Blockbusters, Sequels, and the Nature of Innovation*

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

June 17, 2026

Abstract

Using detailed product- and invention-level data from the pharmaceutical industry, we demonstrate that firms with high-selling "blockbuster" drugs concentrate their development efforts on new drugs that target the same customer segments and are technologically similar to existing offerings. We propose that this pattern can be explained by an expectation of "demand stickiness", whereby firms expect current sales to persist when follow-on products offer similar functionality, making investment in the development of similar technology within that customer segment more likely. Firms' marketing and R&D capabilities reinforce this behavior, but they do not fully explain it. By shifting the focus from firm-level characteristics to product-level drivers of innovation, our findings highlight how the commercial success of individual products shapes the direction of technological change in R&D-intensive industries.

*We would like to thank Ashish Arora, Victor Bennett, Evgenii Fadeev, Dan Gross, Sharique Hasan, Connie Helfat, Felix Poege, David Ridley, and Todd Zenger, as well as seminar participants at Duke University, University of Utah, NBER Productivity Lunch, and Munich Summer Institute for helpful comments and suggestions. 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 support from the Louisiana Board of Regents Endowed Chairs for Eminent Scholars program. Author names are listed in alphabetical order.

†Duke University, Fuqua School of Business & NBER

‡Freeman School of Business, Tulane University, NBER & Max Planck Institute for Innovation and Competition

§Duke University, Fuqua School of Business

1 Introduction

Since Schumpeter (1942), the relationship between the nature of innovation and firm characteristics, particularly size and capabilities, has long preoccupied scholars of technological change (Cohen, 2010). This literature largely treats the firm as the relevant unit of analysis, implicitly assuming that innovation choices reflect firm-level attributes. However, firms typically operate portfolios of products with highly heterogeneous commercial performance, where a small number of "blockbuster" products account for a disproportionate share of sales. This heterogeneity raises the possibility that innovation is shaped not only by firm-level characteristics but also by the success of individual products within the firm.

In the aerospace industry, for example, Boeing has produced four successive generations of 737 aircraft, first introduced in 1967, all characterized by incremental modifications to the same basic short-haul, narrow-body design.[1] Similarly, in the smartphone industry, Apple released the iPhone in 2007, and nearly 20 years later, the company's main focus remains the next generation of its flagship product.[2] In these examples, firms concentrate their R&D efforts around their highest-selling products by developing new offerings that rely on similar technological foundations and serve the same customer segments. This pattern is not obvious. Standard accounts of innovation emphasize firm-level capabilities or market opportunities, yet they offer limited explanations for why innovation within the same firm systematically concentrates around a small number of highly successful products rather than diversifying across alternative technological approaches.

We begin by documenting these patterns in the pharmaceutical industry, a setting that provides unusually rich data on both annual product-level sales and firms' subsequent investment in development activity within narrowly defined customer segments (e.g., Branstetter et al., 2022; Dranove et al., 2022; Higgins et al., 2021; Branstetter et al., 2016). We combine annual product sales with clinical development activity from 1998 to 2010 across the U.S., U.K., Germany, and France, allowing us to track commercial success to firms' later investments within narrowly defined therapeutic segments. We measure follow-on investment as the number of clinical candidates a firm initiates in the same customer segment five years later.[3] We also measure the extent to which new

[1] https://www.nytimes.com/2019/04/08/business/boeing-737-max-.html

[2] https://www.wsj.com/tech/why-apples-ai-success-hinges-on-the-iphone-a0fc26c4

[3] In the pharmaceutical industry, follow-on innovation often refers to incremental product development where first-in-class drugs are refined to improve safety, efficacy, and patient adherence, often by finding new formulations or dosage forms. In our context, we define follow-on innovation more broadly, capturing any and all subsequent product development targeting the same customer segment as an existing branded drug. This broader definition allows us to empirically explore the similarity between follow-on products and existing branded drugs.

drug development projects resemble the firm's existing high-selling products along two technological dimensions: shared biological targets (e.g., Dranove et al., 2022), which capture similarity in therapeutic mechanism, and molecular similarity, measured using Tanimoto scores (e.g., Krieger et al., 2022).

We document two key regularities in the data. First, higher-selling products are followed by substantially greater clinical development within the same therapeutic area. Products with less than $100 million in annual sales are associated with fewer than two subsequent clinical candidates, compared with approximately five for products exceeding $1 billion. Second, follow-on candidates are disproportionately similar to the firm's high-selling products with respect to both therapeutic targets and chemical structure. Taken together, these findings reveal a strong link between individual product sales and reinvestment within the same customer segment using technology that resembles that embodied in the high-selling existing product.

Having established these regularities, we consider the forces that may explain this pattern. It is not immediately clear whether these patterns reflect firms' existing capabilities or their responses to persistent demand conditions. Adopting a Marshallian perspective, we examine both supply- and demand-side influences. On the supply side, firms' innovation may reflect their existing capabilities (Barney, 2001), including R&D (Cohen and Levinthal, 1990; Henderson and Cockburn, 1994) and marketing capabilities (Chan et al., 2007). On the demand side, firms may respond to customer preferences, particularly if demand within a segment is persistent over time (Adner and Snow, 2010; Ahuja et al., 2014; Wang et al., 2020).

First, we consider an exclusively supply-side explanation for why subsequent innovation will target the same customer segment with similar technology. When a product achieves high sales, firms may already possess or invest in marketing, distribution, and other capabilities tailored to a customer segment. Consistent with the resource-based view (RBV) of the firm (Penrose, 1959; Teece, 1986; Helfat, 1997; Chan et al., 2007; Barney, 2001; Henderson and Cockburn, 1994), firms may consequently have an incentive to develop products that target the same customer segment in order to fully utilize these assets. Similarly, with regard to our finding of technological similarity between a high-selling drug and the drug candidates that are under development, high sales within a customer segment may reflect a firm's possession of technological capabilities relevant to that customer segment. If these capabilities are both firm-specific and lumpy, their possession increases the expected returns to developing follow-on products that build on the same technological foundations. Firms will have an incentive to develop technologically similar products in order to more

efficiently utilize such capabilities (Tripsas and Gavetti, 2000).

Second, our demand-side account centers on expectations about future demand. Firms may anticipate that their current product will not sustain its performance indefinitely due to competition or other factors (Grabowski and Vernon, 1992; Aghion et al., 2005; Argente et al., 2018; Argente and Yeh, 2022). At the same time, they may expect demand within a segment to be "sticky," such that a follow-on product addressing similar needs and preferences is likely to achieve comparably strong sales (Dubé et al., 2010; Bronnenberg et al., 2012). These expectations of persistent demand and the associated revenue stream create incentives to reinvest in the same narrowly defined customer segment.[4] Note that we use demand stickiness broadly to refer to persistence in demand within a customer segment, including but not limited to repeat purchases by the same individual consumers.

If firms are incentivized to develop products for the same customer segment as their high-selling product(s), a key question is what technologies will then be pursued. In particular, how might those technologies relate to those embodied in the high-selling product? We argue that, for a follow-on product to achieve comparable sales, it must deliver comparable functionality (Lancaster, 1966; Henderson and Clark, 1990). This objective, in turn, will constrain the set of viable technological approaches to product development. To illustrate, consider that functionality in pharmaceuticals is characterized by a drug candidate's therapeutic efficacy and side effect profile. For example, a new treatment for high cholesterol must rely on chemistry and biological mechanisms relevant to cholesterol metabolism. As a result, products targeting the same customer segment are likely to rely on overlapping technological approaches, even when alternative approaches may exist.

We then turn back to our data to examine evidence for both sets of mechanisms. To test whether demand stickiness can explain these patterns of follow-on development, we consider two potential sources of demand stickiness. The first captures firms' investments in detailing, reflecting the sales-force and physician relationships that may make product demand more persistent (Berndt et al., 1995). The second builds on evidence that adverse side effect events can make physicians reluctant to switch medications, generating persistent prescribing patterns (Janakiraman et al., 2008; Feng and Maini, 2024). We measure this channel using the relative incidence of adverse events among rival products in the same therapeutic class, some fraction of which is plausibly outside the focal firm's own product-development choices.

Our results show that adverse-event-driven demand stickiness is associated with greater follow-

[4] An obvious question, considered below, is the source of demand stickiness. In our view, while demand stickiness may be strictly demand-driven, independent of the actions of the firm, firms may also invest in marketing and related capabilities to create or reinforce stickiness.

on development. When rival products exhibit higher adverse-event incidence, firms initiate more clinical candidates in the same customer segment as high-selling products, and these candidates are more likely to share the focal product's biological target. We find no corresponding increase in molecular similarity when rival products exhibit higher adverse-event incidence. This pattern is consistent with the idea that adverse events often arise from a drug's therapeutic mechanism rather than its precise molecular structure, so drugs acting on the same target may inherit similar safety risks even when their chemical structures differ (Rudmann, 2013; Campillos et al., 2008). Accordingly, safety-driven demand stickiness appears to encourage follow-on development that preserves therapeutic mechanisms without necessarily preserving molecular design. We interpret these findings as reflecting a distinction between targets and means. Firms maintain therapeutic mechanisms while varying the molecular approaches used to achieve them when side-effect concerns are significant. By contrast, we do not find statistically significant effects for detailing, although the correlation between sales and detailing ($r = 0.27$) limits the statistical power of this test.

We also find that existing R&D capabilities are positively associated with investment in technologically similar development projects relative to a firm's existing products, consistent with the RBV. However, the positive relationship between lead product sales and technologically similar follow-on projects remains robust and is not sensitive to the inclusion of these R&D capability measures. This suggests the presence of a meaningful residual relationship that cannot be explained solely by firm-level technical capabilities. Furthermore, when product-level sales and adverse-event-based demand stickiness are included jointly with R&D capabilities and market exclusivity, the interaction between demand stickiness and lead product sales remains predictive of technically similar follow-on development. This indicates that demand stickiness is an important factor in explaining incremental technological advancement following high-selling products.

This paper contributes to at least three conversations in the literature. First, we contribute to the literature on the determinants of incremental innovation by providing a demand-driven explanation that operates at the level of the individual product. In contrast, existing work has primarily emphasized supply-side explanations that focus on firm-level determinants. For example, 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 ability to appropriate returns (Cohen and Klepper, 1996; Arora et al., 2023). Other studies highlight the role of entry and competition, with incumbents tending to invest in incremental innovation while entrants pursue more radical approaches (Schumpeter, 1942; Arrow, 1972; Acemoglu and Cao, 2015; Akcigit and Kerr, 2018).

In this paper, we also show that variation in product-level outcomes within firms systematically shapes the direction of incremental innovation.[5] Specifically, higher sales of a firm's leading product incentivize the development of follow-on products that target the same customer segment, leading firms to pursue technologies that are closely related to those underlying prior successes. Moreover, while prior work typically examines the technical quality of innovation at the firm level, often measured using forward patent citations, we define incremental technical advance with reference to the technology of specific products. This shift highlights the value of product-level analysis for understanding the determinants of the direction of technical change.

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 potentially "disruptive" competitive threats posed by firms entering adjacent but neglected market segments. Unlike our model, which ties the pursuit of a customer segment to incremental technical advance, his analysis does not offer clear 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 firm's main customers.

Finally, our study offers implications for the literature on the decline in R&D productivity, including in the pharmaceutical industry. Prior research has emphasized exogenous supply-side factors, such as longer development timelines and rising costs (Cockburn, 2006; Higgins and Rodriguez, 2006), as well as broader increases in the difficulty of generating new ideas over time (*e.g.*, Bloom et al., 2020; Gordon, 2017). In contrast, we highlight firms' endogenous innovation choices. If firms disproportionately direct innovation toward the same customer segments as their highest-selling products, and in doing so pursue technologies closely related to existing ones, innovation may become increasingly incremental. In the aggregate, such behavior can reduce the rate at which novel technologies are explored, even in the absence of any decline in underlying technological opportunities.

2 The Blockbuster Effect on R&D

We establish two facts that motivate the direction of inquiry in this paper. We first provide a brief background on the pharmaceutical industry, highlighting examples of blockbuster products and subsequent innovations by the firm. We then provide a high-level summary of the data and

[5] In addition to firm-level characteristics, recent work highlights the role of managerial characteristics in shaping whether a firm pursues incremental versus radical innovation (Acemoglu et al., 2022).

measures used to test the prevalence of these phenomena at scale, before showing the descriptive results. An extensive explanation of the data is provided in Section 4.1.

2.1 The Setting: Blockbuster Drugs in the Pharmaceutical Industry

Developing new drugs requires substantial investment and a long time horizon. Estimates of the cost of bringing a new drug to market average $879.3 million (Sertkaya et al., 2024), with upper-bound estimates reaching $2.8 billion (Wouters et al., 2020).[6] Moreover, the full process from molecule discovery to drug approval can take 10 to 15 years (Corr and Williams, 2009; Paul et al., 2010), with clinical trials alone averaging 8.3 years (Brown et al., 2022). Over the past thirty years, the pharmaceutical industry has become increasingly dependent on "blockbuster" drugs, whereby a small number of products generate a disproportionate share of industry revenues (Munos, 2009). Consistent with this trend, the number of drugs with annual sales exceeding $1 billion increased more than four-fold, from 12 in 1998 to 49 in 2010.[7,8,9]

A well-known example of a "blockbuster" pharmaceutical product is AstraZeneca's Prilosec, a proton-pump inhibitor used to treat acid reflux.[10] As shown in Figure 1, in the late 1990s Prilosec generated more than $4 billion in annual sales. As its patents approached expiration, AstraZeneca invested over $150 million in the development of 18 potential successor drugs.[11] These candidates remained close to the original product, including reversible proton-pump inhibitors, liquid and extended-release formulations, and combination therapies with other heartburn treatments. The only drug that emerged as a minor improvement over Prilosec was Nexium. Nexium is an isomer of Prilosec—essentially a refined version of the same compound. Although its therapeutic differences were limited, it was sufficiently distinct to secure new patent protection. Supported by substantial marketing efforts, Nexium ultimately replaced Prilosec's sales and allowed AstraZeneca to retain

[6] These costs have risen over time, in part because the probability that a drug candidate ultimately receives regulatory approval has declined. From 2000 to 2015, Wong et al. (2019) estimate that the probability of eventual approval was 13.8 percent. A more recent analysis covering 2014–2023 finds approval rates closer to 6.7 percent (Citeline, 2024).

[7] This figure is based on our data from IQVIA, which includes branded product sales in the USA, UK, France, and Germany. Figure 3 illustrates these trends.

[8] 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 largely unchanged for over two decades.

[9] Historically, blockbusters treated common conditions affecting large patient populations, such as high cholesterol (e.g., Lipitor). More recently, advances in large molecule biologics have enabled drugs with extremely high prices to achieve blockbuster status even when targeting relatively small patient populations. For example, Stelara reached more than $20 billion in annual sales in 2023 despite treating roughly 150,000 patients in the United States. See `https://www.stelarainfo.com/crohns-disease/about-stelara/`

[10] We discuss additional examples in Appendix H.

[11] `https://www.wsj.com/articles/SB10233263696799108 40?mod=Searchresults&pos=1&page=1`

its customer base in the heartburn market.

In the following subsection, we show that this example reflects a broader pattern in pharmaceutical innovation. That is, the commercial success of a drug is systematically associated with substantial follow-on development efforts focused on the same narrow customer segment and relying on technologically similar approaches. This tendency may channel firms toward innovation trajectories that build on prior commercial successes rather than explore fundamentally different technological approaches.

[Figure 1 about here.]

2.2 Empirical Regularities: Product Sales and Subsequent Drug Development

We document the simple association between product sales and, respectively, the market focus and technical character of drug development. For this purpose, we first provide a brief overview of the data and measures employed, leaving a more detailed description of data sources to Section 4.1.

We combine product-level sales data from IQVIA with clinical trial data from Cortellis. The data are organized at the firm–therapeutic class–year level, where we interpret a therapeutic class as a distinct customer segment, which is defined at either the third or fourth levels of the Anatomical Therapeutic Chemical (ATC) classification system. For each firm–ATC–year observation, we record the sales of the firm's highest-selling product within that therapeutic class. We use the number of drug candidates in clinical trials initiated in the same firm–ATC in year $t + 5$ as a measure of the firm's investment in the same customer segment as an existing product. The five year lag is intended to approximate the initiation of a drug development project.[12]

We use two approaches to determine the technological similarity between a follow-on drug candidate and an existing high-selling product. First, we determine whether the candidate shares the same biological target as the existing product. A drug's biological target is typically a protein whose activity the drug is designed to modulate to produce a therapeutic effect. Second, we calculate the molecular similarity between the candidate and the existing product using Tanimoto scores. The Tanimoto score is a standard metric in scientific research and has recently been adopted to research in the economics of innovation (Krieger et al., 2022). Empirically, we measure these two similarity variables by, respectively, the count of follow-on drug development projects with the same

[12] At scale, entry into clinical trials is the first time a researcher can observe a drug development project in data. The time between initial target selection and the start of Phase 1 trials will vary. For example, Hughes et al. (2011) estimate an average of four years.

biological target as the existing product, and by the count of follow-on projects whose chemistry is at least 85% similar to the existing product.

Figure 2 illustrates the association between product sales and follow-on development. On the x-axis for all three panels we plot the log of lead product sales, and on the y-axis we plot counts of development projects. Panel (a) includes the count of all follow-on drug development projects within the same firm-ATC-3 that have begun clinical testing in year $t + 5$. Panels (b) and (c) then restrict the counts to those follow-on development projects which share similar technological attributes as the existing lead product across both the intended biological target (panel (b)), and chemical composition (panel (c)).

[Figure 2 about here.]

These simple visualizations of the raw data point to a distinct positive association between lead product sales and follow-on development within the same customer segment (*i.e.*, therapeutic market). To quantify the magnitude of the difference, a drug with annual sales less than $100 million is associated with the firm pursuing, on average, less than two follow-on drug candidates in clinical trials. In contrast, products exceeding $1 billion in annual sales are associated with, on average, five follow-on candidates in clinical trials in the same therapeutic area. While the count magnitudes for technologically similar clinical candidates is substantially smaller, we again observe a clear positive trend in both forms of technological similarity as sales of the lead product in a firm-year-ATC-3 increase.

In summary, we observe two empirical patterns. First, as product sales increase, firms invest more heavily in developing new products within the same customer segment as the existing drug. Second, the drugs developed in that segment are increasingly likely to be technologically similar to the original product. In the next section, we develop theoretical explanations for these patterns.

3 Theory

We interpret these empirical patterns through the lens of the Marshallian "scissors", in which outcomes reflect the joint influence of supply- and demand-side forces. In our setting, firms' innovation choices may reflect both the capabilities and co-specialized assets they have accumulated and their expectations about the persistence of demand within a customer segment. These forces are not mutually exclusive and may reinforce one another. Accordingly, we first discuss how existing ca-

pabilities can encourage reinvestment in related products and technologies, and then examine how expectations of sticky demand may independently shape the direction of follow-on innovation.

3.1 Supply-side explanations

One might explain both regularities with a strictly supply-side argument focusing on firm capabilities. First, consider the observation of a positive correlation between product sales and development in the same customer segment. One mechanism relates to the fixed costs associated with building and maintaining a specialized sales force. In the pharmaceutical industry, marketing and sales (i.e., detailing) require highly trained sales representatives in specific disease areas, regulatory constraints, and clinical evidence. Such capabilities are also firm-specific, hard to replace, and difficult to redeploy, and they can constitute an important source of competitive advantage. Consequently, once a firm has built a sales infrastructure targeting physicians within a particular therapeutic class, they may have an incentive to continue to develop products in that particular customer segment in order to efficiently utilize that capability (Chan et al., 2007). Firms may thus prefer to develop additional products within the same segment to leverage these fixed investments.

Next, consider the second empirical regularity that the technology of follow-on drugs tends to resemble the technology of the high-selling drug. An explanation here can also operate through the supply side of the firm's R&D capabilities. High product sales in a customer segment may reflect not only favorable demand conditions but also the firm's possession of segment-specific technological capabilities, such as specialized scientific expertise, proprietary platforms, clinical development experience, and regulatory knowledge. These accumulated capabilities increase the expected returns to projects that build on related technological foundations (Henderson and Cockburn, 1994; Barney, 2001). Developing technologically similar products allows firms to leverage fixed investments in R&D infrastructure, personnel, and accumulated know-how, thereby reducing marginal development costs and technical uncertainty (Tripsas and Gavetti, 2000).

3.2 Demand-side explanations

Another explanation for this pattern could be that, if firms view current sales as an indicator of future demand conditions, then a follow-on product that addresses similar needs and preferences as an existing blockbuster is likely to generate comparable revenues. When R&D involves substantial fixed costs, this persistence in expected sales (i.e., demand stickiness) becomes especially important. Because firms must incur large upfront investments, development is more attractive in

segments where expected sales—and thus profits—are higher due to fixed cost spreading (Cohen and Klepper, 1996). As a result, firms in the pharmaceutical industry have an incentive to concentrate development efforts in the therapeutic areas of their highest-selling products, rather than shifting resources to segments with lower expected demand.

Demand stickiness can arise from demand-side factors that generate persistence in consumption patterns.[13] Demand stickiness can arise, for example, from switching costs tied to customers' complementary investments (Klemperer, 1987).[14] In many industries, consumers also face uncertainty about product quality, and past success serves as a signal of desirability and reliability. For example, in pharmaceuticals, physicians can be reluctant to switch from prescribing drugs that have proven effective, particularly when alternative treatments carry uncertainty about efficacy or side effects. When a high-selling drug within a therapeutic class performs well compared to rival products, firms may expect continued demand.

Demand stickiness may not, however, be purely demand-driven and can be endogenous to the firm. Firms can actively cultivate demand stickiness through their own investments. For instance, pharmaceutical companies invest heavily in marketing and physician detailing to reinforce product adoption and brand recognition. These investments can increase demand stickiness by strengthening both patient inertia and deepening relationships with physicians. If firms anticipate that such investments increase the likelihood that demand will persist, then high sales may trigger both additional commercialization efforts and further R&D in the same therapeutic class.

How might high sales combined with an expectation of sticky demand affect the nature of subsequent technical advance? Our argument is that the firm's decision to pursue the high-selling product's customer segment sets the technological objective (i.e., the agenda) for the follow-on

[13] The idea that demand can persist over time is well established in both the economics and marketing literature. In economics, closely related concepts such as *demand inertia* or *consumer inertia* capture the persistence of customer relationships with specific products or firms (Dubé et al., 2010; Bronnenberg et al., 2012; Gourio and Rudanko, 2014; Hortaçsu et al., 2017; Bornstein, 2025). The marketing literature similarly studies how firms' strategies, such as advertising and product positioning, both shape and respond to this persistence in consumer behavior (Banerjee and Bandyopadhyay, 2003; Chintagunta, 1998). While our notion of demand stickiness is closely related to these ideas, it is conceptually distinct and broader. Specifically, we focus on the persistence of demand at the level of the customer segment, that is, the extent to which the demand conditions supporting strong sales for a given product carry over to subsequent products targeting similar needs and preferences, rather than on repeated purchases by the same individual consumers. In this sense, consumer inertia can be viewed as one mechanism contributing to the broader notion of demand stickiness that we study. In order to address this distinction empirically, we compare settings with high versus low repeat consumption, captured by chronic versus acute disease markets, to assess whether our results are driven primarily by repeated purchases by the same individuals or by persistence in demand at the customer segment level in Table G.16.

[14] For example, in aerospace, airlines incur substantial fixed costs to train pilots and maintain fleets tailored to a specific aircraft model. Such investments create inertia, as switching to a different model entails retraining and operational disruption.

R&D. That objective is to replicate the prior product's functionality in order to address the needs and preferences of the customer segment.

The next question is what means the firm will adopt to satisfy that objective. First, the need to reproduce the successful drug's functionality obviously constrains the set of viable approaches (Peteraf and Bergen, 2003). In pharmaceuticals, for example, a follow-on drug intended to treat the same condition must achieve comparable therapeutic efficacy and side effect profiles. Moreover, firms are more likely to pursue technologies closely related to those embodied in the existing product, even when alternative technological approaches exist. Pursuing entirely distinct technological approaches would typically require additional investment in new scientific capabilities and increase technical uncertainty and development risk, raising the expected gross return required to make the selection of an alternative technology profitable. This implies that the pursuit of sticky demand within a customer segment will bias firms toward approaches proximate to those previously employed (Nerkar and Roberts, 2004). Thus, we are suggesting that an expectation of persistent high demand in a customer segment leads the supply response, both with respect to technological objectives and the means to achieve them.[15]

In our empirical analysis, we accordingly distinguish between two types of technical similarity. The first captures the technological target of innovation, which determines functionality and reflects whether follow-on projects act on the same underlying biological mechanisms. The second captures the means used to achieve this target, reflecting the extent to which projects rely on similar molecular design choices. This distinction allows us to assess whether these two dimensions respond differently to variation in demand stickiness, and to examine patterns consistent with firms defining a target that shapes functionality and then selecting the technological means to achieve it.

3.3 Market-level moderating factors

The above arguments assume that all innovation is internal to the firm. In reality, however, firms may have access to markets for technology. Firms in the pharmaceutical industry may have access to drug candidates via acquisition or licensing. In the pharmaceutical industry, one can think of the thickness of such external technology markets to be reflected by the abundance of biotech firms in a given therapeutic class.

The presence of biotechnology firms can have offsetting impacts on a firm's investment in follow-

[15] Our consideration of the leading role of the demand side responds to Leiponen and Helfat (2010)'s observation that research on innovation has largely neglected the role and determination of R&D objectives in driving the direction of technical advance.

on innovation. On the one hand, a robust supply of biotechnology firms in the same therapeutic market as its high-selling drug could suppress internal investment in development within that customer segment if externally available inventions substitute for technologies the firm could otherwise develop in-house. In this case, firms may prefer to acquire or license promising technologies rather than invest internally, weakening the relationship between product sales and subsequent follow-on innovation. On the other hand, the presence of biotechnology firms within a therapeutic market may provide a form of insurance that encourages firms to double down on internal development related to their high-selling products. Pharmaceutical firms may believe that the risks associated with a blockbuster-focused development strategy are mitigated when a vibrant market for external technologies provides fallback options in the event of late-stage development failures, pipeline gaps, or the emergence of rival technologies (Higgins and Rodriguez, 2006; Danzon et al., 2007; Chan et al., 2007; Hermosilla, 2021). Accordingly, external technology markets may either substitute for or reinforce internally directed follow-on innovation. These countervailing forces yield ambiguous predictions. If acquisition substitutes for internal R&D, we expect a negative interaction between biotechnology presence and the relationship between lead product sales and follow-on development. Conversely, if biotechnology firms provide insurance against pipeline risk, we expect a positive interaction.

Another market-level characteristic to consider is the intensity of direct competition. *A priori*, the influence of competition on the blockbuster effect is not obvious. One might expect that increased competition within a narrowly defined therapeutic class reduces firms' incentives to remain focused on existing customer segments using similar technologies as competitors attempt to capture market share from existing high-selling product lines. Thus, competition may reduce follow-on investment through rivals' business-stealing incentives that encourage differentiation (Arora et al., 2025). On the other hand, more competition within a well-defined market may also signal the attractiveness of an opportunity, causing an incumbent to double down on its investment in its high-selling drugs in that market. Thus competition may increase investment by signaling strong demand and validating the attractiveness of the market.

4 Data, Variables and Methods

In this section, we provide a more comprehensive explanation of the data sources that were briefly introduced in Section 2, and the steps that we take to build our dataset. We then describe the

construction of all variables used in our analysis. Lastly, we detail the empirical specifications we use to test the mechanisms through which individual product sales might explain variation in the customer segment and technological similarity of drug development projects.

4.1 Data

Product level sales and drug development projects are at the core of our dataset. Our product-level sales data come from IQVIA, and they include the sales of all drug products sold in the U.S., U.K., France, and Germany from 1998 to 2010. In this paper, we restrict our focus to the sales of branded therapeutic small molecule (chemical-based) and large molecule (biologic-based) drugs.[16] We also collect drug development data from Cortellis, an extensive dataset covering global drug development activities. The raw data includes 79,309 drug development projects conducted by 13,050 companies from 1980 to 2020.[17]

We match the drug development and sales data across three dimensions: firm, therapeutic class (ATC-3), and year (i, j, t). After merging our data together, our sample consists of 57 corporate parent firms, whose sales of branded therapeutic small molecule and large biologic drugs represent 78.8% of total sales during our sample period. Although all the firms are large, there is still substantial heterogeneity across the firms in our sample. Firms sell, on average, 88.8 branded therapeutic drugs in a year, with annual sales ranging from $505 million to nearly $32.8 billion.[18] On average, firms are actively selling products in 45.3 distinct ATC-3s. Note that Firm–ATC-3 observations are included in our sample if we observe at least one year with non-zero sales or a non-zero count of drug development projects at $t + 5$.[19]

We also collect data from several other sources for control variables and to create variables that test proposed mechanisms. We gather data on marketing expenditures (*i.e.*, detailing) from IQVIA, patent data from the FDA Orange Book, data on severe side effects from the FDA Adverse Event Reporting System, patenting activity and deal data from Cortellis, data on scientific publications matched to therapeutic class from Branstetter et al. (2022), and data from COMPUSTAT for firm

[16] Appendix Figure C.1 shows the distribution of products by their active ingredient classification.

[17] Public registration of clinical trials was not enforced until 2006, so the data are less complete prior to that year. However, as explained in Section 4.3, even though our sales data span from 1998 to 2010, our empirical specification employs forward-looking measures using the drug development data by lagging it five years relative to the sales data, which helps mitigate these data limitations.

[18] We include only firm-years with at least $500 million in sales, as manual inspection revealed that firms below this threshold, though classified as branded therapeutic drug developers, typically sell herbal or homeopathic products. We ran all analyses on a dataset without this exclusion, and found directionally consistent results.

[19] 19.8% of observations have no sales (and only drug development), and 26.7% of observations have no drug development at $t + 5$ (and only sales).

cash flows and R&D expenditures. We report descriptive statistics for all variables used in analysis featured in the main paper in Table 1. See Table D.1 in the Appendix for a correlation matrix.

[Table 1 about here.]

4.2 Variables

4.2.1 Dependent Variables

Customer segment similarity. We measure the customer segment similarity of follow-on drug development by the number of a firm's drug development projects initiated in the same therapeutic area (ATC-3) as the focal product. This measure serves as our best proxy for the dollar amount invested in development and enables us to analyze whether firms with especially high-selling products in an ATC-3 are more likely to invest in new development within the same therapeutic category, *i.e.*, the same customer segment. 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 and t is the year in which the firm's product sales within an ATC-3 are observed.[20] By considering early-stage projects at $t+5$, we intend to capture the projects which were first *initiated* in year t. In other words, we assume that the first time that a drug development project is observed in the data (*i.e.,* phase I trials) is five years after any initial investment decisions, including pre-clinical studies, were made by the firm.[21]

Technical similarity. We adopt two approaches to measuring the technical similarity between products and subsequent development.[22] These measures capture distinct decisions in the development of new drug candidates, occurring at different stages of the drug development process. First, we compare the biological target for firm i's lead product within an 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). Note that an important (and not obvious) feature of our empirical setup is that we focus exclusively on the lead product—that is, the highest-selling product—in each firm, year, ATC-3 observation. We do this for two reasons. First, on average, a firm sells 1.46 branded therapeutic small- or large-molecule drugs in an ATC-3–year (median = 1, st. dev. =

[20] In Appendix G, 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.

[21] Note that our results remain robust to alternative lag structures (see Appendix Table G.1).

[22] See Appendix Figures D.1 and D.2 for illustrated examples of the measure construction.

2.08). Moreover, when multiple drugs are sold in a firm-year-ATC-3, the drugs listed are typically variations in formulation or dosage size on the principal product. Second, 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 set of products. The variable *Target Overlap*$_{t+5}$ is the count of drugs in development with at least one biological target in common with the lead product.[23]

Our second measure of technical similarity uses a Tanimoto score to reflect the molecular similarity between commercialized products and projects under development. 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. We compute 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). We define the variable *Molecularly Similar Projects*$_{t+5}$ as the count of early-stage drug development projects in ATC-1 k that share an 85% or greater overlap in molecular structure with the lead sales product *i.e.,* with a Tanimoto score greater than 0.85.[24] We also create the variable *Same Molecule Projects*$_{t+5}$, which is the count of drug development projects using the same molecule as the lead selling product (i.e., Tanimoto = 1). This measure focuses attention on reformulations of the focal product. We also create a variable to count the number of molecularly similar but distinct projects (i.e., 0.85 < Tanimoto < 1). The main limitation of these variables, however, is that we are restricted to marketed drugs and drug development projects for small molecule therapeutics since chemical formulae do not exist for large molecule biologic drugs.[25]

[23] Our measure captures overlap, since 14.6% of drugs in our development data are designed for *multiple* targets. In additional analysis presented in Appendix Table G.5, we show that our key findings our robust to the more restrictive measure of *same* target projects.

[24] Choosing a cutoff Tanimoto score to classify projects as molecularly similar is a debated topic in the cheminformatics literature (*e.g.*, Maggiora et al., 2014), with the statistical significance of any Tanimoto score depending on the underlying molecular representation itself. Our decision to use 0.85 as the cut-off provides a more conservative estimate compared to related work (*e.g.*, Frankel et al., 2023). In addition, in Appendix G.5 we show how our results are robust to other measure constructs, including the minimum, mean, and maximum Tanimoto scores. The interesting feature of these alternate measures is that they carry a different interpretation. The minimum Tanimoto score reflects the similarity of the *least similar* drug development project. The mean Tanimoto score captures the *average* similarity of drug development projects compared to the focal sales product, and the max Tanimoto score reflects the similarity of the *most similar* drug development project.

[25] These measures capture different aspects of technical similarity but are not exhaustive. For instance, only the target-based action measure includes both small and large molecule drugs, though it reflects similarity in objectives,

These measures reflect two distinct dimensions of technological similarity. Since a drug's binding to a biological target will drive a drug's functionality—its therapeutic efficacy and side effect profile—the binding to a specific biological target will typically constitute the technological objective of a development effort. Thus, when a drug under development has the same target as the existing drug, the firm is essentially trying to replicate (perhaps improve upon) its functionality. Molecular similarity captures a different form of technological similarity. There are typically some number of molecules available internally as well as externally via the technology market that may bind to a given target, yielding potentially different efficacy outcomes. Thus, one might consider a firm's choice of molecules in clinical trials as reflecting the different means for potentially achieving a technological objective–i.e., a given functionality.

4.2.2 Independent Variable

The main independent variable in our analysis is *Lead Product Sales*$_t$, which, for a given firm, refers to the nominal sales (in USD billions) of the highest-selling product in a given ATC-3–year pair (j, t). Note that this can be \$0 if the firm does not sell a product in an ATC-3-year, but is engaged in drug development projects in that ATC-3 in year $t+5$. In additional analyses in Appendix Table G.2, we test our baseline specification with an inflation-adjusted sales variable.

4.2.3 Firm Capabilities

Co-specialized assets. 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).[26] We define *Detailing*$_{t\text{-}3}$ as firm i's total detailing expenditure (in USD billions) in year $t-3$ for ATC-1 j.[27] We lag detailing by three years as contemporaneous sales and detailing expenditure are highly correlated.

not means. Firms may pursue the same targets using technically distinct approaches.

[26] A "lock-in" effect emerges, for example, if the firm decides to invest in an ATC-3 partly to more fully utilize some fixed capacity, such as the detail personnel assigned to that ATC.

[27] We define our detailing variable at the ATC-1 level since 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). In additional analyses presented in Appendix Table G.7, we test other levels of fungibility.

R&D capabilities. Research and development comprise two distinct sets of activities and firm capabilities. On the one hand, research activities focus on the search for and discovery of new drug candidates. On the other hand, development focuses on taking those drug candidates through the clinical trial process. We proxy for a firm's ATC-3-specific research capabilities with the variable *Scientific Publications*$_t$, which is a count of scientific articles published in year t with an author affiliated with firm i that pertain to therapeutic class ATC-3 j. In addition to publications, patents are an important output of research and are critically important in the pharmaceutical industry. As such, we control for a firm's upstream research capabilities and underlying stock of innovations with the count of a firm's patent applications. We define *Patent Applications*$_t$ as the count of patents applied for by firm i in market ATC-3 j in year t.

We proxy for a firm's development capabilities (e.g., experience in drafting clinical trial protocols, 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. As an alternative to all of these measures, we also collect data from COMPUSTAT on the subsample of public firms on *R&D Expenditure*$_t$, which reflects the dollar amount spent on research and development activities reported by the firm in public statements.

Financial resources. We also consider a firm's capability to invest in the development of new products, which will be heavily influenced by the cash that it can draw upon (Hall and Lerner, 2010). We create the variable *Total Other Sales*$_t$, which is the sum of all other product sales—excluding the focal, lead selling product—in a firm-year-ATC-3, to proxy for the firm's cash flows in a given ATC-3 over time.

Patent expiration. Branded pharmaceutical products are protected by patents, which, at their expiration, allow for generic 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 a firm's innovation choices. We collect market exclusivity data from the FDA Orange Book to identify New Chemical Entity (NCE) patents for lead products in our sample. We define the variable *NCE expiration (<5yr)*$_t$ as a dummy equal to one if firm i's lead sales product in ATC-3 j in year t is within 5 years of expiration for its NCE patent, and zero otherwise.

Deals. We also measure the extent to which the focal firm is engaged in the market for technology using deals data from Cortellis. This variable is distinct from our measure of the thickness of the market for technology, which instead measures the potential pool of technologies

a firm could acquire. The majority of deals recorded in these data are drug-specific (76.7%), and covers the funding of early-stage research, licensing and manufacturing agreements. The remaining deals are technology-based (17.3%, i.e., platforms, processes, research tools), patent-related (3.75%, i.e., exclusive or non-exclusive rights) and company-centric (2.28%, i.e., M&A or joint ventures). We construct the variable $Deals_t$, which is a count of all deals the parent company starts in year t.

4.2.4 Market-level factors

Markets for technology. 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). Small, research-intensive biotechnology firms have engaged in an ever-increasing proportion of research and early-stage development. The extent to which we observe pharmaceutical firms engaging in technically similar product development could be moderated by the thickness of these external technology markets. We 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 biotechnology firms to be those firms active in early-stage drug development if they are not a subsidiary of the parent firms in our sample.

Competition. The intensity of branded therapeutic competition may moderate the nature and direction of firms' innovative activities, and, as noted above, the effect of a blockbuster's sales on subsequent development. To account for such 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.

4.2.5 Adverse Events

To examine the role of demand stickiness as a mechanism behind the blockbuster effect, we exploit variation in the intensity of adverse side effects. To the extent that concerns about adverse side effects significantly influence prescribing behavior and contribute to demand stickiness, our argument—that demand stickiness drives firms' investments in product development and, in turn, fosters technically similar product development—becomes more compelling. We collect data from the FDA Adverse Event Reporting System, which is a comprehensive dataset containing all adverse side effects for specific drugs reported to the FDA since 2004.[28] We first match these drugs to their

[28] One might argue that self-reported side effects may not accurately reflect a drug's true side effect profile. However, self-reporting better captures the *perception* of side effect severity, which aligns with our objective of measuring physicians' demand stickiness.

ATC-3 code, and then, for each ATC-3–Year, we count the number of distinct adverse side effect events reported to the FDA.

The variable *Adverse Event Rate$_t$ (excl. focal)* is the sum of all adverse event reports for products excluding the focal product in an ATC-3-year divided by the sum of all units sold for products excluding the focal product in an ATC-3-year. Normalizing by the total units sold helps to make comparisons across ATC-3s of different sizes, and better reflects the likelihood of an adverse side effect event.[29] The distribution of the resulting measure has a heavy right skew, so we take the natural logarithm. Formally, the construction of the measure can be defined as follows:

$$\text{Adverse Event Rate}_{j,t}^{(-p)} = \frac{\sum_{q \in \mathcal{P}_{j,t} \setminus \{p\}} \text{Adverse Event Reports}_{q,j,t}}{\sum_{q \in \mathcal{P}_{j,t} \setminus \{p\}} \text{Units sold}_{q,j,t}} \tag{1}$$

Where p denotes the focal product, q indexes all other products sold in ATC-3 j in year t, and $\mathcal{P}_{m,t}$ is the set of products in market m in year t. We interpret a higher adverse event rate as proxying for higher demand stickiness. Our argument is that in these "dangerous" markets, doctors will be less inclined to switch from the drugs that they typically prescribe.

4.3 Empirical Specification

Our empirical set-up examines how having a high-selling product influences the direction of product development. All of our dependent variables are non-negative count variables, so the appropriate estimation technique is to use the following Poisson fixed-effects regression:

$$\mathbb{E}[\, y_{i,j,t+5} \mid x_{i,j,t}, \gamma_{j,t}, \alpha_i \,] = \exp(\beta\, x_{i,j,t} + \gamma_{j,t} + \alpha_i) \tag{2}$$

Our dependent variable $y_{i,j,t+5}$ represents the similarity of drugs in development at time $t+5$ relative to the existing lead drug on the market for firm i in therapeutic class j during year t, where similarity is measured, respectively, with respect to customer segment and technology. The independent variable $x_{i,j,t}$ is the lead product sales for firm i in ATC-3 class j in year t. The five-year time lag between the dependent variable $y_{i,j,t+5}$ and the independent variable $x_{i,j,t}$ enables us to capture the forward-looking nature of a firm's investment strategy: that is, how a firm's investment strategy to replace or update their current products in the future would depend on current product sales. Estimation is by Poisson pseudo–maximum likelihood (PPML), the standard estimator for

[29] Standard units represent a single dose unit, which allows for comparisons of quantity sold across different drugs. Standard units are calculated by IQVIA to equate a single dose of a pill, capsule, or liquid.

saturated fixed–effect count models and is robust to overdispersion. We cluster standard errors at the firm-level to account for correlation in development efforts within firms.

Time-varying variation in technological opportunity or demand could be correlated with the likelihood for firms to have a blockbuster drug, as well as the quantity and nature of their innovation strategies. These opportunities can include, for example, new basic science research ('R') discoveries or molecules available for development in external technology markets ('D'). Accordingly, in all specifications, we include an *ATC-1*$\times$ *Year* interacted fixed effect $\gamma_{j,t}$ (Branstetter et al., 2022). We also include firm fixed effects α_i to account for unobserved firm-level factors. However, given that we have 13 years of data, and that product sales are sticky, one might wonder if there is enough within-firm variation. In Appendix F, we document the extent of within-firm variation, and in Appendix Table G.14 we show that our baseline effects remain when we drop firm fixed effects.

We make small modifications to this specification to test potential mechanisms. To estimate demand stickiness, we use Equation (3), where the coefficient of interest is β_3. For the firm capability explanations, we test the sensitivity of the coefficient on lead product sales by adding these variables z linearly in Equation (4) to explore the extent to which they may account for the variation explained by lead product sales.

$$\mathbb{E}[\, y_{i,j,t+5} \mid x_{i,j,t}, m_{i,j,t}, \gamma_{j,t}, \alpha_i \,] = \exp(\beta_1 x_{i,j,t} + \beta_2 m_{i,j,t} + \beta_3 x_{i,j,t} m_{i,j,t} + \gamma_{j,t} + \alpha_i) \tag{3}$$

$$\mathbb{E}[\, y_{i,j,t+5} \mid x_{i,j,t}, z_{i,j,t}, \gamma_{j,t}, \alpha_i \,] = \exp(\beta_1\, x_{i,j,t} + \beta_2\, z_{i,j,t} + \gamma_{j,t} + \alpha_i) \tag{4}$$

5 Results

5.1 Robustness of the Empirical Regularities

We begin by subjecting the descriptive results shown in Section 2 to a stringent set of tests. The results of estimating Equation (2) are presented in Table 2. Columns (1) and (2) use *Early-Stage Projects*$_{t+5}$ as the dependent variable; Columns (3) and (4) use *Target Overlap Projects*$_{t+5}$; Columns (5) and (6) use *Molecularly Similar Projects*$_{t+5}$; and Columns (7) and (8) use *Same Molecule Projects*$_{t+5}$. Odd-numbered columns include ATC1–year fixed effects, while even-numbered columns additionally include firm fixed effects.

Beyond supporting the earlier descriptive evidence, these estimates yield economically interpretable magnitudes and enable within-firm comparisons through the inclusion of firm fixed effects.

For example, in column (2), the coefficient of 0.161 on *Lead Product Sales*$_t$ implies that a \$1 billion increase in product sales is associated with a 17.5% increase in the number of drug development projects within the same therapeutic class for a given firm.

The economic magnitude is substantially larger for technologically similar projects. A \$1 billion increase in lead product sales is associated with a 31.0% increase in the number of projects initiated within the same ATC1 that act on the same target as the focal product for a given firm. For molecular similarity, the effect is even larger: a \$1 billion increase in lead product sales is associated with a 82.8% increase in the number of small-molecule development projects with significantly similar chemistry to the focal product for a given firm.

[Table 2 about here.]

In Appendix G, we probe the sensitivity of this result to different measures, units of analysis, and specifications. Table G.1 utilizes different time horizons to count early-stage projects (Columns (1)-(3)). We also split the dependent variable into counts of small-molecule and biologic projects to test for any differences in the main effect by the underlying technology of drug development. Lastly, we explore the effect on the count of late-stage projects (phases 2 or 3) a firm has at $t + 5$, to determine whether the effect is limited to newly initiated projects or also extends to moving projects through later-stage trials. Table G.2 adjusts lead product sales for inflation, and Table G.3 changes the independent variable to the relative share of sales a product occupies compared to the firm's total sales. Table G.4 estimates OLS models with a logged dependent variable, and Table G.5 estimates the relative share of development projects on product sales. The results are mostly consistent with our key finding, with the exception that the OLS model with a logged dependent variable and the model with the share-based dependent variable yield directionally consistent but noisy estimates.

While our unit of analysis is at the firm–ATC-3–year level, ATC-3s vary in their specificity and may actually incorporate a broad range of diseases. For example, L01F is the ATC-3 code for monoclonal antibodies and antibody conjugates. This includes the ATC fourth-level codes L01FD for HER2 (human epidermal growth factor receptor 2) inhibitors (drugs for treating HER2-positive cancers), as well as L01FE which is the fourth level code for EGFR (epidermal growth factor receptor) inhibitors. This suggests that measuring therapeutic areas at the ATC4 level may better align with more narrowly defined customer segments.[30] ATC-4 level associations between lead

[30] We do our analysis at the ATC-3 level for two reasons. First, we suffer from a severe zero problem in an

product sales and product development are reported in Table G.12. Even at this more granular level the positive association persists.[31]

Lastly, to distinguish individual-level consumer inertia from broader demand stickiness, we exploit variation across therapeutic classes in whether treatments are typically used chronically or acutely, using the ATC-3-level classification from Huber et al. (2013). If our results were driven mainly by repeated consumption by the same patients, the relationship between product sales and follow-on development should be substantially stronger in chronic markets. We find that the sales-development relationship is strongest for project counts in chronic markets, but technologically similar development persists in both chronic and acute markets (see Appendix Table G.16). This suggests that while repeated patient consumption may help explain the quantity response, the shift toward technologically similar innovation reflects a broader form of demand stickiness at the customer-segment level, driven by persistent needs, preferences, and prescribing patterns even as individual patients enter and exit the market.

5.2 Potential Mechanisms for Customer Segment Similarity

5.2.1 Demand Stickiness

In Section 3, we propose that one potential driver of the relationship between sales and continued development within the same customer segment is the firm's expectation of demand stickiness—namely, that demand for follow-on products will resemble that of existing products. This implies that when demand stickiness is higher, firms have stronger incentives to invest in new clinical projects within the same customer segment as their high-selling products.

We examine two possible drivers of stickiness. First, firms may cultivate demand stickiness through their own investment in their sales force. Detailing expenditure constitutes the largest portion of pharmaceutical companies' promotional spending, which aims to increase physicians' brand loyalty to a drug line. Therefore, we expect that the higher the detailing expenditure, the

ATC4-level panel due to the sparsity of observations at the ATC4 level of aggregation. Second, we are constrained by our data. In particular, the crosswalk between the WHO and EphMRA ATC systems has the most consistent coverage at the third level, which is needed to merge sales and development data.

[31] We use an LLM-driven approach to predict the ephMRA ATC4 code for drug development projects in the largest three ATC1s (by sales): Cardiology (C), Antineoplastic and immunomodulating agents (L), and Neurology (N). Note that ATC1 L is where the majority of cancer drugs are classified. Our approach involves leveraging all of the data fields provided by Cortellis—including a textual extract describing the history of the drug development project—and using a few shot prompting technique to match projects to the most appropriate ATC4 code. We achieve matches for 35,118 out of 35,879 projects (97.9%). See Appendix E for a comprehensive description of our process, including the prompt and manual validation results. One should treat these results, however, with a degree of caution, as the LLM approach introduces another layer of measurement error.

greater the demand stickiness.

Alternatively, factors outside the boundaries of the firm may help explain the stickiness of sales. We examine variation in one source of demand stickiness by exploiting differences in reported adverse side effects. In the pharmaceutical industry, demand stickiness can arise from the switching costs physicians face due to concerns about the potential side effects of new products. We hypothesize that, within a therapeutic class (i.e., ATC-3) where other products (i.e., excluding the focal firm's product) are associated with a higher incidence of adverse side effects, physicians will be more likely to continue prescribing the focal product. In other words, physicians will be less likely to switch to "potentially more dangerous" alternatives.

We proxy for demand stickiness using the variable *Adverse Event Rate$_t$ (excl. focal)*, defined as the log-transformed ratio of the total number of adverse event reports for products excluding the focal product in an ATC-3-year to the total number of units sold for those same products. We expect that firms will have greater incentives to invest in similar drug development projects related to their focal product when sales are high, particularly when competing products in that market are associated with higher risk.

Columns (3) and (6) of Table 3 separately examine the two sources of sales stickiness, detailing and adverse event rates, while column (7) includes both simultaneously. Here we estimate Equation (3), and we are primarily interested in the coefficients on the interaction terms, which capture whether the relationship between lead product sales and subsequent development activity becomes stronger when demand is more persistent. Although the interaction between product sales and detailing expenditures is not statistically significant, the interaction between lead product sales and adverse event rates is positive and statistically significant in the full specification.[32] This finding is consistent with the argument that firms have stronger incentives to invest in follow-on development within existing therapeutic markets when physicians are less willing to switch away from established treatments. In therapeutic classes where rival drugs are associated with greater adverse event risks, demand for incumbent products may be more persistent, increasing the expected returns to developing successor products within the same customer segment.

[Table 3 about here.]

We acknowledge that adverse event rates can be an imperfect proxy for patients' and physicians'

[32]Separate sample regressions for markets with below- and above-median adverse event rates show this pattern more clearly. In Appendix Table G.17, the coefficient on *Lead Product Sales$_t$* is significantly lower in low adverse event rate markets and higher in high adverse event rate markets, consistent with this interpretation.

reluctance to switch treatments because they may reflect both safety and efficacy. In certain diseases, perhaps less well understood, a higher count of side effects may reflect a more dangerous drug, or it could reflect a drug with high utilization because efficacy is high enough to outweigh the risks of side effects. From this, it is unclear whether our story about an aversion to side effects is correct. Below we describe three steps we take to alleviate this concern.

First, if physicians and patients tolerate adverse events if efficacy is sufficiently high, we would then expect to see high utilization of the drug. Our measure accounts for efficacy in this regard by normalizing by total units sold. Second, we replicate our results with ATC-3 fixed effects. This is the closest we can get to controlling for unobserved disease-specific characteristics. We report results from estimating the specification with ATC-3 fixed effects in column (1) of Appendix Table G.15. The estimated coefficient on the interaction term between lead product sales and adverse event rate (excl focal) remains positive, albeit weakly significant (p=0.072). Third, we examine a subsample of neurological therapeutic areas (ATC1 = N), where the clinical literature suggests that treatment efficacy is often modest and difficult to distinguish across available drugs, while side-effect profiles vary substantially across treatments (Kola and Landis, 2004; Hyman, 2012; Citrome and Ketter, 2013). In such settings, prescribing decisions are more likely to be shaped by differences in tolerability rather than differences in efficacy, making variation in adverse events a more informative proxy for demand stickiness. Consistent with this interpretation, we find patterns resembling our Table 3 column (7) results in the neurology subsample (see column (2) of Appendix Table G.15).

5.2.2 Sales and Marketing Capabilities

Another potential explanation of the blockbuster effect on investment in clinical trials within the same customer segment could be sales and marketing capabilities, independent of the role of demand stickiness. In particular, if firms have a large salesforce capability in a particular customer segment, this could create an incentive for the firm to continue innovating in that area so that they can use this specialized asset in the future. The direction of causality could run the other way, of course. Firms expecting to develop new drugs in a customer segment may invest more heavily in their salesforce capacity. Empirically, we try to mitigate this concern in our specification by lagging detailing by three years. This approach, however, is limited given that detailing has a high degree of serial correlation. In econometric terms, a firm's sales force and marketing capability may be confounding the relationship between product sales and the quantity of subsequent development.

Column (2) of Table 3 shows the sensitivity of the main effect to the inclusion of detailing. The

results indicate no statistically significant relationship between detailing and project development. Moreover, the magnitude of the coefficient on lead product sales changes negligibly with the inclusion of detailing. Appendix Table G.7 shows consistent results when detailing is measured at both the ATC-3 and firm levels. The main exception is that, when detailing is measured at the ATC-3 level, the coefficient on lead product sales declines modestly, which is consistent with detailing explaining variation in product sales as well.[33]

5.3 Potential Mechanisms for Technical Similarity

5.3.1 Demand Stickiness

Following the same format as Section 5.2.1, we examine how demand stickiness moderates the relationship between product sales and the number of technologically similar projects by leveraging proxies for possible drivers of demand stickiness, notably adverse events and firm detailing. The results for both target overlap and molecular similarity are reported in Table 4.

First, with regard to technical similarity measured by target overlap, the results suggest that the adverse events driver of demand stickiness moderates the effect of product sales on technical similarity. In both columns (6) and (7), the coefficient on the interaction term is positive and significant, which suggests that firms are more likely to initiate clinical trials that target the same protein as their high-selling focal product in markets where existing rival products exhibit a higher rate of adverse side effects. [34] In columns (2) and (3), the results do not suggest that investment in detailing plays a role in the relationship between product sales and target overlap.[35]

As for molecular similarity, the moderation results are muted: across both sources of demand stickiness, we do not find any evidence supporting a moderating effect. Despite finding no effect, these results are intuitive when we consider the sequencing of R&D decisions. Adverse events often reflect how drugs exert their effects in the body. When side effects arise from a drug's intended action, they are difficult to eliminate through changes in molecular design (Rudmann,

[33] Another possible explanation for the lack of an effect of detailing is the collinearity (r = 0.27) between product sales and detailing. This interpretation is consistent with Appendix Table G.14, which presents estimates without firm fixed effects. In this specification, detailing has a direct effect on project investment, and the magnitude of the coefficient on lead product sales decreases by approximately one third. Taken together with the demand stickiness results, Table 3 provides evidence that neither mechanism should be discounted.

[34] Columns (4)–(6) in Appendix Table G.17 present separate sample regressions for markets with below- and above-median adverse event rates, which are consistent with this finding.

[35] We note that the large increase in the coefficient on lead product sales in column (3) of both panels may be driven by the correlation between lead product sales and detailing. In particular, the interaction term between lead product sales and detailing closely resembles a squared term in lead product sales. Because the coefficient on this interaction term is negative, it effectively amplifies the positive coefficient on lead product sales.

2013). As a result, drugs that bind to the same biological target tend to exhibit similar safety profiles. Consistent with this, prior evidence shows that drugs with similar side effects often act on the same targets even when their chemical structures differ (Campillos et al., 2008). This suggests that adverse event similarity is driven more by shared mechanisms than by chemical similarity. This interpretation aligns with our findings: while higher product sales are, on average, associated with both greater target overlap and increased molecular similarity in follow-on projects, adverse-event-driven demand stickiness moderates the relationship between lead product sales and target overlap, but not that with molecular similarity.

Taken together, these results suggest that high-selling products in high demand stickiness markets shape the "target" of R&D, where the target refers to achieving similar functionality to the existing high-selling drug. This functionality is driven by efficacy and safety considerations, which are largely determined by the underlying biological target. Conditional on this target, firms then choose molecular structures that best achieve these objectives, with molecular design serving as a means to implement the desired functionality.

[Table 4 about here.]

5.3.2 R&D Capabilities and Market Exclusivity

Table 5 evaluates whether the relationship between lead product sales and technologically similar follow-on development can be explained by firm specific R&D and related capabilities. Across both measures of technological similarity, the coefficient on lead product sales remains positive, statistically significant, and remarkably stable across specifications that control for patenting activity, scientific publications, deal activity, R&D expenditures, portfolio size, other firm level characteristics and market exclusivity. These findings suggest that while selected firm capabilities and market exclusivity may contribute to follow-on innovation, they do not fully account for the tendency of firms to pursue technologically related development around their highest selling products. Rather, the results are consistent with the view that commercially successful products themselves shape the direction of subsequent technological search within firms. Specifically, blockbuster products appear to anchor firms' innovation efforts toward commercially validated technological trajectories, encouraging development activity that remains proximate to existing sources of market success rather than more distant or exploratory forms of search.

In Appendix Table G.13, we include all demand stickiness interaction terms alongside con-

trols for R&D capabilities and market exclusivity. In this full specification, the results remain consistent with the Table 4 and 5 results. Similar to Table 4, the coefficient on the interaction between lead product sales and adverse event rates is positive and statistically significant for *Early-Stage Projects*$_{t+5}$ and *Target Overlap Projects*$_{t+5}$, and remains insignificant for *Molecularly Similar Projects*$_{t+5}$. This last result, suggesting the absence of a link between lead product sales and molecular similarity due to demand stickiness is quite telling. The results suggest that, while lead product sales may set the objective (the biological target) for the follow-on R&D, the relationship between lead product sales and the means to achieve that objective (i.e., the specific molecule) is much weaker. The next section on market level factors sheds some light on this finding.

[Table 5 about here.]

5.4 Market-Level Factors: Markets for Technology and Competition

While this paper's main focus is on the firm and individual products, the extent to which we observe pharmaceutical firms engaging in technologically similar product development may be moderated by characteristics of the markets in which they operate. We explore two such dimensions: the thickness of external technology markets, and competition. We measure the thickness of the market for technology by the abundance of firms in early stage development with no commercial history. We interpret this measure as capturing the supply of external innovation available for acquisition or licensing—presumably created by biotechs. To measure competition in the market, we count the number of firm selling branded therapeutic drugs within the same ATC-3 in a given year.

In Table 6, we examine the moderating role of biotech activity and competition on both customer segment and technological similarity. First, while investment in drug development within the same customer segment as high-selling drugs may be higher in markets with more active biotechs and competitors, neither factor appears to moderate the relationship between a firm's lead product sales and the number of projects initiated within the same customer segment.

However, we find evidence that the abundance of biotech firms shapes the technological direction of post-blockbuster drug development. The negative coefficient on the interaction between lead product sales and biotech activity in columns (10) and (12) is consistent with firms substituting for the internal development of similar molecularly successor drugs toward external options provided by biotechs. This finding is consistent with the fact that the strength of technology markets distinguishes pharmaceuticals from all other manufacturing industries.[36]. Combined with earlier

[36] As reported in Arora et al. (2016), 56% of of pharmaceutical firms rely on licensing and 39% rely on

results suggesting that molecular structure serves as a means to achieve a given therapeutic target, these findings indicate that when external innovation is readily available, firms may rely more on acquiring or licensing molecular solutions to achieve the technological functionality characterizing their existing high-selling products.

[Table 6 about here.]

6 Discussion

This section considers an important question raised by our analysis: Can blockbuster product development strategies partly account for R&D productivity declines within industries? A thorough examination of this question goes well beyond the scope of the current paper. We highlight this question, however, in the hope of stimulating further research.

The declining productivity trend in pharmaceutical innovation has been thoroughly examined in the literature (Cockburn, 2006). Explanations for this phenomenon, however, have generally been limited to analyses of supply-side shocks to the pharmaceutical industry (Krieger et al., 2022; Dranove et al., 2022), or the more economy-wide argument that coming up with novel ideas has become more costly (Bloom et al., 2020). By focusing on individual product sales, this paper proposes a complementary explanation. Specifically, we argue that firms with high-selling products tend to develop new products targeting the same customer segments which are, in turn, *technically* similar to their existing products. If all firms follow this strategy, what would be the implications for industry-wide productivity?

In Figure 3, we present two industry-wide trends from our data. Panel (a) shows that the number of $1 billion (inflation-adjusted) "blockbusters" has increased more than four-fold, from 12 in 1998 to 49 in 2010.[37,38] 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 whose development efforts focus on 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%.

acquisitions as the sources of the most important new products introduced in 2009.

[37] 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 that could generate large sales (Kola and Landis, 2004).

[38] 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.

We acknowledge that these figures are purely descriptive accounts. Nevertheless, these trends are suggestive of the phenomena we examine in this paper: a greater reliance on high-selling products is associated with innovation that is more incremental relative to a firm's dominant product *as well as less novel overall at the industry level*. Importantly, our results complement the research suggesting a productivity decline, particularly in pharmaceuticals (Cockburn, 2006; Bloom et al., 2020). These papers focus on exogenous supply-side factors, arguing that the decline is the consequence of the depletion of opportunities and a concomitant increase in the costs of research per unit output. In contrast, our analysis highlights the importance of a demand-side mechanism; that even if opportunities were not declining, we may still observe a decline in R&D productivity due to firms' endogenous responses to the emergence of especially high-selling drugs. While potentially complementary to the supply-side explanations for the R&D productivity decline noted above, our perspective would also be consistent with the possibility that technological opportunities are not declining at all. Indeed, while extending well beyond the sample period employed in our empirical analysis, the emergence of CRISPR and MRNA-based therapies—arguably general purpose biomedical technologies widely applicable across therapeutic classes—would suggest technological opportunities may even be on the rise, at least in biomedicine.

[Figure 3 about here.]

For policy and management, it is important to distinguish between supply- and demand-side explanations. The former may be addressed, for example, through government support for the upstream science, while the latter would not. If the main explanation is a demand-side one, and it was determined that firms' blockbuster product development strategies were indeed stranding commercializable science, policies more directly acting on firms' incentives would be more effective.

7 Conclusion

This paper documents two empirical regularities in the pharmaceutical industry with broader implications for the direction of technological change. First, products with outsized sales are associated with increased development of new products targeting the same customer segment. In our data from 1998 to 2010, a drug with annual sales below $100 million is associated, on average, with fewer than two drug candidates in clinical trials within its therapeutic area, whereas a product with annual sales exceeding $1 billion is associated with roughly five such candidates. Second, the

follow-on products that firms pursue tend to embody technologies that resemble those underlying their existing high-selling products. Taken together, these patterns indicate that successful products are associated not only with where firms innovate, but also with how they innovate, shaping the technological trajectories firms subsequently pursue.

We then examine two broad classes of explanations for these regularities. One emphasizes supply-side factors, such as firms' marketing, distribution, and R&D capabilities, which may be co-specialized to particular customer segments and technological approaches. The other highlights demand-side forces, specifically the expectation that demand is persistent or "sticky" within a customer segment, creating incentives to reinvest in products that satisfy similar needs and preferences. Our empirical analysis finds support for both channels. At the same time, several findings point to an important role for demand-side mechanisms that has been under-emphasized in prior work. In particular, we show that firms are more likely to concentrate development within a customer segment and to pursue technologies that share the same underlying biological targets as their high-selling products when demand is more sticky, as proxied by factors such as the relative side-effect profiles of rival drugs. By contrast, we find no or little corresponding increase in molecular similarity. This pattern is consistent with the idea that demand stickiness operates through the preservation of therapeutic functionality, which is closely tied to biological targets, while allowing flexibility in the molecular means used to achieve it.

Taken together, our findings point to a channel through which demand stickiness, combined with the presence of high-selling products, shapes the direction of follow-on innovation. When demand is more persistent, firms face stronger incentives to continue targeting the same customer segment and to preserve similar functionality as the high-selling products in subsequent products. In the pharmaceutical context, such functionality is closely tied to the underlying biological target that a drug acts upon. At the same time, firms retain flexibility in how they achieve this functionality, as multiple molecular approaches may be used to act on a given target. Our results are consistent with a pattern in which demand conditions are associated with alignment in customer segments and therapeutic targets that shape functionality, with firms then selecting the technological means used to achieve them. By highlighting this mechanism, our study points to the importance of considering how realized product success can endogenously channel innovation toward more incremental trajectories, with potential implications for the direction and pace of technological change at the firm and industry level.

Our conclusions should be interpreted in light of several limitations. Most importantly, our

empirical analysis does not establish causal identification. Identifying the causal effect of product sales on the direction of subsequent innovation would require an exogenous shock that affects current product sales without independently influencing future product demand or firms' subsequent development choices. In the absence of such a shock, our analysis should be interpreted as documenting systematic relationships that are consistent with both supply- and demand-side mechanisms rather than definitively isolating causal effects. Moreover, while we theorize a sequential mechanism in which firms first target the same customer segment and subsequently pursue similar technologies, we are unable to empirically test the sequencing of each step of this process. Doing so would require exogenous variation that constrains firms' customer-segment choices while leaving technological choices otherwise unaffected. We leave this question for future research.[39]

Additionally, we have explored the role of demand stickiness only in the pharmaceutical industry. That being said, the pharmaceutical industry offers rich data, including product-specific sources of demand stickiness, which allows us to better identify the role of stickiness in shaping firms' product development strategies. We suggest, however, that the phenomenon of demand stickiness itself—as well as its implications for subsequent product development—is broadly generalizable. Indeed we motivated this paper with examples from aviation and smartphones, where firms have concentrated their R&D efforts in developing new products that serve the same customer segment, and in turn these products often resemble many of the technological features of the existing product. What is likely to vary across industries is not the relevance of demand stickiness itself, but rather its sources. In pharmaceuticals, stickiness may arise from physician prescribing habits, side-effect concerns, or detailing investments, whereas in other settings stickiness may stem from customer learning—such as training pilots to fly a new model of aircraft— or from complementarities among other existing products and services—such as those within the Apple ecosystem.

More broadly, our results highlight the importance of understanding how firms form expectations about the persistence of demand. While incremental innovation can generate meaningful improvements in efficacy, safety, manufacturability, and patient adherence, sustained concentration within existing customer segments may also influence the balance between exploratory and exploitative innovation at the industry level. Future research could examine whether similar dynamics arise in other R&D-intensive industries characterized by high development costs, substantial commercialization investments, and persistent customer relationships.

[39] As an example, consider Medicare Part D, which has been used in the literature as an exogenous shock to demand (*e.g.,* Dranove et al., 2022). While this would be a shock to the product sales, it would also affect the expected demand for *new* products in the other Medicare markets.

References

Acemoglu, D., Akcigit, U., and Celik, M. A. (2022). Radical and incremental innovation: The roles of firms, managers, and innovators. *American Economic Journal: Macroeconomics*, 14(3):199–249.

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

Adner, R. and Snow, D. (2010). Old technology responses to new technology threats: Demand heterogeneity and technology retreats. *Industrial and Corporate Change*, 19(5):1655–1675.

Aghion, P., Bloom, N., Blundell, R., Griffith, R., and Howitt, P. (2005). Competition and innovation: An inverted-u relationship. *The Quarterly Journal of Economics*, 120(2):701–728.

Ahuja, G., Lampert, C. M., and Tandon, V. (2014). Paradigm-changing vs. paradigm-deepening innovation: How firm scope influences firm technological response to shocks. *Organization Science*, 25(3):653–669.

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

Argente, D., Lee, M., and Moreira, S. (2018). Innovation and product reallocation in the great recession. *Journal of Monetary Economics*, 93:1–20.

Argente, D. and Yeh, C. (2022). Product life cycle, learning, and nominal shocks. *The Review of Economic Studies*, 89(6):2992–3054.

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., Cohen, W. M., and Walsh, J. P. (2016). The acquisition and commercialization of invention in american manufacturing: Incidence and impact. *Research Policy*, 45(6):1113–1128.

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.

Arora, A., Hasan, S., and Miles, W. D. (2025). If you had one shot: Scale and herding in innovation experiments. *Working Paper*.

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.

Banerjee, B. and Bandyopadhyay, S. (2003). Advertising competition under consumer inertia. *Marketing Science*, 22(1):131–144.

Barney, J. B. (2001). Resource-based theories of competitive advantage: A ten-year retrospective on the resource-based view. *Journal of Management*, 27(6):643–650.

Berndt, E. R., Bui, L., Reiley, D. R., and Urban, G. L. (1995). Information, marketing, and pricing in the us antiulcer drug market. *The American Economic Review*, 85(2):100–105.

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.

Bornstein, G. (2025). Entry and profits in an aging economy: The role of consumer inertia. Working Paper 33820, National Bureau of Economic Research.

Branstetter, L., Chatterjee, C., and Higgins, M. J. (2016). Regulation and welfare: evidence from paragraph iv generic entry in the pharmaceutical industry. *The RAND Journal of Economics*, 47(4):857–890.

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

Bronnenberg, B. J., Dubé, J.-P. H., and Gentzkow, M. (2012). The evolution of brand preferences: Evidence from consumer migration. *American Economic Review*, 102(6):2472–2508.

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

Campillos, M., Kuhn, M., Gavin, A.-C., Jensen, L. J., and Bork, P. (2008). Drug target identification using side-effect similarity. *Science*, 321(5886):263–266.

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.

Chintagunta, P. K. (1998). Inertia and variety seeking in a model of brand-purchase timing. *Marketing Science*, 17(3):253–270.

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

Citeline (2024). Why are clinical development success rates falling?

Citrome, L. and Ketter, T. (2013). When does a difference make a difference? interpretation of number needed to treat, number needed to harm, and likelihood to be helped or harmed. *International Journal of Clinical Practice*, 67(5):407–411.

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

Cohen, W. M. (2010). Chapter 4 - fifty years of empirical studies of innovative activity and performance. In Hall, B. H. and Rosenberg, N., editors, *Handbook of The Economics of Innovation, Vol. 1*, volume 1 of *Handbook of the Economics of Innovation*, pages 129–213. North-Holland.

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

Cohen, W. M. and Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly*, 35(1):128–152.

Corr, P. and Williams, D. (2009). The pathway from idea to regulatory approval: Examples for drug development. *Conflict of Interest in Medical Research Education and Practice*.

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.

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.

Dubé, J.-P., Hitsch, G. J., and Rossi, P. E. (2010). State dependence and alternative explanations for consumer inertia. *The RAND Journal of Economics*, 41(3):417–445.

Feng, J. and Maini, L. (2024). Demand inertia and the hidden impact of pharmacy benefit managers. *Management Science*, 70(12):8940–8961.

Frankel, A. P., Krieger, J. L., Li, D., and Papanikolaou, D. (2023). Evaluation and Learning in R&D Investment. Technical report, National Bureau of Economic Research.

Gordon, R. (2017). *The Rise and Fall of American Growth: The US Standard of Living Since the Civil War*. Princeton University Press.

Gourio, F. and Rudanko, L. (2014). Customer capital. *Review of Economic Studies*, 81(3):1102–1136.

Grabowski, H. G. and Vernon, J. M. (1992). Brand loyalty, entry, and price competition in pharmaceuticals after the 1984 drug act. *The Journal of Law and Economics*, 35(2):331–350.

Hall, B. H. and Lerner, J. (2010). Chapter 14 - the financing of rd and innovation. In Hall, B. H. and Rosenberg, N., editors, *Handbook of The Economics of Innovation, Vol. 1*, volume 1 of *Handbook of the Economics of Innovation*, pages 609–639. North-Holland.

Helfat, C. E. (1997). Know-how and asset complementarity and dynamic capability accumulation: the case of r&d. *Strategic Management Journal*, 18(5):339–360.

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

Henderson, R. M. and Clark, K. B. (1990). Architectural innovation: The reconfiguration of existing product technologies and the failure of established firms. *Administrative Science Quarterly*, pages 9–30.

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.

Higgins, M. J., Yan, X., and Chatterjee, C. (2021). Unpacking the effects of adverse regulatory events: Evidence from pharmaceutical relabeling. *Research Policy*, 50(1):104126.

Hortaçsu, A., Madanizadeh, S. A., and Puller, S. L. (2017). Power to choose? an analysis of consumer inertia in the residential electricity market. *American Economic Journal: Economic Policy*, 9(4):192–226.

Huber, C. A., Szucs, T. D., Rapold, R., and Reich, O. (2013). Identifying patients with chronic conditions using pharmacy data in switzerland: an updated mapping approach to the classification of medications. *BMC Public Health*, 13(1):1030.

Hughes, J., Rees, S., Kalindjian, S., and Philpott, K. (2011). Principles of early drug discovery. *British Journal of Pharmacology*, 162(6):1239–1249.

Hyman, S. E. (2012). Revolution stalled. *Science Translational Medicine*, 4(155):155cm11–155cm11.

Janakiraman, R., Dutta, S., Sismeiro, C., and Stern, P. (2008). Physicians' persistence and its implications for their response to promotion of prescription drugs. *Management Science*, 54(6):1080–1093.

Klemperer, P. D. (1987). *Markets With Consumer Switching Costs (Oligopoly)*. PhD thesis, Stanford University.

Klepper, S. (1996). Entry, exit, growth, and innovation over the product life cycle. *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.

Lancaster, K. J. (1966). A new approach to consumer theory. *Journal of Political Economy*, 74(2):132–157.

Leiponen, A. and Helfat, C. E. (2010). Innovation objectives, knowledge sources, and the benefits of breadth. *Strategic Management Journal*, 31(2):224–236.

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

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.

Nerkar, A. and Roberts, P. W. (2004). Technological and product-market experience and the success of new product introductions in the pharmaceutical industry. *Strategic Management Journal*, 25(8-9):779–799.

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.

Penrose, E. T. (1959). *The Theory of the Growth of the Firm*. Basil Blackwell.

Peteraf, M. A. and Bergen, M. E. (2003). Scanning dynamic competitive landscapes: a market-based and resource-based framework. *Strategic Management Journal*, 24(10):1027–1041.

Rudmann, D. G. (2013). On-target and off-target-based toxicologic effects. *Toxicologic Pathology*, 41(2):310–314.

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

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.

Teece, D. J. (1986). Transactions cost economics and the multinational enterprise an assessment. *Journal of Economic Behavior & Organization*, 7(1):21–45.

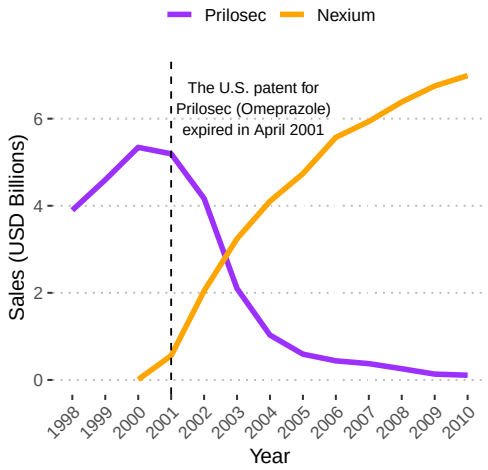
Tripsas, M. and Gavetti, G. (2000). Capabilities, cognition, and inertia: Evidence from digital imaging. *Strategic Management Journal*, 21(10-11):1147–1161.

Wang, T., Aggarwal, V. A., and Wu, B. (2020). Capability interactions and adaptation to demand-side change. *Strategic Management Journal*, 41(9):1595–1627.

Wong, C. H., Siah, K. W., and Lo, A. W. (2019). Estimation of clinical trial success rates and related parameters. *Biostatistics*, 20(2):273–286.

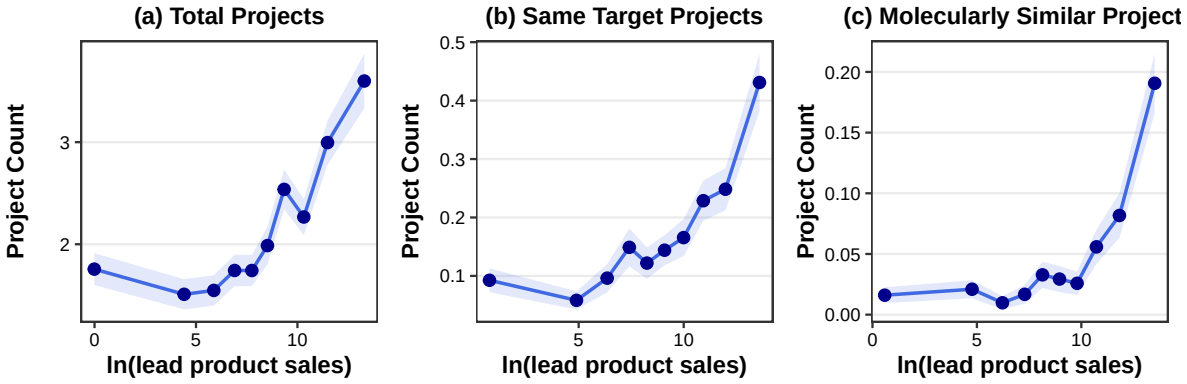
Wouters, O. J., McKee, M., and Luyten, J. (2020). Estimated research and development investment needed to bring a new medicine to market, 2009-2018. *JAMA*, 323(9):844–853.

Figure 1: Annual Sales of Prilosec and Nexium (1998-2010)



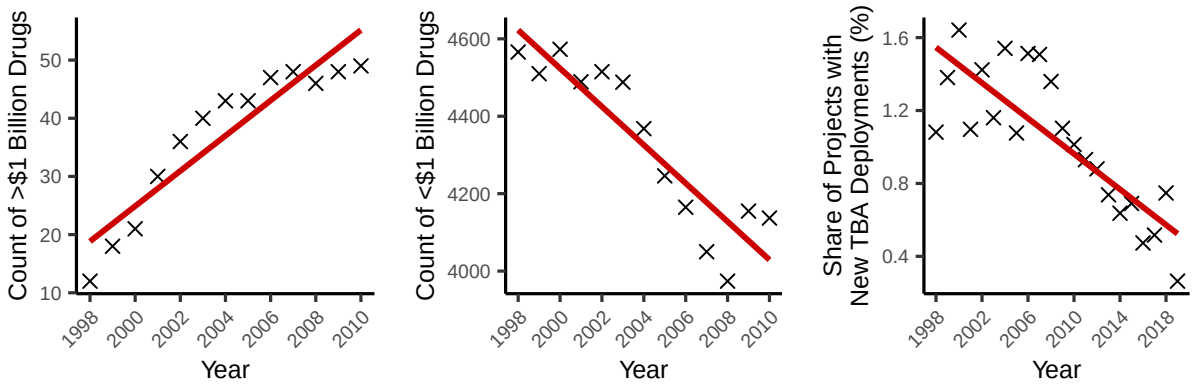
Notes: This figure plots the total annual sales for Prilosec and Nexium sold by AstraZeneca between 1998 and 2010. The dashed line at 2001 indicates the year in which the U.S. patent for Prilosec expired.

Figure 2: The association between product sales, the count of early-stage development projects, and the count of technologically similar projects



Notes: This figure plots the relationship between product sales and the quantity and nature of development. The raw data is organized at the firm-ATC-3-year level. All figures are binned scatter plots, and the shaded banded region represents the 95% confidence interval around the bin mean.

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-ATC-3 Level Descriptive Statistics

	Mean	St. Dev.	Min	Q1	Median	Q3	Max	N
Early-Stage Projects$_{t+5}$	1.998	4.278	0.000	0.000	0.000	2.000	57.000	32515
Target Overlap Projects$_{t+5}$	0.174	0.646	0.000	0.000	0.000	0.000	8.000	16316
Molecularly Similar Projects$_{t+5}$	0.048	0.253	0.000	0.000	0.000	0.000	4.000	14320
Same Molecule Projects$_{t+5}$	0.025	0.160	0.000	0.000	0.000	0.000	2.000	14320
Lead Product Sales$_t$	0.074	0.357	0.000	0.000	0.001	0.011	9.677	32515
Biotechs$_t$	3.519	3.087	0.000	0.970	3.240	5.210	22.290	32515
Deals$_t$	3.662	5.229	0.000	0.000	1.000	6.000	28.000	32515
ATC1 Detailing$_{t-3}$	0.018	0.038	0.000	0.000	0.001	0.016	0.324	25615
Patent Applications$_t$	1.161	2.933	0.000	0.000	0.000	1.000	41.000	32515
Scientific Publications$_t$	2.565	14.818	0.000	0.000	0.000	0.000	617.000	32515
Total Other Sales$_t$	7.762	7.505	0.012	1.405	5.209	13.390	32.803	32515
R&D Expenditure$_t$	7.434	24.871	0.003	0.917	2.506	4.883	196.802	27889
Total Products$_t$	1.560	1.386	0.020	0.700	1.060	1.960	5.160	32515
Adverse Event Rate$_t$	-5.571	2.542	-13.589	-7.244	-5.954	-4.191	8.166	16659
Adverse Event Rate$_t$ (excl. focal)	-2.758	3.600	-12.908	-5.662	-2.905	-0.207	14.140	4441
NCE expiration$_t$ (<5yr)	0.029	0.168	0.000	0.000	0.000	0.000	1.000	32515
Number of Competitors$_t$	12.552	7.112	0.000	7.000	12.000	18.000	31.000	32515

Notes: This table summarizes all variables used in the main analysis. The dataset includes 23,361 Firm–ATC-3–Year (i, j, t) observations, though some variables have missing data for the following reasons: *Target Overlap Count*$_{t+5}$ (N=14,431) requires identifying the lead product's target and a development project in ATC-3 j at $t+5$. *Molecularly Similar Projects*$_{t+5}$ (N=12,996) requires a small molecule lead, its SMILES code, and a matching project at $t+5$. *Adverse Event Rate*$_t$ (N=12,037) requires at least one product sold in a therapeutic class year to be matched to the FAERS database. Note that for products not matched, we cannot distinguish whether this is due to our matching algorithm or whether the product has no adverse event reports in a given year. *Adverse Event Rate*$_t$ *(excl. focal)* (N=4,011) is even more restrictive because it also requires that the focal product sold in firm-year-ATC-3 is matched to FAERS. See Appendix D for full variable definitions.

Table 2: The association between product sales, the count of early-stage development projects, and the count of technologically similar projects

	Early-Stage Projects$_{t+5}$		Target Overlap Projects$_{t+5}$		Molecularly Similar Projects$_{t+5}$		Same Molecule Projects$_{t+5}$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Lead Product Sales$_t$	0.327	0.161	0.412	0.270	0.625	0.603	0.514	0.443
	(0.040)	(0.036)	(0.072)	(0.054)	(0.096)	(0.087)	(0.098)	(0.105)
ATC1 × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	No	Yes	No	Yes	No	Yes	No	Yes
Observations	22,950	20,838	15,819	13,033	13,043	11,820	12,842	9,367
Pseudo R-squared	0.050	0.421	0.068	0.234	0.097	0.174	0.070	0.126

Notes: This table reports Poisson pseudo-maximum-likelihood regression results for the association between product sales, the count of early-stage development projects, and the count of technologically similar projects. For each dependent variable, we report results with and without firm fixed effects. Standard errors are clustered at the firm level and reported in parentheses.

Table 3: Demand Stickiness and Customer-Segment Similarity

	Early-Stage Projects$_{t+5}$						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Lead Product Sales$_t$	0.161	0.168	0.204	0.127	0.129	0.250	0.122
	(0.036)	(0.034)	(0.171)	(0.032)	(0.032)	(0.108)	(0.181)
ATC1 Detailing$_{t-3}$		-1.257	-1.223				-0.430
		(0.916)	(0.924)				(1.193)
Lead Product Sales$_t$ × ATC1 Detailing$_{t-3}$			-0.253				1.614
			(0.999)				(1.040)
Adverse Event Rate$_t$ (excl. focal)					0.012	0.003	-0.004
					(0.008)	(0.012)	(0.008)
Lead Product Sales$_t$ × Adverse Event Rate$_t$ (excl. focal)						0.023	0.043
						(0.015)	(0.010)
ATC1 × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	20,838	15,866	15,866	3,636	3,636	3,636	3,592
Pseudo R-squared	0.421	0.433	0.433	0.392	0.392	0.395	0.398

Notes: This table reports results testing the demand stickiness mechanism and the role of co-specialized detailing assets on investment within the same customer segment. All models are estimated with Poisson pseudo-maximum-likelihood with fixed effects at the Firm and ATC1 by Year levels. Standard errors clustered at the firm level are shown in parentheses.

Table 4: Product sales, technical similarity, and demand stickiness

Panel A: Target overlap

	Target Overlap Projects$_{t+5}$						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Lead Product Sales$_t$	0.270	0.236	0.624	0.201	0.212	0.497	0.455
	(0.054)	(0.046)	(0.131)	(0.064)	(0.062)	(0.118)	(0.157)
ATC1 Detailing$_{t-3}$		2.529	3.200				5.617
		(2.018)	(2.074)				(2.658)
Lead Product Sales$_t$ × ATC1 Detailing$_{t-3}$			-3.143				0.160
			(1.074)				(1.533)
Adverse Event Rate$_t$ (excl. focal)					0.037	0.012	0.013
					(0.026)	(0.023)	(0.026)
Lead Product Sales$_t$ × Adverse Event Rate$_t$ (excl. focal)						0.052	0.056
						(0.017)	(0.025)
ATC1 × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	13,033	10,058	10,058	2,584	2,584	2,584	2,557
Pseudo R-squared	0.234	0.244	0.249	0.201	0.203	0.211	0.225

Panel B: Molecular similarity

	Molecularly Similar Projects$_{t+5}$						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Lead Product Sales$_t$	0.603	0.565	0.971	0.417	0.417	0.523	0.685
	(0.087)	(0.096)	(0.174)	(0.076)	(0.076)	(0.134)	(0.222)
ATC1 Detailing$_{t-3}$		0.205	1.735				-2.652
		(3.588)	(3.290)				(4.237)
Lead Product Sales$_t$ × ATC1 Detailing$_{t-3}$			-3.278				-1.521
			(1.126)				(1.946)
Adverse Event Rate$_t$ (excl. focal)					-0.001	-0.016	-0.004
					(0.032)	(0.041)	(0.039)
Lead Product Sales$_t$ × Adverse Event Rate$_t$ (excl. focal)						0.020	0.005
						(0.020)	(0.029)
ATC1 × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	11,820	8,536	8,536	2,224	2,224	2,224	2,208
Pseudo R-squared	0.174	0.194	0.201	0.235	0.235	0.236	0.240

Notes: This table reports results testing the demand stickiness mechanism and the role of co-specialized detailing assets in driving technical similarity. The technological similarity between a product and development projects is measured with respect to the target-based action (Panel (A)) and–for the subset of small molecule products and development projects–the molecule (Panel (B)). All models are estimated with Poisson pseudo-maximum-likelihood with fixed effects at the Firm and ATC1 by Year levels. Standard errors clustered at the firm level are shown in parentheses.

Table 5: The mediating role of firm-specific capabilities

Panel A: Target overlap

	Target Overlap Projects$_{t+5}$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Lead Product Sales$_t$	0.265	0.266	0.271	0.252	0.295	0.268	0.274	0.275
	(0.054)	(0.055)	(0.054)	(0.050)	(0.054)	(0.056)	(0.056)	(0.051)
Patent Applications$_t$	0.038							0.026
	(0.013)							(0.011)
Scientific Publications$_t$		0.004						0.003
		(0.002)						(0.002)
Deals$_t$			0.001					-0.005
			(0.009)					(0.008)
Total Other Sales$_t$				-0.023				-0.017
				(0.016)				(0.017)
R&D Expenditure$_t$					0.012			0.011
					(0.003)			(0.002)
Total Products$_t$						-0.397		-0.375
						(0.466)		(0.541)
NCE expiration$_t$ (<5yr)							0.774	0.759
							(0.129)	(0.153)
ATC1 × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	13,033	13,033	13,033	13,033	12,104	13,033	13,033	12,104
Pseudo R-squared	0.237	0.235	0.234	0.234	0.238	0.235	0.243	0.249

Panel B: Molecular similarity

	Molecularly Similar Projects$_{t+5}$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Lead Product Sales$_t$	0.602	0.606	0.603	0.592	0.627	0.608	0.629	0.660
	(0.088)	(0.087)	(0.087)	(0.088)	(0.116)	(0.086)	(0.091)	(0.118)
Patent Applications$_t$	0.007							0.011
	(0.024)							(0.026)
Scientific Publications$_t$		-0.002						-0.001
		(0.006)						(0.006)
Deals$_t$			0.003					0.011
			(0.011)					(0.011)
Total Other Sales$_t$				-0.016				0.025
				(0.032)				(0.036)
R&D Expenditure$_t$					0.005			0.004
					(0.006)			(0.007)
Total Products$_t$						0.724		0.630
						(0.354)		(0.556)
NCE expiration$_t$ (<5yr)							1.159	1.141
							(0.175)	(0.197)
ATC1 × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	11,820	11,820	11,820	11,820	10,412	11,820	11,820	10,412
Pseudo R-squared	0.174	0.174	0.174	0.174	0.150	0.175	0.191	0.168

Notes: This table reports results testing the role of firm-specific supply side capabilities in driving technical similarity. The technological similarity between a product and development projects is measured with respect to the target-based action (Panel (A)) and–for the subset of small molecule products and development projects–the molecule (Panel (B)). All models are estimated with Poisson pseudo-maximum-likelihood with fixed effects at the Firm and ATC1 by Year levels. Standard errors clustered at the firm level are shown in parentheses.

Table 6: The Moderating Role of Market-Level Factors

	Early-Stage Projects$_{t+5}$				Target Overlap Projects$_{t+5}$				Molecularly Similar Projects$_{t+5}$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Lead Product Sales$_t$	0.161	0.100	0.098	0.132	0.270	0.271	0.142	0.172	0.603	0.791	0.656	0.874
	(0.036)	(0.046)	(0.082)	(0.050)	(0.054)	(0.135)	(0.105)	(0.120)	(0.087)	(0.173)	(0.188)	(0.229)
Biotechs$_t$		0.222		0.217		0.026		0.041		0.034		0.045
		(0.016)		(0.017)		(0.021)		(0.019)		(0.026)		(0.027)
Lead Product Sales$_t$ × Biotechs$_t$		-0.006		-0.005		-0.001		-0.010		-0.032		-0.036
		(0.005)		(0.005)		(0.018)		(0.019)		(0.015)		(0.015)
Number of Competitors$_t$			0.048	0.010			-0.019	-0.026			-0.013	-0.020
			(0.003)	(0.005)			(0.008)	(0.007)			(0.011)	(0.012)
Lead Product Sales$_t$ × Number of Competitors$_t$			0.002	-0.002			0.009	0.011			-0.003	-0.003
			(0.005)	(0.004)			(0.007)	(0.008)			(0.012)	(0.013)
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	20,838	20,838	20,838	20,838	13,033	13,033	13,033	13,033	11,820	11,820	11,820	11,820
Pseudo R-squared	0.421	0.583	0.445	0.584	0.234	0.235	0.236	0.238	0.174	0.177	0.175	0.179

Notes: This table explores the moderating role of market-level factors, including the number of biotechs and the number of competitors, to test whether market structure moderates the main effect. All specifications report estimates from Poisson Pseudo Maximum Likelihood models with fixed effects at the Firm and ATC1 by Year levels. Standard errors are clustered at the firm level and reported in parentheses.

Appendix:

Blockbuster Products, Sequels, and the Nature of Innovation

A Model

A.1 Setting

We model a firm's decision whether to develop a product that targets a new customer segment or develop a product in the customer segment in which it is currently active. We refer to products targeting new customer segments as *new* products and those in its existing customer segment as *similar* products. While products targeting existing segments are generally more technically similar to the firm's existing offerings, we treat this as an empirical question.

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 budget constraints, with a limit of d in each period allocated towards investment. In practice, these budget constraints arise not only from a lack of cash reserves but also from the limited managerial attention available. Consequently, the firm can invest in at most one type of product per period. 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 a different customer segment or (2) develop a similar product that targets the same customer segment as that of 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 B.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 develops a similar product, it incurs the fixed cost d [40], and the demand for the similar product is *sticky* in the sense that its demand in period two, q_2, is given by:

$$q_2 = \rho q_1 + \epsilon, \tag{5}$$

where the parameter $\rho \in (0, \infty)$ captures the stickiness of the demand for the similar product,[41] and $\epsilon \sim N(0, \sigma_\epsilon^2)$ represents the exogenous shock to demand.[42] 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 customer segment to generate a period two demand that approximates the period one demand realized for the existing product.

[40] For simplicity, we assume that the fixed cost d is identical for both developing new and similar products. Allowing for a lower fixed cost for similar products would not alter our qualitative results; rather, it would strengthen them (see Appendix B.2 for the proof).

[41] Demand stickiness, ρ, can take any positive value in this 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 $(0, 1]$.

[42] Klepper (1996) uses a similar demand schedule, where demand remains relatively stable across periods.

A.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:

$$\pi_2^N = (p - c)\mu - d, \tag{6}$$

$$\pi_2^S = (p - c)\rho q_1 - d. \tag{7}$$

Firms choose 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:

$$\pi_2 = \max_{\{N,S\}} \left[\pi_2^N, \pi_2^S\right], \tag{8}$$

$$\text{s.t. } \pi_2 \geq 0, \tag{9}$$

where Equation (9) 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 Equation (8).

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 in period two:*

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

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

The proof of Proposition 1 can be found in Appendix B. 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, implying that developing similar products becomes more attractive for firms when demand stickiness is higher.

Figure 4 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 choose to develop a similar product expands.

[Figure 4 about here.]

A.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 likely to invest in* similar *products, compared to firms with lower sales from existing products.*

This prediction follows directly from the implication of Proposition 1, which indicates that there exists a sales threshold above which firms are more likely to develop similar products. The next prediction, illustrated in Figure 4, highlights the role of demand stickiness in shaping firms' incentives to invest in products that are similar to their previous offerings.

Prediction 2. *Firms are more likely to invest in similar products when demand stickiness is higher.*

B Proof and Additional Analysis

B.1 Proof

For the following proof, we use the time subscript t for generality. In the two-period model presented in Appendix A, we can specify $t = 2$ and $t - 1 = 1$.

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

$$\pi_t^N \geq \pi_t^S \implies q_{t-1} \geq \frac{\mu}{\rho} > 0. \tag{10}$$

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

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

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

In Appendix A 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:

$$\pi_t^N = (p - c)\mu - d^N, \tag{11}$$

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

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_{t-1} that delineates two distinct regions for the optimal investment decision:*

1. *Firms with an existing product demand $q_{t-1} \in [0, q^*]$ will choose to invest in new product development.*

2. *Firms with an existing product demand $q_{t-1} \in [q^*, \infty)$ 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.

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

In Appendix A 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_{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., $q_t = q_{t-1}$.

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

$$\pi_t^N = (p-c)(\mu + q_{t-1}) - d, \tag{13}$$

$$\pi_t^S = (p-c)(\rho+1)q_{t-1} - d. \tag{14}$$

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 $(p-c)q_{t-1}$ cancels out between π_t^N and π_t^S, leaving the threshold value q^* unaffected.

C Constructing the Sample

C.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.

C.2 Product matching

C.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.

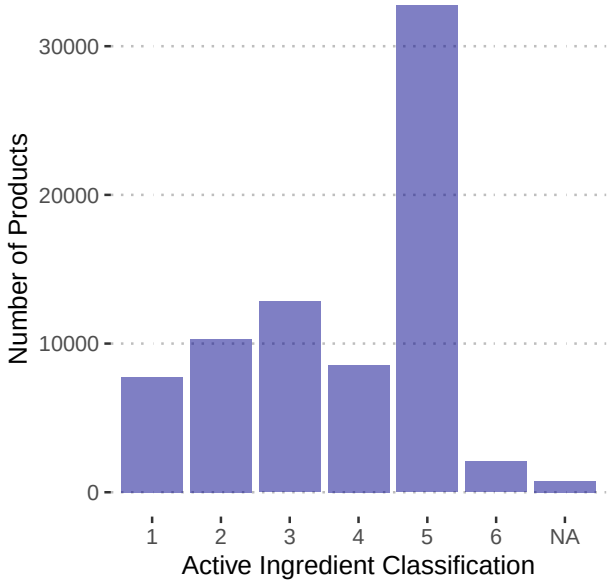
A total of 4,784 active ingredients were classified into six categories. The categories were created through an iterative process as we learned more about the type of active ingredients in the data. The six categories are listed below, and the frequency distribution of products for each classification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

D Data

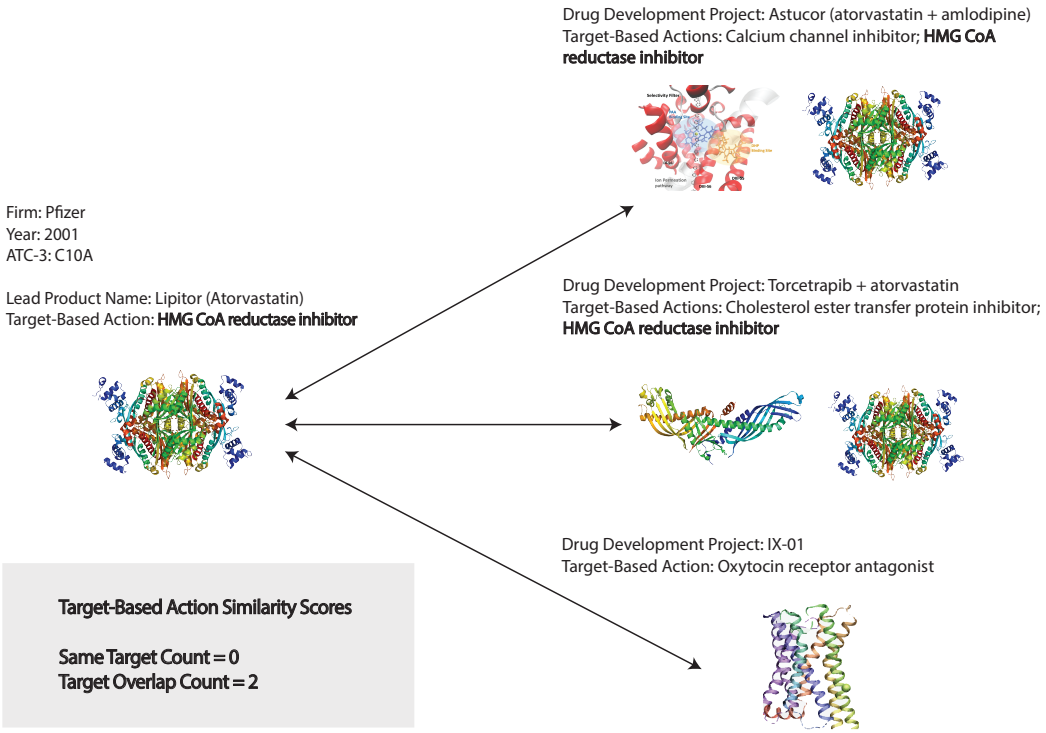
Measure Name	Variable	Description of Measure	Data Source
Early-Stage Project Count$_{t+5}$	Investment intensity in early-stage development	The count of projects in early-stage drug development which firm i has ongoing in ATC-3 j in year $t+5$. We define projects as being in early stages if year t falls in the interval $x-1 < t < x+1$ where x is the project start year.	Cortellis
Late-Stage Project Count$_{t+5}$	Investment intensity in late-stage development	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.	Cortellis
Early-Stage Project Share$_{t+5}$	Relative investment intensity in early-stage development	The share of projects in early-stage drug development which firm i has ongoing in ATC-3 j in year $t+5$ out 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.	Cortellis
Target Overlap Count$_{t+5}$	Target-based action technical similarity of early-stage development projects	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.	Cortellis
Target Overlap Share$_{t+5}$	Relative target-based action technical similarity of early-stage development projects	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.	Cortellis
Same Target Count$_{t+5}$	Target-based action technical commonality of early-stage development projects	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.	Cortellis
Same Target Share$_{t+5}$	Relative target-based action technical commonality of early-stage development projects	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.	Cortellis
Molecularly Similar Projects$_{t+5}$	Molecular technical similarity of early-stage development projects	The count of early-stage drug development projects in ATC-1 k that share an 85% or greater overlap in molecular structure with the lead sales product i.e., with a Tanimoto score greater than 0.85.	Cortellis, ChemMine Tools
Mean Tanimoto Score$_{t+5}$	Average molecular technical similarity of early-stage development projects	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.	Cortellis, ChemMine Tools
Max Tanimoto Score$_{t+5}$	Greatest molecular technical similarity of early-stage development projects	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.	Cortellis, ChemMine Tools
Lead Product Sales$_t$	Absolute product sales	The absolute sales in USD billions of firm i's lead selling product in ATC-3 j in year t.	IQVIA
\$1Bn lead Product$_t$	Especially high absolute product sales	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.	IQVIA
Sales Share Lead Product/Firm$_t$	Relative product sales	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.	IQVIA
\$3Bn lead Product$_t$	Especially high absolute product sales	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.	IQVIA

Lead Product Sales$_t$ (1998 dollars)	Inflation-adjusted product sales	The sales in USD billions of firm i's lead selling product in ATC-3 j in year t adjusted for 1998 dollars.	IQVIA
\$1Bn lead Product$_t$ (1998 dollars)	Inflation-adjusted especially high product sales	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.	IQVIA
Cash Flow$_t$	Absolute cash reserves	Firm i's total cash flow in year t.	COMPUSTAT
R&D Expenditure$_t$	Total expenditure on R&D activity	Firm i's total R&D expenditure in year t.	COMPUSTAT
Number of Competitors$_t$	Intensity of competition	The number of parents firms actively selling branded therapeutic drugs in ATC-3 j in year t.	IQVIA
NCE expiration$_t$ (<5yr)	Proximity to the patent cliff	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.	FDA Orange Book
Lower Relative Incidence of Side Effects$_t$	Extent to which a firm's product is safer than it's direct competitors	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.	FAERS
ATC3-Year Side Effects Count$_t$	The safety of drugs in a therapeutic class	The total count, in thousands, of adverse side effect events reported to the FDA in ATC3-Year j,t.	FAERS
Market Commercial Potential$_t$	The likelihood that science in a therapeutic class can be translated into commercially successful innovations	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).	Masclans et al. (2024)
ATC1 Detailing Expenditure$_{t-3}$	Expenditure on marketing	The sum of firm i's ATC-1 k detailing expenditure in USD billions within year $t-3$.	IQVIA
Patent Applications$_t$	Development capabilities	The number of patents firm i applied for in year t which pertain to ATC-3 j.	FDA Orange Book
Scientific Publications$_t$	Research capabilities	The count of scientific articles published by firm i in year t which pertain to ATC-3 j.	Branstetter et al. (2022)
Total Other Sales$_t$	Firm size	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.	IQVIA
Total Products$_t$	Development Capabilities	Firm i's total number (in hundreds) of products in sales in year t.	IQVIA

Table D.1: Correlation Matrix

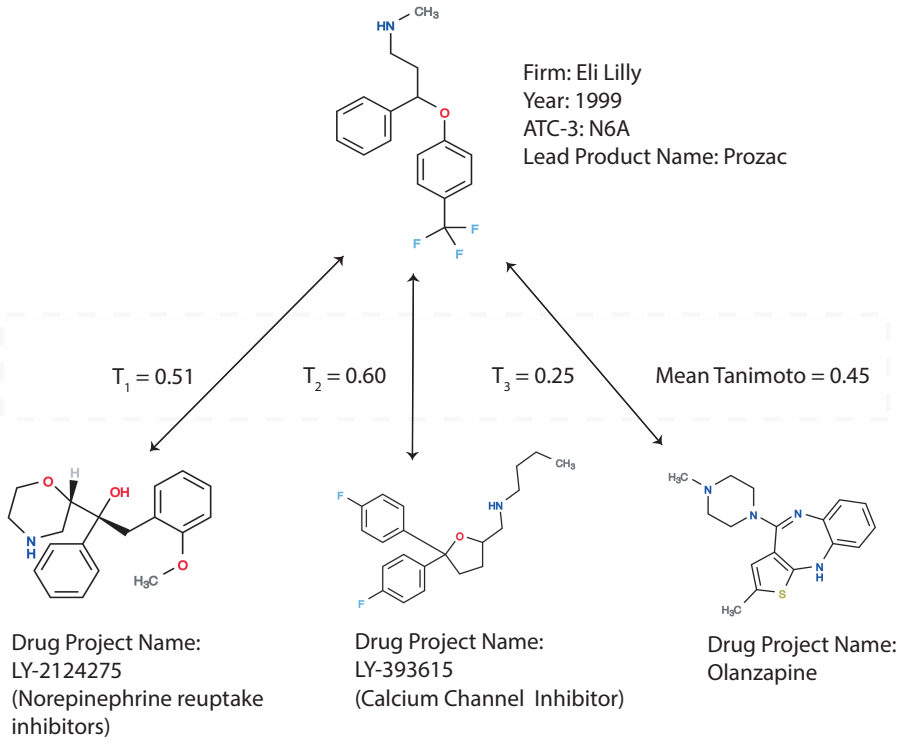
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
Early-Stage Projects$_{t+5}$	1.00																
Target Overlap Projects$_{t+5}$	0.21	1.00															
Molecularly Similar Projects$_{t+5}$	0.07	0.32	1.00														
Same Molecule Projects$_{t+5}$	0.05	0.23	0.67	1.00													
Lead Product Sales$_{t}$	0.14	0.10	0.19	0.13	1.00												
Biotechs$_{t}$	0.38	0.00	0.00	0.01	0.07	1.00											
Deals$_{t}$	0.30	0.12	0.02	0.04	0.06	0.04	1.00										
ATC1 Detailing$_{t-3}$	0.13	0.08	0.04	0.03	0.27	-0.01	0.18	1.00									
Patent Applications$_{t}$	0.37	0.06	0.02	0.03	0.04	0.34	0.31	0.05	1.00								
Scientific Publications$_{t}$	0.21	0.05	0.03	0.02	0.09	0.11	0.16	0.11	0.15	1.00							
Total Other Sales$_{t}$	0.34	0.07	0.01	0.04	0.10	0.03	0.48	0.34	0.10	0.14	1.00						
R&D Expenditure$_{t}$	-0.02	-0.02	0.00	-0.02	-0.01	0.04	-0.02	-0.02	-0.03	-0.01	-0.07	1.00					
Total Products$_{t}$	0.09	-0.00	-0.01	0.00	-0.03	-0.07	0.17	0.18	0.09	0.06	0.32	-0.11	1.00				
Adverse Event Rate$_{t}$	0.01	-0.01	0.00	0.00	0.01	0.02	0.06	-0.12	-0.00	-0.00	0.05	-0.04	0.00	1.00			
Adverse Event Rate$_{t}$ (excl. focal)	0.03	0.01	0.02	0.03	-0.01	0.11	0.01	-0.20	0.02	0.04	-0.00	-0.00	-0.06	0.63	1.00		
NCE expiration$_{t}$ (<5yr)	0.08	0.10	0.10	0.13	0.10	0.07	-0.00	0.08	0.04	0.03	0.03	0.00	-0.01	0.05	0.05	1.00	
Number of Competitors$_{t}$	0.14	-0.03	0.00	-0.00	0.06	0.27	-0.08	0.13	0.09	0.05	-0.09	0.04	-0.11	-0.23	0.13	0.00	1.00

Figure D.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.

Figure D.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.

E ATC-4 Level Data

E.1 Assigning ATC4s to drug development projects

The task is to predict the ephMRA ATC level four code for each drug development project. We design the following classification function which uses a few-shot prompting approach. Within each 'conversation', we first use example input data (three Cortellis projects), and then provide example output data (manually assigned ATC4 codes). We repeatedly query the *ChatGPT-4.0 API* with data for each drug development project. The output is one or more ATC4 codes which are predicted to be applicable to the drug development project. 97.9% of projects that we attempt to classify are matched to at least one ATC4 code.

Classification function:

```
def classify(data, example_input, example_output, ephmra_df):

    system_prompt = f"""
    Please identify the ephMRA atc4 code or codes which most
    closely matches the drug development project.

    ALL of these are {area} drugs.

    Here is a list of {area} ephMRA atc LEVEL FOUR codes
    and their descriptions that I would like you to use:

    {ephmra_df}

    You MUST return a JSON object, which follows this exact format:

    {{
      "Drug.Id": <id>,
      "ephMRA_atc4s": "<semicolon separated list>"
    }}

    DO NOT include ANY OTHER TEXT.
    """

    messages = [
        {"role": "system",
        "content": system_prompt},
        {"role": "user",
        "content": f"""Please identify the ATC4 LEVEL FOUR code
        for this complete list of {area} drugs: {example_input}"""},
        {"role": "assistant",
        "content": example_output},
        {"role": "user",
        "content": f"""Please identify the ATC4 LEVEL FOUR code
        for this complete list of {area} drugs: {data}"""}
```

```
]

response = client.chat.completions.create(
    messages=messages,
    model=[INSERT MODEL NAME HERE]
)

return response.choices[0].message.content
```

Inputs:

- *data*: a chunk of raw Cortellis data, including the Drug Name, list of Active Indications, Target-based Actions (where applicable, SMILES code (where applicable), and Extract.
- *example_input*: three Cortellis drug development projects within *area*.
- *example_output*: manually labelled ATC4 codes for the three Cortellis drug development projects in *example_input*.
- *area*: Broader therapy area (i.e., ATC1).
- *ephmra_df*: List of ATC4 codes and their descriptions, retrieved from IQVIA

E.2 Descriptive Statistics and Validation

Table E.1 summarizes all variables used in the baseline analysis of the ATC4 level dataset. The dataset includes 11,744 Firm–ATC4–Year observations, covering the ATC1s for Cardiology (C), Antineoplastic and immunomodulating agents (L), and Neurology (N). These are the largest ATC1s by sales in our dataset. Note that ATC1 L is where the majority of cancer drugs are classified.

Table E.1: Firm-Year-ATC4 Level Descriptive Statistics

	Mean	St. Dev.	Min	Q1	Median	Q3	Max	N
Early-Stage Projects$_{t+5}$	0.140	0.758	0.000	0.000	0.000	0.000	17.000	11865
Target Overlap Projects$_{t+5}$	0.076	0.334	0.000	0.000	0.000	0.000	5.000	8697
Molecularly Similar Projects$_{t+5}$	0.054	0.293	0.000	0.000	0.000	0.000	4.000	6851
Same Molecule Projects$_{t+5}$	0.029	0.189	0.000	0.000	0.000	0.000	2.000	6851
Lead Product Sales$_t$	0.106	0.442	0.000	0.000	0.003	0.027	9.677	11865
Year	2003.744	3.700	1998.000	2001.000	2004.000	2007.000	2010.000	11865

Table E.2 reports validation results. We take a random sample of 10 projects and inspect the accuracy of their ATC4 assignments. Note: drugs and firms have been anonymized due to data confidentiality.

E.3 Descriptive Results

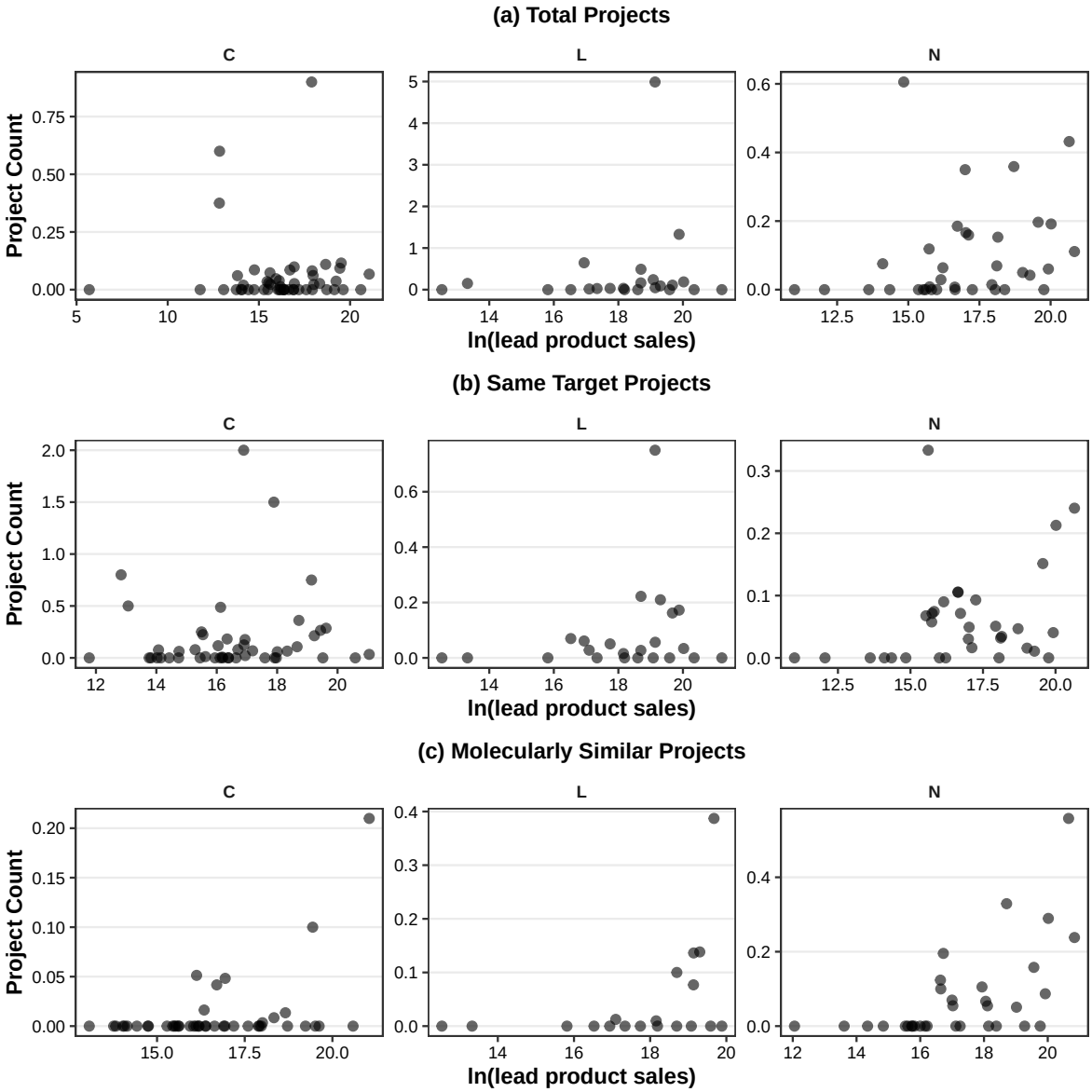
For the three therapeutic class categories with the highest sales volumes–Cardiology (C), Antineoplastic and immunomodulating agents (L), and Neurology (N)–we explore correlations between product sales and developments at the more granular fourth level of the ATC system. In Table E.1 we show the relationship between (logged) lead product sales and early-stage projects (at $t + 5$) for each ATC-4 in each of the three ATC1s, averaged across firms and years. Consistent with the ATC3-level analyses, product sales are positively associated with a higher volume of subsequent

Table E.2: Evaluation of EphMRA ATC4 Classifications

Drug Id	ATC4	Indication	Target	Extract	Evaluation
XX1	N5C0	NA	Sigma opioid receptor agonist	"LU-29253 is a potent, selective sigma-2 opioid agonist ... developed for anxiety ..."	N5C0 (*Anxiolytics, other*) fits a sigma-opioid receptor agonist aimed at anxiety. Mechanism aligns with anxiolytic pharmacology. Other possible classifications: N7X0 (other psycholeptics).
XX2	L4A0	Cancer	NA	"MegaOX40L ... an immunotherapy for prostate cancer and other tumors ... development ongoing ..."	L4A0 (*Antineoplastic agents, other immunomodulators*) correctly classifies an OX40L ligand immunotherapy under oncology. No alternative classification apparent.
XX3	L1X9	NA	p53 tumor suppressor protein modulator	"ALT-802 ... fusion protein of IgG1 and soluble T-cell receptor ... directed against p53 antigen ... for cancer ..."	L1X9 (*Other antineoplastic agents*) fits a TCR-Ig fusion targeting p53 in cancer immunotherapy. Other potential classifications: L4A0 (antineoplastic immunomodulators).
XX4	L4A0	NA	NA	"Cell-based vaccine ... potential treatment of cancer ..."	L4A0 (*Antineoplastic immunomodulators*) is appropriate for a cell-based cancer vaccine. Other possible classification: L1X9 (other antineoplastics).
XX5	N4A0	NA	Catechol O-methyltransferase inhibitor	"COMT inhibitor ... developed for Parkinson's disease ... phase I trial ... discontinued ..."	N4A0 (*Antiparkinson drugs, COMT inhibitors*) matches mechanism and indication. No alternative classification apparent.
XX6	L1X3	NA	NA	"Bi-specific antibody ... redirects T-cells to kill tumor cells ... immunotherapy ..."	L1X3 (*Antineoplastic agents, monoclonal antibodies*) fits a bispecific antibody immunotherapy for cancer. Could also fit L4A0 (immunomodulators).
XX7	N7X0	NA	Opioid receptor antagonist; Sigma receptor antagonist	"CNS-1169 ... opiate antagonist ... developed as potential treatment for schizophrenia ..."	N7X0 (*Other psycholeptics / antipsychotics*) is a reasonable match for an opioid antagonist for schizophrenia, though borderline. Alternative: N7A0 (antipsychotics, other).
XX8	L1X4	Autoimmune disease; Leukemia	Flt3 tyrosine kinase inhibitor	"Flt3 kinase inhibitors ... for AML and autoimmune diseases ... lead optimisation studies ..."	L1X4 (*Protein kinase inhibitors*) is the correct class for Flt3 inhibitors. Alternative: L1X9 (other antineoplastic agents).
XX9	—	NA	Amiloride-sensitive cation channel inhibitor	"ASIC inhibitors ... antagonize calcium ions ... potential treatment for acute stroke ..."	No ATC4 provided. Mechanism suggests N7B (other CNS agents) or C8C (calcium channel blockers); clarification needed.
XX10	L4A0	Colorectal tumor	Indoleamine 2,3-dioxygenase 1 inhibitor	"TQBWX-220 ... potent small-molecule IDO-1 inhibitor ... for colorectal cancer ..."	L4A0 (*Antineoplastic immunomodulators*) suits an IDO-1 inhibitor in immuno-oncology. Other potential: L1X9 (other antineoplastics).

innovation with a narrowly defined therapeutic class, as well as a higher count of technologically similar projects, measured by both biological target and the drug's chemistry.

Figure E.1: ATC4 level correlations between product sales, the count of early-stage development projects, and the count of technologically similar projects



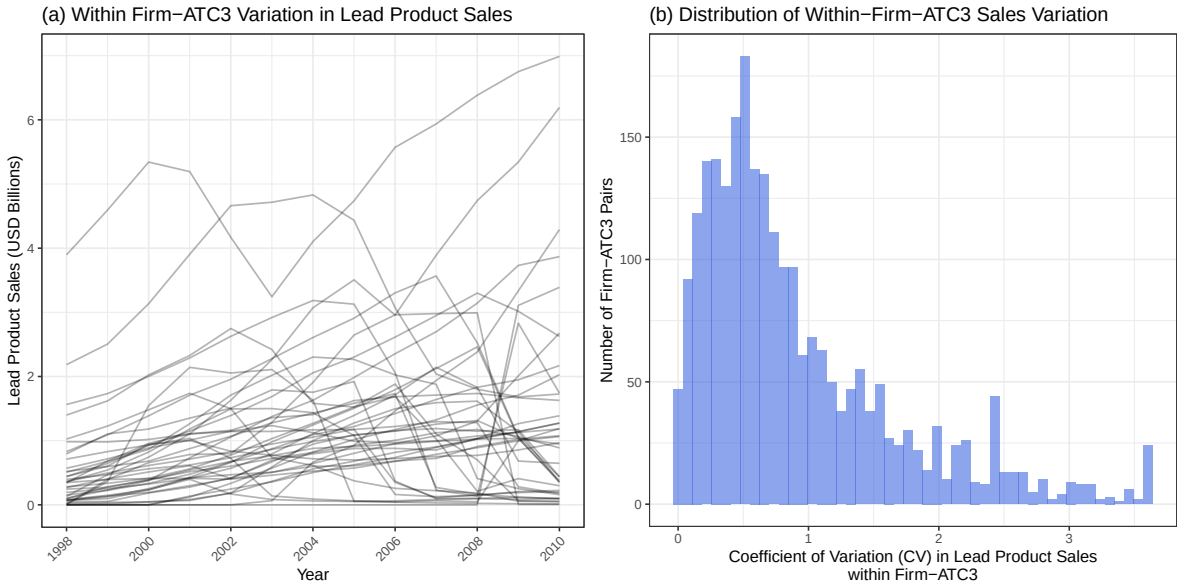
Notes: This figure plots the relationship between (logged) mean lead product sales and mean early-stage projects (at $t + 5$) for each ATC-4. We show a panel for each ATC-1 for which we predict the ATC-4 code of development projects. The ATC1s included are Cardiology (C), Antineoplastic and immunomodulating agents (L), and Neurology (N). These are the largest ATC1s by sales in our dataset. Note that ATC1 L is where the majority of cancer drugs are classified.

F Empirical Specification

While we have an average of 410 ATC3-year observations per firm (mean 8.89 years and 44.6 distinct ATC-3s), one might wonder if there is enough within-firm variation in product sales and product development, especially given the sticky nature of pharmaceutical sales.

On average we observe 1.55 (median=1) different lead products in a firm-ATC3. Thus for most firm-ATC3s, the dominant sales product remains the same during our sample period. However, their sales are not flat. Figure F.1 panel (a) plots within-pair trajectories for a random sample of 40 firm–ATC3 cells. The figure shows that, even conditional on the same lead product, sales move substantially over time. Panel (b) shows the distribution of within–pair coefficients of variation (median CV = 0.67), demonstrating substantial time-series movement even conditional on the same lead product. Firm fixed effects therefore remove only levels and not the identifying within-sales variation.

Figure F.1: Within Firm-ATC3 variation in Lead Product Sales



Notes: This figure illustrates the variation in sales over time for lead products in a firm-ATC3. Panel (a) plots within-pair trajectories for a random sample of 40 firm–ATC3 cells. For ease of exposition and comparison, we sample observations where the lead selling product reaches $1 billion in sales at some point during the sample period. The figure shows that, even conditional on the same lead product, sales move substantially over time. Panel (b) shows the distribution of within–pair coefficients of variation (median CV = 0.67), where the coefficient of variation is the ratio of the standard deviation to the mean.

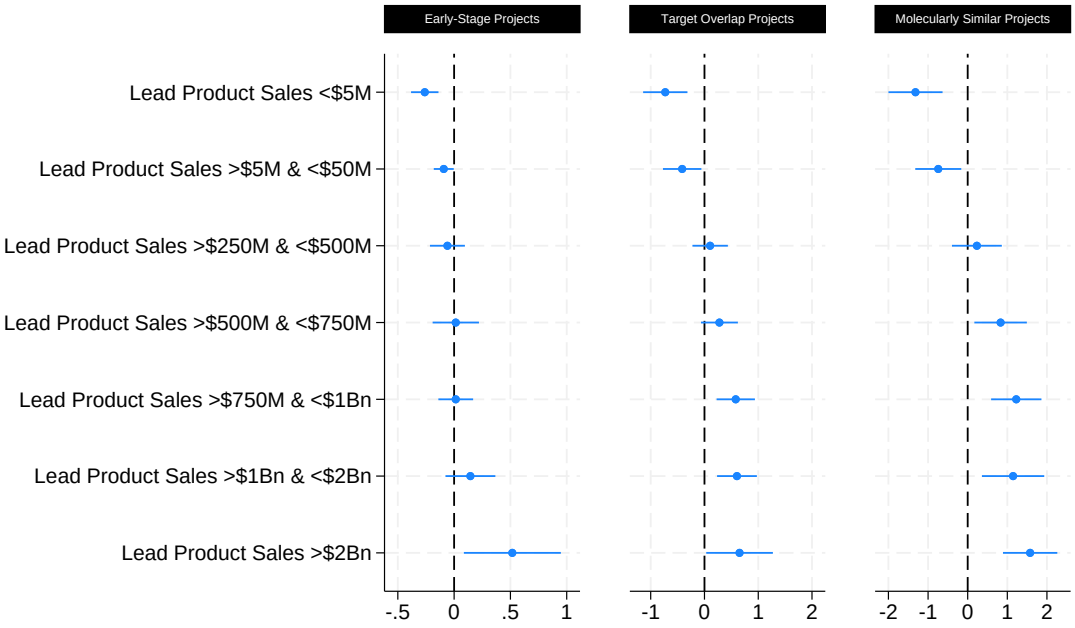
G Robustness Analysis

Table G.1: Baseline Results with Alternate Dependent-Variable Horizons

	(1) Early-Stage Projects$_{t+1}$	(2) Early-Stage Projects$_{t+3}$	(3) Early-Stage Projects$_{t+5}$	(4) Small-Molecule Early-Stage Projects$_{t+5}$	(5) Biologic Early-Stage Projects$_{t+5}$	(6) Late-Stage Projects$_{t+5}$
Lead Product Sales$_t$	0.149	0.155	0.161	0.122	0.115	0.149
	(0.031)	(0.033)	(0.036)	(0.023)	(0.064)	(0.031)
ATC1 × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	20,185	20,838	20,838	8,650	8,302	20,185
Pseudo R-squared	0.415	0.423	0.421	0.225	0.199	0.415

Notes: This table reports Poisson pseudo-maximum-likelihood estimates of the relationship between lead product sales and follow-on project investment using alternative outcome horizons and project-type splits. The coefficients remain positive across the shorter and longer horizons and across the small-molecule, biologic, and late-stage variants, indicating that the baseline relationship is not driven by one particular window or project category. Standard errors are clustered at the firm level and reported in parentheses.

Figure G.1: Coefficient Plot of Absolute Lead Product Sales on Similarity Measures



Notes: This figure plots coefficient estimates from Poisson pseudo-maximum-likelihood regressions that replace continuous lead product sales with indicator variables for different sales bands. The omitted category is lead product sales between $50M and $250M, so each point shows how the outcome changes relative to that benchmark. The figure highlights the nonlinearity in the blockbuster effect across early-stage projects, target-overlap projects, and molecularly similar projects. The underlying regressions include ATC1-year and firm fixed effects, and standard errors are clustered at the firm level.

Table G.2: Baseline Results Using Inflation-Adjusted Lead Product Sales

	Early-Stage Projects$_{t+5}$	Target Overlap Projects$_{t+5}$	Molecularly Similar Projects$_{t+5}$	Same Molecule Projects$_{t+5}$
	(1)	(2)	(3)	(4)
Lead Product Sales$_t$ (1998 dollars)	0.192	0.325	0.741	0.544
	(0.044)	(0.065)	(0.112)	(0.130)
ATC1 × Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
Observations	20,838	13,033	11,820	9,367
Pseudo R-squared	0.421	0.234	0.176	0.127

Notes: This table re-estimates the main specifications using lead product sales expressed in 1998 dollars. The coefficients remain positive across all four outcomes, indicating that the main results are not an artifact of nominal price growth or inflation over time. Estimates are from Poisson pseudo-maximum-likelihood models. Standard errors are clustered at the firm level and reported in parentheses.

Table G.3: Baseline Results Using Lead Product Sales Share

	Early-Stage Projects$_{t+5}$	Target Overlap Projects$_{t+5}$	Molecularly Similar Projects$_{t+5}$	Same Molecule Projects$_{t+5}$
	(1)	(2)	(3)	(4)
Sales Share Lead Product/Firm$_t$	1.218	2.667	4.416	3.651
	(0.706)	(0.787)	(0.858)	(1.115)
ATC1 × Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
Observations	20,838	13,033	11,820	9,367
Pseudo R-squared	0.418	0.239	0.178	0.134

Notes: This table replaces absolute lead product sales with the lead product's share of total firm sales. The positive coefficients show that the relationship between blockbuster products and related follow-on investment is not solely about absolute scale; it also appears when the focal product is measured relative to the firm's overall sales base. Estimates are from Poisson pseudo-maximum-likelihood models. Standard errors are clustered at the firm level and reported in parentheses.

Table G.4: Baseline Results Estimated with OLS

	ln(1 + Early-Stage Projects$_{t+5}$)	ln(1 + Target Overlap Projects$_{t+5}$)	ln(1 + Molecularly Similar Projects$_{t+5}$)	ln(1 + Same Molecule Projects$_{t+5}$)
	(1)	(2)	(3)	(4)
Lead Product Sales$_t$	0.079	0.061	0.063	0.026
	(0.061)	(0.019)	(0.015)	(0.010)
ATC1 × Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
Observations	22,971	16,316	14,316	14,316
Adjusted R-squared	0.457	0.157	0.076	0.040

Notes: This table estimates log-transformed versions of the main count outcomes using OLS rather than Poisson models. The positive coefficients for the technological similarity outcomes (columns 2-4) indicate that the main pattern is not driven purely by the Poisson functional form, although the same ATC3 project count estimate is insignificant in this linear specification (column 1). Standard errors are clustered at the firm level and reported in parentheses.

Table G.5: Baseline Results with Alternative Dependent Variables

	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$_{t+5}$	Mean Tanimoto Score$_{t+5}$	Max Tanimoto Score$_{t+5}$
	(1) OLS	(2) Poisson	(3) Poisson	(4) OLS	(5) OLS	(6) OLS	(7) OLS	(8) OLS
Lead Product Sales$_t$	0.013	0.149	0.388	0.010	0.003	0.007	0.012	0.040
	(0.008)	(0.031)	(0.058)	(0.006)	(0.001)	(0.002)	(0.004)	(0.010)
ATC1 × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	22,971	20,185	11,717	16,316	16,316	14,276	14,276	14,276
Adjusted R-squared	0.088			0.182	0.095	0.284	0.138	0.241
Pseudo R-squared		0.415	0.208					

Notes: This table broadens the outcome space to include project shares, late-stage projects, same-target measures, and alternative summaries of molecular similarity. The estimates are generally positive across these alternative outcomes, suggesting that larger lead products are associated not only with more projects, but also with more concentrated and more technically related follow-on investment. The estimator used in each column is shown in the table. Standard errors are clustered at the firm level and reported in parentheses.

Table G.6: Baseline Results with Adverse Event Rate Interactions

	Early-Stage Projects$_{t+5}$	Target Overlap Projects$_{t+5}$	Molecularly Similar Projects$_{t+5}$	Same Molecule Projects$_{t+5}$
	(1)	(2)	(3)	(4)
Lead Product Sales$_t$	0.453	0.922	0.557	1.108
	(0.211)	(0.207)	(0.402)	(0.438)
Adverse Event Rate$_t$	-0.032	-0.044	0.032	0.060
	(0.011)	(0.041)	(0.069)	(0.089)
Lead Product Sales$_t$ × Adverse Event Rate$_t$	0.046	0.101	0.011	0.119
	(0.029)	(0.029)	(0.061)	(0.063)
ATC1 × Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
Observations	9,337	5,749	5,332	4,099
Pseudo R-squared	0.398	0.202	0.221	0.150

Notes: This table interacts lead product sales with the focal product's market adverse event rate to test whether safety conditions moderate the blockbuster effect. In this table, the adverse event rate includes those from the focal product. The baseline lead-sales coefficients remain positive, while the interaction terms show how that relationship varies with product risk across the four outcomes. Estimates are from Poisson pseudo-maximum-likelihood models. Standard errors are clustered at the firm level and reported in parentheses.

Table G.7: Sensitivity of Early-Stage Projects to Detailing Controls

	Early-Stage Projects$_{t+5}$			
	(1) Baseline	(2) ATC3 Detailing	(3) ATC1 Detailing	(4) Firm-Year Detailing
Lead Product Sales$_t$	0.161	0.125	0.168	0.157
	(0.036)	(0.049)	(0.034)	(0.035)
ATC3 Detailing$_{t-3}$		2.357		
		(1.746)		
ATC1 Detailing$_{t-3}$			-1.257	
			(0.916)	
Firm-Year Detailing$_{t-3}$				0.024
				(0.719)
ATC1 × Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
Observations	20,838	16,062	15,866	16,050
Pseudo R-squared	0.421	0.432	0.433	0.432

Notes: This table adds alternative detailing controls to the early-stage project specification to test whether the baseline result is sensitive to the level at which sales-force intensity is measured. The coefficient on lead product sales remains similar across columns (1), (3), and (4), implying that the effect of lead product sales on early-stage investment is not effected by higher level detailing proxies. The estimate is most sensitive in column (2), where detailing is measured at the ATC3 level. However, at this level, the collinearity with lead product sales (also measured at the ATC3 level) is greatest. Estimates are from Poisson pseudo-maximum-likelihood models. Standard errors are clustered at the firm level and reported in parentheses.

Table G.8: Same-Molecule Projects: Detailing and Adverse Event Mechanisms

	Same Molecule Projects$_{t+5}$						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Lead Product Sales$_t$	0.443	0.422	0.822	0.224	0.225	0.459	0.366
	(0.105)	(0.120)	(0.227)	(0.115)	(0.114)	(0.165)	(0.244)
ATC1 Detailing$_{t-3}$		-0.284	1.453				-7.459
		(3.933)	(4.016)				(4.356)
Lead Product Sales$_t$ × ATC1 Detailing$_{t-3}$			-3.778				1.687
			(1.778)				(1.852)
Adverse Event Rate$_t$ (excl. focal)					0.035	-0.006	-0.009
					(0.041)	(0.040)	(0.041)
Lead Product Sales$_t$ × Adverse Event Rate$_t$ (excl. focal)						0.065	0.071
						(0.023)	(0.026)
ATC1 × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9,367	6,539	6,539	1,883	1,883	1,883	1,826
Pseudo R-squared	0.126	0.122	0.128	0.168	0.170	0.178	0.188

Notes: This table focuses on same-molecule follow-on projects and asks whether detailing intensity and adverse event conditions account for the relationship between blockbuster sales and within-molecule development. The lead-sales coefficient remains positive across specifications, but in columns (6) and (7) the interaction with the adverse event rate measure provides evidence of the demand stickiness mechanism. Estimates are from Poisson pseudo-maximum-likelihood models. Standard errors are clustered at the firm level and reported in parentheses.

Table G.9: Same-Molecule Projects: Supply-Side Capability Correlates

	Same Molecule Projects$_{t+5}$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Lead Product Sales$_t$	0.442	0.459	0.443	0.397	0.455	0.446	0.479	0.467
	(0.107)	(0.104)	(0.105)	(0.100)	(0.132)	(0.107)	(0.109)	(0.118)
Patent Applications$_t$	0.007							0.005
	(0.030)							(0.026)
Scientific Publications$_t$		-0.009						-0.009
		(0.008)						(0.008)
Deals$_t$			-0.014					-0.006
			(0.018)					(0.021)
Total Other Sales$_t$				-0.058				-0.029
				(0.045)				(0.060)
R&D Expenditure$_t$					-0.036			0.018
					(0.088)			(0.108)
Total Products$_t$						0.319		0.364
						(0.505)		(0.532)
NCE expiration$_t$ (<5yr)							1.524	1.511
							(0.212)	(0.240)
ATC1 × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9,367	9,367	9,367	9,367	8,283	9,367	9,367	8,283
Pseudo R-squared	0.126	0.127	0.126	0.127	0.108	0.126	0.159	0.141

Notes: This table adds supply-side capability measures one at a time and jointly for same-molecule follow-on projects. The coefficient on lead product sales remains positive and at a consistent magnitude across models, indicating the importance of product sales in explaining investment in same-molecule projects. Estimates are from Poisson pseudo-maximum-likelihood models. Standard errors are clustered at the firm level and reported in parentheses.

Table G.10: Similar-but-Different Molecule Projects: Detailing and Adverse Event Mechanisms

	Molecularly Similar but Distinct Projects$_{t+5}$						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Lead Product Sales$_t$	1.048	0.873	1.369	1.235	1.233	1.144	1.875
	(0.138)	(0.118)	(0.281)	(0.283)	(0.264)	(0.310)	(0.388)
ATC1 Detailing$_{t-3}$		4.014	6.710				23.518
		(6.395)	(5.594)				(9.290)
Lead Product Sales$_t$ × ATC1 Detailing$_{t-3}$			-4.193				-9.655
			(1.937)				(2.460)
Adverse Event Rate$_t$ (excl. focal)					-0.078	-0.055	0.007
					(0.069)	(0.074)	(0.083)
Lead Product Sales$_t$ × Adverse Event Rate$_t$ (excl. focal)						-0.032	-0.087
						(0.054)	(0.053)
ATC1 × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	8,255	6,145	6,145	1,052	1,052	1,052	1,046
Pseudo R-squared	0.229	0.284	0.291	0.404	0.409	0.410	0.452

Notes: This table examines molecularly similar but chemically distinct follow-on projects (Tanimoto similarity score > 0.85 but < 1.0). The lead-sales coefficient remains large and positive after adding detailing and adverse-event interactions, suggesting that the blockbuster effect for nearby-but-distinct molecules is not fully absorbed by those two channels. Estimates are from Poisson pseudo-maximum-likelihood models. Standard errors are clustered at the firm level and reported in parentheses.

Table G.11: Similar-but-Different Molecule Projects: Supply-Side Capability Correlates

	Molecularly Similar but Distinct Projects$_{t+5}$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Lead Product Sales$_t$	1.048	0.968	0.980	0.964	1.041	0.979	0.981	1.028
	(0.139)	(0.133)	(0.139)	(0.144)	(0.168)	(0.140)	(0.134)	(0.169)
Patent Applications$_t$	-0.002							0.023
	(0.035)							(0.037)
Scientific Publications$_t$		0.005						0.007
		(0.008)						(0.010)
Deals$_t$			0.020					0.034
			(0.021)					(0.024)
Total Other Sales$_t$				-0.021				0.009
				(0.047)				(0.050)
R&D Expenditure$_t$					0.008			0.006
					(0.006)			(0.007)
Total Products$_t$						1.249		0.917
						(1.043)		(1.233)
NCE expiration$_t$ (<5yr)							0.595	0.584
							(0.347)	(0.362)
ATC1 × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	8,255	9,126	9,126	9,126	7,807	9,126	9,126	7,568
Pseudo R-squared	0.229	0.236	0.236	0.236	0.215	0.238	0.238	0.218

Notes: This table relates molecularly similar but distinct follow-on projects to the same set of supply-side capability controls used above. The lead-sales coefficient stays consistent after adding these controls, which suggests that the relationship is not merely reflecting broader innovative capacity or lifecycle variables. Estimates are from Poisson pseudo-maximum-likelihood models. Standard errors are clustered at the firm level and reported in parentheses.

Table G.12: ATC4 Validation Sample Baseline Results

	Early-Stage Projects$_{t+5}$	Target Overlap Projects$_{t+5}$	Molecularly Similar Projects$_{t+5}$	Same Molecule Projects$_{t+5}$
	(1)	(2)	(3)	(4)
Lead Product Sales$_t$	0.339	0.259	0.551	0.401
	(0.125)	(0.131)	(0.086)	(0.131)
ATC1 × Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
Observations	5,740	5,295	4,226	3,215
Pseudo R-squared	0.235	0.138	0.227	0.160

Notes: This table reproduces the baseline specifications in the finer ATC4 validation sample. The positive coefficients across all four outcomes indicate that the main results are not driven by aggregation at the ATC3 level. Estimates are from Poisson pseudo-maximum-likelihood models. Standard errors are clustered at the firm level and reported in parentheses.

Table G.13: Putting It All Together: Supply-side and demand-side mechanisms

	Early Stage Projects$_{t+5}$	Target Overlap Projects$_{t+5}$	Molecularly Similar Projects$_{t+5}$
	(1)	(2)	(3)
Lead Product Sales$_t$	0.030	0.597***	0.945***
	(0.136)	(0.167)	(0.281)
ATC1 Detailing$_{t-3}$	-1.189	9.021***	3.165
	(1.406)	(2.706)	(3.850)
Lead Product Sales$_t$ × ATC1 Detailing$_{t-3}$	1.064	-1.226	-2.794
	(0.861)	(1.425)	(1.898)
Adverse Event Rate$_t$ (excl. focal)	-0.008	0.017	0.004
	(0.007)	(0.027)	(0.040)
Lead Product Sales$_t$ × Adverse Event Rate$_t$ (excl. focal)	0.015*	0.047*	0.013
	(0.008)	(0.023)	(0.030)
Patent Applications$_t$	0.088***	0.031	-0.033
	(0.013)	(0.021)	(0.029)
Scientific Publications$_t$	0.008***	0.000	-0.003
	(0.002)	(0.004)	(0.009)
Deals$_t$	-0.009	-0.019	-0.045
	(0.012)	(0.026)	(0.036)
Total Other Sales$_t$	0.010	-0.046	-0.029
	(0.017)	(0.042)	(0.061)
R&D Expenditure$_t$	-0.005**	-0.001	-0.006
	(0.002)	(0.004)	(0.011)
Total Products$_t$	-0.152	-0.194	1.047
	(0.474)	(0.862)	(0.584)
NCE expiration$_t$ (<5yr)	0.364***	0.452	0.208
	(0.091)	(0.286)	(0.311)
ATC1 × Year FE	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes
Observations	3,302	2,383	1,933
Pseudo R-squared	0.478	0.225	0.184

Notes: This table estimates a full specification that stacks the main demand-side and supply-side mechanisms in one model. In this kitchen-sink specification, the coefficient on lead product sales becomes small for early-stage project counts but remains large and positive for target-overlap and molecularly similar projects, perhaps suggesting that the technical-similarity channel is the most robust part of the blockbuster effect. Estimates are from Poisson pseudo-maximum-likelihood models. Standard errors are clustered at the firm level and reported in parentheses.

Table G.14: Understanding the Blockbuster Effect Across Firms

	Early-Stage Projects$_{t+5}$						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Lead Product Sales$_t$	0.327	0.222	0.288	0.243	0.245	0.300	0.196
	(0.040)	(0.038)	(0.123)	(0.038)	(0.038)	(0.099)	(0.133)
ATC1 Detailing$_{t-3}$		6.750	6.853				4.805
		(2.604)	(2.588)				(2.294)
Lead Product Sales$_t$ × ATC1 Detailing$_{t-3}$			-0.504				0.557
			(0.757)				(0.726)
Adverse Event Rate$_t$ (excl. focal)					0.016	0.011	0.013
					(0.009)	(0.011)	(0.010)
Lead Product Sales$_t$ × Adverse Event Rate$_t$ (excl. focal)						0.011	0.020
						(0.015)	(0.015)
ATC1 × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	22,950	17,542	17,542	4,434	4,434	4,434	4,367
Pseudo R-squared	0.050	0.076	0.077	0.075	0.076	0.077	0.097

Notes: This table removes firm fixed effects, so the coefficients are identified from across-firm as well as within-market-year variation. The estimates on lead product sales remain positive, indicating that the blockbuster effect is also visible in cross-sectional comparisons across firms, although these are not the preferred within-firm specifications. Here we do not find support for the demand-channel stickiness mechanisms. Estimates are from Poisson pseudo-maximum-likelihood models. Standard errors are clustered at the firm level and reported in parentheses. All specifications include the fixed effects listed in the table.

Table G.15: Separating safety from efficacy

	Early-Stage Projects$_{t+5}$	
	(1)	(2)
		ATC1 = Neuro
Lead Product Sales$_t$	0.161	0.167
	(0.056)	(0.077)
ATC1 Detailing$_{t-3}$	0.403	0.521
	(0.910)	(2.380)
Lead Product Sales$_t$ × ATC1 Detailing$_{t-3}$	-0.267	0.805
	(0.314)	(0.718)
Adverse Event Rate$_t$ (excl. focal)	0.002	-0.042
	(0.010)	(0.021)
Lead Product Sales$_t$ × Adverse Event Rate$_t$ (excl. focal)	0.010	0.063
	(0.006)	(0.010)
ATC1 × Year FE	Yes	Yes
Firm FE	Yes	Yes
ATC3 FE	Yes	Yes
Observations	3,110	524
Pseudo R-squared	0.617	0.484

Notes: This table removes firm fixed effects, so the coefficients are identified from across-firm as well as within-market-year variation. The estimates on lead product sales remain positive, indicating that the blockbuster effect is also visible in cross-sectional comparisons across firms, although these are not the preferred within-firm specifications. Here we do not find support for the demand-channel stickiness mechanisms. Estimates are from Poisson pseudo-maximum-likelihood models. Standard errors are clustered at the firm level and reported in parentheses. All specifications include the fixed effects listed in the table.

Table G.16: Baseline Results in Chronic and Acute Markets

	Early-Stage Projects$_{t+5}$		Target Overlap Projects$_{t+5}$		Molecularly Similar Projects$_{t+5}$		Same Molecule Projects$_{t+5}$	
	(1) Chronic	(2) Acute	(3) Chronic	(4) Acute	(5) Chronic	(6) Acute	(7) Chronic	(8) Acute
Lead Product Sales$_t$	0.158	0.065	0.230	0.268	0.524	1.385	0.293	2.119
	(0.034)	(0.086)	(0.063)	(0.190)	(0.079)	(0.284)	(0.122)	(0.485)
ATC1 × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9,734	10,846	6,256	5,737	5,894	3,660	4,720	2,186
Pseudo R-squared	0.446	0.441	0.212	0.314	0.217	0.169	0.160	0.195

Notes: This table splits the sample between chronic and acute markets to compare how the blockbuster effect varies across treatment settings. Standard errors are clustered at the firm level and reported in parentheses.

Table G.17: Split Sample Regressions

	Early-Stage Projects$_{t+5}$			Target Overlap Projects$_{t+5}$			Molecularly Similar Projects$_{t+5}$		
	(1) All	(2) High AE Rate	(3) Low AE Rate	(4) All	(5) High AE Rate	(6) Low AE Rate	(7) All	(8) High AE Rate	(9) Low AE Rate
Lead Product Sales$_t$	0.127	0.247	0.102	0.201	0.629	0.113	0.417	0.587	0.376
	(0.032)	(0.142)	(0.029)	(0.064)	(0.161)	(0.066)	(0.076)	(0.148)	(0.073)
Sample	All nonmissing	Above cutoff	Below cutoff	All nonmissing	Above cutoff	Below cutoff	All nonmissing	Above cutoff	Below cutoff
ATC1 × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,636	1,784	1,695	2,584	1,092	1,215	2,224	672	995
Pseudo R-squared	0.392	0.390	0.415	0.201	0.250	0.244	0.235	0.259	0.231

Notes: This table splits the sample by the median of the adverse event rate measure, and reports the coefficient estimate on lead product sales. All models are estimated with Poisson Pseudo Maximum Likelihood. Standard errors are clustered at the firm level and reported in parentheses.

H Examples of Blockbuster Drugs and Subsequent Innovation

Pfizer & Lipitor. An instructive example of a firm's efforts to target the same customer segment with similar technology as one of their blockbusters is Pfizer's investment of nearly $1 billion in the development of torcetrapib, a cholesterol drug intended to improve upon its blockbuster statin Lipitor. Although the drug caused a higher fatality rate and ultimately failed in later stage clinical trials, the case illustrates the scale of investment firms are willing to undertake, even when the success of follow-on therapies is highly uncertain.[43]

Abbvie & Humira. Another example concerns AbbVie (and previously Abbott) and the drug Humira. Humira is a TNF (tumor necrosis factor) blocker used for a host of autoimmune conditions, including psoriasis, arthritis, and Crohn's disease. At its peak, the drug exceeded $20 billion in annual revenue, becoming the highest-grossing drug in history to date. With patent protection due to expire in 2023, Abbvie began searching for replacements early. Clinical trial data shows that Abbvie internally worked on the development of at least seven distinct immunology drugs, three of which were also TNF-inhibitors. The successful product from this pipeline was upadacitinib (Rinvoq), a JAK-1 inhibitor. In addition, in 2016 AbbVie purchased risankizumab from Germany-based Boehringer Ingelheim for $595 million, which later became Skyrizi.[44]

Novo Nordisk & GLP-1. The phenomenon remains as relevant as ever with the rapid growth of GLP-1 drugs for obesity. The Danish firm Novo Nordisk pioneered this class of drugs with treatments originally developed for Type II diabetes, Ozempic and Wegovy, which later achieved blockbuster status as weight-loss drugs. Faced with the added pressure of increasing competition—most notably from Eli Lilly—Novo Nordisk is doubling down on its efforts to produce the next generation of weight-loss treatments.[45] These efforts include oral formulations, drugs with fewer side effects, and therapies requiring less frequent dosing.[46]

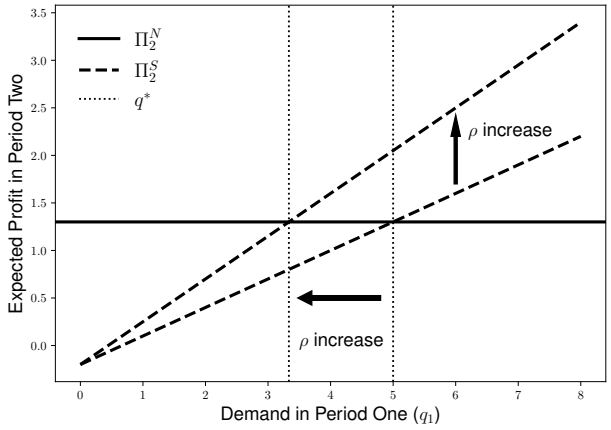
[43] https://www.nytimes.com/2006/12/04/health/04pfizer.html?pagewanted=print

[44] https://www.biospace.com/business/just-how-steep-is-the-drug-patent-cliff-ask-these-pharmas

[45] see https://www.wsj.com/health/pharma/novo-nordisk-united-laboratories-say-new-obesity-drug-helps-patients-lose-weight-in-trial-99d136d5?mod=Searchresults&pos=4&page=1

[46] https://www.wsj.com/health/pharma/the-day-pharmas-weight-loss-gold-rush-intensified-c5c32c11?mod=article_inline

Figure 4: 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.