooses $D^{t+1} = 0$ when $\sigma^t = b$. Hence, the cost of choosing $D^1 = 1$ is equal to $C^1 > 0$, and the lifetime benefit is at most equal to second-period benefit from development of $N\beta\mu_Q\alpha\Delta_2^N$ (the benefit can be less than this value if the firm pays additional development costs in periods prior to $t$). That is, if $N = 0$, then the benefit is zero, and if $N = 1$, the benefit is $\beta\mu_Q\alpha\Delta_2^N$. Applying Assumption 1, the benefit is less than the cost even if $N = 1$, and so the firm chooses $D^1 = 0$. Therefore this history is in fact not on path, yielding the desired contradiction. ■

## A.3 Proofs

**Proof of Proposition 1.**

Recall that the prior probability of $S = 1$ is $\pi$, regardless of drug novelty; the probability of $\sigma^t = g$ conditional on $S = 1$ is 1; and the probability of $\sigma^t = g$ conditional on $S = 0$ is $1 - q^t$.

For novel drugs ($N = 1$), Bayes' Rule tells us that the probability of $S = 1$ given the arrival at development phase $t$ (supposing that the firm follows a strategy consistent with Lemma 2) is

$$Pr(S = 1|D^t = 1, N = 1)\frac{\pi}{\pi + (1 - \pi)\prod_{s=1}^{t-1}(1 - q^t)} \tag{13}$$

For incremental drugs ($N = 0$), there is an additional signal $\sigma^0$, and so the corresponding probability of $S = 1$ given the arrival at development phase $t$ is

$$Pr(S = 1|D^t = 1, N = 0) = \frac{\pi}{\pi + (1 - \pi)\prod_{s=0}^{t-1}(1 - q^t)}. \tag{14}$$

We can see that $Pr(S = 1|D^t = 1, N)$ is larger for $N = 0$ than $N = 1$. Plugging in $t = M$ implies the second statement of the Proposition.

Furthermore, under the strategy described by Lemma 2, the probability of proceeding to development phase $t+1$ when $t < M$ – that is, $Pr(D^{t+1} = 1|D^t = 1, N)$ – is the probability that $\sigma^t = g$:

$$Pr(D^{t+1} = 1|D^t = 1, N) = Pr(S = 1|D^t = 1, N) + (1 - Pr(S = 1|D^t = 1, N))(1 - q^t). \quad (15)$$

This expression is increasing in $Pr(S = 1|D^t = 1, N)$, implying the first statement of the proposition. ■

**Proof of Proposition 2.** The likelihood of developing a drug of type $N$ conditional on entering pre-clinical testing is $Pr(D^1 = 1|P = 1, Y) = 1 - F_R(\overline{R}^N)$, which is weakly decreasing in $\overline{R}_N$. The revenue conditional on successful development for a drug of type $N$ is $\mathbb{E}_{R\sim F_R}[R|R \geq \overline{R}^N]$, which is weakly increasing in $\overline{R}^N$. ■

**Proof of Proposition 3.** In equilibrium, a drug of type $N$ is developed conditional on entering pre-clinical testing if and only if $R \geq \overline{R}^N$; this probability weakly decreases in $\overline{R}^N$. Moreover, the distribution of revenues conditional on success is the truncation of $R \sim F_R$ to values for which $R \geq \overline{R}^N$; the expectation $\mathbb{E}_{R\sim F_R}[R|R \geq \overline{R}^N]$ is weakly increasing in $\overline{R}^N$. So, for both parts of the Proposition, the results on probability of development conditional on entering pre-clinical testing and on expected revenue conditional on successful development will follow from the comparative static on $\overline{R}^N$.[28]

Recall the decomposition (6) of the firm's expected profit as a function of the choice $D^1$, where the three parameters do not appear in $V_1^N(R)$ or $\Delta_2^{N=1}$. The threshold $\overline{R}^N$ is the value of $R$ at which the expression (6) is constant over $D^1 \in \{0, 1\}$. Hence, $\overline{R}^N$ is defined by

$$V_1^N(\overline{R}^N) - C^1 + N\beta\mu_Q\alpha \cdot \Delta_2^{N=1} = 0. \quad (16)$$

Moreover, recall that for each $N$, it holds that $V_1^N(R)$ is continuous, and is strictly increasing in $R$ when $V_1^N(R) > 0$; $V_1^N(0) = 0$; the range of $V_1^N(R)$ given $R \sim F_R$ extends to strictly above $C^1$; and $\Delta_2^{N=1} > 0$. Assumption 1 further states that $C^1 + N\beta\mu_Q\alpha \cdot \Delta_2^{N=1} < 0$. So for each $N$, there exists a unique $\overline{R}^N \in \mathbb{R}_{++}$ satisfying the above equation.

1. For $N = 0$, Equation (16) is independent of $\beta$, $\mu_Q$, and $\alpha$, and thus there is a unique solution in $\mathbb{R}_{++}$ that is independent of these parameters.

[28] If $\overline{R}^N$ is above or below the support of the distribution, or is in a "hole" in the distribution, then the comparative statics on this probability and expected revenue hold weakly rather than strictly.

2. For $N = 1$, the equation (16) depends on $\beta$, $\mu_Q$, and $\alpha$ only through the product $\beta\mu_Q\alpha$. Because the LHS is strictly increasing in $\beta\mu_Q\alpha$, it holds that $\overline{R}^N$ must strictly decrease in this term in order to maintain (16).

■