Blockbusters, Sequels, and the Nature of Innovation*

Wesley M. Cohen† Matthew J. Higgins‡ William D. Miles§
Yoko Shibuya§

March 20, 2025

Abstract

In this paper, we provide a demand-driven explanation for the determinants of incremental innovation. Using detailed product- and invention-level data from the pharmaceutical industry, we demonstrate that firms with particularly high-selling "blockbuster" drugs tend to develop new products that target customer segments similar to their existing blockbusters. Consequently, these new products are more likely to be technically similar to the blockbuster products. This behavior, driven by the stickiness of demand for existing product offerings, limits firms' incentives to invest in entirely new products targeting different customer segments. Our findings offer insights into how blockbuster products shape firms' customer and innovation choices, with implications for understanding the dynamics of technological change in R&D-intensive industries.

*Author names are listed in alphabetical order. We acknowledge with gratitude IQVIA Incorporated for their generous support and access to their data. The statements, findings, conclusions, views, and opinions contained and expressed herein are not necessarily those of IQVIA Incorporated or any of its affiliated or subsidiary entities. The statements, findings, conclusions, views, and opinions contained and expressed in this article are based in part on data obtained under license from the following IQVIA Incorporated or affiliate information service(s): IQVIA Midas™, IQVIA National Disease and Therapeutic Index™, IQVIA National Prescription Audit™, IQVIA Health Incorporated or its affiliates. Higgins acknowledges funding from the Sorenson Center for Discovery and Innovation Studies.

†Duke University Fuqua School of Business & NBER

‡The University of Utah David Eccles School of Business, Max Planck Institute for Innovation and Competition & NBER

§Duke University Fuqua School of Business

1 Introduction

Since Schumpeter (1942), the relationship between the nature of innovation and the characteristics of markets or firms has preoccupied scholars of technological change (Cohen, 2010). In this paper, we shift the focus to a more granular level—that of the individual product—to consider the relationship between the sales of individual products and the nature of innovation. Specifically, we argue that when a firm's products achieve very high sales (or "blockbuster" status) they will subsequently invest in the development of similar products (*i.e.*, products that appeal to buyers with the needs and preferences of the buyers of the original product). When a firm's product has outsized sales, the incentive to develop products targeting the same customer segment is very high because the product generates significant profit. We further argue that appealing to the same customer segment—distinguished by the buyer needs and preferences specific to the blockbuster product—can also lead the firm to develop products that are also technologically similar.[1]

The logic behind our argument is simple. First, we assume that in R&D intensive industries a firm with a blockbuster product typically has an incentive to replace or update the product to sustain its sales performance.[2] We also assume that, due to the R&D and other up-front fixed costs incurred to develop a new product, the returns on those investments increase with the level of output of that product (Cohen and Klepper, 1996). Thus, the greater the sales of such products, the higher the returns. We further argue that to sustain the superior performance of its existing blockbuster product, firms will subsequently target the same customer segment as that served by that product. The rationale for this behavior is an assumption that firms believe that demand for their leading product is "sticky," meaning that if they develop (or acquire) a successor product that appeals to the same customer segment as the prior product, comparable

[1]For example, for many years Intel's development efforts focused on successive generations of chips based on their very high-selling X86 architecture.

[2]For example, Apple introduces successive generations of its iPhone yearly in order to maintain sales.

sales will ensue.[3] We also argue that, to appeal to the customer segment served by their high selling product, the firm will tend to develop a product that is technologically similar in order to approximate the same functionality as their blockbuster product. More generally, we are arguing that as the sales of a firm's product increases, the firm becomes more likely to invest in the development of products that resemble that product technologically; or, more starkly, blockbusters spawn incremental technical advance.[4]

In addition to offering implications for the nature of technical advance within firms' customer segments, we suggest that firms' incentives to capitalize on sticky demand offer implications for firms' portfolios of product development efforts across segments. To the extent that a firm faces budget constraints (whether due to limited financial resources or managerial attention), having a blockbuster product in a given customer segment will increase the opportunity cost of investing in the development of products in other segments and consequently dampen that investment.

To empirically support our argument, we require detailed firm-specific, product-level sales data. Even more challenging is the need for product-level data on firms' development efforts, which would allow for an assessment of whether the firms (1) target similar customer segments and (2) develop technically similar products to the firm's existing products. To our knowledge, such comprehensive data exist for only one industry: pharmaceuticals.

We collect product-level sales data for all human therapeutic drugs sold in the U.S., U.K., France, and Germany between 1998 and 2010 from IQVIA. We then match these data to drug development data from Cortellis. The resulting dataset enables us to examine the relationship between a firm's product's sales in a therapeutic class and its drug development activity in the following years. We systematically measure

[3]While pervasive across industries, demand stickiness can, as discussed below, have different endogenous (e.g., spending on marketing or sales) and/or exogenous sources.

[4]In this paper, incremental advance means specifically that the technology under development is similar to the technology of the blockbuster product.

drug development projects with respect to the firm's existing, high-selling products in two ways. First, we examine the relationship between the sales of a lead product and whether the firm invests in clinical development trials for drugs within the same therapeutic function as the lead product (*i.e.*, reflecting the development of drugs in the same customer segment). Second, we examine the relationship between lead product sales and whether the firm invests in clinical development trials for drugs that are technologically similar to the lead product.

Our empirical analysis confirms our core hypotheses. As the sales of a given product increase, firms are more likely to invest in developing products that target the same customer segments as the existing products and, at the same time, are more likely to develop products that are technically similar to their existing high-sales product. Although our core argument is supported, questions arise. First, could the apparent sales stickiness underpinning our results be endogenous with respect to firms' own marketing and sales efforts (also known as "detailing")? To address this issue, we include controls that measure a pharmaceutical firms' detailing expenditures. We find little sensitivity of our main effects with respect to the inclusion of detailing expenditure.

Next, we test the sensitivity of our main effect to variations in an exogenous source of buyer-side demand stickiness. Specifically, the literature documents the existence of demand inertia in pharmaceuticals, where physicians may be reluctant to switch medications due to concerns over side effects. Using a measure of adverse side effect reports, we quantify both product-level and market-level (*i.e.*, customer segment) side effects. First, our analysis offers evidence that side effects importantly account for demand stickiness in pharmaceuticals. Second, our analysis further reveals that when a product's side effect report counts are lower than the average side effect reports in a customer segment for which adverse side effect counts are high, firms are more likely to invest in products targeting similar customer segments and those in the development

of drugs that are technically similar to their existing products.

The interpretation of this result is that when a product has a better-than-average side effect profile in a market where other products experience severe side effects, physicians have a stronger incentive to remain loyal to the product line. Consequently, demand stickiness is higher, and firms are more likely to invest in products targeting the same customer segments. These findings highlight an exogenous source of demand stickiness as a key mechanism behind firms' incentives to invest in products targeting similar customers and technologies. While side effects may not be the only source of stickiness in pharmaceuticals, it reinforces the argument that demand stickiness plays a significant role in shaping the nature of innovation for pharmaceutical firms.

Thus far, we have explored a demand-driven explanation for why a product's outsized sales may incentivize a firm to pursue incremental technical advance as reflected by the development of technology that is similar to that of its "blockbuster" product. Yet there may also be supply side explanations. For example, perhaps a firm develops similar products to efficiently utilize what may be, to some degree, fixed capacities or expertise reflected by its R&D personnel. Accordingly, we control for a firm's R&D capabilities and observe relatively little impact on our core result regarding the effects of lead product sales.

In our analysis, we also investigate three other factors that may moderate or reinforce firms' incentives to develop products therapeutically or technically similar to their high-sales ones: the number of biotechnology firms, the intensity of competition within their market(s), and patent cliffs. The effect of these variables is uncertain a priori and remains an empirical question. For example, having more biotechnology firms in the same market segment could reduce an incumbent's investment in technically similar drugs by providing more outside opportunities for acquiring innovative technologies. Conversely, firms may be more likely to continue developing drugs technically similar to their blockbuster if they believe that, should this strategy fail, they can limit the

downside through an acquisition of a biotechnology firm.

We acknowledge that our empirical analysis cannot establish causality between product sales and the firm's innovation due to the lack of instruments for product sales. To address this limitation, we take several steps. First, we introduce time lags for the dependent variable to mitigate reverse causality with product sales. Additionally, we include firm fixed effects to control for unobserved firm capabilities and time-varying fixed effects at the broad therapeutic class group (ATC-1) to account for technological opportunities and expected demand. Our core findings regarding the relationship between product sales and either the firm's investment in the development of products serving the same customer segment, or in the development of technologically similar products remain robust across all specifications.

This paper contributes to the literature on the determinants of incremental innovation. A large body of research provides supply-driven explanations, such as firm characteristics. For instance, large firms are more likely to invest in incremental innovation due to risk aversion (Munos and Chin 2011; Krieger et al. 2022; Dranove et al. 2022) or their superior ability to appropriate returns (Arora et al. 2023). Other studies focus on the relationship between innovation and entry order (Schumpeter 1942; Arrow 1972; see Acemoglu and Cao 2015), with empirical evidence suggesting that incumbents tend to invest in incremental innovation, while entrants favor "radical" innovation (Akcigit and Kerr 2018). Our work diverges from these prior studies by, first, focusing at the level of the product rather than that of the firm. Accordingly, while these prior studies speak to drivers of the overall technical quality of innovation, our study defines the incremental character of technical advance with reference to the technology of a specific product. Second, we also offer a demand-driven mechanism: the sales of a firm's leading product(s), especially those with outsized sales, incentivize firms to develop products that target the needs and preferences specific to their leading product(s), and, in the course of doing so, develop products that are technologically

similar. Thus, in our study, firms invest in incremental technical advance to sustain sales, while in Arora et al. (2023), for example, firm sales (i.e., size) are taken as a given.

Our paper also relates to the work of Christensen (1997) who argues that industry incumbents focus on serving their current customers and consequently tend to disregard "disruptive" competitive threats posed by entrants as long as those entrants do not initially attempt to compete in their core market segments.[5] Unlike our model that ties the pursuit of a customer segment to incremental technical advance, his analysis does not offer implications for the nature of technical advance. His framework would support, for example, the notion that an incumbent may well introduce altogether new products and technologies as long as they appeal to the firms' main customers.

Our study also offers implications for the literature on the decline in R&D productivity in the pharmaceutical industry, as well as overall. Previous research has identified exogenous factors, such as longer drug development timelines and rising costs (Cockburn 2006; Higgins and Rodriguez 2006) as accounting for declining R&D productivity in the drug industry. Other studies that consider the economy more broadly, suggest that the increasing difficulty of generating new ideas over time (e.g., Bloom et al., 2020; Gordon, 2017) may account for declining R&D productivity across the economy. In contrast to these studies, we highlight the possibility that mechanisms underlying the stickiness of demand for specific, high-selling products, can also account for firms endogenously choosing to move toward more incremental innovation by developing products that are technically similar to their leading products.

[5]Christensen's key distinction is between "sustaining" versus "disruptive" innovation, where the latter is innovation that appeals to the firms' main customers, while the latter refers to innovation that does not, typically introduced in market segments that are more marginal from the firm's perspective. In his view, it is these marginal market segments that may then serve as launching pads for entrants that, once established, can then subsequently threaten (i.e., "disrupt") the leading incumbents in their core markets.

2 Model

2.1 Setting

We develop a decision-making model for a firm to determine whether to invest in developing a product that targets new customer segments or a product that targets customer segments similar to an existing one. We refer to products targeting new customer segments as *new* products and those targeting customer segments similar to an existing one as *similar* products. In practice, targeting similar customer segments often entails the development of technically similar products. However, for the sake of model simplicity, we do not account for technical similarity in the model. The relationship between targeting the same customers and technical similarity will be tested and confirmed empirically in Section 5.

Consider a two-period framework for a firm's investment decisions. In the first period, the firm has no existing products and thus must decide whether to invest in developing a new product. If the firm chooses to proceed, it incurs a fixed cost $d > 0$ and faces stochastic, non-negative product demand, which is drawn from a distribution $q_1 \sim F(\mu, \sigma^2)$, where q_1 denotes the product demand in period one.

The firm operates under financial constraints, with a budget limit of d in each period allocated towards investment. Consequently, the firm can invest in at most one type of product per period.[6] The firm is risk-neutral and a price-taker in the market, with a constant price p applicable to all products. The marginal cost of production is also constant and denoted by c. We assume that $(p - c)\mu > d$ to ensure that there is a viable investment opportunity.

Given that the firm has developed a new product in period one, it faces two investment options in period two: (1) develop a new product that targets different customer segment or (2) develop similar products that target the customer segments similar to

[6]These budget constraints, in practice, arise not only from a lack of cash reserves but also from the limited managerial attention available.

the existing product developed in period one. For simplicity, we assume that the firm cannot sell products developed in period one in period two. Relaxing this assumption does not change the model's results (see Appendix A.3).

If the firm chooses to develop a new product in period two, it incurs the fixed cost d and draws stochastic demand from the same distribution as in period one. Alternatively, if the firm opts for development of similar products, it incurs the fixed cost d [7], and the demand for the similar product is *sticky* in the sense that its demand in period two, q_2, is given by the following expression:

$$q_2 = \rho q_1 + \epsilon, \tag{1}$$

where the parameter $\rho \in (0, \infty)$ captures the stickiness of the demand for the similar product,[8] and $\epsilon \sim N(0, \sigma_\epsilon^2)$ represents the exogenous shock to demand.[9] The value of ϵ is revealed only after firms make their investment decisions.

Demand stickiness, denoted by ρ, refers to the tendency of (the same) customers to maintain a similar or greater level of demand when a company introduces new products that are similar to previous ones. This stickiness can arise from both endogenous and exogenous factors affecting firms. In our model, we treat demand stickiness as an exogenous factor, without delving into the specific mechanisms that contribute to it. In the empirical analysis, we focus on controlling for endogenous sources of demand stickiness, evaluating the robustness of our results, and leveraging exogenous variations of demand stickiness to test the underlying mechanism.

[7] For simplicity, we initially assume that the fixed cost d is identical for both developing new and similar products. While it is reasonable to argue that developing similar products might be less costly, allowing for a lower fixed cost for similar products would not alter our qualitative results; rather, it would strengthen them by making incremental innovation more appealing to firms. Appendix A.2 proves this by relaxing the assumption and exploring the model under different fixed-cost conditions.

[8] Demand stickiness, ρ, can take any positive value in this two-period model. However, when extending this model to an infinite period framework, the rate at which stickiness grows from period to period would increase exponentially. In such a case, it becomes necessary to bound ρ within the interval $(0, 1]$.

[9] Klepper (1996) uses a similar demand schedule over time, implying that product demand remains relatively stable across periods.

2.2 Analysis

Let π_2^N and π_2^S denote the *expected* profit for a firm investing in a new product and a similar product, respectively, in period two. The profit expressions are given by:

$$\pi_2^N = (p - c)\mu - d, \tag{2}$$

$$\pi_2^S = (p - c)\rho q_1 - d. \tag{3}$$

Firms make investment decisions, whether to invest in a new or similar product, to maximize their expected profit at the beginning of period two. The optimization problem can be expressed as:

$$\pi_2 = \max_{\{N,S\}} \left[\pi_2^N, \pi_2^S\right], \tag{4}$$

$$\text{s.t. } \pi_2 \geq 0, \tag{5}$$

where Equation (5) represents the participation constraint, ensuring that firms achieve non-negative profits from their investment decisions. The following proposition summarizes the firm's optimal investment choice for the problem defined in (4), based on the demand for the existing product in period one.

Proposition 1. *There exists a threshold value q^* for the demand of the existing product q_1 that delineates two distinct regions for the optimal investment decision:*

1. *Firms with an existing product demand $q_1 \in [0, q^*]$ will choose to invest in a new product in period two.*
2. *Firms with an existing product demand $q_1 \in [q^*, \infty)$ will opt to invest in developing a similar product in period two.*

The threshold value q^* is given by $q^* = \frac{\mu}{\rho}$.

The proof of Proposition 1, along with other proofs, can be found in Appendix A.

When demand for a firm's existing product is low in period one, the firm expects low demand for a similar product in period two due to demand stickiness, leading it to invest in new product development. Conversely, when the previous product's demand is high, the firm opts to develop a similar product. As shown in Proposition 1, the threshold value q^* that determines the firm's choice depends on demand characteristics. As q^* decreases, the region for similar product development, $q_1 \geq q^*$, expands. The threshold falls with ρ, the demand stickiness, implying that the similar product development becomse more attractive for firms when demand stickiness is higher.

Figure 1 shows the expected profits from developing similar versus new products as a function of previous demand q_1. In the range $q_1 \in [0, q^*]$, the firm chooses to develop a new product ($\pi_2^N \geq \pi_2^S$), while for $q_1 \geq q^*$, the optimal choice shifts to a similar product ($\pi_2^S \geq \pi_2^N$). As demand stickiness (ρ) increases, the expected profit from developing a similar product (π_2^S) rises, which causes the threshold sales value q^* to decrease. Consequently, with higher demand stickiness, the range where firms should opt to develop a similar product over a new product expands.

[Figure 1 about here.]

2.3 Empirical Predictions

We conclude this section by deriving empirical predictions based on our model results. Our first empirical prediction focuses on the relationship between the sales of existing products and the type of investment made by firms.

Prediction 1. *Firms with higher sales from existing products are more inclined to invest in* similar *products rather than in new (dissimilar) products, compared to firms with lower sales from existing products.*

This prediction directly stems from the implication of Proposition 1, which indicates that there is a sales threshold above which firms tend to engage in developing similar

products. The next prediction, illustrated in Figure 1, highlights the mechanism behind firms' incentives to invest in similar products to their previous high-sales products, demand stickiness.

Prediction 2. *The relationship between high-sales existing products and firms' incentives to invest in similar products, rather than in new (dissimilar) products, strengthens with demand stickiness.*

3 Institutional Setting

Our model is general and applicable to any industry where product development involves fixed costs and demand is sticky. To test our predictions, we use product development and sales data from the pharmaceutical industry, which offers distinct advantages for this analysis. First, the sector provides comprehensive data on product development and sales, unlike other industries where such information is often limited. Second, we can systematically measure the similarity between existing and developing products, a process that is more subjective in many other sectors. This section links our model to the empirical analysis by offering institutional context for the pharmaceutical industry, incorporating industry-specific factors such as drug approval processes, patents, and product sales exclusivity. It also highlights the growing reliance of the sector on business models centered around high-sales products.

3.1 Research and Development (R&D) Process

The new product (i.e., drug) development process in the pharmaceutical industry is lengthy and involves significant fixed costs. The drug development process is divided into research and development, where research focuses on discovering drug candidates, while development involves regulatory-mandated clinical trials, which are divided into three phases: Phase I, Phase II, and Phase III. The journey from molecule discovery to

eventual drug approval ranges from 10-15 years (Corr and Williams, 2009; Paul et al., 2010) with development times (i.e., clinical trials) alone averaging 8.3 years (Brown et al., 2022). Recent cost estimates for the development of a new drug average $879.3 million (Sertkaya et al., 2024) with upper bound estimates of $2.8 billion (Wouters et al., 2020). Importantly, these costs have increased over time (Office, 2021).

During clinical trials, drugs target specific indications, which are publicly reported and used to define distinct product markets in our analysis. These indications are categorized in the Anatomical Therapeutic Chemical (ATC) classification system, controlled by the WHO, which focuses on substances. The European Pharmaceutical Market Research Association (EPHMRA) developed a similar system, focusing on product classification by indication. The EPHMRA system has four levels of disaggregation, which we use in our analysis: ATC1 (broad anatomical group, e.g., cardiovascular), ATC2 (therapeutic subgroup, e.g., antihypertensives), ATC3 (pharmacological subgroup, e.g., plain antihypertensives), and ATC4 (chemical/pharmacological subgroup, e.g., centrally acting antihypertensives).

3.2 Factors Affecting the Rate and Direction of Innovation

Patents play a crucial role in the pharmaceutical industry, rewarding firms' R&D efforts with sales exclusivity. When a company develops a new drug, it can apply for a patent to protect the drug's formula, manufacturing process, or method of use. In addition to patent protection, FDA approval grants the drug five years of regulatory protection (data exclusivity). Certain exceptions apply, such as seven years of data exclusivity for orphan drugs, three years for reformulations, and additional exclusivity for pediatric drugs or GAIN Act antibiotics. For biologics, data exclusivity lasts twelve years, during which no generics can enter the market, regardless of patent status. After data exclusivity ends, the drug is protected only by patents, providing market exclusivity until patent expiration. Recent estimates show that chemical drugs have a median

combined data and market exclusivity of 14.4 years, while biologics last 21.5 years (Rome et al., 2021). Once exclusivity expires, generic manufacturers can enter the market.

Over the past 30 years, the biotechnology (biotech) sector has rapidly evolved as a significant source of new drug candidates for pharmaceutical companies. The number of strategic alliances between pharmaceutical and biotech firms dramatically increased (Rothaermel, 2001; Deeds and Hill, 1996). Whether through acquisition (e.g., Krieger et al., 2022; Higgins and Rodriguez, 2006; Chan et al., 2007; Danzon et al., 2007) or licensing (Hermosilla, 2021; Arora et al., 2009; Schafer, 2002), the pharmaceutical industry has become increasingly dependent on externally acquired molecules to fuel its R&D efforts (Ceccagnoli et al., 2010).

3.3 Sources of Demand Stickiness

Drug demand can exhibit significant stickiness, arising from various sources. First, substantial product-switching costs may arise due to concerns about potential side effects of new products. Unlike other consumer goods, prescription drugs are not directly purchased by consumers; instead, physicians largely determine which drug a patient will use. Due to concerns about severe unintended side effects, physicians may be reluctant to switch to a new product and may continue prescribing the same or similar drugs from the same pharmaceutical companies (King, 2002; Manchanda and Chintagunta, 2004).

Another source of demand stickiness is pharmaceutical companies' marketing activities. To influence physicians' product choices, companies engage in promotional payments for speaking fees, educational seminars, and providing meals (Agha and Zeltzer, 2022), as well as detailing (Larkin et al., 2017). Detailing is a sales process where representatives meet with physicians and provide free samples, making up the largest

portion of pharmaceutical promotional expenditures.[10] These activities foster brand loyalty (Costa-Font et al., 2014) and demand inertia at the drug level (Manchanda and Honka, 2005; Lee, 2016; Feng, 2024).[11]

3.4 An Increasing Reliance on "Blockbuster" Products

Over the past 30 years, the pharmaceutical industry has increasingly relied on "blockbuster" products—drugs that achieve very high sales, often exceeding $1 billion annually.[12] This reliance grew as the R&D pipeline for new drugs began to slow in the early 2000s, with many companies struggling to develop novel therapies capable of generating significant sales (Kola and Landis, 2004). Historically, blockbusters were often characterized by their broad market appeal, treating common conditions like high cholesterol (e.g., Lipitor). These drugs benefit from economies of scale, large marketing campaigns, and intellectual property protections (e.g., data and market exclusivity).

This growing dependency on blockbuster products has significant implications for pharmaceutical firms' strategies. First, it influences their R&D investment decisions. The fear of failure and the substantial investment required to bring a blockbuster drug to market often make firms hesitant to invest in more novel, high-risk R&D programs (Munos, 2009). With chemical-based drugs, this has led to a focus on widespread diseases with larger markets, often at the expense of treatments for rare diseases (Cockburn, 2006). However, with the rise of biologic-based drugs (e.g., Stelara), which have very high prices, companies can now target smaller markets.[13] Additionally, the dependency on blockbusters has prompted firms to adopt aggressive patenting strategies

[10]For instance, in 2016, $20.3 billion was spent on pharmaceutical promotion, with $19.1 billion allocated to detailing (Schwartz and Woloshin, 2019).

[11]In the U.S., pharmaceutical companies are also allowed to engage in direct-to-consumer advertising (e.g., TV commercials). While studies show that such advertising boosts patient demand (Dave and Saffer, 2012), Brekke and Kuhn (2006) find it complements detailing efforts.

[12]Discovering new drugs that reach blockbuster status is rare. Munos (2009) found that the probability of a drug reaching blockbuster status was around 21 percent in an analysis of 329 approved drugs, a success rate that had remained unchanged for over 20 years.

[13]The cost of Stelara, a treatment for psoriatic arthritis, is around $28,000 for one 90mg syringe (`https://www.goodrx.com/stelara`), with treatments needed every 12 weeks.

to extend product lifecycles. One common strategy is evergreening, where companies seek to extend patent life and market exclusivity for a blockbuster drug through incremental innovations, such as new formulations or delivery systems (Hemphill and Sampat, 2012).

4 Data and Methods

4.1 Data

We gather product-level sales data from IQVIA for all drug product sales in the U.S., U.K., France, and Germany from 1998 to 2010. We restrict our focus to the sales of branded therapeutic small molecule (chemical-based) and large molecule (biologic-based) drugs.[14] Crucially, for each product, these data detail the parent firm and the EPHMRA ATC code. We aggregate product-level sales to the ATC-3 level. In this paper, we consider the ATC-3 level to represent a distinct product submarket within a broader ATC-1 industry.[15]

We then collect drug development data from Cortellis, which is an exhaustive dataset on global drug development covering 79,309 global drug development projects conducted by 13,050 companies from 1980 to 2020.[16] Public registration of clinical trials was not enforced until 2006, so the data are less complete prior to this year.[17] However, even though our sales data are from 1998-2010, we employ forward-looking

[14]To identify these drugs, we classify active ingredients—the principal compound or component of a drug—into one of six categories: common elements or metals, plant-derived substances (e.g., homeopathic remedies), ingredients commonly found in over-the-counter drugs, medical devices or apparatus (e.g., bandages), small molecules for human therapeutic drugs, and biologics such as proteins and enzymes.

[15]For example, we consider the ATC-3 codes N01A (General Anaesthetics) and N01B (Local Anaesthetics) as product submarkets within ATC-1 N (Neurology).

[16]We focus on the firm, which is the originator of a drug, rather than the firm or organization that actually carries out a clinical trial. Often large pharmaceutical firms will outsource the clinical trial process to other firms, hospitals, or universities. However, it is still the originator firm funding the project and making decisions about the lead compound.

[17]In 2006, the WHO stated that all clinical trials should be registered, and it identified a minimum trial registration dataset.

measures with the drug development data, and this mitigates the data issues. 29,028 projects (approximately 37% of our data) have a biologic lead compound, and the remaining drug projects are small molecule therapeutics. We match the drug development data to the product sales data across all three dimensions: firm–therapeutic class (ATC-3)–year (i, j, t). After merging our data together, our sample consists of 66 corporate parent firms.[18] Albeit large, there is still substantial heterogeneity in the firms in our sample. Firm-Year level descriptives in Table 1 show that firms sell, on average 95.7 branded therapeutic drugs in a year, with annual sales range from $501 million to nearly $39 billion[19].

[Table 1 about here.]

We also collect data from several other sources, which we use to create variables that test proposed mechanisms and control for important features. We gather data on marketing expenditure (i.e., detailing) from IQVIA, patent exclusivity data from the FDA Orange Book, data on severe side effects from the FDA Adverse Event Reporting System[20], patenting activity from Cortellis, data on scientific publications matched to therapeutic class from Branstetter et al. (2022), and data from COMPUSTAT for firm cash flows and R&D expenditures.

[18]While 66 is a relatively small number of firms, it is still representative of the industry, with their sales of branded therapeutic small molecule and large biologic drugs accounting for 89% of total sales during our sample period.

[19]We set a lower threshold of $500 million in sales at which a firm-year is included in our sample. The reason for doing this is that, upon manual inspection of the data, firms with annual sales of less than $500 million who are classified as selling and developing branded therapeutic drugs are, in fact, not. Instead, these drugs appear to be herbal medicines or other homeopathic remedies. Although not shown in the present paper, all of our results do not change if we remove this exclusion criterion.

[20]For more details, see the FDA website https://www.fda.gov/drugs/fdas-adverse-event-reporting-system-faers/fda-adverse-event-reporting-system-faers-public-dashboard

4.2 Variables

4.2.1 Dependent Variables

Customer Similarity. We measure the therapeutic similarity of follow-on drug development by the number of drug projects initiated in a specific therapeutic area relative to a firm's products in the market. This measure serves as our best proxy for the dollar amount invested in development and enables us to analyze whether firms with especially high-sales products in an ATC-3 are more likely to invest in new development within the same therapeutic categories. The variable *Early-Stage Project Count*$_{t+5}$ reflects the number of early-stage projects that firm i has ongoing in ATC-3 j in year $t+5$, where projects are defined as early-stage if they are within 1 year of their identified start date.[21]

Technical Similarity. To measure the technical similarity between products and subsequent development, we compare the intended biological target for firm i's lead product within ATC-3 j and year t to the biological targets of drugs under development within the same ATC-1 five years later at $t+5$ (Dranove et al., 2022). A complication is that 14.6% of drugs in our development data are designed for *multiple* targets. Our measure of target-based-action similarity—*Target Overlap*$_{t+5}$—counts the number of drugs in development with at least one target in common with the lead product. We also measure the molecular similarity between products and development projects with Tanimoto scores. The Tanimoto score is widely used in chemical informatics to compute the chemical similarity between compounds and has recently been used to assess the novelty of innovations in the pharmaceutical industry (Krieger et al., 2022). We use ChemMine Tools to compute Tanimoto scores between products and drug development projects (Backman et al., 2011), which range from 0 to 1, where a score of 1 indicates that the two chemicals are structurally identical.[22] We compute

[21] In additional analyses reserved for Appendix D, we also examine later-stage projects, as this measure captures larger investments (e.g., phase II and III clinical trials), where firms push drugs toward commercial outcomes.

[22] A limitation is that Tanimoto scores do not reflect differences in stereosymmetry.

the Tanimoto score between the chemical compound of a firm's lead product in an ATC-3 at year t and all of the firm's development projects in the same ATC-1 at $t+5$ (426,295 lead product–project pairs). The variable *Mean Tanimoto Score*$_{t+5}$ is the average of these Tanimoto scores computed for each set of lead product-development project pairs.[23] The main limitation of this measure, however, is that we are restricted to marketed products and drug projects that are small molecule therapeutics since chemical formulae do not exist for large biologic drugs.

4.2.2 Independent Variables

The main independent variable in our analysis is *Lead Product Sales*$_t$, which is the nominal sales (in USD billions) of firm i's lead marketed product in an ATC-3–year (j,t). We focus our attention on only the highest selling product in a firm-year-ATC3 since, on average, a firm will sell 1.80 branded therapeutic small or large molecule drugs in an ATC-3-year (median = 1, Q3 = 2).[24] We also create the variable *$1Bn Lead Product*$_t$, which is a dummy variable set to 1 if firm i's lead-selling product in ATC-3–year (j,t) exceeds $1 billion in sales.[25] This variable captures the traditional notion of a "blockbuster" product.

[Table 2 about here.]

[Table 3 about here.]

[23] In Appendix D, we show how our results are robust to other measure constructs, including the minimum and maximum Tanimoto scores.

[24] Additionally, for empirical tractability and interpretation of the results, we prefer making comparisons between *one* product and a set of development projects, rather than some aggregate of a set of products. This is especially relevant when we are interested in the sales of products. Typically, when multiple drugs are sold in a firm-year-atc3, the drugs listed are variations in formulation or dosage size on the principal product.

[25] In Appendix D, we conduct various robustness analyses, including adjusting for inflation and sales relative to the total sales for firm i in year t.

4.2.3 Control Variables

We control for a number of variables that the literature has shown to influence a pharmaceutical firm's innovation strategy. These include controls at the firm level (*i.e.*, firm capabilities, downstream co-specialized assets, threats from patent expiration, size and financial resources) as well as market level (*i.e.*, the availability of external technology, level of market competition, technological opportunities, and commercial potential of markets).

R&D Capabilities Reserch and development comprise two distinct sets of activities and firm capabilities. On the one hand, research activities focus on the search and discovery of new drug candidates. While, on the other hand, development focuses on taking those drug candidates through the clinical trial process. Patents are an important output of research and are critically important in the pharmaceutical industry. As such, we control for a firm's research capabilities and underlying stock of innovations by their patent applications. We define $Patent\ Applications_t$ as the the count of patents applied for by firm i in market ATC-3 j in year t. Next, we control for a firm's development capabilities (*e.g.*, experience in drafting clinical trial protocols, placing and running clinical trials, and navigating the underlying regulatory processes) by their prior drug approvals. We define $Total\ Products_t$ as the total number of approved drugs sold by firm i in year t.

Detailing. Pharmaceutical firms make significant investments in downstream co-specialized assets such as marketing, detailing, and manufacturing. As noted previously, detailing is by far the largest component of a firm's total marketing and promotion efforts and have been shown to influence physician behavior. Importantly, these investments can create a "lock-in" effect into focal markets, which could potentially influence the innovation strategy of a firm (Chan *et al.*, 2007). As such, we control for detailing and define $ATC1\ Detailing\ Expenditure_{t\text{-}3}$ as firm i's total detailing expendi-

ture in year t for ATC-1 k.[26] We lag detailing share by three years as contemporaneous sales and detailing expenditure are highly correlated.

Patent Expiration. Branded pharmaceutical products are protected by patents, which, at their expiration, allow for generic (or biosimilar) entry into the market. The expiration of these patents—often referred to as "patent cliffs"—are significant events for firms. As these patent cliffs get closer, the need to replace revenue streams will become more acute and could influence the a firm's innovation choices. In order to control these effects we collect market exclusivity data from the FDA Orange Book to identify New Chemical Entity (NCE) patents for lead sales products in our sample. We define the variable *NCE expiration (<5yr)*$_t$ as a dummy equal to 1 if firm i's lead sales product in ATC-3 j in year t is within 5 years of expiration for its NCE patent.

Firm Characteristics. Along with firm fixed effects, we also include observable firm characteristics as controls in our specifications. We create the variable *Total Other Sales*$_t$ which correspond to a firm's total sales (excluding their lead sales product in ATC3-Year j, t). We also collect data for all of the public firms in our sample from Compustat for annual *Cash Flow*$_t$ and *R&D Expenditure*$_t$.

Biotechs. The markets for technology are critically important in the pharmaceutical industry as a supplier of innovation (e.g., Arora *et al.*, 2009; Ceccagnoli *et al.*, 2010; Hermosilla, 2021). Since the early 1990s, small, research-intensive biotechlogy firms have engaged in an ever increasing proportion of research and early-stage development. The extent to which we observe pharmaceutical firms engaging in incremental innovation could be moderated by the thickness (or thinness) of these external technology markets. We therefore define the variable *Biotechs*$_t$ as a count of the number of biotechnology firms active in a specific ATC-3 j market in year t. We consider biotech-

[26]We define our detailing variable at the ATC-1 level, as the training of sales forces in detailing for broad therapeutic categories (ATC-1) should be fungible across various markets (ATC-3s) contained within that broad therapeutic market (ATC-1). For example, detailing personnel trained in neurology should have little difficulty detailing to physicians about epilepsy or depression drugs. Those same skills, however, are less fungible between neurology (ATC-N) and anti-infectives (ATC-J), for example.

nology firms to be active in early-stage drug development if they are not affiliated with the parent firms in our dataset.

Competition. Pharmaceutical companies do not operate in every market. As such, the nature of branded therapeutic (*i.e.*, non-generic) competition that they face could moderate the nature and direction of innovation they pursue, regardless of whether they have a blockbuster drug. To account for these possible effects from branded competition, we create the variable *Number of Competitors*$_t$, defined as the number of firms actively selling branded therapeutic drugs in each ATC-3 j market in year t

Technological Opportunity. Time-varying, industry-specific variation in technological opportunity could be correlated with the likelihood for firms to have a blockbuster drug, as well as the quantity and nature of their innovation strategies. Accordingly, in all specifications, we include an *ATC-1* $\times$ *Year* interacted fixed effect (Branstetter et al., 2022).

Commercial Potential. We match our publication data to measures of commercial potential developed by Masclans et al. (2024) to calculate a time-varying market-level measure of commercial potential for each ATC-3. We match commercial potential scores to the subset of articles published during or after 2000 and where the authors are affiliated with an R1 U.S. University. The variable *Market Commercial Potential*$_t$ is the mean commercial potential score computed from all articles published in an ATC-3–Year (jt).

4.2.4 Mechanism

In testing our proposed mechanism, we seek to identify the extent to which demand is sticky for existing products in a therapeutic class. We leverage the notion that one potential exogenous source of demand stickiness arises from significant product switching costs stemming from concerns about the potential side effects of new products. Unlike

other consumer goods, prescription drugs are not directly purchased by consumers; instead, physicians largely determine which drug a patient will use. If one assumes that physicians seek to maximize patient utility, then they may be reluctant to switch to a new product. Instead, they might prefer to continue prescribing the same or similar drugs from the same pharmaceutical companies (King, 2002; Manchanda and Chintagunta, 2004). Accordingly, we expect that physicians' preference to prescribe the same or similar products is stronger when side effect profiles are more severe for direct product competitors within a product submarket, implying that demand stickiness is greatest when the relative incidence of side effects for your drug is low. This should be especially true in therapeutic classes with a greater tendency for adverse side effects.

We collect data from the FDA Adverse Event Reporting System,[27] which is a comprehensive dataset containing all events of adverse side effects for specific drugs reported to the FDA since 2004.[28,29] We first match these drugs to their ATC-3 code, and then for each ATC-3–Year we count the number of distinct adverse side effect events reported to the FDA.[30] The variable *ATC3-Year Side Effects Count*$_t$ represents the total count of adverse side effect events reported to the FDA in ATC3-Year j, t. The variable *Lower Relative Incidence of Side Effects*$_t$ is a dummy variable equal to 1 if Firm i's lead sales product in ATC3-Year j, t has a lower incidence of adverse side effect reports compared to other products sold in that ATC3-Year, and 0 otherwise. The relative incidence of side effects is calculated accordingly:

[27]For more details, see https://www.fda.gov/drugs/fdas-adverse-event-reporting-system-faers/fda-adverse-event-reporting-system-faers-public-dashboard

[28]These data are also hosted by the NBER; https://www.nber.org/research/data/fda-adverse-event-reporting-system

[29]A limitation of these data are that they rely on self-reporting of adverse side effects. Thus the true count of adverse side effect events is likely much higher than what we observe. Econometrically, however, this is only of concern if one expects that the tendency to report side effects is correlated with some unobserved variable that influences the firm's incentive to invest in similar or different (with respect to customer segment or the technology) drug development projects.

[30]Although there is variation in the severity of the outcomes of side effect events (i.e., the most severe being death), by definition, all side effect events are adverse.

$$\text{Relative Incidence of Side Effects}_t = \frac{\frac{\text{Focal products side effects count}}{\text{Focal product standard units sold}}}{\frac{\text{All other products side effects count}}{\text{All other product standard units sold}}} \tag{6}$$

Where standard units (SUs) represent a single dose unit, which allows for comparisons of quantity sold across different drugs.[31] We interpret a lower (higher) incidence of adverse side effects as high (low) demand stickiness: if your product has a lower (greater) incidence for adverse side effects compared to your competitor products, then doctors will be more (less) willing to prescribe your drug again in the future, and will instead prefer to prescribe the drugs with fewer/less harmful side effects.

4.3 Empirical Specification

We examine how having a blockbuster product influences the direction of future innovation and product development. To investigate this relationship, we estimate the following baseline regression:

$$y_{i,j,t+5} = \beta x_{i,j,t} + \alpha_1 c_{i,j,t} + \alpha_2 c_{i,t} + u_i + u_{j,t} + \epsilon_{i,j,t} \tag{7}$$

In this equation, our dependent variable $y_{i,j,t+5}$ represents the similarity of drugs in development at $t+5$ compared to the existing drug on the market for firm i in therapeutic class j during year t. The independent variable $x_{i,j,t}$ represents the existence of blockbuster products for firm i in ATC-3 j during year t. If our empirical prediction holds, then having blockbuster products increases the incentive for similar product development; therefore, we expect the coefficient β to be positive. The variable $c_{i,j,t}$ accounts for firm-ATC-3-year-level controls, including detailing expenditure and patent applications. Finally, the variable $c_{i,t}$ controls for characteristics at the firm-year level, which includes total sales and the number of products. We also include therapeutic class-year fixed effect $u_{j,t}$ and firm fixed effect u_i for unobserved heterogeneity across

[31] SU is an IQVIA designation that represents a single dose unit such as a pill, capsule, or equal amount of liquid.

therapeutic-years as well as firms.[32]

5 Results

We present our results in four parts. First, we report findings from our baseline regression (Equation 7), demonstrating that having blockbuster products encourages firms to develop similar products (what we refer to as the blockbuster effect). Second, we show that demand stickiness is a key mechanism behind the blockbuster effect. Third, we explore variables that may moderate the blockbuster effect. Lastly, we consider the opportunity cost of firms investing in incremental innovations to their existing blockbuster products.

5.1 Is There a Blockbuster Effect on Innovation?

Table 4 presents the results of our baseline regression (Equation 7) across our full set of dependent variables: *Early-Stage Projects*$_{t+5}$ (Columns 1 and 2), *Target Overlap*$_{t+5}$ (Columns 3 and 4), and *Mean Tanimoto Score*$_{t+5}$ (Columns 5 and 6). The odd-numbered columns use nominal sales (USD billions) of the lead product in the therapeutic class as the independent variable, while the even-numbered columns use a dummy variable indicating whether the lead product's sales exceed \$1 billion.[33]

[Table 4 about here.]

Our results are consistent across all specifications; high-sales products are associated with a higher likelihood of product development for the same target customer, and in turn, development which is technologically similar to the firm's exisitng product. For example, in Column 2, the coefficient for the dummy variable, *\$1Bn Lead Product*$_t$, is

[32]For the remainder of the paper, we do not cluster the standard errors at the firm level, as there is limited justification for the residuals being correlated at the firm level. However, we find that the empirical results remain qualitatively the same when we cluster the standard errors at the firm level.

[33]All regression models include controls for detailing expenditure, patent applications, and fixed effects.

positive and significant. This implies that if the lead product in a therapeutic class generates more than one billion dollars in sales, the firm is likely to develop more products in the same therapeutic class, compared to therapeutic classes with lead products generating less than one billion dollars in sales.[34]

Similarly, the positive coefficients on *Lead Product Sales*$_t$ and *$1Bn Lead Product*$_t$ in Columns 3 - 6 indicate that having high-sales products induce firms to invest in products that have similar target-based actions (Columns 3 and 4) as well as molecularly similar drugs (Columns 5 and 6) with respect to the high-sales products. Collectively, all of these results are consistent with Prediction 1.

Figure 2 illustrates the threshold effect of *Lead Product Sales*$_t$ on incremental innovation measures. The coefficient plots display results from the baseline regression using dummy variables for different lead product sales bands. We observe that the coefficient on *Lead Product Sales*$_t$ gradually increases as the lead product sales bands increase, indicating a continuous relationship between product sales and incremental innovation.

[Figure 2 about here.]

In Appendix D, we perform robustness checks regarding the relationship between lead product sales *share* relative to overall firm sales and the nature of innovation instead of absolute sales. Results suggest that thresholds for absolute product sales have more substantial implications for the nature of firms' innovation than threseholds for

[34]In Appendix Table D.11 we explore the sensitivity of the main effect of *Lead Product Sales*$_t$ on the similarity of subsequent development projects to each of the control variables *Patent Application*$_t$, *ATC1 Detailing Expenditure*$_{t\text{-}3}$, and *Total Products*$_t$. We also include an alternate measure for the firm's research capabilities—*Scientific Publications*$_t$. However, such data are on avaible for approximately half of observations, which is why we exclude this measure from the results reported in the main paper. Overall, this analysis reveals that for *Early-Stage Projects*$_{t+5}$ the main effect is sensitive to both the inclusion of patents and publications. Regarding *Target Overlap*$_{t+5}$, the coefficient on *Lead Product Sales*$_t$ is insensitive to the inclusion of any of the controls. However, in the full models (columns 6 and 12), the coefficient on *Lead Product Sales*$_t$ is significantly smaller than in the baseline models with no controls included (columns 1 and 6). We also acknowledge that these controls do not comprehensively capture all firm capabilities concerning research, development, and downstream co-specialized assets.

product sales share, consistent with our theory.[35] Also, given that the dependent variables *Early-Stage Projects*$_{t+5}$ and *Target Overlap*$_{t+5}$ are count variables, we estimate these relationships using Poisson and Negative Binomial models, and find qualitatively similar results.

5.2 Mechanism: Demand Stickiness

So far, we have found evidence for the presence of the "blockbuster effect," which suggests that firms are more likely to invest in drugs that are similar to their high-sales products. A source of the blockbuster effect is that the demand is *sticky*—expected demand for products that are similar to existing high-sales products remains high. This implies that in markets with higher demand stickiness, we should observe a greater incentive for firms to invest in drug development projects that serve the same target customer segments, as implied by Prediction 2.

To test this hypothesis, we examine variations in both exogenous and endogenous sources of demand stickiness. For the exogenous source, we focus on the revelation of post-marketing adverse side effects across therapeutic classes. As discussed in Section 3, a key source of demand stickiness in the pharmaceutical industry stems from the significant switching costs that physicians face, particularly due to concerns about the potential side effects of new products. We hypothesize two things. First, if a therapeutic class exhibits products with severe side effects, physicians are more likely to stick with current products, increasing the demand stickiness at the market level, and firms will have higher incentives to continue investing in similar products. Second, if a firm has products with better-than-average side effect profiles in therapeutic areas with a higher incidence of adverse side effect events, demand stickiness will be even higher for the products with better-than-average side effect profiles.

We regress three similarity measures, *Early-Stage Projects*$_{t+5}$, *Target Overlap*$_{t+5}$,

[35] Note that, though conceptually different, absolute sales and sales share are highly correlated in our data (Pearson correlation coefficient = 0.59).

and *Mean Tanimoto Score*$_{t+5}$, onto the ATC3-level side effect report counts, *ATC3-Year Side Effects Counts*$_t$, and a dummy variable for products with an incidence of adverse side effects lower than average, *Lower Relative Incidence of Side Effects*$_t$ (see Section 4.2.4 for details on the measure). The coefficient on *ATC3-Year Side Effects Counts*$_t$ reveals whether firms with products in ATC3s with high side effect counts are more likely to invest in similar products. Furthermore, we include an interaction term between *ATC3-Year Side Effects Counts*$_t$ and *Lower Relative Incidence of Side Effects*$_t$ to determine whether firms with products that have better-than-average side effect profiles in ATC3s with severe side effects have even greater incentives to invest in similar products.

Table 5 shows the results of the regressions. In columns 1, 4, and 7, we include only *Lead Product Sales*$_t$. In columns 2, 5, and 8, we add *ATC3-Year Side Effects Counts*$_t$ and *Lower Relative Incidence of Side Effects*$_t$. In columns 3, 6, and 9, we include the interaction term between *ATC3-Year Side Effects Counts*$_t$ and *Lower Relative Incidence of Side Effects*$_t$. Three key observations emerge. First, in columns 2 and 5, when the dependent variables are *Early-Stage Projects*$_{t+5}$ and *Target Overlap*$_{t+5}$, the coefficient on *ATC3-Year Side Effects Counts*$_t$ is positive and significant, implying that firms are more likely to invest in products targeting the same customer segments and using similar target-based actions in therapeutic classes where the side effect report count is higher. Moreover, the coefficient on *Lead Product Sales*$_t$ is sensitive to the inclusion of *ATC3-Year Side Effects Counts*$_t$, implying that the market-level side effect profile plays a significant role in explaining firms' incentives to invest in products targeting the same customer segments and using similar target-based actions.

Second, in columns 3 and 6, with the dependent variables *Early-Stage Projects*$_{t+5}$ and *Target Overlap*$_{t+5}$, the coefficient on the interaction term between *Lower Relative Incidence of Side Effects*$_t$ and *ATC3-Year Side Effects Counts*$_t$ is positive and significant. This implies that firms with products having lower-than-average side effect

report counts in therapeutic areas with high side effect report counts are more likely to develop products targeting the same customer segments, as well as products using similar target-based actions to existing products. These results are consistent with Prediction 2; the blockbuster effect is stronger when demand stickiness is higher.

Finally, with the dependent variable *Mean Tanimoto Score*$_t$, we find no effect on the interaction term as well as the ATC3-level side effect report counts. In other words, demand stickiness has a direct effect on the firm's incentive to invest in innovations that serve similar target customers, but this does not necessarily translate into molecular similarity. One reason we do not find an effect on our measures of molecular similarity is due to a potential trade-off firms may face in markets characterized by adverse side effects. When adverse side effects are more likely, it suggests that there is substantial room for improvement in that therapeutic class. Firms may therefore be more willing to develop a different molecule, with the goal of reducing the incidence of adverse side effects.

[Table 5 about here.]

As for endogenous sources of demand stickiness, we consider variations in detailing expenditure. As explained in Section 3.3, detailing expenditure constitutes the largest portion of pharmaceutical companies' promotional spending, which aims to increase physicians' brand loyalty to a line of drugs. Therefore, we assume that the higher the detailing expenditure, the greater the stickiness of product demand. In Table D.9 in the Appendix, we present the results of the main regressions both with and without detailing expenditure controls. The coefficients on *Lead Product Sales*$_t$ are not significantly sensitive to the inclusion of detailing expenditure as a control, suggesting that, while detailing expenditure does have some effect, its impact is not as strong as that of the exogenous factor of demand stickiness. Furthermore, the coefficients on the interaction term between lead product sales and detailing expenditure are mixed, with some showing positive effects and others negative. A very large coefficient (either

positive or negative) for the interaction terms could be due to the correlation between the two variables (correlation of 0.26 shown in Table 3).

5.3 Moderating Variables

In this subsection, we investigate whether certain factors may be moderating—either positively or negatively—the blockbuster effects we observed. Specifically, we focus on two market-level variables, the presence of biotechnology firms and the degree of branded drug competition, as well as one firm-level variable, the expiration of important patents. Table 6 shows the results of our baseline regression with the interaction between *Lead Product Sales*$_t$ and these moderating variables. Panel (a) uses *Early-Stage Projects*$_{t+5}$ as the dependent variable, while Panel (b) uses *Target Overlap*$_{t+5}$ as the dependent variable.[36]

Number of biotech firms (*Biotechs*$_t$). Small, research-intensive biotechnology firms have increasingly become an important supplier of external technology for pharmaceutical firms (Ceccagnoli *et al.*, 2010). Their presence within a specific therapeutic market can act like a type of insurance or hedge. For example, if a pharmaceutical firm experiences a late-stage product failure or a gap in their pipeline, a vibrant supply of biotechnology firms become viable acquisition targets (*e.g.*, Higgins and Rodriguez, 2006; Danzon *et al.*, 2007; Chan *et al.*, 2007; Hermosilla, 2021). This suggests that a larger supply of biotechnology firms within a specific market j, in year t, should positively influence firm *i's* early-stage project count in year *t+5*. This is precisely what we see in Table 6, Panel (A), Column (1), the coefficient on *Biotechs*$_t$ is positive and significant.[37]

Blockbuster drugs are highly valuable for firms. Thus, we should expect to see the

[36]Table D.5 in the Appendix shows our baseline regression results with all moderating and control variables for all four dependent variables: early-stage project counts, target overlap, mean Tanimoto score, and patent extensions.

[37]From the biotechnology firm's perspective, it is entirely rationale for them to endogenously position themselves in markets (especially blockbuster markets) where they may be viewed as an acquisition target.

level of insurance that a firm may want in a market to increase with the value of that market. For example, in the unrelated arts market, the level of insurance an owner would want will increase as the value of a painting increases. In our context, this logic suggests that the number of biotechnology firms (*i.e.*, the level of insurance) within the same market as a valuable lead product should positively moderate the number of early-stage products. Our results support this conjecture. In Table 6, Panel (a), Column 2, we find that the coefficient on the interaction term between *Lead Product Sales*$_t$ and *Biotechs*$_t$ is positive and significant. The importance of this result is that if a pharmaceutical firm pursues incremental innovation and experiences a development failure, it should be able to respond to that shock through the external technology markets.

There is a traditional notion that biotechnology firms work on novel research at the frontiers of science. While this may be true, our results suggest they appear to be doing so in highly valuable markets (*i.e.*, markets where we see blockbuster products). Given that most biotechnology firms do not have the capabilities or downstream co-specialized assets to take products all the way to market, they remain dependent upon the pharmaceutical industry to partner, license, or acquire their technology. It is plausible, therefore, that another impact of the blockbuster model is that it is distorting research markets by pulling biotechnology firms into markets where their partnering prospects are higher (and away from markets where their research may have a greater social impact).

Finally, while the presence of biotechnology firms appears to influence the *number* of early-stage projects, does their presence influence the underlying *nature* of innovation? We explore this question in Table 6, Panel (b), where the dependent variable is *Target Overlap*$_{t+5}$. In Column 2, we find that the cofficient on the interaction term between *Lead Product Sales*$_t$ and *Biotechs*$_t$ is negative. While this suggests that the presence of biotechnology firms negatively moderates target overlap (*i.e.*, incremental innovation),

the overall blockbuster effect (*i.e.*, the relationship between *Lead Product Sales*$_t$ and *Target Overlap*$_{t+5}$) remains positive and significant; firms with blockbuster products still pursue incremental innovation.

Market competition (*Number of Competitors*$_t$). Whether more market competition strengthens or moderates a firm's incentive to invest in similar products with respect to high-sales products is also nuanced. Concerning similarity with respect to developing products for the same target *customer*, the interaction term in column (4) of panel (a) suggests that competition positively moderated the firm's incentive to develop new products for the customers of their exisiting blockbuster products.

As for technical similarity, column (4) in panel (b) suggests a different implication of competition on the firms technological trajectory. The negative coefficient on the interaction term suggests that increased market competition increases the threat to the demand for high-sales products from competitors' new innovations, which on average increases the firms incentive to pursue technically different—and presumably "better"—innovations.

NCE patent expiration (*NCE Expiration*$_t$). Lastly, we investigate whether firms' incentives to invest in similar drugs change as patents near expiration. An NCE patent is granted for a new chemical entity that has not been previously marketed, typically providing exclusive rights to sell the drug for around 20 years from the filing date. As the patent nears expiration, firms face the threat of generic entry. Firms can either abandon the drug and invest in new drugs or attempt to extend exclusivity by introducing a new but similar drug for the same condition and launching a promotional campaign to switch users to the product. The first mitigates the blockbuster effect, while the second strengthens it.

Columns 5 and 6 in panel (a) of Table 6, using early-stage project count as the dependent variable show a positive and significant coefficient on the interaction be-

tween *Lead Product Sales*$_t$ and *NCE expiration*$_t$, indicating that as patent expiration nears, firms increase investment in similar drugs. This suggests that the latter effect dominates overall. In addition, Columns 5 and 6 in panel (b), which use target overlap as the dependent variable, indicate that these innovation projects are also *technically* similar.

[Table 6 about here.]

5.4 Is There An Opportunity Cost of Incremental Innovation?

The opportunity cost of firms investing in incremental innovations to existing blockbusters are innovations for *other* customer segments. It is interesting to ask what are these potentially under-served markets. To begin answering this question, we are motivated by the intuition that markets vary (and perhaps over time, too) in the extent to which they are *commercially promising*. We consider a commercially promising market to be an ATC-3 therapeutic class where there is a high probability for useful inventions to be developed into drugs. To measure this, we draw on the commercial potential measure developed by Masclans et al. (2024).[38] We categorize our sample into two groups based on the ATC3-year-level commercial potential: high and low commercial potential ATC3s.[39]

With this framework in place, we address the following question: When firms have high-sales products in *low* commercial potential therapeutic areas, do they reduce investment in *high* commercial potential therapeutic areas? Table 7 presents evidence by regressing the number of early-stage projects in high commercial potential ATC3s (at the ATC1 level) on a dummy variable indicating whether the firm has any $1 billion blockbuster products in low commercial potential ATC3s within the same ATC1. Column 1 shows a negative relationship between the presence of blockbuster products in

[38]This measure scores scientific publications by their commercial potential. We thus link publications to therapeutic class and year, and aggregate the commercial potential of articles.

[39]Where "high commercial potential" is market-years in the top quartile of commercial potential.

low commercial potential areas and the number of projects invested in high commercial potential areas within the same ATC1. Further still, column 2 illustrates that this relationship persists even when the firm has a scientific capability—proxied for by the count of publications—in high commercial potential markets. Overall, these results indicate that firms reduce investment in high commercial potential areas when they have blockbuster products in low commercial potential areas, suggesting the presence of potential forgone or delayed innovations.

[Table 7 about here.]

5.5 Is There A Rotation of Science?

Consistent with our theory, we have shown that firms with high-sales (*i.e.*, blockbuster) products are associated with a higher likelihood of developing similar products targeting the same customer base as the blockbuster product. Our results also suggest that these follow-on investments pursue similar target-based actions and are molecularly similar. We now consider whether firms change the nature of the science they utilize in their follow-on investments. Recall that pharmaceutical innovation can either focus on small-molecule (*i.e.*, chemical-based) or large-molecule (*i.e.*, biologic-based) drugs. Thus, our interest is in whether a firm, for example, with a small-molecule (*i.e.*, chemical-based) blockbuster would continue their follow-in investments within small-molecules or whether they might use those revenues to rotate into large-molecule (*i.e.*, biologic-based) development.

While the underlying scientific and firm capabilities between small- and large-molecule drug development differ, there are reasons why a firm may want to undertake these investments. For example, in a different context, Branstetter et al. (2022) found that firms that experienced significant generic penetration in small-molecule markets rotated away from small-molecule innovation towards large-molecule innovation within the same therapeutic market. They argued that the rotation is rational for firms given

that small- and large-molecules face different regulatory processes and, as a result, different incentive structures. These differences include longer data exclusivities and the lack of early generic entry mechanism (*i.e.*, Paragraph IV challenges) for large-molecule drugs. Combined these make it possible for a large-molecule blockbuster to maintain that status for a longer period of time relative to a small-molecule blockbuster.

Appendix Table D.10 reports results from a seemingly unrelated regression (SUR) that considers two separate linear models predicting small-molecule and large-molecule projects, respectively, as a system. In Panel (A), the dependent variable, *Early-Stage Projects*$_{t+5}$, is now restricted and defined as the count of small-molecule, early-stage projects that firm i has ongoing in ATC-3 j in year $t+5$. In Panel (B), the dependent variable is likewise restricted and defined as the count of large-molecule, early-stage projects that firm i has ongoing in ATC-3 j in year $t+5$. We make two other refinements. In Column (1) we limit *Lead Product Sales*$_t$ to include only sales from small-molecule drugs and in Column (2) we limit *Lead Product Sales*$_t$ to include only sales from large-molecule drugs. With this specification, a positive (negative) coefficient on *Lead Product Sales*$_t$ would imply a positive (negative) relationship between those particular sales and the type of innovation (*i.e.*, small- or large molecule) being considered.

In Panel (A), Column (1) we find a positive relationship between *Lead Product Sales*$_t$ and the count of small-molecule, early-stage projects. These results suggest that firms with leading products that are small-molecule are making further investments in new early-stage projects *within* the same type of science (*i.e.*, small-molecule projects). In contrast, in Panel B, Column (1) the coefficient on *Lead Product Sales*$_t$ is not significant. Recall, in Panel (B) the dependent variable is defined as the count of large-molecule, early-stage projects. This result suggests that firms with leading products that are small-molecule are not making investments in large-molecule, early-stage projects within that therapeutic market. In this case, firms do not appear to be

changing their type of science.

Additionally, in Panel (B), Column (2) we find a positive relationship between *Lead Product Sales*$_t$ and the count of large-molecule, early-stage projects. Simililar to the above, these results suggest that firms with leading products that are large-molecule are making further investments in new early-stage projects *within* the same type of science (*i.e.*, large-molecule projects). In contrast to the above, however, in Panel (A), Column (2) we find a positive relationship between *Lead Product Sales*$_t$ from large-molecule drugs and the count of small-molecule early-stage projects. This suggests that these firms are expanding investments in *both* types of science. They continue investments in large-molecule early-stage projects (Panel B, Column 2) and at the same time make investments in small-molecule early-stage projects (Panel A, Column 2). Importantly, these investments in different types of science are within the same therapeutic market as the lead product.

6 Discussion and Conclusion

In this paper, we examine profit-maximizing firms for which pursuing innovations that are incremental to their blockbuster products is optimal. This raises the question: Could the adoption of this strategy lead to social inefficiencies? It is important to clarify that, in this paper, we are not making any theoretical claims about this possibility but rather posing the question for further exploration.

6.1 Is There An Impact On Consumer Welfare?

Another potential inefficiency created by firms pursuing a Blockbuster Strategy is the delayed entry of generics for blockbuster drugs. The impact of blockbuster drugs on consumer welfare is significant, as these medications often serve a large patient population, providing effective treatments for various health conditions. For example, As-

traZeneca's drug Prilosec, a proton-pump inhibitor drug used to treat acid reflux, was highly beneficial for many patients and had a substantial market presence. However, the true benefit to consumer welfare could be maximized if generic alternatives were available. Generic drugs tend to drive down prices, making medications more affordable and accessible to a larger number of patients, thus improving overall healthcare outcomes.

In contrast, AstraZeneca's introduction of Nexium, a drug that was very similar to Prilosec (an isomer of the former), created a delay in generic entry. Nexium was an improvement for only a small proportion of patients, yet its release extended the period during which the brand-name drug dominated the market. This strategy, while beneficial to the firm in terms of extended profits, often does not align with the broader public interest. Consumers face higher prices and delayed access to more affordable generic alternatives, potentially leading to decreased overall welfare. Therefore, the impact of such a strategy may be a double-edged sword, benefiting drug manufacturers but potentially harming consumers by limiting access to cheaper options.

6.2 Implications for Productivity Trends

The declining productivity trend in pharmaceutical innovation has been thoroughly examined in the literature (Cockburn, 2006). However, explanations for this phenomenon have generally been limited to analyses of demand and supply-side shocks (Krieger et al., 2022; Dranove et al., 2022), or the broader notion that fewer novel ideas remain to be discovered (Bloom et al., 2020). This paper proposes a different explanation for *why* by focusing on individual product sales beneath the proverbial hood of the firm. Specifically, firms with high-selling products tend to develop new products targeting the same customers, which are *technically* similar to their existing successful products. If all firms follow this strategy of developing technically similar products to their existing ones, what would be the implications for industry-wide productivity?

In Figure 3, we present two industry-wide trends from our data. Panel (a) displays a simple count of the number of distinct $1 billion drugs sold each year from 1998 to 2010 for the companies in our analysis. According to our data, the number of "blockbusters" has increased more than four-fold, from 13 in 1998 to 80 in 2010.[40] In contrast, Panel (b) shows that the number of non-blockbusters has declined over the same period. Importantly, Panel (c) illustrates the share of all early-stage drug development projects for the firms in our data that utilize a new target-based action—one that has never before been used in a drug development project in the pharmaceutical industry. The share of projects with a new target has *decreased* by approximately four-fold, from 1.6% to 0.4%[41]. We acknowledge that these figures are purely descriptive accounts of the raw data.[42] Nevertheless, these trends are suggestive of the phenomena we examine in this paper: a greater reliance on high-sales products is associated with innovation that is more incremental relative to a firm's dominant product, and, in fact, less novel overall at the industry level.

[Figure 3 about here.]

6.3 Generalizing to other Industries

The present study lacks generalizability, as our findings only consider pharmaceutical firms. Yet it is important to note that we observe blockbuster products and incremental innovation with respect to those products in other industries. For example, in the movie industry, filmmakers tend to invest in sequels following the success of a highly popular film rather than creating entirely new movies. Similarly, in the tech industry,

[40]Note that our sales data are limited to the U.S., U.K., France, and Germany, so the global number of blockbuster products is likely higher. Additionally, the data focus on therapeutic drugs for human use, excluding over-the-counter medications. A comprehensive analysis of these and other factors is beyond the scope of this paper.

[41]While novelty is inherently linked to previous discoveries, the number of biological targets is finite, so the likelihood of discovering new targets should decrease over time. However, scientific advances, such as the Human Genome Project, have accelerated target discovery.

[42]There is a selection issue, as the most novel, radical drugs are less likely to be reported if they fail early, especially in the earlier years of our sample when reporting enforcement was weaker. As a result, our data likely underestimates novelty, particularly in those early years when it was highest.

companies like Apple release a flagship product, such as the iPhone, followed by sequels. In the aerospace sector, firms like Boeing often focus on incremental improvements to existing models, like the Boeing 737.

We believe that demand stickiness is an important mechanism behind strategies that emphasize incremental innovation for high-sales products. However, the sources of demand stickiness may vary across industries. In the movie industry, filmmakers continue investing in sequels because they expect similar or higher sales based on demand stickiness. This expectation stems from consumer uncertainty regarding movie quality, which is reduced when a sequel is likely to mirror a previous successful film. Here, demand stickiness arises from reduced uncertainty about new products.

In the aerospace industry, demand stickiness is driven by the significant training investments airlines make when purchasing aircraft. When airlines buy an aircraft model, they incur substantial costs to train pilots and maintain the fleet. Switching to a new aircraft type would require retraining, leading to additional costs and delays. This creates demand stickiness, encouraging manufacturers to invest in incremental improvements to existing models.

In some other industries, demand stickiness may be absent, limiting the effectiveness of the blockbuster strategy. In the toy industry, for example, demand stickiness is often lower due to the potential for imitation. High-sales products in this sector are often not protected by patents, making them easy for competitors to replicate. As a result, companies do not expect similar demand for incrementally improved versions, reducing the incentive to invest in similar products.

6.4 Concluding Remarks

Why do some firms continue to invest in similar products rather than pursuing entirely new, innovative products? In this paper, we argue that product sales are a key. Specifically, firms with high-sales products (i.e., "blockbuster products") are more likely to

invest in products that target the same customer segments, and consequently, develop products that are technically similar to those blockbuster products, thus driving incremental innovation.

The key mechanism driving firms' incentives to engage in product development that targets similar customer segments as high-sales products is demand stickiness. When demand is sticky, firms expect that the same customer segements will maintain a similar demand level for a new product that targets the same customer segments. As a result, industries with high demand stickiness are more likely to see firms focusing on the same customer segments. The causes of demand stickiness vary across industries and can include demand-side factors, such as high product switching costs resulting from the risks of using a new product, as well as substantial costs associated with training employees to use a new product. Other factors could be endogenous to firms, such as extensive marketing campaigns designed to encourage customers to stick with similar products, also contribute to demand stickiness.

The conclusions from this paper are not without limitations. Principaly, we lack causal identification in our empirical analyses due to a lack of a valid instrument. The challenge arises due to the conditions one would need to satisfy. First, the instrument would need to affect a firm's current product sales at time t. Second—and most challenging—is that the instrument must *only* affect the similarity of subsequent development projects through the sales of the existing product. This means the instrument must have no effect on the demand for future products, and no effect on demand stickiness. As an example, consider Medicare Part D, which has been used in the literature as an exogenous shock to demand that varied across therapeutic classes (e.g., Dranove et al., 2022). While this would be a shock to the sales of current product sales, Medicare Part D would also have an effect on both (i) the expected demand for *new* products in Medicare markets, and (ii) the stickiness of demand for existing products, given the expansion in coverage.

Nonetheless, his paper raises important questions for future research. How do firms navigate the trade-off between exploiting existing demand stickiness and exploring new product development opportunities? What strategies allow companies to sustain innovation when demand stickiness erodes? While the blockbuster strategy has proven effective in many industries, our results suggest that its success is far from universal. Understanding the evolving nature of demand stickiness will be critical for firms seeking to balance short-term profitability with long-term innovation.

References

Acemoglu, D. and Cao, D. (2015). Innovation by entrants and incumbents. *Journal of Economic Theory*, 157:255–294.

Agha, L. and Zeltzer, D. (2022). Drug diffusion through peer networks: The influence of industry payments. *American Economic Journal: Economic Policy*, 14(2):1–33.

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

Arora, A., Cohen, W., Lee, H., and Sebastian, D. (2023). Invention value, inventive capability and the large firm advantage. *Research Policy*, 52(1):104650.

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

Arrow, K. J. (1972). *Economic welfare and the allocation of resources for invention*. Springer.

Backman, T. W., Cao, Y., and Girke, T. (2011). Chemmine tools: an online service for analyzing and clustering small molecules. *Nucleic acids research*, 39(suppl_2):W486–W491.

Bloom, N., Jones, C. I., Van Reenen, J., and Webb, M. (2020). Are ideas getting harder to find? *American Economic Review*, 110(4):1104–1144.

Branstetter, L., Chatterjee, C., and Higgins, M. J. (2022). Generic competition and the incentives for early-stage pharmaceutical innovation. *Research Policy*, 51(10):104595.

Brekke, K. R. and Kuhn, M. (2006). Direct to consumer advertising in pharmaceutical markets. *Journal of health Economics*, 25(1):102–130.

Brown, D. G., Wobst, H. J., Kapoor, A., Kenna, L. A., and Southall, N. (2022). Clinical development times for innovative drugs. *Nat Rev Drug Discov*, 21(11):793–794.

Ceccagnoli, M., Graham, S. J., Higgins, M. J., and Lee, J. (2010). Productivity and the role of complementary assets in firms' demand for technology innovations. *Industrial and corporate change*, 19(3):839–869.

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

Christensen, C. M. (1997). *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail*. Harvard Business School Press.

Cockburn, I. M. (2006). Is the pharmaceutical industry in a productivity crisis? *Innovation policy and the economy*, 7:1–32.

Cohen, W. M. (2010). Fifty years of empirical studies of innovative activity and performance. *Handbook of the Economics of Innovation*, 1:129–213.

Cohen, W. M. and Klepper, S. (1996). A reprise of size and r & d. *The Economic Journal*, 106(437):925–951.

Corr, P. and Williams, D. (2009). The pathway from idea to regulatory approval: examples for drug development.

Costa-Font, J., Rudisill, C., and Tan, S. (2014). Brand loyalty, patients and limited generic medicines uptake. *Health policy*, 116(2-3):224–233.

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

Dave, D. and Saffer, H. (2012). Impact of direct-to-consumer advertising on pharmaceutical prices and demand. *Southern Economic Journal*, 79(1):97–126.

Deeds, D. L. and Hill, C. W. (1996). Strategic alliances and the rate of new product development: An empirical study of entrepreneurial biotechnology firms. *Journal of business venturing*, 11(1):41–55.

Dranove, D., Garthwaite, C., and Hermosilla, M. (2022). Does consumer demand pull scientifically novel drug innovation? *The RAND Journal of Economics*, 53(3):590–638.

Feng, J. (2024). History dependence in drug demand: Identification and implications for entry incentives. *Review of Economics and Statistics*, 106(2):455–469.

Gordon, R. (2017). *The rise and fall of American growth: The US standard of living since the civil war*. Princeton university press.

Hemphill, C. S. and Sampat, B. N. (2012). Evergreening, patent challenges, and effective market life in pharmaceuticals. *Journal of health economics*, 31(2):327–339.

Hermosilla, M. (2021). Rushed innovation: Evidence from drug licensing. *Management Science*, 67(1):257–278.

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

King, C. (2002). Marketing, product differentiation and competition in the market for antiulcer drugs.

Klepper, S. (1996). Entry, exit, growth, and innovation over the product life cycle. *The American economic review*, pages 562–583.

Kola, I. and Landis, J. (2004). Can the pharmaceutical industry reduce attrition rates? *Nature reviews Drug discovery*, 3(8):711–716.

Krieger, J., Li, D., and Papanikolaou, D. (2022). Missing novelty in drug development. *The Review of Financial Studies*, 35(2):636–679.

Larkin, I., Ang, D., Steinhart, J., Chao, M., Patterson, M., Sah, S., Wu, T., Schoenbaum, M., Hutchins, D., Brennan, T., et al. (2017). Association between academic medical center pharmaceutical detailing policies and physician prescribing. *Jama*, 317(17):1785–1795.

Lee, C.-Y. (2016). Brand loyalty and learning in pharmaceutical demand. *Academia Economic Papers*, 44(3):333.

Manchanda, P. and Chintagunta, P. K. (2004). Responsiveness of physician prescription behavior to salesforce effort: An individual level analysis. *Marketing Letters*, 15:129–145.

Manchanda, P. and Honka, E. (2005). The effects and role of direct-to-physician marketing in the pharmaceutical industry: an integrative review. *Yale J. Health Pol'y L. & Ethics*, 5:785.

Masclans, R., Hasan, S., and Cohen, W. M. (2024). The commercial potential of science and its realization: Evidence from a measure using a large language model. *NBER working paper*.

Munos, B. (2009). Lessons from 60 years of pharmaceutical innovation. *Nature reviews Drug discovery*, 8(12):959–968.

Munos, B. H. and Chin, W. W. (2011). How to revive breakthrough innovation in the pharmaceutical industry. *Science translational medicine*, 3(89):89cm16–89cm16.

Office, C. B. (2021). Research and development in the pharmaceutical industry.

Paul, S. M., Mytelka, D. S., Dunwiddie, C. T., Persinger, C. C., Munos, B. H., Lindborg, S. R., and Schacht, A. L. (2010). How to improve r&d productivity: the pharmaceutical industry's grand challenge. *Nature reviews Drug discovery*, 9(3):203–214.

Rome, B. N., Lee, C. C., and Kesselheim, A. S. (2021). Market exclusivity length for drugs with new generic or biosimilar competition, 2012–2018. *Clinical Pharmacology & Therapeutics*, 109(2):367–371.

Rothaermel, F. T. (2001). Incumbent's advantage through exploiting complementary assets via interfirm cooperation. *Strategic management journal*, 22(6-7):687–699.

Schafer, D. P. (2002). In-licensing as a business model. *Nature biotechnology*, 20(Suppl 6):BE36–BE39.

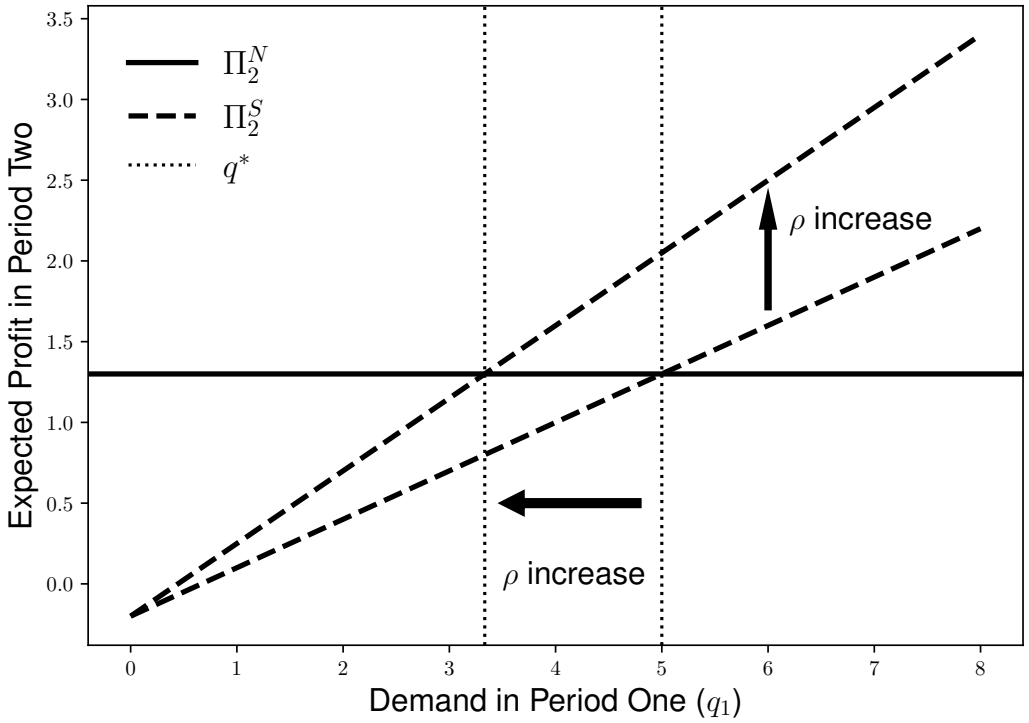
Schumpeter, J. (1942). *Capitalism, Socialism and Democracy*. Harper Row.

Schwartz, L. M. and Woloshin, S. (2019). Medical marketing in the united states, 1997-2016. *Jama*, 321(1):80–96.

Sertkaya, A., Beleche, T., Jessup, A., and Sommers, B. D. (2024). Costs of drug development and research and development intensity in the us, 2000-2018. *JAMA Network Open*, 7(6):e2415445–e2415445.

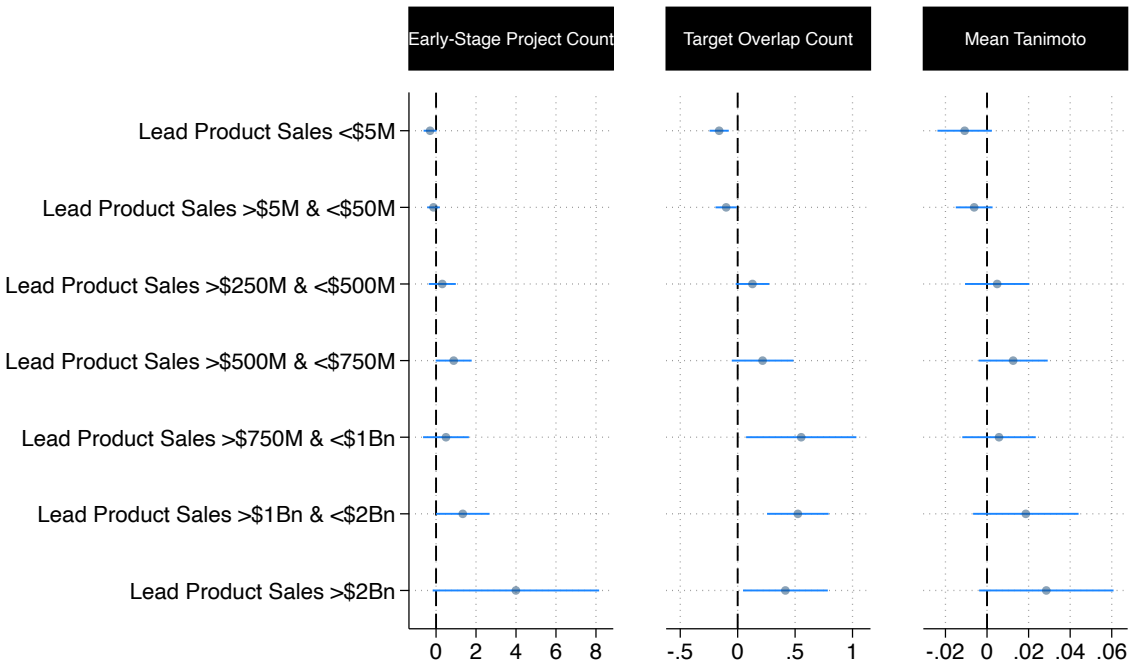
Wouters, O. J., McKee, M., and Luyten, J. (2020). Estimated research and development investment needed to bring a new medicine to market, 2009-2018. *Jama*, 323(9):844–853.

Figure 1: Illustration of Proposition 1



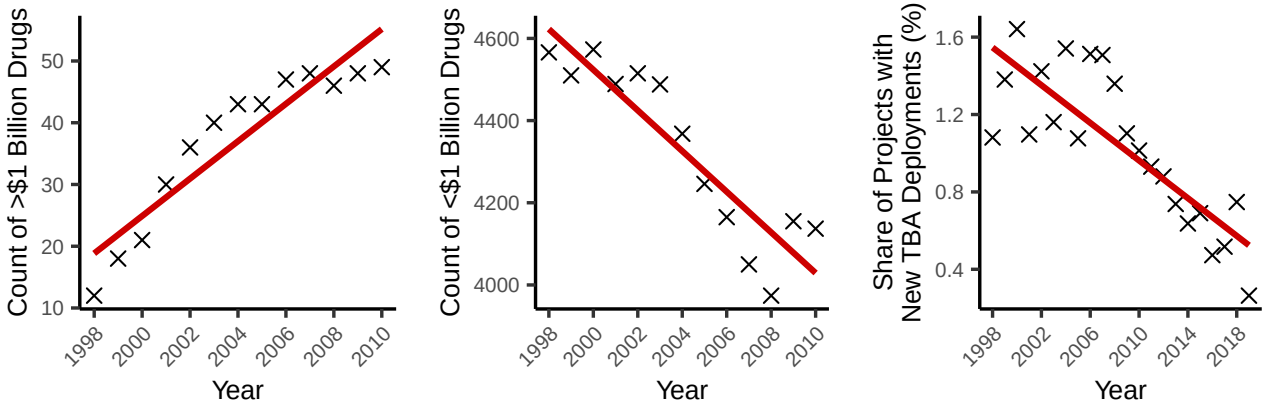
Notes: This figure illustrates Proposition 1, showing the expected profits from developing similar products versus investing in a new product, as a function of previous product sales q_1. In the figure, the solid line represents π_2^N (profit from developing a new product), the dashed line represents π_2^S (profit from developing a similar product), and the dotted line marks the threshold sales value q^*. The parameter values used to create the figure are $\mu = 3$, $p = 0.5$, $c = 0.3$, $d = 0.1$. In the figure, we change the value of ρ from 0.6 to 0.9. A key constraint on these parameters is $(p - c)\mu > d$, ensuring that the development of a new product remains financially viable.

Figure 2: Coefficient Plot of Absolute Lead Product Sales on Incremental Innovation Measures



Notes: This figure illustrates the threshold effect of lead product sales on our measures of incremental innovation. The coefficient plots show results from the baseline regression using dummy variables representing different bands of absolute sales. The baseline band is lead product sales between $50M and $250M. We use the same specification in all models that includes the controls for detailing expenditure, patent applications, total other sales, total products sold, and fixed effects at the ATC1-Year and Firm levels.

Figure 3: Time Trends for the Number of Blockbuster Products, Non-Blockbuster Products, and Development Projects using a Novel Target-Based Action



(a) Blockbuster Drugs (b) Non-Blockbuster Drugs (c) Target-Based Action Novelty

Notes: This figure plots non-parametric, descriptive time trends from the raw data. All panels use data for only the 66 parent firms in our sample. Panel (a) plots the count of $1 billion branded therapeutic products sold per year between 1998 and 2010, adjusted for 1998 dollars. Panel (b) plots the count of branded therapeutic products sold per year between 1998 and 2010 with sales *less than* $1 billion, adjusted for 1998 dollars. Panel (c) classifies the novelty of the TBA of drug development with respect to all previous projects in Cortellis, not just those initiated by the sample of 66 firms.

Table 1: Firm-Year Level Descriptive Statistics

	Mean	St. Dev.	Min	Q1	Median	Q3	Max	N
Total Sales_t (USD Billions)	4.955	6.156	0.501	0.865	1.937	6.633	32.803	576
Total Products_t	95.696	119.718	2.000	23.000	65.500	108.000	572.000	576
Total Projects_{t+5}	80.571	80.273	1.000	23.000	52.000	104.000	368.000	576

Notes: This table reports some key descriptive statistics at the Firm-Year level. *Total Sales*$_t$ represent the nominal sales, in USD billions, of only branded therapeutic small molecule or biologic drugs sold in the U.S., U.K., Germany, and France. *Total Products*$_t$ is a count of distinct products that make up the total sales figure. *Total Projects*$_{t+5}$ *is the count of all early-stage drug development projects (pre-clinical or phase 1 clinical trials) that a firm is engaged in at* $t + 5$.

Table 2: Firm-Year-ATC3 Level Descriptive Statistics

	Mean	St. Dev.	Min	Q1	Median	Q3	Max	N
Early-Stage Project Count_{t+5}	3.976	5.430	0.000	0.000	2.000	5.000	62.000	29291
Target Overlap Count_{t+5}	0.204	0.694	0.000	0.000	0.000	0.000	17.000	17014
Mean Tanimoto Score_{t+5}	0.374	0.099	0.027	0.308	0.378	0.442	1.000	14024
Lead Product Sales_t	0.076	0.361	0.000	0.000	0.001	0.013	9.677	29291
\$1Bn Lead Product_t	0.018	0.134	0.000	0.000	0.000	0.000	1.000	29291
Biotechs_t	3.607	2.608	0.000	1.820	3.230	4.930	19.630	29291
Number of Competitors_t	17.618	8.606	0.000	11.000	18.000	24.000	37.000	29291
NCE expiration_t (<5yr)	0.034	0.182	0.000	0.000	0.000	0.000	1.000	29291
Higher Relative Incidence of Side Effects_t	0.072	0.258	0.000	0.000	0.000	0.000	1.000	8246
ATC3-Year Side Effects Count_t	6.956	8.957	0.001	1.109	3.581	9.213	62.207	16964
Market Commercial Potential_t	0.530	0.120	0.062	0.453	0.536	0.617	0.805	13520
ATC1 Detailing Expenditure_{t-3}	0.024	0.041	0.000	0.000	0.006	0.027	0.324	28161
Patent Applications_t	2.310	3.946	0.000	0.000	1.000	3.000	44.000	29291
Total Other Sales_t	7.219	7.440	0.012	1.123	4.537	12.087	32.803	29291
Total Products_t	1.791	1.612	0.020	0.730	1.120	2.290	5.720	29291

Notes: This table reports key descriptive statistics for all variables featured in analyses in the main paper. In total, our dataset contains 29,291 Firm-ATC3-Year (i, j, t) observations. However, some observations have missing data. 17,014 observations have a value for *Target Overlap Count*$_{t+5}$, since this measure requires (i) identifying the target-based-action of the lead sales product, and (ii) Firm i having at least one development project in ATC3 j in year $t + 5$. For similar reasons, only 14,024 observations have a value for *Mean Tanimoto Score*$_{t+5}$, since this measure requires (i) the lead sales product being a small molecule therapeutic, (ii) identifying the molecule formula (SMILES code) of the drug, and (iii) Firm i having at least one small molecule therapeutic development project in ATC3 j in year $t + 5$. For the variable *Lower Relative Incidence of Side Effects*$_t$, the lead selling product in 8,246 observations is matched to FAERS to obtain the count of adverse side effect events for that product in year t. 16,964 observations have a value for *High Side Effects Count*$_t$, since these data are only available from 2004 onwards, and a measure can only be constructed for an ATC-3–Year if at least one product is matched and has an adverse side effect event reported. *Market Commercial Potential*$_t$ is available for 13,520 observations since we do not have publications mapped to every ATC3-Year. 28,161 observations have data for *ATC1 Detailing Expenditure*$_{t-3}$, since this is a lagged measure and our detailing data begins in 1998 (therefore observations in 1998, 1999, and 2000 are missing detailing data). See Appendix Table C.1 for a detailed description of each variable.

Table 3: Correlation Matrix

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Early-Stage Project Count$_{t+5}$	1.00														
Target Overlap Count$_{t+5}$	0.22	1.00													
Mean Tanimoto Score$_{t+5}$	0.06	0.11	1.00												
Lead Product Sales$_t$	0.17	0.18	0.09	1.00											
\$1Bn Lead Product$_t$	0.13	0.18	0.07	0.79	1.00										
Biotechs$_t$	0.47	0.03	0.02	0.06	0.05	1.00									
Number of Competitors$_t$	0.07	0.01	0.02	0.07	0.06	0.14	1.00								
NCE expiration$_t$ (<5yr)	0.09	0.12	0.08	0.09	0.06	0.04	-0.02	1.00							
Higher Relative Incidence of Side Effects$_t$	-0.05	0.00	0.03	-0.01	0.00	-0.04	-0.18	0.02	1.00						
ATC3-Year Side Effects Count$_t$	0.14	0.05	0.02	0.19	0.17	0.28	0.47	0.04	-0.08	1.00					
Market Commercial Potential$_t$	0.07	-0.03	-0.01	0.03	0.02	0.20	0.11	0.02	-0.01	0.20	1.00				
ATC1 Detailing Expenditure$_{t-3}$	0.22	0.16	0.07	0.26	0.22	-0.02	0.11	0.07	-0.05	0.07	0.02	1.00			
Patent Applications$_t$	0.40	0.12	0.03	0.06	0.07	0.35	0.07	0.04	-0.03	0.09	0.03	0.13	1.00		
Total Other Sales$_t$	0.42	0.11	0.07	0.11	0.09	0.04	-0.09	0.03	-0.06	-0.02	-0.00	0.40	0.17	1.00	
Total Products$_t$	0.03	-0.06	-0.03	-0.06	-0.04	-0.04	-0.12	-0.01	0.12	-0.04	0.00	0.05	-0.01	0.13	1.00

Notes: This table reports the pairwise Pearson correlation coefficients between all variables featured in analyses in the main paper. In total, our dataset contains 29,291 Firm-ATC3-Year (i, j, t) observations. However, some observations have missing data. See the notes in the Descriptives Statistics in Table 2 for an explanation on missingness. See Appendix Table C.1 for a detailed description of each variable.

Table 4: Is There a Blockbuster Effect?

	Early-Stage Projects$_{t+5}$		Target Overlap$_{t+5}$		Mean Tanimoto Score$_{t+5}$	
	(1)	(2)	(3)	(4)	(5)	(6)
Lead Product Sales$_t$	1.252***		0.187***		0.013***	
	(0.072)		(0.012)		(0.002)	
\$1Bn Lead Product$_t$		2.481***		0.507***		0.026***
		(0.193)		(0.032)		(0.005)
Patent Applications$_t$	0.498***	0.499***	0.017***	0.016***	-0.000	-0.000
	(0.007)	(0.007)	(0.001)	(0.001)	(0.000)	(0.000)
ATC1 Detailing Expenditure$_{t-3}$	0.341	1.383*	1.100***	1.146***	0.024	0.039*
	(0.785)	(0.782)	(0.158)	(0.157)	(0.023)	(0.023)
Total Other Sales$_t$	-0.105***	-0.123***	-0.013***	-0.015***	0.001*	0.000
	(0.013)	(0.013)	(0.003)	(0.003)	(0.000)	(0.000)
Total Products$_t$	0.835***	0.823***	0.057**	0.055**	-0.001	-0.001
	(0.113)	(0.113)	(0.023)	(0.023)	(0.005)	(0.005)
ATC1 × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	28,158	28,158	16,360	16,360	13,420	13,420
Adj R-squared	0.415	0.413	0.133	0.134	0.146	0.144

Notes: This table presents the results of our baseline regression (Equation 7) across our full set of dependent variables: *Early-Stage Projects*$_{t+5}$ (Columns 1 and 2), *Target Overlap*$_{t+5}$ (Columns 3 and 4), and *Mean Tanimoto Score*$_{t+5}$ (Columns 5 and 6). The odd-numbered columns use nominal sales (in USD billions) of the lead product in the therapeutic class as the independent variable, while the even-numbered columns use a dummy variable indicating whether the lead product's sales exceed \$1 billion. The number of observations included in each model is less than the available data due to observations that are either singletons or separated by a fixed effect. Standard errors are shown in parentheses. Significance levels: * p<0.10, ** p<0.05, *** p<0.01

Table 5: Testing the Stickiness Mechanism

	Early-Stage Projects$_{t+5}$			Target Overlap$_{t+5}$			Mean Tanimoto Score$_{t+5}$		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Lead Product Sales$_t$	1.252***	0.720***	0.720***	0.187***	0.102***	0.102***	0.013***	0.015***	0.015***
	(0.072)	(0.097)	(0.097)	(0.012)	(0.014)	(0.014)	(0.002)	(0.002)	(0.002)
Lower Relative Incidence of Side Effects$_t$		-0.133	-0.426*		-0.062**	-0.099***		-0.012**	-0.016**
		(0.193)	(0.238)		(0.030)	(0.037)		(0.006)	(0.008)
ATC3-Year Side Effects Count$_t$		0.093***	0.043*		0.003***	-0.003		-0.000	-0.001
		(0.006)	(0.024)		(0.001)	(0.004)		(0.000)	(0.001)
Lower Relative Incidence of Side Effects$_t$ × ATC3-Year Side Effects Count$_t$			0.051**			0.006*			0.001
			(0.025)			(0.004)			(0.001)
Patent Applications$_t$	0.498***	0.566***	0.565***	0.017***	0.014***	0.014***	-0.000	0.000	0.000
	(0.007)	(0.015)	(0.015)	(0.001)	(0.002)	(0.002)	(0.000)	(0.000)	(0.000)
ATC1 Detailing Expenditure$_{t-3}$	0.341	3.588***	3.590***	1.100***	1.293***	1.292***	0.024	-0.038	-0.038
	(0.785)	(1.352)	(1.352)	(0.158)	(0.206)	(0.206)	(0.023)	(0.033)	(0.033)
Total Other Sales$_t$	-0.105***	-0.093***	-0.093***	-0.013***	-0.005	-0.005	0.001*	0.001	0.001
	(0.013)	(0.034)	(0.034)	(0.003)	(0.005)	(0.005)	(0.000)	(0.001)	(0.001)
Total Products$_t$	0.835***	0.079	0.084	0.057**	0.048	0.049	-0.001	-0.014	-0.014
	(0.113)	(0.231)	(0.231)	(0.023)	(0.036)	(0.036)	(0.005)	(0.009)	(0.009)
ATC1 × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	28,158	8,214	8,214	16,360	6,653	6,653	13,420	4,977	4,977
Adj R-squared	0.415	0.462	0.462	0.133	0.154	0.154	0.146	0.228	0.228

Notes: This table seeks to provide empirical evidence for the sales stickiness mechanism. The variable *Lower Relative Incidence of Side Effects*$_t$ is a dummy variable equal to 1 if Firm i's lead sales product in ATC3-Year j,t has a higher incidence of adverse side effect reports compared to other products sold in that ATC3-Year, and 0 otherwise. The incidence of side effects is calculated from $\frac{\text{focal products side effects count}}{\text{focal product sales}}$ divided by $\frac{\text{all other products side effects count}}{\text{all other product sales}}$. We interpret a lower (higher) incidence of adverse side effects as high (low) demand stickiness: if your product has a lower (greater) incidence for adverse side effects compared to your competitor products, then doctors will be more (less) willing to prescribe your drug again in the future, and will instead prefer to prescribe the drugs with fewer/less harmful side effects. *ATC3-Year Side Effects Count*$_t$ represents the total count of adverse side effect events reported to the FDA in ATC3-Year j,t. The coefficient of interest is on the interaction between *Lower Relative Incidence of Side Effects*$_t$ and *ATC3-Year Side Effects Count*$_t$. The positive coefficient on the interaction term in columns (3) and (6) suggests that if a product has a lower incidence of adverse side effects compared to its direct competitors within the same ATC3-Year, *and* that ATC3-Year is categorized by a high count of adverse side effect events, then there is a greater incentive for the firm to invest in drug development projects for that same customer segment, and greater incentive to invest in development projects that drug the same target. Standard errors are shown in parentheses. Significance levels: * p<0.10, ** p<0.05, *** p<0.01

Table 6: Moderators and Correlates

(a) Early-Stage Project Count

	Early-Stage Projects$_{t+5}$					
	(1)	(2)	(3)	(4)	(5)	(6)
Lead Product Sales$_t$	1.006***	-0.628***	1.078***	-1.017***	1.124***	1.069***
	(0.064)	(0.101)	(0.076)	(0.199)	(0.076)	(0.078)
Biotechs$_t$	1.019***	0.978***				
	(0.010)	(0.010)				
Lead Product Sales × Biotechs$_t$		0.318***				
		(0.015)				
Number of Competitors$_t$			0.063***	0.056***		
			(0.004)	(0.004)		
Lead Product Sales × Number of Competitors$_t$				0.094***		
				(0.008)		
NCE expiration$_t$ (<5yr)					0.987***	0.757***
					(0.140)	(0.157)
Lead Product Sales × NCE expiration$_t$ (<5)						0.958***
						(0.290)
Patent Applications$_t$	0.214***	0.211***	0.484***	0.482***	0.496***	0.496***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
ATC1 Detailing Expenditure$_{t-3}$	1.960***	2.306***	1.058	1.000	0.089	0.009
	(0.662)	(0.657)	(0.782)	(0.780)	(0.785)	(0.785)
Total Other Sales$_t$	-0.066***	-0.070***	-0.105***	-0.104***	-0.102***	-0.101***
	(0.011)	(0.011)	(0.013)	(0.013)	(0.013)	(0.013)
Total Products$_t$	0.707***	0.729***	0.833***	0.840***	0.827***	0.828***
	(0.095)	(0.094)	(0.112)	(0.112)	(0.113)	(0.113)
Constant	-1.161***	-1.001***	0.891***	1.014***	1.939***	1.939***
	(0.204)	(0.202)	(0.246)	(0.245)	(0.239)	(0.239)
ATC1 × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	28,158	28,158	28,158	28,158	28,158	28,158
Adj R-squared	0.584	0.590	0.422	0.424	0.416	0.417

(b) Target Overlap

	Target Overlap$_{t+5}$					
	(1)	(2)	(3)	(4)	(5)	(6)
Lead Product Sales$_t$	0.174***	0.217***	0.175***	0.272***	0.175***	0.144***
	(0.013)	(0.020)	(0.013)	(0.033)	(0.013)	(0.013)
Biotechs$_t$	0.006***	0.008***				
	(0.002)	(0.002)				
Lead Product Sales × Biotechs$_t$		-0.008***				
		(0.003)				
Number of Competitors$_t$			-0.002***	-0.002**		
			(0.001)	(0.001)		
Lead Product Sales × Number of Competitors$_t$				-0.004***		
				(0.001)		
NCE expiration$_t$ (<5yr)					0.263***	0.118***
					(0.023)	(0.026)
Lead Product Sales × NCE expiration$_t$ (<5)						0.586***
						(0.047)
Patent Applications$_t$	0.015***	0.015***	0.017***	0.017***	0.016***	0.016***
	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
ATC1 Detailing Expenditure$_{t-3}$	1.108***	1.090***	1.080***	1.080***	1.021***	0.964***
	(0.158)	(0.158)	(0.158)	(0.158)	(0.157)	(0.157)
Total Other Sales$_t$	-0.012***	-0.012***	-0.013***	-0.013***	-0.011***	-0.010***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Total Products$_t$	0.056**	0.055**	0.057**	0.057**	0.055**	0.055**
	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.022)
Constant	0.075	0.068	0.136***	0.123**	0.075	0.077
	(0.050)	(0.050)	(0.052)	(0.052)	(0.050)	(0.050)
ATC1 × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	16,360	16,360	16,360	16,360	16,360	16,360
Adj R-squared	0.134	0.134	0.134	0.134	0.140	0.148

Notes: This table shows the results of our baseline regression, Equation (7), with interaction variables between *Lead Product Sales*$_t$ and *Biotechs*$_t$, *Number of Competitors*$_t$, and *NCE Expiration*$_t$. Panel (a) uses *Early-Stage Projects*$_{t+5}$ as the dependent variable, while Panel (b) uses *Target Overlap*$_{t+5}$ as the dependent variable. Standard errors are shown in parentheses. Significance codes: * p<.1, ** p<.05, *** p<.01

Table 7: Blockbuster Products in Low-CP Markets are Associated with Reduced Investment in High-CP Markets

	High-CP Early-Stage Project Share$_{t+5}$	
	(1)	(2)
Low-CP Sales Share$_t$	-0.051***	-0.049***
	(0.007)	(0.007)
High-CP Publication Count$_t$ (100s)		0.067***
		(0.006)
Total Detailing$_{t-3}$ (USD Billions)	0.002	-0.009
	(0.009)	(0.009)
Total Patent Applications$_t$ (100s)	0.008	0.003
	(0.014)	(0.014)
Total Firm Sales$_t$ (USD Billions)	0.002*	0.002
	(0.001)	(0.001)
Total Firm Products$_t$ (100s)	0.009	0.017
	(0.013)	(0.012)
ATC1 × Year FE	Yes	Yes
Firm FE	Yes	
Observations	6,066	6,066
Adjusted R-squared	0.576	0.585

Notes: . Standard errors are shown in parentheses. Significance codes: * p<.1, ** p<.05, *** p<.01

Appendix:

Blockbuster Products, Sequels, and the Nature of Innovation

A Proof and Additional Analysis

A.1 Proof

Proof of Proposition 1. We compare a firm's expected profit under new product and similar product development. Firms choose to invest into developing new product than similar product if and only if $\pi_t^N \geq \pi_t^S$. The condition $\pi_t^N \geq \pi_t^S$ can be rearranged to

$$\pi_t^N \geq \pi_t^S \implies q_{k,t-1} \geq \frac{\mu}{\rho} > 0. \tag{8}$$

Let $q^* = \frac{\mu}{\rho}$ denote the threshold value such that firms with $q_{k,t-1} \leq q^*$ chooses to invest in new product development. Firms with $q_{k,t-1} \geq q^*$ chooses to invest in similar innovation.

Lastly, we check whether each firm makes non-negative profit. Firms with $q_{k,t-1} \in [0, q^*]$ makes non-negative profit because $\pi_t^N = (p-c)\mu - d \geq 0$ by assumption of viable new product development opportunity. Firms with $q_{k,t-1} \in [q^*, \infty]$ makes non-negative profit because $0 \leq \pi_t^N \leq \pi_t^S$ in this region and $0 \leq \pi_t^N$ because of the assumption of vaiable new product development opportunity. □

A.2 Additional analysis: Smaller fixed cost of developing a similar product

In Section 2 of the main paper, we assumed that the fixed cost of developing a new product is denoted by d and that this cost is the same for developing a similar product. However, one might argue that developing a similar product is less expensive for the firm. Therefore, in this section, we relax the assumption that the fixed costs for developing a new product and a similar product are identical.

Let d^S and d^N represent the fixed costs of developing a similar product and a new product, respectively, with the condition $\frac{d^N}{2} < d^S \leq d^N$. We assume that each firm is financially constrained with a budget of d^N. As a result, firms can develop at most one product, either new or similar.

When a firm chooses to invest in a similar product, it will have a remaining budget of $d^N - d^S$. We assume that this remaining budget can be invested in other opportunities that provide a deterministic return of r. Thus, the return on investing the remaining budget in these opportunities is given by $(1 + r)(d^N - d^S)$.

With this revised setting, we can rewrite firms' expected profits when investing in a new and similar product as:

$$\pi_t^N = (p - c)\mu - d^N, \tag{9}$$

$$\pi_t^S = (p - c)\rho q_{k,t-1} - d^N + (1 + r)(d^N - d^S). \tag{10}$$

The results in Proposition 1 will be revised as follows.

Proposition 2. *There exists a threshold value q^* for the demand of the existing product $q_{k,t-1}$ that delineates two distinct regions for the optimal investment decision:*

1. *Firms with an existing product demand $q_{k,t-1} \in [0, q^*]$ will choose to invest in new product development.*
2. *Firms with an existing product demand $q_{k,t-1} \in [q^*, \infty)$ will opt to invest in developing a similar product.*

The threshold value q^* is given by $q^* = \frac{\mu}{\rho} - \frac{(1+r)(d^N - d^S)}{(p-c)\rho}$.

The proof of Proposition 2 follows the proof of Proposition 1. The intuition behind Proposition 2 is that when the fixed cost of developing a similar product is lower than the cost of developing a new product, the range of conditions under which firms optimally choose to invest in a similar product broadens. This is because the reduced cost of developing a similar product provides an additional incentive for firms to prefer it over a new product.

A.3 Additional analysis: Allowing firms to continue selling their product from period $t-1$

In Section 2 of the main paper, we assumed that firms cannot sell products developed in period $t-1$ during period t. In this subsection, we relax this assumption and allow firms to sell multiple products in period t.

Suppose a firm develops a new product in period $t-1$ and the demand for this product is given by $q_{k,t-1}$. If the firm can sell product k alongside a new or similar product developed in period t, we assume that the demand for the previously developed product in period t remains the same as in period $t-1$, i.e., $q_{k,t} = q_{k,t-1}$.

With this revised setting, the firms' expected profits when investing in a new or similar product can be expressed as:

$$\pi_t^N = (p-c)(\mu + q_{k,t-1}) - d, \tag{11}$$

$$\pi_t^S = (p-c)(\rho+1)q_{k,t-1} - d. \tag{12}$$

Here, firms always have an incentive to sell previously developed products due to the positive profit margin $p > c$.

Under this revised scenario, the results in Proposition 1 remain unchanged. This is because the term $(p-c)q_{k,t-1}$ cancels out between π_t^N and π_t^S, leaving the threshold value q^* unaffected.

B Constructing the Sample

B.1 Firm Name Matching

In Cortellis, each drug development project is assigned to an originator firm. Our task is to match each project to a parent firm in IQVIA. First, we created a time-varying corporate tree using IQVIA data. Product sales are assigned to the firm selling the product and the parent firm, so from these data we can infer corporate trees for each year 1998 through 2010. In other words, for each parent firm–year, we have a list of subsidiaries. A limitation of this approach is that we are only including subsidiaries that have non-zero sales. A parent firm may have a subsidiary in year t which *only* performs R&D.

We trialed various matching techniques, and ultimately, we decided on an approach that minimizes false positive matches. This approach involved a detailed cleaning process which included removing punctuation, phrases in parentheses, and white spaces. We also identify and exclude universities in Cortellis from the match. Lastly, we take a step to standardize firm names in both sources by removing common suffixes (e.g., *Pharma, Group, Gmbh, Corp, Pharmaceutical, Therapeutic, Inc, Labs, Co, Ltd, Llc*). We then perform an exact match on the resulting cleaned names.

B.2 Product matching

B.2.1 Matching IQVIA products to chemical compounds in Cortellis

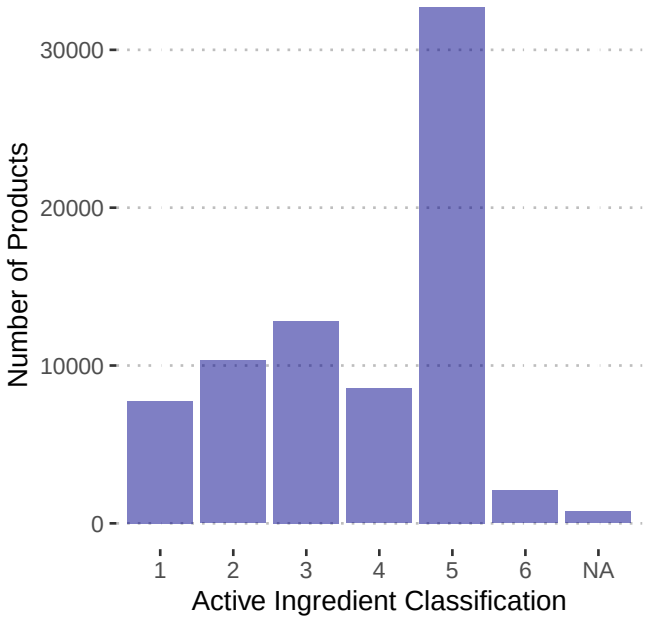
To create our measures of target-based and molecular similarity, we need to find the molecule and target for each lead sales product. To do this, we need to match our product sales data to their drug development project entry in Cortellis by *drug name*. In a similar vein to our firm-name matching, we employ a fuzzy-matching algorithm to match product names to their drug development entry in Cortellis. A limitation here is that for products in sales in our sample—particularly in earlier years—their drug development will have started 10-15 years before. As previously mentioned, coverage of drug development projects prior to 2006 is sparse. However, we are looking for drug development projects that turned into launched products, so we are confident our coverage of these drugs is good. This would be more of an issue if, for example, we were looking for drug development projects in the 1980s that failed in the pre-clinical or phase 1 clinical trial phase.

Our approach began with classifying the active ingredients of products. The goal of this step is to identify which products should in theory have a chemical formula. Since our sales data contain biologic drugs, homeopathic medicines, and medical devices, many products should not have a SMILES code. Restricting the set of drugs that we then search for a SMILES code to those which are relevant will improve the accuracy of our matching.

A total of 4,784 active ingredients were classified into six categories. The categories were created through an iterative process as we learned more about the type of active ingredients in the data. The six categories are listed below, and the frequency distribution of products for each classification is plotted in Figure B.1.

1. Common elements or metals
2. Plants (possibly the latin name)
3. Common ingredients in over-the-counter drugs e.g., paracetamol
4. Not an ingredient for a human therapeutic drug e.g., bandages, test kits, supplements, baby milk, antiseptics, insecticides, disinfectant, or cosmetics
5. Important ingredient for a small molecule human therapeutic drug
6. Biologic compound, such as proteins and enzymes

Figure B.1: Count of Products by each Active Ingredient Classification



2,453 ingredients were identified as small molecules utilized in human therapeutic drugs, corresponding to over 30,000 products in our data. This classification also allows us to restrict our focus on small molecule and biologic drugs when we define the lead sales products. Out of 4,378 distinct lead products (in a firm–year–ATC-3 (i, j, t) observation), 3,939 products have all of their active ingredients classified as important for small molecule human therapeutic drugs.

We then sought to match these 3,939 products to Cortellis for when they were a drug development project. This match would allow us to then infer the chemical formulae (SMILES code) and intended therapeutic target that the drug acts upon. We performed standard text cleaning to both product and project drug names (i.e., removing punctuation). We first sought a match on the product name, and failing that we then looked to match on the active ingredient name. When multiple matches were returned, we selected the SMILES code that was matched most frequently (in the case of a tie, we selected one at random). In total, we identified a SMILES code for 2,859 products (73%).

C Data

C.1 Description of Variables

Variable Name	Description of Measure
Early-Stage Project Count$_{t+5}$	The count of projects in early-stage drug development which firm i has ongoing in ATC-3 j in year t+5. We define projects as being in early stages if year t falls in the interval $x-1 < t < x+1$ where x is the project start year.
Late-Stage Project Count$_{t+5}$	The count of projects in later-stage drug development which firm i has ongoing in ATC-3 j in year t+5. We define projects as being in early stages if year t falls in the interval $x+3 < t < x+5$ where x is the project start year.
Early-Stage Project Share$_{t+5}$	The count of projects in early-stage drug development which firm i has ongoing in ATC-3 j in year $t+5$ as a share of all projects ongoing in the same ATC-1 k.We define projects as being in early stages if year t falls in the interval $x-1 < t < x+1$ where x is the project start year.
Target Overlap Count$_{t+5}$	The count of early-stage development projects in year $t+5$ that share at least one target-based action with firm i's lead selling product in ATC-3 j in year t. We only look at early-stage development projects within the same ATC-1 as ATC-3 j.
Target Overlap Share$_{t+5}$	The share of early-stage development projects in year $t+5$ which share at least one target-based action with firm i's lead selling product in ATC-3 j in year t. We only look at early-stage development projects within the same ATC-1 as ATC-3 j.
Same Target Count$_{t+5}$	The count of early-stage development projects in year $t+5$ with the same target-based-action combination as firm i's lead selling product in ATC-3 j in year t. We only look at early-stage development projects within the same ATC-1 as ATC-3 j.
Same Target Share$_{t+5}$	The share of early-stage development projects in year $t+5$ which have the same target-based action as firm i's lead selling product in ATC-3 j in year t. We only look at early-stage development projects within the same ATC-1 as ATC-3 j.
Mean Tanimoto Score$_{t+5}$	The mean Tanimoto score computed between firm i's lead selling product in ATC-3 j in year t and all drugs in development within the same ATC-1 k at $t+5$. This variable measures the average molecular similarity between all drugs in early-stage development at $t+5$ to the firm's lead product at time t in ATC4 j.
Min Tanimoto Score$_{t+5}$	The minimum Tanimoto score computed between firm i's lead selling product in ATC-3 j in year t and all drugs in development within the same ATC-1 k in year $t+5$. This variable measures the molecular similarity of the least similar drug in early-stage development at $t+5$ to the firm's lead product at time t in ATC-3 j.
Max Tanimoto Score$_{t+5}$	The maximum tanimoto score computed between firm i's lead selling product in ATC-3 j in year t and all drugs in development within the same ATC-1 k at $t+5$. This variable measures the molecular similarity of the most similar drug in early-stage development at $t+5$ to the firm's lead product at time t in ATC4 j.
Patent Extension$_t$	A cumulative count of all patents applied for firm i's lead product in ATC-3 j as of year t since the first year of sales for the product. This variable is only created for products which did not have sales in 1998 (i.e., we can identify the year 1999 or later in which sales began).
Lead Product Sales$_t$	The absolute sales in USD billions of firm i's lead selling product in ATC-3 j in year t.
\$1Bn lead Product$_t$	A dummy variable equal to 1 if firm i's lead selling product in ATC-3 j in year t exceeds \$1 billion in sales in year t; 0 otherwise.
Sales Share Lead Product/Firm$_t$	The share of sales for firm i's lead selling product in ATC-3 j's with respect to the firm's total sales in year t.
\$3Bn lead Product$_t$	A dummy variable equal to 1 if firm i's lead selling product in ATC-3 j in year t exceeds \$3 billion in sales in year t; 0 otherwise.

Lead Product Sales$_t$ (1998 dollars)	The sales in USD billions of firm i's lead selling product in ATC-3 j in year t adjusted for 1998 dollars.
\$1Bn lead Product$_t$ (1998 dollars)	A dummy variable equal to 1 if firm i's lead selling product in ATC-3 j in year t exceeds \$1 billion in sales in year t, adjusted for 1998 dollars; 0 otherwise.
Cash Flow$_t$	Firm i's total cash flow in year t.
R&D Expenditure$_t$	Firm i's total R&D expenditure in year t.
Biotechs$_t$	A count of the distinct number of biotech firms, in hundreds, active in early-stage clinical trials in year t in ATC-3 j.
Number of Competitors$_t$	The number of parents firms actively selling branded therapeutic drugs in ATC-3 j in year t.
NCE expiration$_t$ (<5yr)	A dummy variable equal to one if Firm i's leading sales product in ATC-3 j in year t has an NCE patent which is within 5 years of expiry; 0 otherwise.
Lower Relative Incidence of Side Effects$_t$	A dummy variable equal to 1 if Firm i's lead sales product in ATC3-Year j,t has a lower incidence of adverse side effect reports compared to other products sold in that ATC3-Year, and 0 otherwise. The relative incidence of side effects is calculate from $\frac{\text{focal products side effects count}}{\text{focal product standard units so}}$ divided by $\frac{\text{all other products side effects count}}{\text{all other standard units sold}}$. Standard units (SUs) represent a single dose unit, which allows for comparisons of quantity sold across different drugs. SU is an IQVIA designation that represents a single dose unit such as a pill, capsule, or equal amount of liquid. We interpret a higher (lower) incidence of adverse side effects as low (high) demand stickiness: if your product has a greater (lower) incidence for adverse side effects compared to your competitor products, then doctors will be less (more) willing to prescribe your drug again in the future, and will instead prefer to prescribe the drugs with fewer/less harmful side effects.
ATC3-Year Side Effects Count$_t$	The total count of adverse side effect events reported to the FDA in ATC3-Year j,t.
Market Commercial Potential$_t$	The Commercial Potential of ATC-3 j in year t. We compute the mean commercial potential from the scores of all scientific articles published in year t by the top 120 U.S. R1 Universities using the measure developed by Masclans et al. (2024).
ATC1 Detailing Expenditure$_{t\text{-}3}$	The sum of firm i's ATC-1 k detailing expenditure with in year $t-3$.
Patent Applications$_t$	The number of patents firm i applied for in year t which pertain to ATC-3 j.
Total Other Sales$_t$	Firm i's total product sales in Billions of Dollars in year t, minus the sales of their leading selling product in ATC3-Year j,t.
Total Products$_t$	Firm i's total number (in hundreds) of products in sales in year t.

C.2 Sales Stickiness

Our argument rests on the assumption that product sales are *sticky* i.e., sales in period t are expected to look similar to sales in period $t-1$. In our data, then, a time series of product sales should exhibit a high autocorrelation.

In Figure C.3, we present descriptive evidence of sales stickiness using data from selected ATC3 categories and individual products. Panel C.3a shows the annual sales of human therapeutic small molecule and biological drugs for all firms in our data from 1998 to 2010. Panel C.3b presents the autocorrelation functions for 10 lags for each ATC3. At lag 1, the autocorrelation—a comparison of sales at t and $t-1$—is positive and statistically significant, with an average coefficient of 0.79 across all ATC3s.

At the product level, high autocorrelation is also evident, as shown in Panel C.3c for a selection of blockbuster drugs in ATC3 N6A. However, patent expirations introduce

Table C.1: Descriptive Statistics for all Variables

	Mean	St. Dev.	Min	Q1	Median	Q3	Max	N
Early-Stage Project Count$_{t+5}$	3.976	5.430	0.000	0.000	2.000	5.000	62.000	29291
Early-Stage Project Share$_{t+5}$	0.074	0.150	0.000	0.000	0.000	0.098	3.000	29291
Late-Stage Project Count$_{t+5}$	2.701	4.758	0.000	0.000	0.000	4.000	57.000	29291
Target Overlap Count$_{t+5}$	0.204	0.694	0.000	0.000	0.000	0.000	17.000	17014
Target Overlap Share$_{t+5}$	0.008	0.049	0.000	0.000	0.000	0.000	1.000	17014
Same Target Count$_{t+5}$	0.091	0.389	0.000	0.000	0.000	0.000	5.000	17014
Same Target Share$_{t+5}$	0.004	0.034	0.000	0.000	0.000	0.000	1.000	17014
Mean Tanimoto Score$_{t+5}$	0.374	0.099	0.027	0.308	0.378	0.442	1.000	14024
Min Tanimoto Score$_{t+5}$	0.218	0.105	0.000	0.149	0.207	0.273	1.000	14024
Max Tanimoto Score$_{t+5}$	0.547	0.160	0.027	0.446	0.547	0.641	1.000	14024
Patent Extension$_t$	1.597	1.603	0.000	1.000	1.000	2.000	11.000	516
Lead Product Sales$_t$	0.076	0.361	0.000	0.000	0.001	0.013	9.677	29291
\$1Bn Lead Product$_t$	0.018	0.134	0.000	0.000	0.000	0.000	1.000	29291
Sales Share Lead Product/Firm$_t$	0.015	0.061	0.000	0.000	0.000	0.004	0.986	29291
\$3Bn Lead Product$_t$	0.004	0.059	0.000	0.000	0.000	0.000	1.000	29291
Lead Product Sales$_t$ (1998 dollars)	0.064	0.300	0.000	0.000	0.001	0.011	7.926	29291
\$1Bn Lead Product$_t$ (1998 dollars)	0.014	0.119	0.000	0.000	0.000	0.000	1.000	29291
Cash Flow$_t$	10.210	40.605	0.017	0.754	2.038	4.368	374.062	22361
R&D Expenditure$_t$	7.795	25.661	0.003	0.918	2.563	5.125	196.802	22297
Biotechs$_t$	3.607	2.608	0.000	1.820	3.230	4.930	19.630	29291
Number of Competitors$_t$	17.797	8.463	0.000	11.000	18.000	24.000	37.000	28995
NCE expiration$_t$ (<5yr)	0.034	0.182	0.000	0.000	0.000	0.000	1.000	29291
High Side Effects Rate$_t$	0.213	0.409	0.000	0.000	0.000	0.000	1.000	16964
Market Commercial Potential$_t$	0.530	0.120	0.062	0.453	0.536	0.617	0.805	13520
ATC1 Detailing Expenditure$_{t-3}$	0.024	0.041	0.000	0.000	0.006	0.027	0.324	28161
Patent Applications$_t$	2.310	3.946	0.000	0.000	1.000	3.000	44.000	29291
Total Other Sales$_t$	7.219	7.440	0.012	1.123	4.537	12.087	32.803	29291
Total Products$_t$	1.791	1.612	0.020	0.730	1.120	2.290	5.720	29291

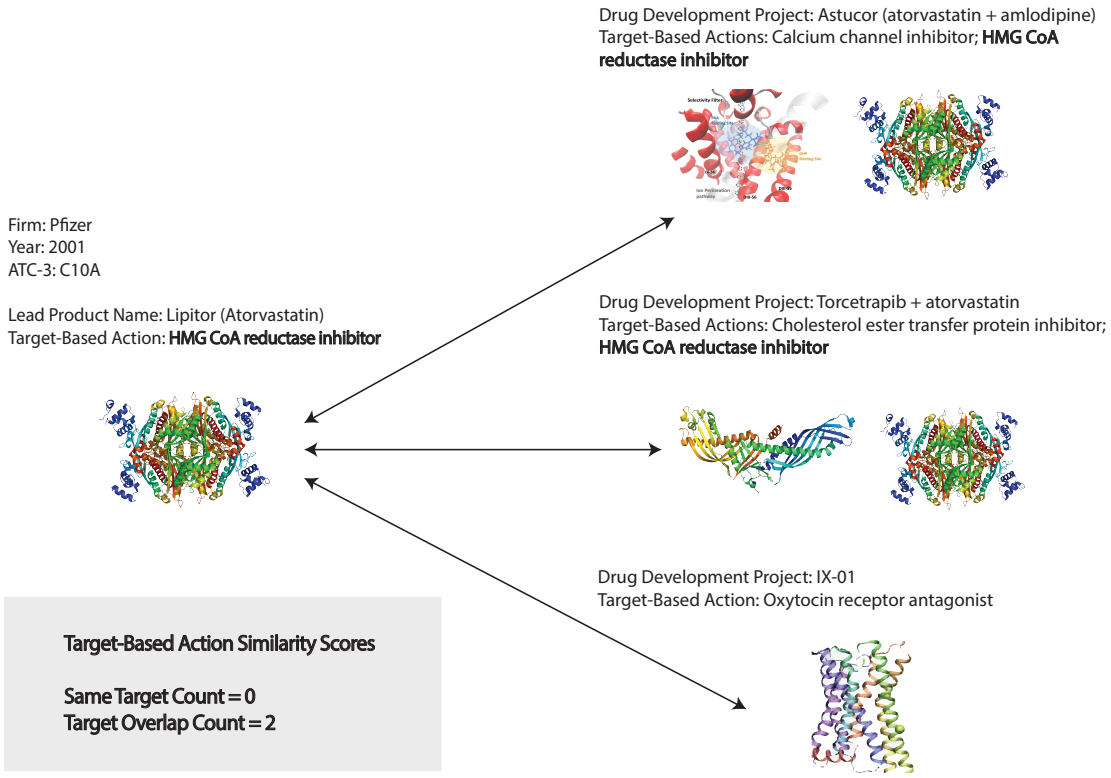
Notes: This table reports descriptive statistics for all variables, including those that only feature in the Appendix.

Table C.2: Correlation Matrix for all Variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)	(25)	(26)	(27)	(28)
Early-Stage Project Count$_{t+5}$	1.00																											
Early-Stage Project Share$_{t+5}$	0.33	1.00																										
Late-Stage Project Count$_{t+5}$	0.74	0.18	1.00																									
Target Overlap Count$_{t+5}$	0.22	0.06	0.22	1.00																								
Target Overlap Share$_{t+5}$	0.00	0.04	0.01	0.50	1.00																							
Same Target Count$_{t+5}$	0.18	0.05	0.18	0.70	0.35	1.00																						
Same Target Share$_{t+5}$	-0.00	0.02	-0.00	0.28	0.78	0.43	1.00																					
Mean Tanimoto Score$_{t+5}$	0.06	0.06	0.06	0.11	0.22	0.07	0.16	1.00																				
Min Tanimoto Score$_{t+5}$	-0.19	0.03	-0.18	-0.03	0.18	-0.03	0.15	0.68	1.00																			
Max Tanimoto Score$_{t+5}$	0.23	0.04	0.23	0.36	0.20	0.32	0.15	0.70	0.18	1.00																		
Patent Extension$_t$	-0.06	-0.05	-0.05	0.01	0.09	0.06	0.07	0.09	0.05	0.08	1.00																	
Lead Product Sales$_t$	0.17	0.09	0.18	0.18	0.07	0.15	0.04	0.09	-0.01	0.18	0.14	1.00																
\$1Bn Lead Product$_t$	0.13	0.07	0.14	0.18	0.08	0.16	0.05	0.07	-0.01	0.16	0.15	0.79	1.00															
Sales Share Lead Product/Firm$_t$	-0.00	0.07	0.00	0.17	0.23	0.16	0.12	0.12	0.12	0.13	0.18	0.58	0.50	1.00														
\$3Bn Lead Product$_t$	0.08	0.03	0.09	0.04	0.01	0.04	0.01	0.04	0.00	0.08	0.04	0.71	0.44	0.32	1.00													
Lead Product Sales$_t$ (1998 dollars)	0.17	0.09	0.18	0.19	0.08	0.16	0.04	0.08	-0.01	0.19	0.13	1.00	0.79	0.58	0.70	1.00												
\$1Bn Lead Product$_t$ (1998 dollars)	0.13	0.06	0.14	0.17	0.07	0.14	0.04	0.06	-0.02	0.15	0.12	0.80	0.89	0.48	0.49	0.80	1.00											
Cash Flow$_t$	-0.06	0.03	-0.07	-0.02	0.02	-0.03	-0.00	0.04	0.05	-0.01	0.08	-0.02	-0.02	0.00	-0.00	-0.02	-0.02	1.00										
R&D Expenditure$_t$	-0.04	0.05	-0.04	-0.02	-0.01	-0.03	-0.01	0.03	0.04	-0.02	0.09	-0.02	-0.01	-0.00	-0.00	-0.02	-0.01	0.81	1.00									
Biotechs$_t$	0.47	0.32	0.40	0.03	-0.00	-0.01	-0.02	0.02	0.02	-0.01	-0.10	0.06	0.05	-0.00	0.03	0.05	0.05	0.01	0.04	1.00								
Number of Competitors$_t$	0.08	0.09	0.08	0.01	0.02	0.01	0.02	0.02	0.00	0.02	-0.06	0.07	0.06	0.04	0.03	0.07	0.06	0.05	0.03	0.14	1.00							
NCE expiration$_t$ (<5yr)	0.09	0.09	0.08	0.12	0.04	0.09	0.02	0.08	0.02	0.14	-0.20	0.09	0.06	0.11	-0.00	0.10	0.06	-0.00	-0.00	0.04	-0.02	1.00						
High Side Effects Rate$_t$	0.09	0.05	0.06	-0.01	-0.01	-0.04	-0.03	0.07	0.04	0.03	-0.10	-0.02	-0.01	-0.03	-0.01	-0.02	-0.01	-0.03	-0.03	0.12	-0.24	0.02	1.00					
Market Commercial Potential$_t$	0.07	0.03	0.07	-0.03	-0.05	-0.04	-0.05	-0.01	-0.00	-0.04	0.08	0.03	0.02	0.00	0.02	0.02	0.02	0.00	0.04	0.20	0.11	0.02	0.01	1.00				
ATC1 Detailing Expenditure$_{t-3}$	0.22	0.02	0.26	0.16	0.03	0.10	0.01	0.07	-0.16	0.22	0.05	0.26	0.22	0.09	0.13	0.26	0.21	-0.07	-0.03	-0.02	0.11	0.07	-0.08	0.02	1.00			
Patent Applications$_t$	0.40	0.15	0.34	0.12	-0.00	0.10	-0.01	0.03	-0.11	0.10	-0.00	0.06	0.07	-0.00	0.03	0.07	0.06	0.09	0.17	0.35	0.07	0.04	0.00	0.03	0.13	1.00		
Total Other Sales$_t$	0.42	-0.01	0.46	0.11	-0.05	0.08	-0.04	0.07	-0.24	0.28	0.01	0.11	0.09	-0.10	0.03	0.11	0.08	-0.15	-0.09	0.04	-0.08	0.03	0.04	-0.00	0.40	0.17	1.00	
Total Products$_t$	0.03	-0.02	0.00	-0.06	-0.09	-0.02	-0.06	-0.03	-0.08	0.02	0.03	-0.06	-0.04	-0.13	-0.03	-0.06	-0.04	-0.15	-0.12	-0.04	-0.12	-0.01	0.02	0.00	0.05	-0.01	0.13	1.00

Notes: This table reports the pairwise correlations for all variables, including those that only feature in the Appendix.

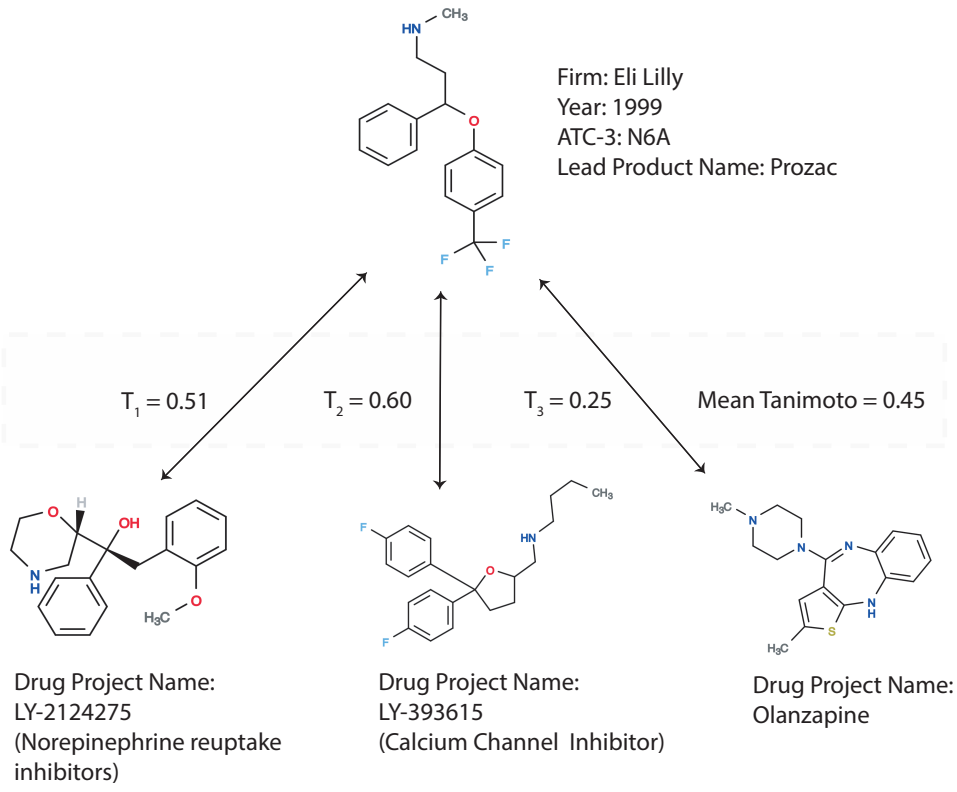
Figure C.1: Calculating Target-Based Action Similarity Of Drug Development Projects to the Firm's Lead Product in an ATC-3–Year



Notes: This figure describes our approach to calculating the target-based action similarity between a firm's lead product in an ATC-3–Year to their ongoing early-stage development projects within the same ATC-1 in year $t + 5$. In this example, we identify Pfizer's lead-selling product in the ATC-3 C10A in 2001, which is Lipitor. We then compare the target-based action of Lipitor to all early-stage drug development projects Pfizer has ongoing in 2006 within the same ATC-1 as Lipitor, which would be C (Cardiovascular). For this illustration, we only show three projects. We then compare the target-based action of Lipitor to all drug development projects and calculate two measures. The first is a simple count of the number of projects that have the same target or target combination as the lead product. In this example, this is 0. Our second measure looks at target overlap, which counts the number of drug development projects that share at least one common target-based action with the lead product. In this example, target overlap would equal 2 as two projects also target HMG CoA reductase.

abrupt changes, as observed with Prozac in 2001, Zoloft in 2006, and Effexor in 2008. The average lag-1 autocorrelation coefficient at the product level is 0.61, lower than at the ATC3 level, reflecting the disruptive impact of such events.

Figure C.2: Calculating Molecular Similarity Of Drug Development Projects to the Firm's Lead Product in an ATC-3–Year



Notes: This figure illustrates our approach to calculating the molecular similarity between a firm's lead product in an ATC-3–Year to their ongoing early stage development projects within the same ATC-1 in year $t+5$. In this example, we identify Eli Lilly's lead-selling product in the ATC-3 N6A in 1999, which is Prozac. We then compare the chemical formulae of Prozac to all early-stage drug development projects Eli Lilly has ongoing in 2004 within the same ATC-1 as Prozac, which would be N (Neurology). For this illustration, we only show three projects. We compute the Tanimoto score (T_i) between all Prozac–project pairs. Our measure of molecular similarity is a simple average across these Tanimoto scores.

Table C.3: Validating the Measure of Commercial Potential with Forward Citations

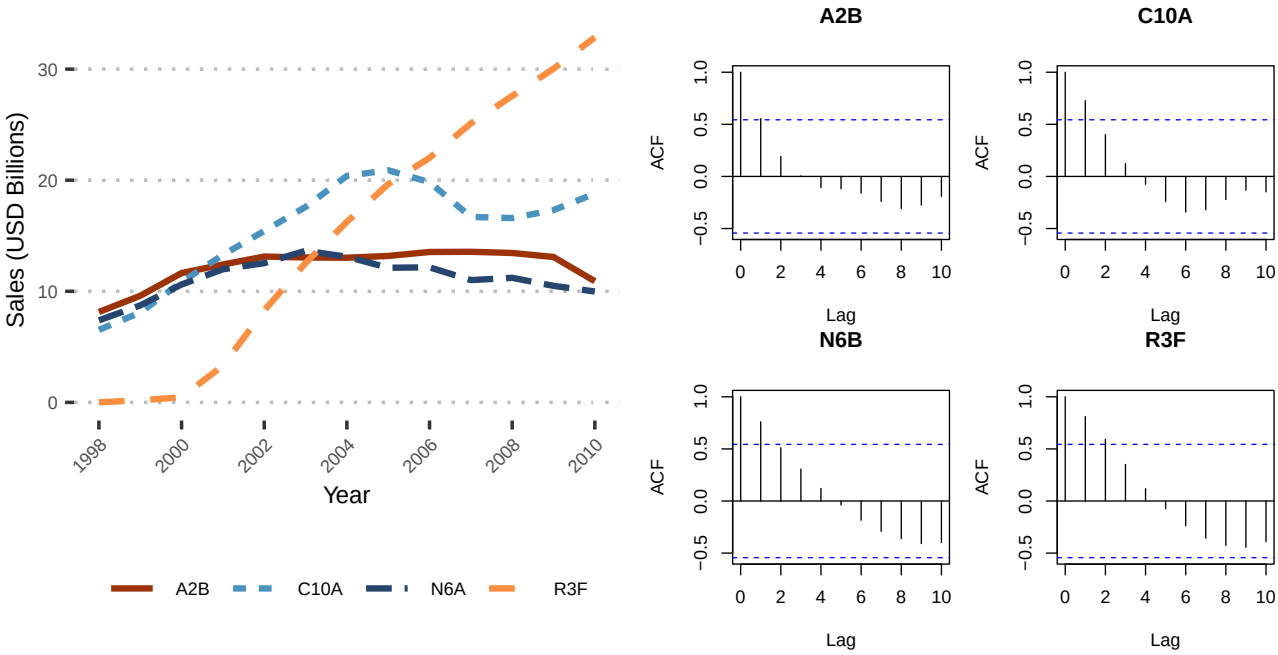
	Count of Forward Cites								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Mean Publication CP	6.586***	23.103***	43.341***						
	(0.638)	(1.103)	(1.475)						
Max Publication CP				6.890***	22.275***	42.295***			
				(0.661)	(1.111)	(1.489)			
Count Publication CP > 0.90							3.303***	15.863***	20.206***
							(0.134)	(0.370)	(0.444)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
ATC1 FE	No	Yes	No	No	Yes	No	No	Yes	No
ATC3 FE	No	No	Yes	No	No	Yes	No	No	Yes
Observations	480,642	255,539	201,742	480,642	255,539	201,742	480,642	255,539	201,742
Adj R-squared	0.013	0.027	0.032	0.013	0.027	0.032	0.014	0.032	0.038

Notes: This table reports validation results for the measure of commercial potential created by (Masclans et al., 2024). Here, we create a patent-level dataset using data from Dimensions. For each patent we count the number of forward citations (other patents that cite the focal patent), and identify all scientific publications that the patent cites. We restrict our attention to only those patent-paper pairs for which the paper is included in our paper-therapeutic class data. We compute the Masclans et al. (2024) commercial potential score for each publication, and then for each patent we aggregate these commercial potential scores in three different ways, by computing: the mean, the max, and the count of papers that have a commercial potential score greater than 0.90. We then regress forward cites on each aggregate measure of commercial potential. In columns (1), (4), and (7), we only include year fixed effects. In columns (2), (5), and (8), we include ATC-1 fixed effects for those patents whose publications only map to one ATC-1. We thus drop any patent observations where more than one distinct ATC-1 is attributed to a patent. Columns (3), (6), and (9) are similar but replace ATC-1 with ATC-3 fixed effects, where one and only unique code is attributed to a patent.

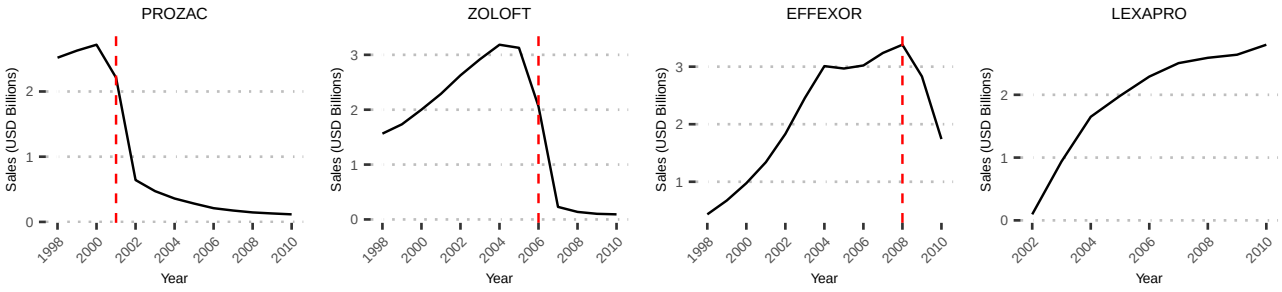
Figure C.3: Sales Stickiness at the Market and Product Levels

(a) Annual ATC3 Sales

(b) Autocorrelation Functions (ACFs)



(c) Product-level Annual Sales (ATC3 N6A)



Notes: In panel C.3c, the dashed red lines represent the years in which the main patent for each drug expired. The main patent for Prozac expired in August 2001, Zoloft in June 2006, and Effexor in 2008. Lexapro went off patent in March 2012, but this falls outside of our sample period.

D Robustness Analysis

Table D.1: Baseline results for Early-Stage Project Count$_{t+5}$ with alternate independent variables and controls.

	Early-Stage Projects$_{t+5}$				
	(1)	(2)	(3)	(4)	(5)
Sales Share Lead Product/Firm$_t$	1.980***				
	(0.450)				
\$3Bn Lead Product$_t$		3.395***			
		(0.426)			
Lead Product Sales$_t$ (1998 dollars)			1.545***		
			(0.087)		
\$1Bn Lead Product$_t$ (1998 dollars)				2.968***	
				(0.216)	
Lead Product Sales$_t$					1.384***
					(0.084)
Cash Flow$_t$					-0.006**
					(0.003)
R&D Expenditure$_t$					0.017***
					(0.004)
Patent Applications$_t$	0.503***	0.503***	0.498***	0.498***	0.519***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.009)
ATC1 Detailing Expenditure$_{t-3}$	2.986***	2.673***	0.241	1.388*	-0.433
	(0.775)	(0.773)	(0.785)	(0.780)	(0.930)
Total Other Sales$_t$	-0.136***	-0.130***	-0.103***	-0.121***	-0.101***
	(0.013)	(0.013)	(0.013)	(0.013)	(0.016)
Total Products$_t$	0.811***	0.820***	0.835***	0.824***	0.635***
	(0.113)	(0.113)	(0.113)	(0.113)	(0.150)
ATC1 × Year FE	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes
Observations	28,158	28,158	28,158	28,158	21,424
Adj R-squared	0.410	0.410	0.416	0.413	0.379

Table D.2: Baseline results for Target Overlap$_{t+5}$ with alternate independent variables and controls.

	Target Overlap$_{t+5}$				
	(1)	(2)	(3)	(4)	(5)
Sales Share Lead Product/Firm$_t$	1.534***				
	(0.078)				
\$3Bn Lead Product$_t$		0.063			
		(0.069)			
Lead Product Sales$_t$ (1998 dollars)			0.237***		
			(0.015)		
\$1Bn Lead Product$_t$ (1998 dollars)				0.519***	
				(0.035)	
Lead Product Sales$_t$					0.186***
					(0.014)
Cash Flow$_t$					0.000
					(0.001)
R&D Expenditure$_t$					0.001
					(0.001)
Patent Applications$_t$	0.017***	0.017***	0.016***	0.016***	0.016***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
ATC1 Detailing Expenditure$_{t-3}$	1.254***	1.714***	1.062***	1.233***	1.386***
	(0.154)	(0.155)	(0.158)	(0.156)	(0.181)
Total Other Sales$_t$	-0.013***	-0.021***	-0.012***	-0.015***	-0.012***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Total Products$_t$	0.060***	0.045**	0.057**	0.053**	0.091***
	(0.023)	(0.023)	(0.023)	(0.023)	(0.030)
ATC1 × Year FE	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes
Observations	16,360	16,360	16,360	16,360	12,294
Adj R-squared	0.141	0.120	0.135	0.132	0.123

Table D.3: Baseline results for Mean Tanimoto Score$_{t+5}$ with alternate independent variables and controls.

	Mean Tanimoto Score$_{t+5}$				
	(1)	(2)	(3)	(4)	(5)
Sales Share Lead Product/Firm$_t$	0.140***				
	(0.012)				
\$3Bn Lead Product$_t$		0.044***			
		(0.010)			
Lead Product Sales$_t$ (1998 dollars)			0.015***		
			(0.002)		
\$1Bn Lead Product$_t$ (1998 dollars)				0.025***	
				(0.005)	
Lead Product Sales$_t$					0.010***
					(0.002)
Cash Flow$_t$					0.000
					(0.000)
R&D Expenditure$_t$					-0.000
					(0.000)
Patent Applications$_t$	-0.000	-0.000	-0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
ATC1 Detailing Expenditure$_{t-3}$	0.024	0.055**	0.026	0.045*	0.062***
	(0.023)	(0.023)	(0.023)	(0.023)	(0.024)
Total Other Sales$_t$	0.001**	0.000	0.001	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Total Products$_t$	-0.001	-0.001	-0.001	-0.001	0.000
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
ATC1 × Year FE	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes
Observations	13,420	13,420	13,420	13,420	10,977
Adj R-squared	0.151	0.144	0.146	0.144	0.126

Table D.4: Baseline results with alternative estimation techniques and dependent variables

	Early-Stage Project Count$_{t+5}$		Target Overlap$_{t+5}$		Early-Stage Project Share$_{t+5}$	Late-Stage Project Count$_{t+5}$	Same Target Count$_{t+5}$	Target Overlap Share$_{t+5}$	Same Target Share$_{t+5}$	Min Tanimoto Score$_{t+5}$	Max Tanimoto Score$_{t+5}$
	(1) Poisson	(2) Negative Binomial	(3) Poisson	(4) Negative Binomial	(5) OLS	(6) OLS	(7) OLS	(8) OLS	(9) OLS	(10) OLS	(11) OLS
main											
Lead Product Sales$_t$	0.107***	0.103***	0.271***	0.475***	0.036***	1.112***	0.094***	0.007***	0.003***	0.010***	0.037***
	(0.011)	(0.015)	(0.027)	(0.049)	(0.002)	(0.063)	(0.007)	(0.001)	(0.001)	(0.002)	(0.003)
Patent Applications$_t$	0.059***	0.080***	0.032***	0.041***	0.007***	0.352***	0.007***	0.000**	0.000	-0.001***	0.000
	(0.001)	(0.002)	(0.005)	(0.006)	(0.000)	(0.007)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
ATC1 Detailing Expenditure$_{t-3}$	0.086	0.357**	3.160***	3.134***	0.053**	0.973	0.280***	0.032***	0.011	-0.104***	0.180***
	(0.159)	(0.173)	(0.652)	(0.656)	(0.027)	(0.684)	(0.090)	(0.011)	(0.008)	(0.022)	(0.036)
Total Other Sales$_t$	-0.016***	-0.021***	-0.031***	-0.037***	-0.002***	-0.052***	-0.008***	-0.000	-0.000	-0.001*	0.000
	(0.003)	(0.003)	(0.012)	(0.012)	(0.000)	(0.012)	(0.002)	(0.000)	(0.000)	(0.000)	(0.001)
Total Products$_t$	0.299***	0.306***	0.302**	0.372***	0.008**	0.003	0.033***	-0.001	-0.001	0.023***	0.022***
	(0.029)	(0.028)	(0.151)	(0.129)	(0.004)	(0.098)	(0.013)	(0.002)	(0.001)	(0.004)	(0.007)
ATC1 × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	28,139	28,161	13,596	16,362	28,158	28,158	16,360	16,360	16,360	13,420	13,420
Pseudo R-squared	0.351	0.130	0.176	0.154							
Adj R-squared					0.095	0.416	0.094	0.156	0.092	0.327	0.257

Table D.5: Full Model

	Early-Stage Projects$_{t+5}$	Target Overlap$_{t+5}$	Mean Tanimoto Score$_{t+5}$
	(1)	(2)	(3)
Lead Product Sales$_t$	1.066***	0.186***	0.013***
	(0.061)	(0.012)	(0.002)
Biotechs$_t$	1.021***	0.007***	-0.002***
	(0.010)	(0.002)	(0.000)
Number of Competitors$_t$	-0.010***	-0.002***	0.001***
	(0.003)	(0.001)	(0.000)
NCE expiration$_t$ (<5yr)	0.641***	0.254***	0.030***
	(0.119)	(0.023)	(0.004)
Patent Applications$_t$	0.213***	0.014***	0.000
	(0.007)	(0.002)	(0.000)
ATC1 Detailing Expenditure$_{t-3}$	1.722**	1.013***	0.021
	(0.669)	(0.157)	(0.023)
Total Other Sales$_t$	-0.065***	-0.011***	0.001*
	(0.011)	(0.003)	(0.000)
Total Products$_t$	0.688***	0.054**	-0.001
	(0.095)	(0.023)	(0.005)
Constant	-1.008***	0.097*	0.358***
	(0.210)	(0.052)	(0.008)
ATC1 × Year FE	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes
Observations	27,877	16,360	13,420
Adj R-squared	0.584	0.141	0.153

Notes: Standard errors shown in parentheses. Significance codes: * p<.1, ** p<.05, *** p<.01

Table D.6: The Blockbuster Effect in "Small" and "Large" Firms

	Annual Sales > $10Bn.$			Annual Sales ≤ $10Bn.$		
	(1) Early-Stage Projects$_{t+5}$	(2) Target Overlap$_{t+5}$	(3) Mean Tanimoto Score$_{t+5}$	(4) Early-Stage Projects$_{t+5}$	(5) Target Overlap$_{t+5}$	(6) Mean Tanimoto Score$_{t+5}$
Lead Product Sales$_t$	1.478***	0.187***	0.011***	0.757***	0.230***	0.016***
	(0.125)	(0.019)	(0.002)	(0.095)	(0.019)	(0.003)
Patent Applications$_t$	0.710***	0.017***	-0.000	0.344***	0.015***	-0.000
	(0.015)	(0.003)	(0.000)	(0.008)	(0.002)	(0.000)
ATC1 Detailing Expenditure$_{t-3}$	-0.736	1.125***	0.047	2.042*	1.669***	0.111**
	(1.535)	(0.284)	(0.029)	(1.059)	(0.245)	(0.045)
Total Other Sales$_t$	-0.107***	0.006	-0.001	-0.038	-0.007	0.004***
	(0.034)	(0.006)	(0.001)	(0.028)	(0.006)	(0.001)
Total Products$_t$	-0.686**	0.153***	-0.008	0.488***	-0.080***	-0.007
	(0.303)	(0.058)	(0.008)	(0.133)	(0.030)	(0.008)
ATC1 × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	8,240	4,657	4,061	19,916	11,691	9,355
Adj R-squared	0.338	0.155	0.100	0.351	0.129	0.166

Table D.7: The Blockbuster Effect across ATC-1s: DV = Early-Stage Project Count$_{t+5}$

	Alimentary tract and Metabolism	Blood and Blood forming Organs	Cardiovascular System	Dermatologicals	Hormonal Preparations	Anti-Infectives	Antineoplastic and Immunomodulating Agents	Muscular-Skeletal System	Nervous System	Respiratory System	Sensory Organs
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Lead Product Sales$_t$	-0.709***	2.025***	2.499***	5.525***	0.439	1.896***	0.873*	2.193***	1.054***	0.010	2.413*
	(0.170)	(0.332)	(0.131)	(1.757)	(1.137)	(0.301)	(0.501)	(0.259)	(0.149)	(0.196)	(1.347)
Patent Applications$_t$	0.513***	0.499***	0.455***	0.272***	0.397***	0.083***	1.184***	0.324***	0.396***	0.436***	0.689***
	(0.023)	(0.034)	(0.020)	(0.021)	(0.048)	(0.019)	(0.042)	(0.029)	(0.018)	(0.022)	(0.032)
ATC1 Detailing Expenditure$_{t-3}$	-8.901**	-43.082*	13.891***	69.329***	82.837	13.820***	-248.008**	-6.676	11.267***	0.861	-41.322***
	(3.838)	(22.564)	(2.689)	(12.264)	(55.259)	(4.067)	(107.436)	(4.303)	(3.821)	(3.775)	(13.738)
Total Other Sales$_t$	0.009	-0.100*	-0.236***	-0.016	-0.028	0.047	-0.155	-0.119***	-0.201***	-0.133***	-0.217***
	(0.034)	(0.060)	(0.038)	(0.036)	(0.059)	(0.031)	(0.097)	(0.045)	(0.037)	(0.039)	(0.073)
Total Products$_t$	0.388	1.109*	1.210***	1.194***	-0.721*	0.532*	-0.663	1.018***	0.519*	0.802**	2.861***
	(0.269)	(0.588)	(0.313)	(0.290)	(0.420)	(0.288)	(0.777)	(0.348)	(0.297)	(0.349)	(0.582)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,551	1,340	3,375	2,583	762	2,507	1,814	1,448	3,878	2,269	1,347
Adj R-squared	0.426	0.484	0.535	0.540	0.418	0.539	0.422	0.456	0.437	0.524	0.506

Table D.8: The Blockbuster Effect across ATC-1s: DV = Target Overlap$_{t+5}$

	Alimentary tract and Metabolism	Blood and Blood forming Organs	Cardiovascular System	Dermatologicals	Hormonal Preparations	Anti-Infectives	Antineoplastic and Immunomodulating Agents	Muscular-Skeletal System	Nervous System	Respiratory System	Sensory Organs
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Lead Product Sales$_t$	0.118***	0.046	0.117***	-0.972**	0.533***	0.496***	0.122***	0.451***	0.314***	0.179***	0.497**
	(0.042)	(0.033)	(0.017)	(0.412)	(0.178)	(0.054)	(0.044)	(0.044)	(0.024)	(0.046)	(0.195)
Patent Applications$_t$	0.015**	0.009*	0.005	0.019***	0.021**	0.007*	0.024***	0.002	0.006*	0.040***	0.020***
	(0.007)	(0.005)	(0.003)	(0.006)	(0.009)	(0.004)	(0.003)	(0.005)	(0.003)	(0.006)	(0.005)
ATC1 Detailing Expenditure$_{t-3}$	4.013***	8.718***	-0.246	1.756	13.222	3.050***	-12.299	-3.294***	1.034	-4.221***	1.952
	(1.259)	(3.037)	(0.407)	(3.707)	(9.607)	(0.854)	(9.182)	(0.843)	(0.686)	(1.059)	(2.041)
Total Other Sales$_t$	-0.012	-0.001	-0.004	0.010	0.023**	-0.013*	-0.016*	-0.014	-0.033***	0.008	-0.043***
	(0.010)	(0.007)	(0.006)	(0.011)	(0.011)	(0.007)	(0.009)	(0.009)	(0.007)	(0.011)	(0.012)
Total Products$_t$	0.001	0.142*	-0.097**	0.071	-0.045	0.068	-0.023	0.135*	0.076	0.353***	0.298***
	(0.080)	(0.081)	(0.047)	(0.079)	(0.073)	(0.061)	(0.069)	(0.069)	(0.054)	(0.098)	(0.093)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,788	762	2,154	1,413	445	1,474	1,236	834	2,379	1,327	786
Adj R-squared	0.138	0.164	0.267	0.265	0.357	0.262	0.191	0.288	0.373	0.469	0.213

Table D.9: The sensitivity of the Blockbuster Effect to Detailing

	Early-Stage Projects$_{t+5}$			Target Overlap$_{t+5}$			Mean Tanimoto Score$_{t+5}$		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Lead Product Sales$_t$	1.259***	1.252***	0.523***	0.211***	0.187***	0.343***	0.013***	0.013***	0.031***
	(0.070)	(0.072)	(0.139)	(0.012)	(0.012)	(0.023)	(0.002)	(0.002)	(0.004)
ATC1 Detailing Expenditure$_{t-3}$		0.341	-0.543		1.100***	1.382***		0.024	0.052**
		(0.785)	(0.797)		(0.158)	(0.161)		(0.023)	(0.024)
Lead Product Sales$_t$ × ATC1 Detailing Expenditure$_{t-3}$			6.632***			-1.398***			-0.150***
			(1.075)			(0.176)			(0.028)
Patent Applications$_t$	0.500***	0.498***	0.499***	0.017***	0.017***	0.016***	-0.000	-0.000	-0.000
	(0.007)	(0.007)	(0.007)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)
Total Other Sales$_t$	-0.103***	-0.105***	-0.105***	-0.012***	-0.013***	-0.013***	0.001*	0.001*	0.001*
	(0.013)	(0.013)	(0.013)	(0.003)	(0.003)	(0.003)	(0.000)	(0.000)	(0.000)
Total Products$_t$	0.812***	0.835***	0.835***	0.057**	0.057**	0.058**	-0.001	-0.001	-0.001
	(0.111)	(0.113)	(0.113)	(0.023)	(0.023)	(0.023)	(0.004)	(0.005)	(0.005)
ATC1 × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	29,291	28,158	28,158	17,014	16,360	16,360	14,020	13,420	13,420
Adj R-squared	0.415	0.415	0.416	0.128	0.133	0.137	0.143	0.146	0.148

Table D.10: Seemingly Unrelated Regression (SUR)

	Small Molecule Sales	Large Biologic Sales
	(1)	(2)
	Count of Small Molecule Projects	Count of Small Molecule Project
Lead Product Sales$_t$	1.132***	1.002***
	(0.076)	(0.308)
Patent Applications$_t$	0.362***	0.704***
	(0.009)	(0.036)
ATC1 Detailing Expenditure$_{t-3}$	4.288***	-13.685*
	(0.964)	(7.206)
Total Other Sales$_t$	-0.091***	0.061
	(0.017)	(0.060)
Total Products$_t$	1.074***	0.733
	(0.151)	(0.680)
	Count of Large Biologic Projects	Count of Large Biologic Projects
Lead Product Sales$_t$	0.051	1.189***
	(0.036)	(0.172)
Patent Applications$_t$	0.086***	0.349***
	(0.004)	(0.020)
ATC1 Detailing Expenditure$_{t-3}$	-0.933**	11.910***
	(0.461)	(4.020)
Total Other Sales$_t$	-0.070***	-0.044
	(0.008)	(0.034)
Total Products$_t$	-0.373***	0.784**
	(0.072)	(0.379)
ATC1 × Year FE	Yes	Yes
Firm FE	Yes	Yes
Observations	14,192	913
R-squared	0.396	0.581

Notes: Standard errors are shown in parentheses. Significance codes: * p<.1, ** p<.05, *** p<.01

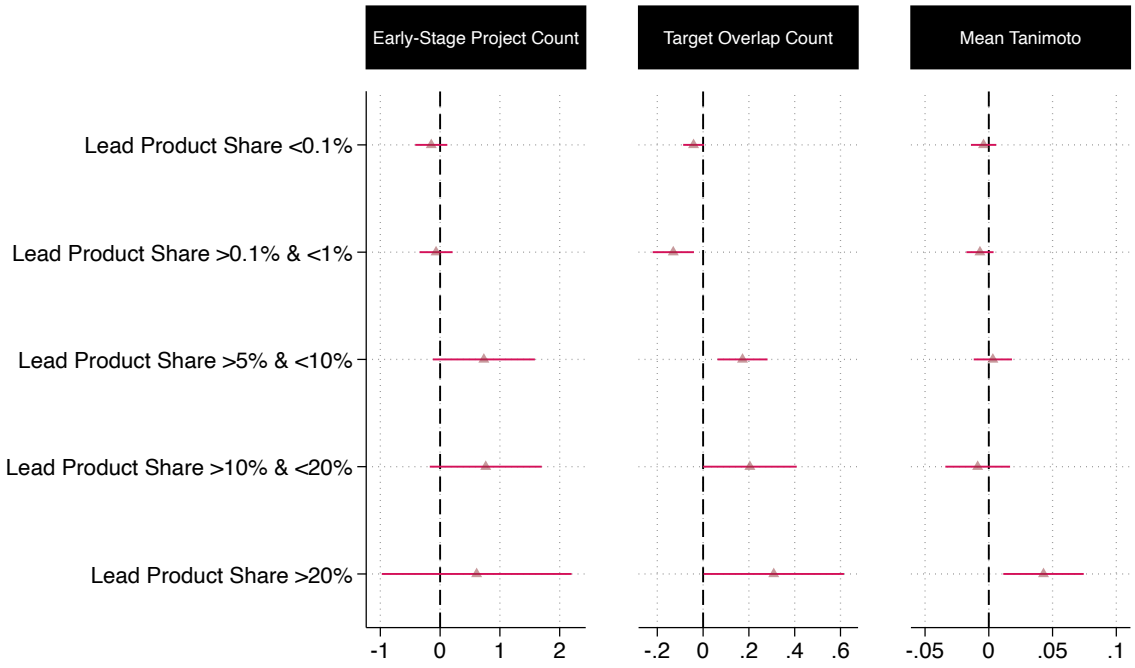
Table D.11: The main effect result from Table 4 where the key controls are added sequentially

	Early-Stage Projects$_{t+5}$						Target Overlap$_{t+5}$					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Lead Product Sales$_t$	1.496***	1.251***	1.291***	1.456***	1.502***	1.050***	0.217***	0.210***	0.186***	0.192***	0.218***	0.155***
	(0.075)	(0.070)	(0.092)	(0.078)	(0.075)	(0.091)	(0.012)	(0.012)	(0.013)	(0.012)	(0.012)	(0.014)
Patent Applications$_t$		0.498***				0.491***		0.017***				0.019***
		(0.007)				(0.011)		(0.001)				(0.002)
Scientific Publications$_t$			0.023***			0.020***			-0.000			-0.000
			(0.002)			(0.002)			(0.000)			(0.000)
ATC1 Detailing Expenditure$_{t-3}$				1.946**		1.902				1.171***		1.323***
				(0.845)		(1.223)				(0.158)		(0.209)
Total Products$_t$					0.581***	1.065***					0.047**	0.071**
					(0.119)	(0.181)					(0.023)	(0.032)
Total Other Sales$_t$	-0.059***	-0.116***	-0.075***	-0.060***	-0.050***	-0.132***	-0.010***	-0.012***	-0.013***	-0.011***	-0.009***	-0.015***
	(0.014)	(0.013)	(0.023)	(0.014)	(0.014)	(0.022)	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.004)
ATC1 × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	29,291	29,291	14,451	28,158	29,291	14,197	17,014	17,014	9,403	16,360	17,014	9,237
Adj R-squared	0.321	0.414	0.363	0.321	0.322	0.446	0.121	0.128	0.136	0.126	0.121	0.150

Notes: Standard errors are shown in parentheses. Significance codes: * p<.1, ** p<.05, *** p<.01

Figure D.1: Coefficient Plots of Sales Share on Incremental Innovation Measures

(a) Full Sample



(b) Absolute Lead Product Sales are less than $1 Billion

