Developing Novel Drugs[*]

Joshua Krieger[†]   Danielle Li[‡]   Dimitris Papanikolaou[§]

November 21, 2017

Abstract

We analyze firms' decisions to invest in incremental and radical innovation, focusing specifically on pharmaceutical research. We develop a new measure of drug novelty that is based on the chemical similarity between new drug candidates and existing drugs. We show that drug candidates that we identify as ex-ante novel are riskier investments, in the sense that they are subsequently less likely to be approved by the FDA. However, conditional on approval, novel candidates are, on average, more valuable—they are more clinically effective; have higher patent citations; lead to more revenue and to higher stock market value. Using variation in the expansion of Medicare prescription drug coverage, we show that firms respond to a plausibly exogenous cash flow shock by developing more molecularly novel drug compounds. This pattern suggests that, at the margin, novel drugs are perceived to be as more valuable ex-ante investments than me-too drugs.

[†]Harvard Business School, jkrieger@hbs.edu
[‡]MIT Sloan and NBER, d_li@mit.edu
[§]Kellogg School of Management and NBER, d-papanikolaou@kellogg.northwestern.edu

Over the past 40 years, the greatest gains in life expectancy in developed countries have come from the development of new therapies to treat conditions such as heart disease, cancer, and vascular disease.[1] At the same time, the development of new–and often incremental–drug therapies has played a large role in driving up health care costs, with critics frequently questioning the true innovativeness of expensive new treatments (Naci, Carter, and Mossialos, 2015). This paper contributes to our understanding of drug investment decisions by developing a measure of drug novelty and subsequently exploring the economic tradeoffs involved in the decision to develop novel drugs.

Measuring the amount of innovation in the pharmaceutical industry is challenging. Indeed, critics argue that "pharmaceutical research and development turns out mostly minor variations on existing drugs, and most new drugs are not superior on clinical measures," making it difficult to use simple drug counts as a measure of innovation (Light and Lexchin, 2012).[2] To overcome this challenge, we construct a new measure of drug novelty for small molecule drugs, which is based on the molecular similarity of the drug with prior drug candidates.[3] Thus, our first contribution is to develop a new measure of pharmaceutical innovation.

Our definition of a novel drug candidate is one that is molecularly distinct from previously tested candidates. Specifically, we build upon research in modern pharmaceutical chemistry to compute a pair-wise chemical distance (similarity) between a given drug candidate and any prior candidates in our data. This similarity metric is known as a "Tanimoto score" or "Jaccard coefficient," and captures the extent to which two molecules share common chemical substructures. We aggregate these pairwise distance scores to identify the maximum similarity of a new drug candidate to all prior candidates. Drugs that are sufficiently different to their closest counterparts are novel according to our measure. Since our metric is based on molecular properties observed at the time of a drug candidate's initial development, it improves upon existing novelty measures by not conflating ex-ante measures of novelty with ex-post measures of success such as receiving priority FDA review.

---

[1]In the United States, the sharpest decline in death rates from the period 1981 to 2001 come from the reduction in the incidence of heart disease. See Life Tables for the United States Social Security Area 1900-2100. https://www.ssa.gov/oact/NOTES/as120/LifeTables_Body.html See also Lichtenberg (2013), which estimates explicit mortality improvements associated with pharmaceuticals.

[2]One of the more vocal critics is Marcia Angell, a former editor of the *New England Journal of Medicine*. She argues that pharmaceutical firms increasingly concentrate their research on variations of top-selling drugs already on the market, called 'me-too' drugs. She concludes: "There is very little innovative research in the modern pharmaceutical industry, despite its claims to the contrary." http://bostonreview.net/angell-big-pharma-bad-medicine. Indeed, empirical evidence appears to be consistent with this view; Naci et al. (2015) survey a variety of studies that show a declining clinical benefit of new drugs.

[3]Small molecule drugs, synthesized using chemical methods, constitute over 80% of modern drug candidates (Ralf Otto, Alberto Santagostino, and Ulf Schrader, 2014). We will discuss larger drugs based on biological products in Section 3.6.

Our novelty measure based on molecular similarity has sensible properties. Pairs of drug candidates classified as more similar are more likely to perform the same function—that is, they share the same indication (disease) or target-action (mechanism). Further, drugs we classify as more novel are more likely to be the first therapy of its kind. In terms of secular trends, our novelty measure indicates a decline in the innovativeness of small molecule drugs: both the number, as well as the proportion, of novel drug candidates has declined over the 1999 to 2014 period.

We next examine the tradeoffs involved with developing innovative drugs. We begin by exploring how the novelty of a drug candidate relates to its (private and social) return from an investment standpoint. Since measuring value added is challenging, we rely on several metrics. First, we use the clinical value added assigned from the French health system, following Kyle and Williams (2017). Since this measure is only available for a subset of approved drugs, we also examine the number of citations to patents associated with the drugs, which the literature on innovation has long argued is related to estimates of economic value added (see, e.g. Hall, Jaffe, and Trajtenberg, 2005). We also use drug revenues as a more direct estimate of economic value. However, since mark-ups may vary systematically between novel and 'me-too' drugs—that is, drugs that are similar to existing drugs—we also rely on estimates of their contribution to firm stock market values. Specifically, we follow Kogan, Papanikolaou, Seru, and Stoffman (2017) and examine the change in the firm's market valuation following either the drug's FDA approval or a successful patent application associated with the drug.

Conditional on being approved by the FDA, novel drugs are on average more valuable. Specifically, relative to drugs in their same cohort that treat the same disease (indication), a one-standard deviation increase in our measure of novelty is associated with a 33 percent increase in the likelihood that a drug is classified as 'highly important' by the French health system; a 10 to 33 percent increase in the number of citations for associated patents; a 15 to 35 percent increase in drug revenues; and a 2 to 8 percent increase in firm valuations.

However, novel drugs are also riskier investments, in that they are less likely to receive regulatory approval. Relative to comparable drugs, a one-standard deviation increase in novelty is associated with a 29 percent decrease in the likelihood that it is approved by the FDA. Thus, novel drugs are less likely to be approved by the FDA, but conditional on approval, they are on average more valuable. Given these estimates, it is challenging to assess directly whether investments in novel drug candidates have higher *ex-ante* social or private value. However, an indirect way of determining whether firms perceive novel drugs to be ex-ante preferable to me-too drugs is to examine how the novelty of a firm's research pipeline

responds to a shock to their financial ability to develop additional drug candidates. Indeed, if firms develop more novel (as opposed to 'me-too') drugs following a positive shock to the firm's cashflows, this suggests that, at the margin, firms perceive novel drugs to be more valuable.

We construct shocks to firm profitability using the introduction of Medicare Part D, which expanded US prescription drug coverage for the elderly. This policy change increased profits of those firms with more drugs that target conditions common among the elderly (Lichtenberg and Sun, 2007; Friedman, 2009). A challenge with this strategy, however, is that the expansion of Medicare affected not only the profitability of the firm's existing assets (i.e. profits coming from drugs already developed) but also the firm's investment opportunities (i.e., the future value of developing new drugs for the elderly). To isolate the causal impact of the increase in cashflow from the increase in investment opportunities, we exploit a second source of variation: affected drugs with longer exclusivity periods (patent life plus additional exclusivity granted by the FDA) are likely to generate more profits than drugs whose patents will soon expire. Since firms with longer exclusivity periods may have different development strategies than firms with patents with shorter exclusivity periods, we also control for the remaining exclusivity periods across their entire drug portfolio.

In essence, our estimation methodology compares firms with the same share of drugs sold to the elderly and the same remaining exclusivity periods across all their overall drug portfolio, but that differ only in how their remaining patent exclusivity is distributed across drugs of varying elder shares. This strategy allows us to identify differences in the intensity of increases in cashflow among firms with similar improvements in investment opportunities. We find that treated firms develop a larger number of new drug candidates. Importantly, this effect is driven by an increase in the number of chemically novel candidates, as opposed to me-too candidates. Further, these new candidates are aimed at a variety of conditions, not simply ones with a high share of elderly patients, implying that our identification strategy is at least partially successful in isolating a shock to cashflows from a shock to investment opportunities.

In addition, we find some evidence that firm managers have a preference for diversification. The marginal drug candidates that treated firms pursue often include drugs that focus on different diseases, or operate using a different mechanism (target) relative to the drugs that the firm has previously developed. These findings suggest that firms use marginal increases in cash to diversify their portfolios and undertake more exploratory development strategies, a fact consistent with models of investment with financial frictions (Froot, Scharfstein, and

3

Stein, 1993), or poorly diversified managers (Smith and Stulz, 1985). Last, our point estimates imply sensible returns to R&D. A one standard deviation increase in Part D exposure leads to an 11 percent increase in subsequent drug development, relative to less exposed firms. For the subset of firms for which we are able to identify cash flow, this translates into an elasticity of the number of drug candidate with respect of R&D expenditure of about 0.75. We obtain a higher elasticity for the most novel drugs (1.01 to 1.59) and a lower elasticity for the most similar drugs (0.02 to 0.31). For comparison, estimates of the elasticity of output with respect to demand (or cashflow) shocks in the innovation literature range from 0.3 to 4 (Henderson and Cockburn, 1996; Acemoglu and Linn, 2004; Azoulay, Graff-Zivin, Li, and Sampat, 2016; Blume-Kohout and Sood, 2013; Dranove, Garthwaite, and Hermosilla, 2014).

We conclude that financial frictions are likely to play a role in limiting the development of novel drug candidates. The ability to observe the returns associated with individual projects is an important advantage of our setting that allows us to make a distinct contribution to the literature studying the impact of financial frictions on firm investment decisions.[4] Specifically, existing studies typically observe the response of investment (or hiring) aggregated at the level of individual firms or geographic locations. By contrast, our setting allows us to observe the risk and return of the marginal project being undertaken as a result of relaxing financial constraints. We find that, relaxing financing constraints leads to more innovation, both at the extensive margin (i.e., more drug candidates) but also at the intensive margin (i.e., more novel drugs). Given that novel drugs are less likely to be approved by the FDA, the findings in our paper echo those in Metrick and Nicholson (2009), who document that firms that score higher in terms of a Kaplan-Zingles index of financial constraints are more likely to develop drugs that pass FDA approval.

By providing a new measure of novelty, our work contributes to the literature focusing on the measurement and determinants of innovation. Our novelty measure is based on the notion of chemical similarity (Johnson and Maggiora, 1990), which is widely used in the process of pharmaceutical discovery. Chemists use molecular similarity calculations to help them search chemical space, build libraries for drug screening (Wawer, Li, Gustafsdottir, Ljosa, Bodycombe, Marton, Sokolnicki, Bray, Kemp, Winchester, Taylor, Grant, Hon, Duvall, Wilson, Bittker,

---

[4]Several studies have documented a causal impact of financing frictions on firm investment decisions. Recent work on this topic examines the response of physical investment (for instance, Lin and Paravisini, 2013; Almeida, Campello, Laranjeira, and Weisbenner, 2011; Frydman, Hilt, and Zhou, 2015); employment decisions (Benmelech, Bergman, and Seru, 2011; Chodorow-Reich, 2014; Duygan-Bump, Levkov, and Montoriol-Garriga, 2015; Benmelech, Frydman, and Papanikolaou, 2017); and investments in R&D (see e.g. Bond, Harhoff, and van Reenen, 2005; Brown, Fazzari, and Petersen, 2009; Hall and Lerner, 2010; Nanda and Nicholas, 2014; Kerr and Nanda, 2015). These frictions may be particularly severe in the case of R&D: Howell (2017) shows that even relatively modest subsidies to R&D can have a dramatic impact on ex-post outcomes.

Dank, Narayan, Subramanian, Winckler, Golub, Carpenter, Shamji, Schreiber, and Clemons, 2014), quantify the "drug-like" properties of a compound (Bickerton, Paolini, Besnard, Muresan, and Hopkins, 2012), and expand medicinal chemistry techniques (Maggiora, Vogt, Stumpfe, and Bajorath, 2014). In parallel work, Pye, Bertin, Lokey, Gerwick, and Linington (2017) use chemical similarity measures to measure novelty and productivity in the discovery of natural products.

Our measure of innovation is based on ex-ante information—the similarity of a drug's molecular structure to prior drugs—and therefore avoids some of the truncation issues associated with patent citations (Hall et al., 2005). Further, since our measure is based only on ex-ante data, it does not conflate the ex-ante novelty of an idea with measures of ex-post success or of market size. By contrast, existing work typically measures 'major' innovations using metrics based on ex-post successful outcomes, which may also be related to market size. Examples include whether a drug candidate gets FDA Priority Review status (Dranove et al., 2014), or whether a drug has highly-cited patents (Henderson and Cockburn, 1996). A potential concern with these types of measures is that a firm will be credited with pursing novel drug candidates only if these candidates succeed and not when—as is true in the vast majority of cases—they fail. Similarly, outcomes such as whether a drug is first in class or is an FDA orphan drug (Dranove et al., 2014; DiMasi and Faden, 2011; Lanthier, Miller, Nardinelli, and Woodcock, 2013; DiMasi and Paquette, 2004) may conflate market size with novelty and may fail to measure novelty of candidates within a particular class. For example, it is easier to be the first candidate to treat a rare condition than a common condition because fewer firms have incentives to develop treatments for the former. Further, measuring novelty as first in class will label all subsequent treatments in an area as incremental, even if they are indeed novel.

Further, our paper relates to work that examines how regulatory policies and market conditions distort the direction of drug development efforts (Budish, Roin, and Williams, 2015); and how changes in market demand affect innovation in the pharmaceutical sector (Acemoglu and Linn, 2004; Blume-Kohout and Sood, 2013; Dranove et al., 2014). Similar to us, Blume-Kohout and Sood (2013) and Dranove et al. (2014) exploit the passage of Medicare Part D, and find more innovation in markets that receive a greater demand shock (drugs targeted to the elderly). Even though we use the same policy shock, our work additionally exploits differences in market exclusivity (patent length) for specific drugs to identify the effect of cash flow shocks separately from changes in product demand that may increase firm investment opportunities. Indeed, we find that treated firms invest in new drugs across

different categories—as opposed to those that only target the elderly—strongly suggesting that our identification strategy effectively isolates cashflow shocks from improvements in investment opportunities.

Last, our measure of novelty can help shed light on several debates in the innovation literature. For instance, Jones (2010); Bloom, Jones, Reenen, and Webb (2017) argue for the presence of decreasing returns to innovation. Consistent with this view, we find that drug novelty has decreased over time. An important caveat is that our measure cannot be computed for biologics, which represent a vibrant research area.

1   Measuring Drug Novelty

A key contribution of our paper is to construct an ex-ante measure of drug novelty that is broadly applicable. To do so, we rely on the "Similarity Property Principle," a central tenant of modern pharmaceutical chemistry, which states that structurally similar molecules are more likely to have similar functional properties (Johnson and Maggiora, 1990). Our measure of drug novelty uses this notion of chemical adjacency, but applies it historically: for a given drug candidate, we define its novelty based on how similar it is to all previously tested drug candidates.

1.1   Data Overview

To conduct our analysis, we construct a panel dataset that tracks firm–quarter level drug development outcomes. We combine data from a number of sources.

The primary data used to construct drug output and novelty measures come from Thomson Reuters Cortellis's Investigational Drugs database, which contains detailed development histories for over 64,067 drug candidates (as of May 2016). Cortellis assembles the data on drug candidates from public records (e.g., patent filings, company press releases, financial filings, clinical trial registries, FDA submissions) and then further cleans the data to assign the proper classifications (e.g., therapeutic indications and drug targets).[5] Hence, the earliest point of entry for a given drug candidate would be the first time a patent is filed or it appears as a developmental drug in documents describing a firm's research pipeline or on its website.

---

[5]In our sample, we see the number of reported molecules increase sharply in the late 1990s; this increase is likely due to an improvement in the reporting of molecules. This is likely due to the Food and Drug Administration Modernization Act, passed in late 1997 and enacted in 1999, which required the reporting of clinical trials. Even though we observe some drug candidates pre-1999, we believe that our data provides fuller coverage post 1999.

Though we have no way of establishing the extent to which the data is representative of all existing drug candidates, we might expect that there is some survivorship bias in the data, since Cortellis sometimes backfills information on drug development. That said, we believe this survivorship bias is more likely to occur in the very early stages of drug discovery. Compounds which show no promise in the earliest screening experiments may never be patented and leave no paper trail for Cortellis to pick up. However, once a company files the Investigational New Drug (IND) Application for the drug and starts clinical trials, then Cortellis will almost always have that drug in the data. Since our measure of novelty relies on comparing a given drug candidate to drugs in Phase 1 clinical trials, we believe this bias is unlikely to significantly affect our measure.

We supplement the data using a variety of sources. We use ChemMine Tools to compute similarity scores. We obtain accounting information for a subset of the companies (those that we can match based on their name) from Compustat. We link approved drugs to their key patents and exclusivity dates using the FDA Orange Book and information from the Federal Register. We obtain patent value information from Kogan et al. (2017). Last, we use the Medical Expenditure Panel Survey (MEPS) to estimate drug revenue and Medicare market share (MMS).

1.2 Similarity Based on Chemical Structure

We create a measure of novelty at the drug candidate level based on its chemical structure. The first step in constructing our novelty indicators is measuring the similarity of two molecules. To identify an appropriate approach, we follow the chemical informatics literature, where scientists have developed tools for measuring the differences between chemical structures.

The most common method for creating these similarity scores is calculating the Tanimoto distance (Jaccard coefficient) between two sets of chemical fragments (Nikolova and Jaworska, 2003). The calculation returns the proportion of features shared by the two chemicals when divided by their union, yielding a score between 0 and 1. It is defined as follows:

$$T_{A,B} \equiv \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}. \tag{1}$$

A Tanimoto distance of 0 implies that the pair of drugs have no common fragments; a score of 1 means they have the same set of atoms and bonding. However, a Tanimoto score of 1 does not necessarily mean that the two chemicals are identical because the Tanimoto score does not take into account a structure's stereosymmetry (its orientation in space). For

example, consider a classic example of a me-too drug, Nexium, and its antecedent, Prilosec. Prilosec is a "racemic mixture," meaning that it is a mixture of two versions of the same molecule, differently oriented, whereas Nexium is comprised of only one orientation of this same molecule—despite their differing orientation, we record the pair as having a score of 1.

In computing the distance metric (1), we rely on ChemMine Tools, an open source program for chemical-informatics.[6] After calculating the Tanimoto distance $T_{i,j}$ across all pairs of drug candidates i and j, we define a drug's overall similarity as

$$\text{maxsim}_i \equiv \max_{j \in P_i} T_{i,j}, \tag{2}$$

where P_i is the set of drug candidates that have reached Phase 1 clinical trials prior to the introduction of candidate i.[7] A novel drug is one with low values of maxsim in (2), since that it means it is sufficiently different to the current set of drug candidates.

Figure 1 illustrates an example of how our novelty measure works, as applied to the drug class of HMG-CoA reductase inhibitors—more commonly known as "statins," used to treat heart disease. In September of 1987, Mevacor (Lovostatin) became the first statin to be approved by the FDA; it's similarity score to prior candidates is 0.25. In October of 1991, a second statin, Pravochol (Pravastatin), was approved. Pravochol's similarity to priority candidates is 0.61, and Mevacor was its closest prior candidate. Next, in December of 1991, a third statin, Zocor, was approved. As one can see from Figure 1, Zocor (Simvastatin) is quite similar to Mevacor and, indeed, its maximum similarity score is 0.82 (0.52 similarity to Pravochol and 0.82 similarity to Mevacor).

1.3   Caveats and Limitations

There are several important caveats to keep in mind regarding our proposed novelty measure.

First, and most importantly, there is no perfect correspondence between structural and functional similarity. Similar molecules may have divergent properties: the drug thalidomide,

---

[6]Appendix B provides more detail about the construction of similarity scores using the simplified molecular-input line-entry system (SMILES) and ChemMine Tools.

[7]We restrict the comparison set in this way to avoid mistakenly labeling a drug candidate as derivative if it was developed at approximately the same time as another candidate. See (DiMasi and Faden, 2011) for evidence on how simultaneously developed drugs might be mislabeled as copycats. When studying firm's diversification decisions in Section 3.3, we use a similar methodology to define a drug's maximum similarity to prior candidates developed by the same firm. Using a more inclusive set of previous drugs—all small molecule drugs that entered any stage of development prior to the focal drug's entry date—yields qualitatively similar results as our main analyses.

for instance, is comprised of two mirror image molecules, one of which is a safe sedative, the other of which causes birth defects. Conversely, chemically dissimilar compounds may have similar biological effects: Crestor and Lipitor have different structural profiles, but are often prescribed interchangeably by doctors. Despite these examples, however, chemical informatics research has shown that Tanimoto similarity measures are nonetheless useful for identifying drug qualities and novelty on average (OHagan, Swainston, Handl, and Kell, 2015; Baldi and Nasr, 2010; Bickerton et al., 2012; Pye et al., 2017). In Section 1.4, we discuss the results of several validation exercises for our measure.

Second, we are only able to compare molecules against prior molecules in the Cortellis data. As such, our measure of maximum similarity is a lower bound for true similarity because we may be missing earlier drugs with similar properties. This is especially true for drugs with similarity scores near 0, which are disproportionately candidates for which we had few previous candidates to compare to. To address these concerns, all of our drug and firm-level analysis will be done with fixed effects for the quarter of a candidate's earliest development date.

Lastly, our approach does not allow us to study the similarity of more complicated drug therapies whose chemical structure is more difficult to characterize. While most drugs are chemically synthesized with known structures, a growing class of new therapies, known as biologics, are based on biological products (proteins, cells, tissues, etc.) that cannot be compared with Tanimoto scores. Although biologics make up for only 20% of drug development, their share is increasing and are often considered to be a source of innovation in the drug industry (Ralf Otto et al., 2014). In Section 3.6 we show that increased cashflows also lead to greater development of biologics.

1.4 Validation

Before describing the characteristics of drugs we identify as novel, we first independently verify that our measure of chemical similarity captures a sense of functional similarity. Conducting this type of validation is difficult because there are no gold standard measures of functional similarity. However, we perform several intuitive checks. First, we would expect drugs that act on the same biological target (e.g., a protein drugs attempt bind to or otherwise impact) to be more similar than drugs that do not. Similarly, though perhaps less crucially, we would expect drugs sharing the same indication (i.e., developed to treat the same medical condition) to be more similar than those with different indications. Appendix Table A.3 reports regressions of pairwise similarity scores on indicator variables for a given pair of drugs

sharing the same biological target or the same disease indication.[8] The coefficients indicate that sharing the same target-action more than doubles pairwise molecular similarity while sharing the same indication increases this by over 25%.

Drugs that are labeled more novel should be also more likely to be the first in their target class. The binned scatterplot presented in Figure 2 shows that this is the case: there is a strong negative relationship between a drug's chemical similarity score and its likelihood of being the first drug candidate for a given target. Comparing two drugs treating the same indication that enter development in the same quarter, we find that a one standard deviation increase in novelty (-0.21) increases a drug's chances of being the first in its broad target class by over 40%. Appendix Table A.4 shows that these results are robust to other specifications and controls.

2 Novelty Descriptives

In this section, we explore relationship between novelty and project risk and return. We use the likelihood of FDA approval as a measure of its riskiness. We measure the drug candidate's value added using several proxies, including, measures of clinical effectiveness, patent citations, and its contribution to revenue and stock market value.

2.1 Descriptive Facts

We begin by describing the characteristics of our novelty measure. Panel A of Figure 3, and Table 1, show the distribution of our similarity metric, maximum Tanimoto distance from prior candidates, denoted as $maxsim_i$. Recall that $maxsim_i$ measures the drug's maximum similarity, so lower values imply higher novelty. We see that the distribution of our ex-ante novelty score is somewhat bi-modal; the vast majority of drugs have maximum similarity scores in excess of 0.2, and most fall in the 0.3 to 0.6 range. However, there is a second peak close to 1 (zero novelty). Approximately 10% of our sample candidates share the same structure as a prior candidate that has also entered development. These include molecules that are stereoisomers, meaning that they differ only in orientation, as well as combination therapies that involve multiple compounds that were previously developed as separate therapies.

Panel B of Figure 3 shows that the novelty of drug candidates has declined over time. This can be see be seen by both the increase in the average similarity of new drug candidates

---

[8]There are almost 1 billion such pairs because each drug may be act on more than one target or be used for more than one indication. We construct pairs for every possible combination.

with existing drugs (Panel B) as well as by the increase in the fraction of new drug candidates that are very similar to prior candidates in Panel C, those with maximum Tanimoto scores of over 0.9. Going forward, we refer to such candidates as can be 'me-too' drugs because they represent only a small modification from existing drugs. This secular decline in drug novelty is consistent with the view that the average level of innovativeness in the pharmaceutical sector has declined over time (Light and Lexchin, 2012; Naci et al., 2015) and is also consistent with the presence of decreasing returns to scale in innovative activity (Jones, 2010; Bloom et al., 2017).

Table 1 provides further descriptive statistics related to all candidate compounds in the sample, as well as for the subset of compounds that are associated with those firms for which we can calculate a measure of exposure to Medicare Part D. We see that the vast majority of drug candidates are not approved by the FDA: 356 out of the 6,374 drug candidates developed by our sample firms in the 1999–2014 period were approved by the FDA by 2015. This attrition occurs at several stages in the drug approval process, including Phase 1 through 3. Naturally, since the approval process is lengthy (the median drug spends 5 years in development) this comparison understates the unconditional probability of success. Focusing on drugs developed prior to 2002 (a reasonable cutoff, since 95% of successful drugs have had development cycles of 13 years or less) we see that the baseline probability that a drug at any stage of development obtains FDA approval is 18%. By contrast, the likelihood of FDA approval conditional on reaching Phase 3 is about 50%.

2.2 Drug Novelty and Likelihood of FDA approval

So far, our findings suggest that drug development decisions are risky, since the unconditional likelihood of successful development (FDA approval) is quite low. We next examine how this success rate varies with the novelty of the underlying candidate. To do so, we estimate linear probability models that relate whether the drug was approved by the FDA ($S_i = 1$) to the drug candidate's ex-ante novelty maxsim_i,

$$S_i = a + b\,\mathrm{maxsim}_i + c\,Z_i + \varepsilon_i. \tag{3}$$

We saturate our specification with a battery of controls, including quarter of development, and disease (ICD-9 indication) fixed effects. We cluster the standard errors by indication.

We estimate equation (3) for all drug candidates, but also report results separately conditioning on different stages in development. We find that chemically novel drugs are

significantly less likely to be approved by the FDA, as we can see in column (1) of Table 2 and Panel A of Figure 4. Compared to drugs of similar age and targeting the same disease (ICD-9 indication), a one standard deviation increase in drug novelty is associated with a 5.2 percentage point decrease in the likelihood of FDA approval. Given that the unconditional likelihood of FDA approval is 18%, this estimate represents a 29% decrease in the likelihood of developing a successful drug candidate.

Further, this negative relationship between novelty and approval persists throughout the development pipeline, though the magnitude of the association attenuates. Figure A.2 and Table A.5 in the Online Appendix report the full set of results. Conditional on reaching Phase 1 or Phase 2, a one standard deviation increase in novelty is associated with an approximately 25% decrease in the likelihood of ultimate approval; conditional on reaching Phase 3, a one standard deviation increase in novelty decreases the likelihood of approval by 10%.

2.3 Drug Novelty and Measures of Value

Even if more novel drugs are riskier investments on the part of the firm, in the sense that they are less likely to be approved by the FDA, it is possible that they are more valuable conditional on approval. Given that measuring private or social value of an innovation is challenging, we base our analysis on several measures of value added. Specifically, we examine measures of social value, measured either by the drug's clinical effectiveness, or by the number of citations the patent receives. Since firm investment decisions are also likely to be guided by considerations about private values, we also examine the relation between chemical novelty and revenue generated by successful drug candidates and their contribution to firm value.

In general, we relate the different measures of value added X_i to our novelty measure using the following specification,

$$X_i = a + b \max\text{sim}_i + c\, Z_i + \varepsilon_i. \tag{4}$$

Depending on the specification and the value added metric, X_i takes either binary values (to identify whether the drug is deemed clinically important), or consists of the logarithm of revenues, citation counts, or estimated contributions to firm value. To ensure that we are comparing otherwise similar drugs, we continue to saturate our specifications with a battery of fixed effects and controls. Whenever possible, we control for a drug's age (development quarter or year) and disease (ICD9 indication). Since most of these metrics are only available for drugs that are approved by FDA, we restrict attention to approved drugs.

12

Drug Novelty and Effectiveness

We begin by examining how (ex-ante) drug novelty correlates with measures of drug quality or effectiveness. To do so, we use the data from the French Haute Autorité de Santé (HAS) health system, which assigns scores based on their clinical added benefits. These value-added (Amélioration du Service Medical Rendu, or ASMR) scores range from one to five (I to V), with V indicating no value added, while I indicates the highest improvement relative to existing drugs. See the Appendix for additional details. Following Kyle and Williams (2017) we can separate the ASMR values into low importance (IV/V) and high importance (I/II/III).We merge our drug-level data using both established drug naming conventions and manual matching.[9] The ASMR scores are assigned only to approved drugs that are available for reimbursement from the French Government health system. After limiting our attention to the first approved indication for drugs covered in both data sets, and for which we can compute novelty scores, we are left with 385 drugs.[10]

We next estimate equation (4), where now the definition of the dependent variable is either the raw ASMR score, or a binary variable that takes the value of one if the drug has been deemed of adding sufficient clinical value (ASMR scores below IV or V). Column (2) of Table 2 reports results using our baseline specification, which examines whether a drug is assigned a score less than V (denoting it has some clinical benefit) and controls for the age of the drug, as measured by the launch year, and indication fixed effects. Panel B of Figure 4 provides a scatter plot. Appendix Table A.6 reports results using additional specifications.

Novel drugs are more likely to be clinically effective. Specifically, comparing drugs of the same age that treat the same indication, a one-standard deviation increase in novelty is associated with a 7 percentage point increase in the likelihood that a drug is classified as adding any value (ASMR<V), and a 3 percentage point increase in the likelihood that it is classified as high importance (ASMR < IV). These magnitudes are substantial, given that the baseline probability that a drug falls into these categories is 20% and 9%, respectively. These results suggest that drugs that are more novel are also more likely to have greater clinical value added, and are therefore likely to be more socially valuable.

---

[9]Specifically, we first merge the Cortellis drugs to HAS drug identifiers (CIP7 codes) using the Anatomical Therapeutic Chemical (ATC) drug codes associated with the CIP7 codes in the French HAS. Next we use the HAS product names to merge to Cortellis drug names. We include exact name matches and manually reviewed the results of a "fuzzy" name matching algorithm to identify additional matches. Finally, we limited the matched set to a drug's earliest entry in the HAS data.

[10]In total, our data from Cortellis contains roughly 1,000 small molecule drugs that achieved regulatory approval in the period of the French data coverage (2008–2013). We only match 385 to the French data due to conservative name matching (with language differences) and because not all drugs achieve regulatory approval in the European Union at the same time as they reach the market in other countries.

Drug Novelty and Patent Citations

Another way to assess the value of drug novelty is to examine measures of the scientific value of the drug candidate. Hall et al. (2005) argue that the number of forward citations to a patent is a useful indicator of the 'quality' of the patent. We link drug candidates to patents using the crosswalk provided by Cortellis. We focus our analysis on the main patents, defined as the first set of patents filed prior to FDA approval that are related to the particular drug candidate. Since a drug may be associated with multiple patents, our analysis is at the drug-indication-patent level. The merged dataset has 3,539 observations. The Online Appendix contains additional details on the match between drug candidates and patents.

We estimate equation (4), where now the dependent variable is the logarithm of (one plus) the number of citation a patent receives. Column (3) of Table 2 reports the results from our baseline specification, which includes controls for the year the patent is granted; the indication (ICD9) treated by the drug; company and drug age (year of development) fixed effects. Panel C of Figure 4 provides a scatter plot of the results. Appendix Table A.7 examines how the choice of controls impacts our results.

We find that patents associated with novel drugs (lower maximum similarity) are more likely to receive a larger number of forward citations. The estimated effects of maximum similarity on the number of patent citations range from a low of -0.367 in Column (3) of Table 2 to a high of -1.33 in Column (4), after adding drug development year fixed effects. The correlation between our measure is both statistically and economically significant. Our estimates imply that a one-standard deviation increase in drug novelty is associated with a increase of 2 to 8 patent citations, when evaluated at the median number of citations a drug-related patent receives (23).

These results imply that patents associated with novel drugs are more likely to have higher scientific value. Given that the innovation literature tends to find a positive relation between patent citations and measures of economic value (Harhoff, Narin, Scherer, and Vopel, 1999; Hall et al., 2005; Moser, Ohmstedt, and Rhode, 2011; Kogan et al., 2017), it is also likely that novel drugs also lead to higher economic profits. The next section explores this possibility further.

Drug Novelty and Revenues

We next study the relation between our measure of the novelty of a drug candidate and revenues. An advantage of our data is that we have estimates of revenue at the drug level, which allows us to isolate the direct effect of novel drugs on firm revenue from indirect effects

due to spillovers or cannibalization of older drugs. To obtain data on drug revenue, we use the expenditures reported in the Medicare Expenditure Panel Survey (MEPS), which cover the 1996–2012 period. To match the names to Cortellis, we employ a name-matching procedure. The Online Appendix B provides further details on the data construction and matching procedure. The data is at the drug-indication-calendar year level. After restricting attention to drugs for which we can compute a similarity score, we are left with 21,755 observations.

We relate our measure of novelty to the drug's log revenues using a panel version of equation (4), which now includes calendar year fixed effects. Column (4) of Table 2 reports the estimated coefficient b from our baseline specification, which also includes controls for drug age; indication; and company effects. Panel D of Figure 4 provides a binned scatter-plot of the results, and Appendix Table A.8 reports results using different combinations of controls.

Conditional on FDA approval, chemically novel drugs generate more revenue for the firm than less novel drugs. In our baseline specification, a one standard deviation increase in novelty is associated with an increase in annual revenue of approximately 0.15 log points. The magnitude varies from 0.15 to 0.35 log points, depending on the choice of controls, as we see in Appendix Table A.8. Given that the unconditional standard deviation of log revenues is approximately 2.1 log points, our estimates imply that novelty can account for a non-trivial fraction of this variation.

As a measure of private value, drug revenue has the advantage of being (relatively) easily measurable. However, it also has some obvious short-comings. First, it ignores the costs of production. It is quite possible that markups are systematically related to the novelty of a drug candidate. If firms charge higher markups for more novel drug candidates, these estimates would understate the relation between novelty and private value. Second, it ignores spillovers on other drugs in the firm's portfolio. These spillovers can be positive, if the firm markets some drugs jointly, or negative, if the new drug cannibalizes older drugs. Hence, we next turn to additional measures of private value that exploit information contained in stock market valuations.

Drug Novelty and Private Value: Stock Market Reactions to FDA Approvals

We first exploit the information contained in the stock market reaction to news about the drug's success—that is, upon obtaining FDA approval. To construct an estimate of the drug's private value, we closely follow the methodology of Kogan et al. (2017). We focus on the firm's idiosyncratic return defined as the firm's return minus the return on the market portfolio, for up to 5 trading days following FDA approval. This window provides time

for the market to incorporate this information, while also reducing the possibility that this return also incorporates unrelated events. Similar to Kogan et al. (2017), we also allow for the possibility that this return window also incorporates stock price movements that are unrelated to the value of the approved drug.[11] We focus our attention on the first approval date for each drug. After restricting the sample to drugs with similarity scores that we can match to the CRSP dataset, we are left with 34 firms and 462 announcement days.

As before, we estimate a version of equation (4), where now the dependent variable is the logarithm of the estimated contribution to firm value. In addition to the usual set of controls, we also control for the year the patent is issued, as well as time-varying firm characteristics that may introduce noise in our estimate of added value (the firm's market capitalization and its volatility). Column (5) of Table 2 reports the estimated coefficient b from our preferred specification that includes the full set of controls. Panel E of Figure 4 provides the associated scatter plot; Appendix Table A.9 reports estimates using different combination of controls.

Our estimates imply that novel drugs contribute more to market values than less novel drugs. In terms of magnitudes, a one standard deviation increase in novelty is associated with a 12% larger stock price increase. This correlation is robust to controlling for year and firm fixed effects, firm size (market capitalization) and indication (ICD9). Panel E of Figure 4 shows the associated scatter plot (with all controls); we see that this relation appears to be monotonic across the full distribution of drug similarity.

However, this result by itself does not necessarily imply that more novel are also more likely to be (privately) valuable. The firm's stock market change following the drug's FDA approval is a composite of both the contribution of the drug to the firm's market value and the likelihood that the FDA approval was a surprise to the market. Specifically, suppose that the ex-ante likelihood of FDA approval is q. Following the approval of the drug by the FDA, the value of the firm should increase by $\Delta V = (1 - q)\,\xi_j$, where ξ_j is the private value of the drug (in dollars). Hence, it is important to adjust these estimates by the differential likelihood that a more novel drug is approved by the FDA (conditional on having reached Phase 3).

---

[11]Specifically, we closely follow Kogan et al. (2017) and assume that the cumulative return of the firm in that 5-day window equals $R_j = v_j + \varepsilon_j$, where $v_j \sim N^+(0, \sigma_v^2)$ denotes the value of drug j – as a fraction of the firm's market capitalization – and $\varepsilon_j \sim N(0, \sigma_\varepsilon^2)$ denotes the component of the firm's stock return that is unrelated to the patent. To calibrate the noise-to-signal ratio $\sigma_v^2/\sigma_\varepsilon^2$ we compare the return volatility of the firm on days with drug approvals to days without drug approvals. Since the distribution of v_j is likely to depend on the drug's novelty, we estimate the signal-to-noise ratio separately across drug novelty categories. We find that, on days in which drugs are approved, the variance of returns is approximately 11–36% larger, depending on their novelty. Our final estimate of the value of drug j is then equal to $E[v_j|r_j]\, M_j$, where M_j is the firm's stock market capitalization at the end of the day prior to the FDA approval.

The point estimates from the last column in Table A.5, imply that a one-standard deviation increase in novelty is associated with approximately a 10% increase in the likelihood of an unsuccessful FDA application, conditional on reaching Phase 3. Hence, much of the 12% larger stock price increase for novel drugs can be attributed to the fact that their approval was more unexpected for these drugs relative to me-too drugs. We therefore conclude that a one-standard deviation increase in novelty is likely associated with a $12 - 10 = 2\%$ higher private value for the approved drug., which is still a sizeable effect.

Drug Novelty and Private Value: Stock Market Reactions to Patents

As further evidence that novel drugs are more privately valuable, we also examine the correlation between novelty and measures of patent values for the patents associated with each drug obtained through the methodology of Kogan et al. (2017). Kogan et al. (2017) provide an estimate of the contribution of individual patents to firm value based on the change in the firm's market value following a successful patent grant. Since their measure is only available for publicly traded firms, we restrict attention to successful patent applications to publicly listed US companies that appear in CRSP. This restriction reduces the sample to 1,785 drug-patent-indication observations, corresponding to 83 firms and 183 drugs.

Similar to the previous section, we estimate a version of equation (4), where now the dependent variable is the logarithm of the estimated contribution to firm value. We employ the same set of controls as before. Column (6) of Table 2 reports the estimated coefficient b from our preferred specification that includes the full set of controls. Panel F of Figure 4 shows the associated scatter plot; Appendix Table A.10 reports estimates using different combination of controls.

We find that patents of novel drugs are likely to contribute more to firm value than patents associated with me-too drugs. The economic magnitude of the estimated effects is substantial: a one-standard deviation increase in novelty is associated with an approximately 8% increase in the (estimated) value of associated patents. To the extent that patent applications associated with novel drugs are also less likely to be successful, this increase represents an upper bound on the relationship between drug novelty and the value of its underlying patents, for the same reasons discussed in the last paragraph of the previous section. Nevertheless, the fact that this estimate is comparable in magnitude to that obtained in the previous sections is suggestive that we are capturing an underlying correlation with private values, as opposed to merely variation in the likelihood of a successful patent application.

2.4 Discussion and Caveats

So far, we have documented that novel drugs are riskier investments, as they are less likely to be approved by the FDA. However, conditional on approval, novel drugs are on average more valuable, regardless of how value is measured. Importantly, the magnitudes are substantial: a one standard deviation increase in novelty is associated with a 29% decrease in the likelihood of approval but a 2–35% increase in the average value conditional on approval. These estimates, however, do not necessarily imply that novel drugs are better (or worse) investments than me-too drugs from an ex-ante standpoint for several reasons.

First and foremost, we have little information on the relative development costs of novel drugs relative to me-too drugs. In general, assessing the costs of development is difficult because we do not have access to internal investment data and, furthermore, a large part of R&D spending is on scientific staff, who may work on multiple projects. One potential (though noisy) proxy for development costs are the number of patients enrolled in clinical trials and the number of trials associated with drugs. Since trials are so expensive, recruiting patients and running trials constitutes a substantial proportion of a drug's development cost. In Table A.1 and Figure A.10 in the Appendix, we consider how the number of patients and number of trials associated to a compound vary by its chemical novelty. We find no consistent relationship between these proxies of development cost and drug novelty. The left hand side panels of Figure A.10, for instance, plot bin scatters of the relationship between drug novelty and number of patients or trials for our full set of drug candidates. We find no relationship between novelty and the number of enrolled patients across all trials. We find a weakly negative relationship between similarity and the total number of trials; however, these appear to be driven by the set of drug candidates with similarity scores of exactly 1, which may include extended release formulations that should require fewer additional trials. Restricting to the set of candidates with similarity strictly less than 1, we find, if anything, that more similar drugs are more expensive, though the relation is not statistically significant.

Second, we do not know the value of failed drugs. One possible assumption is that the (private) value of non-approved drugs is zero, since they generate no revenues for the developing firm. However, this may not be a good assumption. Even if a particular drug candidate is not successful, eliminating this particular possibility implies that some scientific progress has been made (see, e.g. Magazzini, Pammolli, and Riccaboni, 2012), which may lower the cost of future drug development. Along these lines, even if the particular drug candidate is not successful, firms (and key talent) may gain expertise, which again may lower future costs. Third, it is possible that the likelihood that a drug candidate is successful

18

is not independent from the success rate of other candidates. It may be the case that novel candidates have less correlated risk than me-too drugs, which might make them more attractive from a diversification standpoint.

In sum, extrapolating from the estimates in Section 2.2 and 2.3 to make inferences about the ex-ante value of novel drugs is challenging. Another way to arrive at this answer, however, is to examine what firms actually do when faced with the ability to develop more drugs. The next section explores this indirect approach to estimating the ex-ante value of novelty by examining the response of drug development decisions to (plausibly exogenous) cashflow shocks.

3 The Effect of Cashflow Shocks on the Development of Novel Drugs

As a starting point, we examine how the novelty of drug candidates correlates with firm characteristics. The Cortellis data contains scarce information on firm characteristics. However, we observe the history of their development decisions, which can be used to construct a measure of the number of drug candidates developed so far. We refer to this measure interchangeably as firm size or experience, recognizing that it is difficult to separate them without more detailed accounting data. Later on, we construct a more precise measure of firm size (prior to the treatment) defined as the total revenue generated by approved drugs prior to 2003. Since, this measure is only available for a subset of firms (those that existed prior to 2003), we do not use it as our primary metric.

We find that larger (or more experienced) firms are more likely to undertake the development and testing of novel compounds. Specifically, a one standard deviation increase in drug novelty is associated with an approximately 15 to 20 percent increase in the previous experience (as measured by number of past candidates in development) of the firm associated with the drug—as we can see in the top panel of Appendix Figure A.9. Appendix Table A.11 shows results from additional specifications.

The correlation between firm size/experience and novelty does not appear to be driven by acquisitions of novel drugs by large firms. Indeed, a popular view is that smaller firms do cutting edge research in the hopes of being acquired by larger firms, who then follow through with clinical trials and development. The data are not fully consistent with this view. As we see in the bottom panel of Appendix Figure A.9, the relation between drug novelty and firm size/experience is similar when we match candidates with their originating firms. However,

this evidence is only suggestive, because reporting for drugs in preclinical discovery stages is not mandated. We may therefore miss early stage acquisitions made before a drug candidate first appears in our data.

The fact that firm size is correlated with the likelihood to develop novel drugs suggests that financial constraints may play a role in constraining investments in novelty. Indeed, prior findings in the literature suggest that financing constraints may be an important factor limiting innovation in the pharmaceutical industry. For instance, DiMasi, Grabowski, and Hansen (2016) estimate that R&D costs are high (around $1.4 to $2.6 billion dollars per drug). Moreover, returns to developing new drug candidates are skewed and uncertain (since fewer than 20% of drug candidates ever reach approval). Further, intellectual property—even when protected by patents—is difficult to value or collateralize, making access to external finance more difficult. Berndt, Nass, Kleinrock, and Aitken (2015) find that sales of newly-approved drugs are so volatile that returns are often insufficient to cover R&D costs.

The remainder of this section examines how the novelty of a firm's marginal drug investments responds to a shock to firm cashflows. Whether firms respond to an increase in cashflows by developing more drug candidates is indicative of financial constraints. In addition, whether the marginal drug candidate being developed is a novel or a me-too drug is indicative of which of the two types of drugs is more likely to be ex-ante more valuable. This indirect approach of measuring the ex-ante value of novelty, which is based on firms' revealed preference, sidesteps many of the issues discussed in Section 2.4.

3.1   Empirical Strategy

To identify the causal impact of a shock to firm cashflows on drug development, we exploit the introduction of Medicare Part D. The Medicare Modernization Act expanded insurance for elderly Americans to include coverage for prescription drugs taken at home. This Act was passed in 2003 and implemented in 2006. Previous work has shown that it led to an increase in drug sales for some drugs, a decrease in price, and an overall increase in the value of the firms that market these drugs (Duggan and Scott Morton, 2010; Lichtenberg and Sun, 2007; Friedman, 2009). The extent to which a firm benefits from the introduction of Part D depends on two factors: the types of drugs that the firm sells and the amount of market exclusivity remaining on those drugs. Our empirical strategy makes use of both these sources of variation, and therefore allows us to isolate shocks to the firm's cashflows from an improvement in their investment opportunities—that is, an increase in demand for new drugs to the elderly.

First, the extent to which a firm receives increased expected cash flows depends on the share of its customers that are in the Medicare population. A firm that sells only drugs for pediatric conditions should not expect to see an increase in sales except possibly through secondary factors such as wealth effects. By contrast, a firm with drugs for osteoporosis would expect an increase in sales because Part D ensures that its potential customers will now be reimbursed for their purchase of its products. Following previous work (Blume-Kohout and Sood, 2013; Duggan and Scott Morton, 2010; Dranove et al., 2014), we use the notion of a "Medicare Market Share" (MMS) to quantify a drug's exposure to the Part D policy shock, which is a function of the fraction of sales to elderly customers. Throughout the paper, we use the terms MMS and elderly share interchangeably. To construct drug MMS, we match approved drugs in our primary Cortellis dataset to the Medical Expenditure Panel Survey (MEPS), which contains drug–level information on sales, by patient demographics. The Online Appendix describes the matching process. We define a drug's MMS as the share of revenues generated by patients over 65 in 2003. We then construct a firm–level Medicare exposure by aggregating these drug–specific MMS values into Firm $\text{MMS}_{f,2003}$, which is the firm-average of drug level MMS, as of 2003, just prior to the introduction of Part D.

Second, the extent to which a firm receives increased expected cash flows as a result of Part D also depends on the amount of market exclusivity remaining on its current drug portfolio. Firms whose drugs have longer remaining exclusivity at the time of Part D would expect greater increased future cash flow because of their longer horizon for charging monopoly prices. To determine remaining exclusivity for each firm's drugs, we match drugs approved prior to 2004 to their associated patents and, where possible, link the drugs to their key patent expiration dates and FDA exclusivity extensions, relative to the base year 2003.[12] We then aggregate these drug-level measures to the firm level by defining a firm's overall drug life, Overall Drug $\text{Life}_{f,2003}$, as the proportion of its approved drugs with long remaining exclusivity as of 2003. Since our data on exclusivity periods is somewhat noisy, to minimize measurement error we use a cutoff rule. In our baseline results we define long exclusivity as 5, or more, years, which is close to the median remaining life in our sample. Our results are robust to alternative cutoffs, as shown in Appendix Table A.22.

---

[12]The FDA will grant extensions on a drug's market exclusivity period, beyond the relevant patent expiration date, in a number of scenarios. For example, the Orphan Drug Act of 1983 incentivizes the development of drugs for rare ("orphan") diseases through different provisions, including a guarantee of seven years of market exclusivity. Other legislation also sets aside additional market exclusivity for other special drug designations (e.g., five years for New Chemical Entities, and six months for Pediatric Exclusivity). For more information on our drug-to-patent data and patent expiration dates see the Online Appendix, Section 6.3

Relying on either source of variation alone does not identify the impact of expected cash flow. Firms with high MMS may change their investment behavior following Part D for two reasons: first, because they may expect greater cashflow due to increased demand for their existing drugs (this is the effect we would like to identify), and second, because they may also experience an increase in their investment opportunities—that is, an increased demand for new drugs in the areas in which they have more R&D experience. Indeed, previous work has employed a similar strategy—constructing an MMS aggregated to the level of a therapeutic market—in order to study the role of demand in stimulating drug development (Blume-Kohout and Sood, 2013; Dranove et al., 2014). To isolate the impact of cashflows from an increase in the demand for new drugs (i.e., an improvement in investment opportunities), we need a measure of Medicare cashflow exposure that compares firms within the *same* therapeutic market, so that they receive a similar shock to the demand for new drug development. Conversely, firms with high overall drug life may change their investment behavior both because they expect increased cash flows and because these firms have a younger portfolio of drugs in general. For example, firms with more recent drugs may focus more on early stage experimentation while firms with older drugs may focus more on pushing new candidates through clinical trials. If this were the case, we may observe a difference in their innovative investments after Part D, simply because these firms are at different points in the product development cycle.

To isolate the impact of changing expected cash flows, our firm–specific measure of exposure to Part D, defined below, takes into account both drug MMS and market exclusivity:

$$\text{Medicare Drug Life}_{f,2003} \;=\; \sum_{i \in A_f} \left[\frac{\text{Drug MMS}_{i,2003}}{\sum_{j \in A} \text{Drug MMS}_{j,2003}} \mathbb{I}(\text{on patent in } X \text{ yrs})_{i,2003} \right] \quad (5)$$

Here, firm f's Medicare Drug Life in 2003 is defined as the proportion of its approved drugs ($i \in A_f$) with long remaining exclusivity as of 2003, weighted by their drug–level MMS. Firms with higher Medicare Drug Life are those with longer exclusivity on their high MMS drugs.

To see why our definition of treatment (5) helps us isolate the cashflow effect from an increase in future investment opportunities and from differences in product life cycle, consider a simple example. There are two firms, A and B, both with two approved drugs, one with a high MMS of 0.75 (drug H) and another with a low MMS of 0.50 (drug L). Both firms have one drug that will expire soon and another that will not. Since both firms have the same Firm MMS and the same overall drug life, they are predicted to experience similar demand-induced

increases in their incentive to develop drugs for the elderly and they are at the same part of their drug development cycle, as proxied by remaining exclusivity on their approved drugs. However, suppose now that at Firm A, drug H_A will remain on patent, while drug L_A will expire. Meanwhile, suppose that at Firm B the opposite is true: the high MMS drug, H_B, will expire while L_B will remain on patent. Despite their other similarities, we would intuitively expect Firm A to receive a higher cashflow shock as a result of Part D, because its high MMS drug will remain on patent. This is what our measure delivers: Firm A's Medicare Drug Life is $\frac{75}{75+50} \times 1 + \frac{50}{75+50} \times 0 = 0.6$, while Firm B's is $\frac{75}{75+50} \times 0 + \frac{50}{75+50} \times 1 = 0.4$.

Table 3 describes the distribution of our main treatment variable. The median firm has a Medicare Drug Life of 0.54 but most firms have a value of either zero or one. This is because many firms have only one approved drug on the market as of 2003, so that their treatment values can only be 0 or 1. Appendix Figure A.8 shows a smoother distribution of Medicare Drug Life for firms with non extremal values and we show in Appendix Tables A.25 and A.21 that our results are robust to restricting to this subsample, or to using a binary treatment measure.

We estimate the causal impact of a financial shock to drug development using the following linear specification:

$$
\begin{aligned}
\text{New Drug Candidates}_{ft} \;=\; & a_0 + a_1 \text{Post} \times \text{Medicare Drug Life}_{f,2003} \qquad\qquad (6)\\
& + a_2 \text{Post} \times \text{Overall Drug Life}_{f,2003}\\
& + a_3 \text{Post} \times \text{Firm MMS}_{f,2003} + \delta_f + \delta_t + e_{ft}
\end{aligned}
$$

Our main coefficient of interest is a_1, which captures the *cash flow* impact of our main treatment variable defined in Equation (5). We control for an interaction between Overall Drug Life and the post Part D period. This helps alleviate concerns that our estimates of the impact of Part D are not picking up differences in firm development cycles, since firms with longer patent life remaining may choose to pursue different development strategies. We also control for the overall firm's Medicare market share to ensure that we are not picking up a product demand effect. In our baseline specification we include firm- and quarter-dummies to account for unobservable firm differences and aggregate trends in drug development. In addition, we also estimate a specification with company-specific linear time trends (see Table A.19 in the Appendix), to ensure that our results are not driven by pre-existing trends. To account for possible serial correlation in unobservables, we cluster standard errors at the firm level.

3.2 The Effect of Cashflow on Drug Development

Table 3 contains summary statistics of our dataset at the company–quarter level. The average firm in our sample has 0.55 new drug candidates per quarter, but the data are highly skewed: most firms do not have a new drug candidate under development every quarter. This implies that the outcome variables for our analysis will be zero in most company–quarters. We therefore use the logarithm of one plus the number of new, or the number of novel drugs, as our primarily outcome measures. In the Appendix, we show that our findings are robust to using alternative specifications, including count models.

Number of New Candidates

Table 4 examines the causal impact of a financial shock, as described in Equation (6), on the total number new drug candidates under development by our sample firms. Columns (1)-(3) focus on the count of new candidates; Columns (4) to (6) focus on the logarithm of one plus the number of new candidates, which is our preferred outcome measure. Column (4) presents our estimates with only the main treatment variable and the company and time fixed effects. The estimated coefficient a_1 is equal to 0.06 and statistically significant. Looking at Columns (5) and (6), we find that controlling for overall drug life and firm MMS increases the overall magnitude of our estimate (0.268 and 0.263, respectively). The negative coefficient on Post $\times$ Overall Drug Life$_{f,2003}$ indicates that firms with a newer set of drugs as of 2003 proceed to introduce fewer new candidates into development in the post Part D period, suggesting that controlling for differences in firm development cycles is important. Perhaps surprisingly, the inclusion of Post $\times$ Firm MMS$_{f,2003}$ in Column (6) does not materially affect our point estimates, suggesting that (in our sample) demand effects do not appear to increase development separately from cash flow effects.[13] For the remainder of our analysis, we use Column (6) as our baseline specification.

The estimated magnitudes are economically substantial. Focusing on Column 6, we can infer that a one standard deviation (0.41) increase in the main treatment variable leads to an 11% increase in the number of new drug candidates. This corresponds to an elasticity of output to treatment of 0.40.[14] In Section 3.5, we translate these magnitudes in terms of dollars for a subset of our firms.

---

[13]This finding may differ from drug market level estimates of the impact of demand on innovation because our firm–level analysis excludes the possibility of entry by new firms.

[14]To arrive at this figure, we note that for a regression of the form $log(1 + y) = bx + e$, the elasticity is given by $b \times \frac{1+y}{xy}$, where we evaluate at the mean of Medicare exposure in 2003 (0.54) and at the mean of drug output overall (0.55).

Novelty of New Candidates

Next, we examine the novelty of the new drug candidates. Panel A of Figure 5 reports estimates of equation (6) where the outcome variable is the number of drug candidates with a given similarity score. We see that the greatest increase in new candidates comes from an increase in candidates with maximum similarity scores between 0.3 and 0.6. Importantly, we see no increase in very similar (me-too) candidates, defined as those with chemical similarity greater than 0.9. We also do not see increases in the number of drugs with similarity below 0.3, perhaps because fewer than 8 percent of candidates have novelty scores in that range (see Table 1).

Since the number of drugs in each bin does vary, we also report the estimates in terms of elasticities in Panel B of Figure 5. We see that novel drugs (i.e., drugs with a maximum similarity score less than 0.6) are the ones that exhibit the largest elasticity with response to the cashflow shock. The elasticities for me-too drugs are smaller in magnitude and not statistically significant.[15]

In sum, these results imply that the marginal drugs that firms develop in response to a cashflow shock tend to be more novel. One interpretation of these results is that, on the margin, firms perceive novel drugs to have a higher ex-ante value than me-too drugs, and hence direct the higher cashflows in response to the Medicare expansion towards developing more novel drugs.

Event studies

One potential source of concern is that the differences in responses among the treatment and control group reflect pre-existing trends. To address this concern, Figures 6 and 7 show how the estimated effect of the cashflow shock on the number of new and novel drugs, respectively, vary over time. Focusing on Figure 6, we see that firms with different values of Medicare Drug Life$_{f,2003}$ appear to be on parallel trends prior to the introduction of Part D. Following that, firms with high exposure begin to increase their drug output relative to firms with lower exposure starting in 2004, and this increase in drug development appears persistent, at least through the end of our data in 2014. Similarly, Figure 7 shows that the number of drugs in the bottom three quartiles of similarity (shown in the top two panels and bottom left panel) increases following the introduction of Part D. By contrast, we see no

---

[15]Table A.14 in the Appendix repeats the analysis, except that it bins the drugs in terms of novelty deciles. We see the same pattern, namely a larger effect among more novel drug candidates and no significant increase in me–too candidates.

such increase in output for the most chemically derivative drugs. To address any remaining concerns about preexisting trends, Appendix Table A.19 also shows that our main results are robust to including company-year-quarter linear trends.

The persistence of the effect should not be surprising because Medicare Part D was a permanent shock to firm cash flows. What is perhaps surprising is that we observe a small increase in the number of new and novel drug candidates starting in 2004, even though Part D did not go into effect until 2006. Put differently, firms appear to have responded to an increase in expected cash flows in 2003, even though these cashflows did not materialize until 2006. We believe that this quick reaction likely appears in our data for several reasons. First, because we focus on established firms with at least one approved drug in 2003, our sample consists of firms that very likely have a stock of drug candidates in the discovery phase at that time, making it easier for them to respond quickly. A change in expected cash flows may make firms more willing to support more of these candidates as they move through the development pipeline. Second, as discussed in Appendix Section 5.2, the costs of drug discovery are backloaded. Pre–clinical and Phase 1 trials, while expensive compared to investments in R&D in other industries, are less costly than Phase 2 or Phase 3 trials. A firm that anticipates greater revenue in 2 to 3 years may choose to push its more novel discovery-stage candidates into the preclinical testing knowing that it would not need to pay for the bulk of development costs for another few years and, moreover, would only need to make these payments if the drug candidate actually shows promise. Third, it is possible that firms could increase their borrowing in response to this increase in expected cashflows. We explore this idea further in Section 3.5.

3.3 What types of drugs do firms develop?

A natural next step is to further examine the types of drugs that firms develop, and how these new drugs fit into the firms' existing portfolios. Table 5 shows that firms increase development for a wide variety of drugs, not just those that are more likely to be covered by Part D insurance. We split the outcome variable (log of one plus number of new compounds) by the quartile of Medicare market share (MMS) that the new drugs fall into. Comparing elasticities across Columns (1) through (4), we see that firms are equally responsive in developing drugs across all MMS quartiles.

The fact that firms are expanding their portfolios more broadly, and not just in the areas that experience a demand shock as a result of Medicare Part D, strongly indicates that our identification strategy is at least partially successful in isolating a cashflow shock from a

shock to firms' investment opportunities. Were we simply identifying the impact of Part D on demand for the development of drugs for seniors, we would expect the marginal candidates that firms develop to be in high MMS indications.

In Table 6 we examine how these new drugs relate to the firm's existing portfolio of drug investments. Columns (1) to (4) focus on how new candidates compare to a firm's existing candidates on the basis of what disease indication they focus on. Column (1) shows that increased resources lead firms to develop drugs for indications for which they have not developed candidates in the past. A one standard deviation (0.41) increase in Medicare Drug Life increases the number of candidates in indications new to a firm by about 7 percent. Similarly, Column (2) shows that firms receiving a larger Medicare shock reduce the concentration of indications that they focus on, as measured by a decreasing indication–specific within–firm Herfindahl. Columns (3) and (4) show that this same pattern applies when considering candidates that act on new biological targets, as opposed to treating new disease, indications. In sum, we find some evidence that treated firms use some of these additional cashflows to diversify their existing portfolio of drugs, which is consistent with the idea that they behave in a risk averse manner.

3.4   Firm Heterogeneity

We next examine how the effect of cashflows on drug development decisions varies across firms. One possibility is that firms that are more financially constrained are more likely to be affected. It is important to note, however, that identifying which firms are likely to be financially constrained has proven challenging (Farre-Mensa and Ljungqvist, 2016). That said, we do expect that a given dollar increase in cashflows is likely to be more relevant for firms that had low prior cashflows than for firms with high prior cashflows. As a proxy for prior cashflows, we create a measure of the firms' total revenues generated by drug candidates that are approved prior to 2003. We refer to this measure here as firm size. We then estimate equation (6) separately across the firms that are below or above the median firm size in 2003.

Table 7 presents the results. We see that the estimated coefficient a_1 on the main treatment effect is statistically significant for the smaller firms (column (3)). For the larger firms (column (2)), the point estimates are larger, but less precisely estimated. In terms of elasticities, smaller firms display a larger response: a one percent increase in the main treatment variable is associated with a 0.64 vs 0.30 percentage increase in the number of drug candidates across small and large firms, respectively.

By contrast, we find no meaningful differences in the impact of cashflows between large and small firms on their propensity to develop novel vs me-too drugs—see Appendix Table A.11. This lack of differential response across firm size is not particularly surprising. Our measure of firm size conflates prior revenue with firm experience and there is simply not enough variation in the data to separate the two. It is possible that more experienced firms have more opportunities to develop novel drugs than less experienced firms, consistent with the discussion in the beginning of this Section. Even if somehow the development of novel drugs were more constrained in lower-revenue firms (and hence we might expect to see a greater response), these firms might have fewer potential novel candidates to invest in.

In sum, we find that smaller firms are more sensitive to the treatment than large firms, where size is measured by pre-2003 revenue. However, given that this measure is likely correlated with other (unobservable) characteristics that might differentially affect development, such as firm experience, these results should be interpreted with caution.

3.5   Discussion

In this section, we discuss the interpretation of our findings and assess the quantitative significance of our findings relative to the existing literature.

Interpretation of our findings

Taken together, our findings are consistent with the presence of financing frictions in drug development. Models with financing frictions predict that, all else equal, an increase in internal funds will lead to more investment (see, e.g. Froot et al., 1993; Kaplan and Zingales, 1997). In our setting, this implication translates into firms developing a larger number of drug candidates following a shock to the availability of internal funds.

Further, the fact that firms tilt their development towards more novel candidates could be interpreted as evidence for decreasing risk aversion, which is also a direct implication of models with financing frictions, even when the risks associated with the investment projects are idiosyncratic. Specifically, firms anticipate the possibility that, in they future, they will need to raise external funds to finance new investment projects. As long as the costs of external finance are convex, firms will be reluctant to invest in risky projects today (see, e.g. Froot et al., 1993). A positive shock to their current or future cashflows will lower the amount that needs to be raised externally, and therefore make firms more risk tolerant.[16] A

---

[16]A different version of the financial constraints hypothesis is that more novel drugs are more expensive to finance using external funds. One reason why this might be the case is that the degree of information

related possibility is that having more internal funds may allow the firm to invest in a more diversified portfolio of projects, which will lower the average riskiness of the firm's cashflows. This prediction is consistent with the results in Table 6.

The fact that firms increase development of novel drugs in response to cashflow shocks can also be interpreted as indirect evidence that the perceived ex-ante value of novel drugs is higher than me-too drugs. Naturally, this conclusion is subject to several caveats, since it assumes that firm executives that are responsible for drug development decisions have the same objective function as firm shareholders (or a benevolent social planner). That need not be true for several reasons. For instance, firm executives might derive private benefits from developing novel drugs, since this may help strengthen their human capital and reputation, which may lead to an increase in their lifetime income. It is also possible that firm managers who are in charge of drug development decisions overestimate either the likelihood of success, or the drugs revenues upon FDA approval. This possibility is in line with the literature discussing the existence, and persistence, of managerial overconfidence (e.g., DeBondt and Thaler, 1995; Bernardo and Welch, 2001; Malmandier and Tate, 2005; Goel and Thakor, 2008; Malmandier, Tate, and Yan, 2011; Galasso and Simcoe, 2011).

In sum, the results of this section support the view that, on the margin, firm executives perceive the ex-ante value of novel drugs to be higher than me-too drugs. Whether this conclusion translates into novel drugs having higher private or social value than me-too drugs depends on assumptions (and priors) about the extent of agency frictions inside pharmaceutical firms.

Interpreting the Magnitudes

Our analysis so far has been qualitative in nature. Our central finding is that a one standard deviation change in pre-Part D Medicare drug life leads to an 11% percent increase in the development of new and novel drugs. To assess the magnitude of this effect and benchmark it to the existing literature, we need to express our estimates in terms of the implied elasticity of drug development with respect to firm R&D spending. Hence, we need a measure of how much firm resources increase as a result of this policy.

---

asymmetries between the firm and external investors regarding the success probability of a novel drug candidate may be too large. Indeed, this may be the case if the average likelihood of success for a novel drug is sufficiently low (as we see in Section 2), but there is considerable heterogeneity in the *ex-ante* likelihood of approval—and more importantly, firm managers have some information about this likelihood that they cannot credibly share with outside investors. In this case, we would expect to see under-investment in novel drugs by 'high type' firms that need to access external markets due to adverse selection. An increased availability of internal funds will lead these firms to develop more of these novel drugs.

To assess the response of R&D investment to our main treatment variable, we match the public firms in our data to Compustat North America and Compustat Global. We are able to match approximately 50% of our sample firms. For these firms, we estimate our main specification, as defined by Equation (6), but with the log of firm profits and R&D spending as dependent variables. These results are reported in Table 8. Columns (1) and (2) show that firms with higher Medicare Drug Life in 2003 experienced higher growth in R&D and operating cashflows in the years following treatment.

These results can be used to compute the elasticity of drug development with respect to firm R&D spending. Focusing on Column (1), we estimate an elasticity of treatment exposure to R&D expenditure of 0.53.[17] If a one percent increase in treatment leads to both a 0.53 percent increase in R&D and a 0.40 percent increases in drug output, this suggests an elasticity of output to R&D of 0.75. If we apply this same calculation to our analysis by novelty bins, we find an elasticity of output to R&D of about 1.01 and 1.59 for drugs in the top 1 and 2 deciles of novelty, respectively, compared to an elasticity of 0.02 and 0.31 for the top 1 and 2 deciles of the most similar drugs, respectively.[18]

With those considerations in mind, our benchmark elasticity estimate appears sensible, given the range of estimates that exist in the literature. One set of estimates comes from Henderson and Cockburn (1996), who examine determinants of research productivity in the pharmaceutical sector. They find elasticities of R&D with respect to "important" patents of about 0.4 to 0.5. If firms are more responsive to their own spending, we would expect private elasticities to be greater than public elasticities. More recently Azoulay et al. (2016) estimate the casual impact of *public* investments in biomedical research on patenting and drug development by private firms and find elasticities of approximately 0.4–0.6. Other studies document larger elasticities (2.8 to 4) but their estimates are not directly comparable to ours (Acemoglu and Linn, 2004; Blume-Kohout and Sood, 2013).

Last, we examine whether firms responded to an increase in expected cashflows by increasing their borrowing, which might explain why firms responded to the treatment a few years before the increase in cashflows was actually realized. Further, if the expansion of Medicate Part D benefitted treated firms not only with an expansion of the flow in operating

---

[17]This is taken from the coefficient, 0.975, multiplied by 0.54, the mean of treatment exposure in the pre-period.

[18]There are several caveats to this analysis. Because some of our firms include large conglomerates (for instance firms such as Dow Chemical), our R&D figures include spending on sectors that may not be related to pharmaceuticals. More generally, we caution that while we estimate a causal impact of Medicare exposure on drug output, we cannot say that we estimate the associated productivity of R&D spending because lags between R&D expenditure and final commercial output are difficult to predict when it comes to drug innovation. Put differently, we cannot identify which part of R&D spending is used to finance which drugs.

profits, but also via an expansion in their ability to borrow, then the effect on profits may underestimate the relaxation of financing constraints for the treated firms which would affect our elasticity calculations. To assess this possibility, Columns (3) and (4) of Table 8 examine the response of firm-level borrowing in response to treatment. Though the point estimates suggest this might be the case, the coefficients are too noisily estimated to conclude that the response is different from zero; we conjecture that the noisiness of these findings is also driven by the fact that pharmaceutical firms are less likely than other firms to use debt financing (see Section 5.2 in the Online Appendix for a discussion). Panel B shows that these results are robust in limiting our sample to years prior to the financial crisis in 2008.

3.6   Additional Results and Robustness Checks

Here, we provide a brief description of additional results and robustness checks. We refer the reader to the Online Appendix for more details.

In Section 3.2 we shows that firms that experienced an increase in cashflows developed more novel drugs. A natural question is whether these new candidates were developed in-house or acquired by another firm. We find that firms primarily use marginal resources to push forward the development of their own earlier stage drug candidates, rather than by acquiring candidates from other firms (Table A.12 and Figure A.13 in the Appendix). In addition, the majority of new drugs that are developed were originated by that same firm, rather than acquired from another firm (Appendix Table A.12). Further, we find that larger firms account for the majority of the marginal drugs developed as a result of a finance shock, but, in looking at elasticities we find that smaller firms are just as, if not more, responsive (Table A.15 and Figure A.14 in the Appendix). We caution that our results cannot be fully extrapolated to the smallest pharmaceutical and biotechnology firms because our sample is limited to firms that had an approved drug on the market in 2003 (these are the firms for which we can calculate our key variable, Medicare Drug Share).

Next, we examine the robustness of our main results across several dimensions. First, our measures focus on chemical similarity as measured by Tanimoto scores. A limitation of this approach is that it can only be applied to small molecule drugs, and not to more complex biological entities, known as biologics, which has been a growing area of R&D focus in pharmaceuticals. Biologics include a broad class of biological products, usual derived or synthesized using tools from bioengineering rather than pharmaceutical chemistry: vaccines and insulins are both examples of biologics. Remicade (infliximab) for arthritis and Avastin (bevacizumab) for cancer are also examples of modern biologics. Even though biologics

make up a smaller fraction of overall pharmaceutical R&D, understanding the impact of our shock on biologics is potentially important because they have been a source of breakthrough innovation in drug development in recent years and are a priority research area for many pharmaceutical firms. If we were to find that our shock leads to decreases in biologic output, this would complicate our finding that access to financial resources increase novelty. In Table A.18 in the Online Appendix, we show that this is not the case: more treated firms, especially those who have developed biologics prior to Part D, increase their biologic output more relative to less treated firms.

In addition, we find that our results are not driven by pre-existing firm-specific trends (Appendix Table A.19), and are robust to alternative definitions of novelty, specifically novelty with respect to prior candidates for the same indication (Appendix Table A.17). Further, our results are robust to different empirical specifications: Table A.20 in the Appendix considers Poisson Count models, Table A.23 considers a binary outcome variable (based on whether the firm have any new drugs), and Table A.21 considers a binary treatment. Our results are also robust to different definitions of treatment: Table A.22 shows that we can define Medicare Drug Life based on proportion of drugs with more than 7 and 10 years of remaining exclusivity, weighted by drug MMS. In Appendix Table A.24 we estimate alternative specifications wherein we control for the total years of remaining patent life times the post period indicator, as a proxy for both development cycle and firm size, in lieu of controlling for the overall *proportion* of drugs on patent. Last, our results are not driven by the extreme values in the Medicare market share variable shown in Figure A.8; Table A.25 shows that are results are similar if we exclude these firms.

4   Conclusion

This paper develops a new measure of drug novelty based on the chemical similarity of a drug's active ingredients. We show that our measure novelty is significantly positively correlated with measures of private and social value. However, novel drugs are less likely to be approved by the FDA, hence they represent a riskier investment.

To further understand whether the perceived ex-ante value of novel drugs is higher than me-too drugs, we study the types of drugs firms decide to develop in response to a cash flow shock as a result of the Medicare Part D expansion. We show that firms that experience a positive shock respond by both increasing the number of drug candidates that they develop and by pursuing more novel drug candidates. The marginal drugs that are developed as a

result tend to diversify the firm's drug portfolio by pushing firms to develop candidates for different indications and acting on different biological targets.

We conclude that, on the margin, pharmaceutical firms are more likely to develop novel drugs in response to an increase in cashflows. Given that the overall level of me-too drugs is high and increasing, this is consistent with a development strategy in which firms spend their first research dollars on less risky, less novel compounds, and take risks on more novel compounds when given additional resources.

References

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

Almeida, H., M. Campello, B. Laranjeira, and S. Weisbenner (2011). Corporate Debt Maturity and the Real Effects of the 2007 Credit Crisis. *Critical Finance Review 1*, 3–58.

Azoulay, P., J. Graff-Zivin, D. Li, and B. Sampat (2016). Public r&d investments and private-sector patenting: Evidence from nih funding rules.

Backman, T. W. H., Y. Cao, and T. Girke (2011, 05). Chemmine tools: an online service for analyzing and clustering small molecules. *Nucleic Acids Research 39*(suppl_2), W486–W491.

Baldi, P. and R. Nasr (2010, 07). When is chemical similarity significant? the statistical distribution of chemical similarity scores and its extreme values. *Journal of Chemical Information and Modeling 50*(7), 1205–1222.

Benmelech, E., N. K. Bergman, and A. Seru (2011, June). Financing labor. Working Paper 17144, National Bureau of Economic Research.

Benmelech, E., C. Frydman, and D. Papanikolaou (2017, March). Financial frictions and employment during the great depression. Working Paper 23216, National Bureau of Economic Research.

Bernardo, A. E. and I. Welch (2001). On the evolution of overconfidence and entrepreneurs. *Journal of Economics & Management Strategy 10*(3), 301–330.

Berndt, E. R., I. M. Cockburn, and K. A. Grépin (2006). The impact of incremental innovation in biopharmaceuticals. *PharmacoEconomics 24*(2), 69–86.

Berndt, E. R., D. Nass, M. Kleinrock, and M. Aitken (2015). Decline in economic returns from new drugs raises questions about sustaining innovations. *Health Affairs 34*(2), 245–252.

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

Bloom, N., C. I. Jones, J. V. Reenen, and M. Webb (2017). Are ideas getting harder to find? working paper, Stanford University.

Blume-Kohout, M. E. and N. Sood (2013). Market size and innovation: Effects of medicare part d on pharmaceutical research and development. *Journal of Public Economics 97*, 327 – 336.

Bond, S., D. Harhoff, and J. van Reenen (2005). Investment, r&d and financial constraints in britain and germany. *Annals of Economics and Statistics* (79-80), 433–460.

Brown, J. R., S. M. Fazzari, and B. C. Petersen (2009). Financing innovation and growth: Cash flow, external equity, and the 1990s r&d boom. *The Journal of Finance 64*(1), 151–185.

Budish, E., B. N. Roin, and H. Williams (2015, July). Do firms underinvest in long-term research? evidence from cancer clinical trials. *American Economic Review 105*(7), 2044–85.

Chodorow-Reich, G. (2014). The employment effects of credit market disruptions: Firm-level evidence from the 2008-09 financial crisis. *Quarterly Journal of Economics 129*(1), 1–59. Lead article.

DeBondt, W. F. M. and R. H. Thaler (1995). *Handbooks in Operations Research and Management Science*, Volume 9, Chapter Financial Decision-Making in Markets and Firms: A Behavioral Perspective. Elsevier Press.

DiMasi, J. and R. Chakravarthy (2016). Competitive development in pharmacologic classes: Market entry and the timing of development. *Clinical Pharmacology and Therapeutics 100*(6), 754–760.

DiMasi, J. and C. Paquette (2004). The economics of follow-on drug research and development: Trends in entry rates and the timing of development. *PharmacoEconomics 22*(2), 1–14.

DiMasi, J. A. and L. B. Faden (2011, 01). Competitiveness in follow-on drug r&d: a race or imitation? *Nat Rev Drug Discov 10*(1), 23–27.

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

Dranove, D., C. Garthwaite, and M. Hermosilla (2014). Breakthrough or me-too? the impact of medicare part d on biotech innovation.

Duggan, M. and F. Scott Morton (2010, March). The effect of medicare part d on pharmaceutical prices and utilization. *American Economic Review 100*(1), 590–607.

Duygan-Bump, B., A. Levkov, and J. Montoriol-Garriga (2015). Financing constraints and unemployment: Evidence from the Great Recession. *Journal of Monetary Economics 75*(C), 89–105.

Farre-Mensa, J. and A. Ljungqvist (2016). Do measures of financial constraints measure financial constraints? *The Review of Financial Studies 29*(2), 271–308.

Friedman, J. (2009). The incidence of the medicare prescription drug benefit: Using asset prices to assess its impact on drug makers.

Froot, K. A., D. S. Scharfstein, and J. C. Stein (1993, December). Risk Management: Coordinating Corporate Investment and Financing Policies. *Journal of Finance 48*(5), 1629–58.

Frydman, C., E. Hilt, and L. Y. Zhou (2015). Economic effects of runs on early "shadow banks": Trust companies and the impact of the panic of 1907. *Journal of Political Economy 123*(4), 902 – 940.

Galasso, A. and T. S. Simcoe (2011). Ceo overconfidence and innovation. *Management Science 57*(8), 1469–1484.

Goel, A. M. and A. V. Thakor (2008). Overconfidence, ceo selection, and corporate governance. *The Journal of Finance 63*(6), 2737–2784.

Hall, B. H., A. B. Jaffe, and M. Trajtenberg (2005). Market value and patent citations. *The RAND Journal of Economics 36*(1), pp. 16–38.

Hall, B. H. and J. Lerner (2010). Chapter 14 - the financing of r&d and innovation. In B. H. Hall and N. Rosenberg (Eds.), *Handbook of The Economics of Innovation, Vol. 1*, Volume 1 of *Handbook of the Economics of Innovation*, pp. 609 – 639. North-Holland.

Harhoff, D., F. Narin, F. M. Scherer, and K. Vopel (1999). Citation frequency and the value of patented inventions. *The Review of Economics and Statistics 81*(3), 511–515.

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

Hochberg, Y. V., C. J. Serrano, and R. H. Ziedonis (2016). Patent collateral, investor commitment, and the market for venture lending.

Howell, S. T. (2017, April). Financing innovation: Evidence from r&d grants. *American Economic Review 107*(4), 1136–64.

Johnson, M. A. and G. M. Maggiora (1990). *Concepts and applications of molecular similarity.* Wiley.

Jones, B. F. (2010). Age and great invention. *The Review of Economics and Statistics 92*(1), 1–14.

Kaplan, S. N. and L. Zingales (1997). Do investment-cash flow sensitivities provide useful measures of financing constraints? *The Quarterly Journal of Economics 112*(1), 169–215.

Kerr, W. R. and R. Nanda (2015). Financing innovation. *Annual Review of Financial Economics 7*(1), 445–462.

Kogan, L., D. Papanikolaou, A. Seru, and N. Stoffman (2017). Technological innovation, resource allocation, and growth*. *The Quarterly Journal of Economics 132*(2), 665–712.

Kyle, M. and H. L. Williams (2017, January). Is american health care uniquely inefficient? evidence from prescription drugs. Working Paper 23068, National Bureau of Economic Research.

Lanthier, M., K. L. Miller, C. Nardinelli, and J. Woodcock (2013). An improved approach to measuring drug innovation finds steady rates of first-in-class pharmaceuticals, 1987–2011. *Health Affairs 32*(8), 1433–1439.

Lichtenberg, F. (2013). The effect of pharmaceutical innovation on longevity: Patient level evidence from the 1996–2002 medical expenditure panel survey and linked mortality public-use files. *Forum for Health Economics and Policy 16*(1), 1–33.

Lichtenberg, F. R. and S. X. Sun (2007). The impact of medicare part d on prescription drug use by the elderly. *Health Affairs 26*(6), 1735–1744.

Light, D. W. and J. R. Lexchin (2012). Pharmaceutical research and development: what do we get for all that money? *BMJ 345*.

Lin, H. and D. Paravisini (2013). The effect of financing constraints on risk. *Review of Finance 17*(1), 229–259.

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

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

Malmandier, U. and G. Tate (2005). Ceo overconfidence and corporate investment. *The Journal of Finance 60*(6), 2661–2700.

Malmandier, U., G. Tate, and J. Yan (2011). Overconfidence and early-life experiences: The effect of managerial traits on corporate financial policies. *The Journal of Finance 66*(5), 1687–1733.

Mann, W. (2016). Creditor rights and innovation: Evidence from patent collateral.

Metrick, A. and S. Nicholson (2009). Do financial constraints have a real effect on pharmaceutical and biotech investment?

Moser, P., J. Ohmstedt, and P. Rhode (2011). Patents, citations, and inventive output - evidence from hybrid corn.

Myers, S. C. and N. S. Majluf (1984). Corporate financing and investment decisions when firms have information that investors do not have. *Journal of Financial Economics 13*(2), 187–221.

Naci, H., A. W. Carter, and E. Mossialos (2015). Why the drug development pipeline is not delivering better medicines. *BMJ 351*.

Nanda, R. and T. Nicholas (2014). Did bank distress stifle innovation during the great depression? *Journal of Financial Economics 114*(2), 273 – 292.

OHagan, S., N. Swainston, J. Handl, and D. B. Kell (2015). A 'rule of 0.5'for the metabolite-likeness of approved pharmaceutical drugs. *Metabolomics 11*(2), 323–339.

Pye, C. R., M. J. Bertin, R. S. Lokey, W. H. Gerwick, and R. G. Linington (2017). Retrospective analysis of natural products provides insights for future discovery trends. *Proceedings of the National Academy of Sciences 114*(22), 5601–5606.

Ralf Otto, Alberto Santagostino, and Ulf Schrader (2014, December). Rapid growth in biopharma: Challenges and opportunities. *McKinsey & Company*.

Smith, C. W. and R. M. Stulz (1985). The determinants of firms' hedging policies. *The Journal of Financial and Quantitative Analysis 20*(4), 391–405.

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

Tables and Figures

Figure 1: SIMILARITY FOR STATINS

Mevacor
(Similarity Score=0.25)

Pravachol
(Similarity Score=0.61)

Zocor
(Similarity Score=0.82)
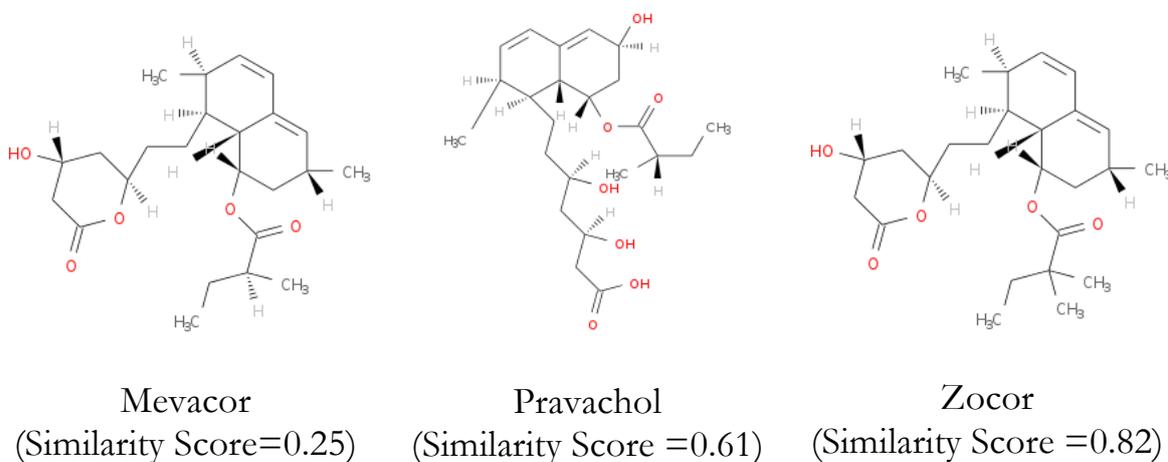
NOTES: Figure 1 provides the molecular structure and maximum similarity score of three early statins. Mevacor (Lovostatin) was the first FDA approved statin (approved in September 1987) and its Tanimoto similarity to prior molecules is 0.25. Pravachol (Pravastatin) is was the second such statin, approved in October 1991; its pair-wise similarity to Mevacor is 0.61 and its overall maximum similarity is also 0.61. Finally, Zocor (Simvastatin) was the third such statin, approved December 1991: its pair-wise similarity to Mevacor is 0.82 and its pairwise to Pravachol is 0.52. Zocor's overall maximum similarity to prior molecules is 0.82.

Figure 2: PROPORTION FIRST-IN-TARGET, BY DRUG SIMILARITY
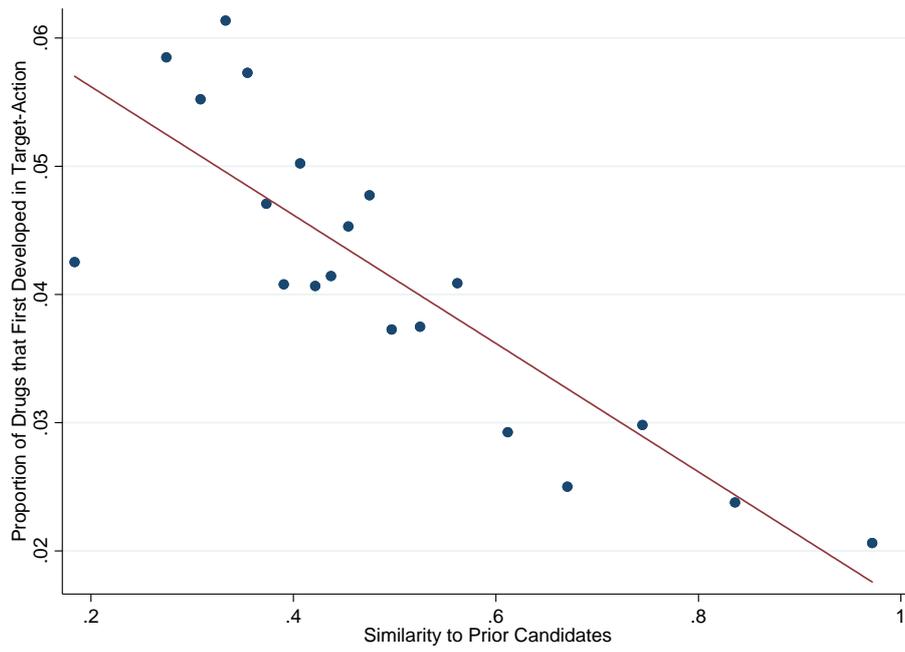
NOTES: Figure 2 presents a binned scatterplot of drug-level similarity against whether a drug is the first developed in its target-action. Each dot represents the proportion of candidates that are the first to be developed in their target-action, among all candidates within a given similarity score bin, conditional on disease (ICD9) and quarter of development fixed effects.

Figure 3: DRUG NOVELTY, DESCRIPTIVE STATISTICS

A. Distribution of Maximum Similarity to Prior Candidates
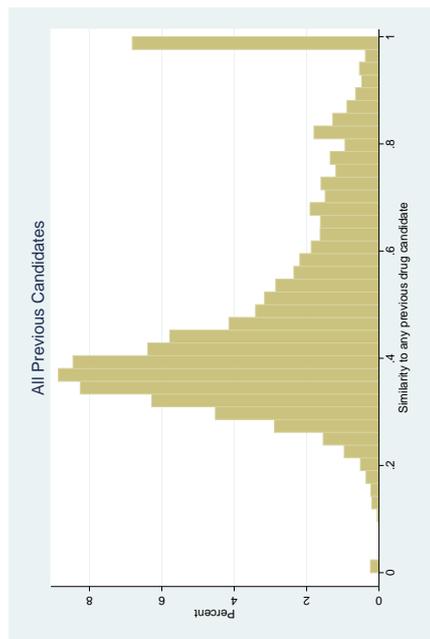
B. Average Similarity over time
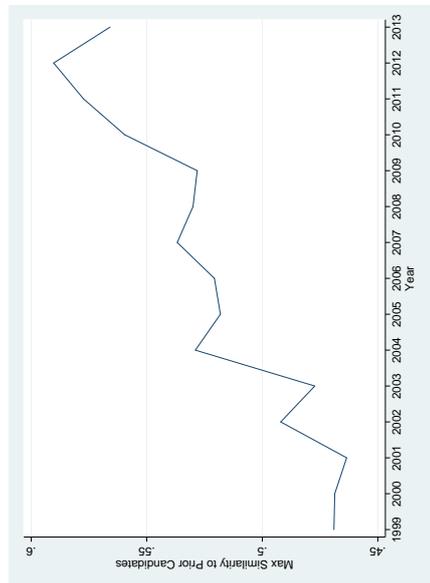
C. Proportion Very Similar Drugs
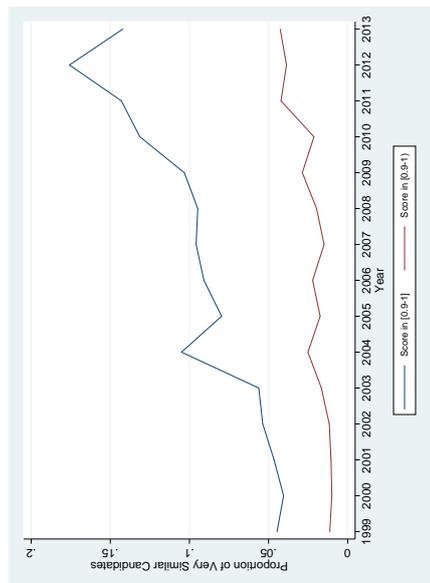
NOTES: Figure 3 displays descriptive statistics of our novelty measure. Panel A displays the distribution of our drug similarity measure. A drug's similarity is measured as its similarity to the most similar drug candidate that had previously entered Phase 1 clinical trials. For more details on this similarity measure, see Section 1.2. Panel B plots the trend in average drug candidate similarity over time. The line represents the average value of new drug candidates' maximum similarity to previously developed drugs, by year. Panel C displays the proportion of new drugs that are very similar. The blue line represents drugs with similarity scores greater than 0.9, which indicates over 90% overlapping chemical structures. The red line plots the same pattern, excluding drugs with similarity equal to one; this is to avoid counting combination therapies which may use the same molecule in conjunction with another molecule. Although our sample includes drug output in 2014, we plot up to 2013 in Panels B and C because our 2014 data do not include the entire year.

Figure 4: Drug Novelty and Risk, Social and Private Values
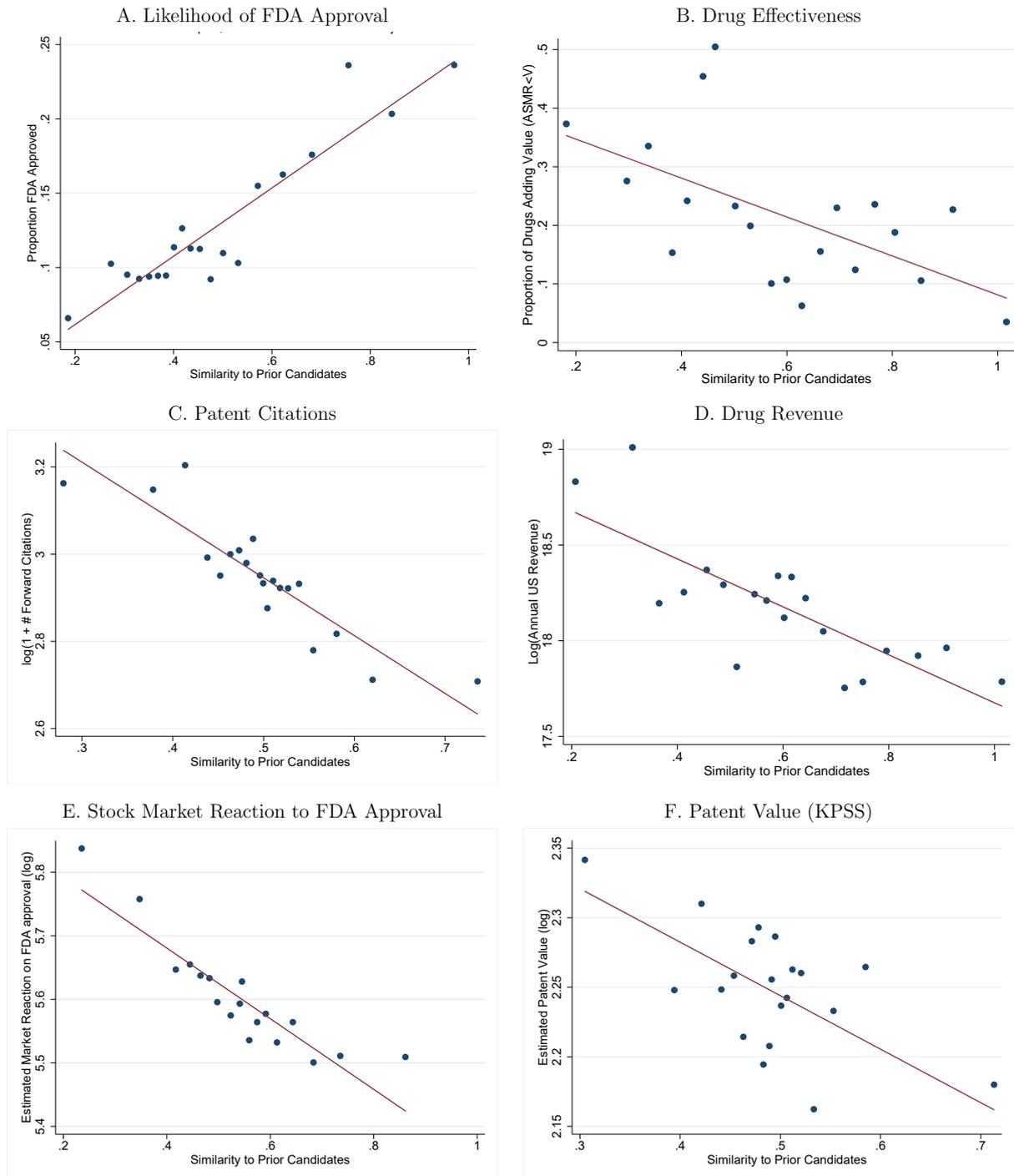
NOTES: Figure presents binned scatterplots of drug-level similarity against several drug characteristics. Panel A examines whether a drug is FDA approved. Panel B examines the drug's added benefit, which is derived from the French health system's clinical added benefits scores (ASMR), which range from one to five (I to V), with V indicating no value added. Panel C examines the logarithm of one plus the number of forward citations the patent receives. Panel D examines drug revenue. Panel E examines the logarithm of the estimated dollar reaction on the (first) approval of the drug by the FDA. Last, Panel F examines the logarithm of the **?** estimated patent values. See Notes to Appendix Figures A.2–A.7 for more details

Figure 5: IMPACT OF ADDITIONAL RESOURCES ON NOVELTY OF DRUG INVESTMENTS

(a) COEFFICIENTS

(b) ELASTICITIES
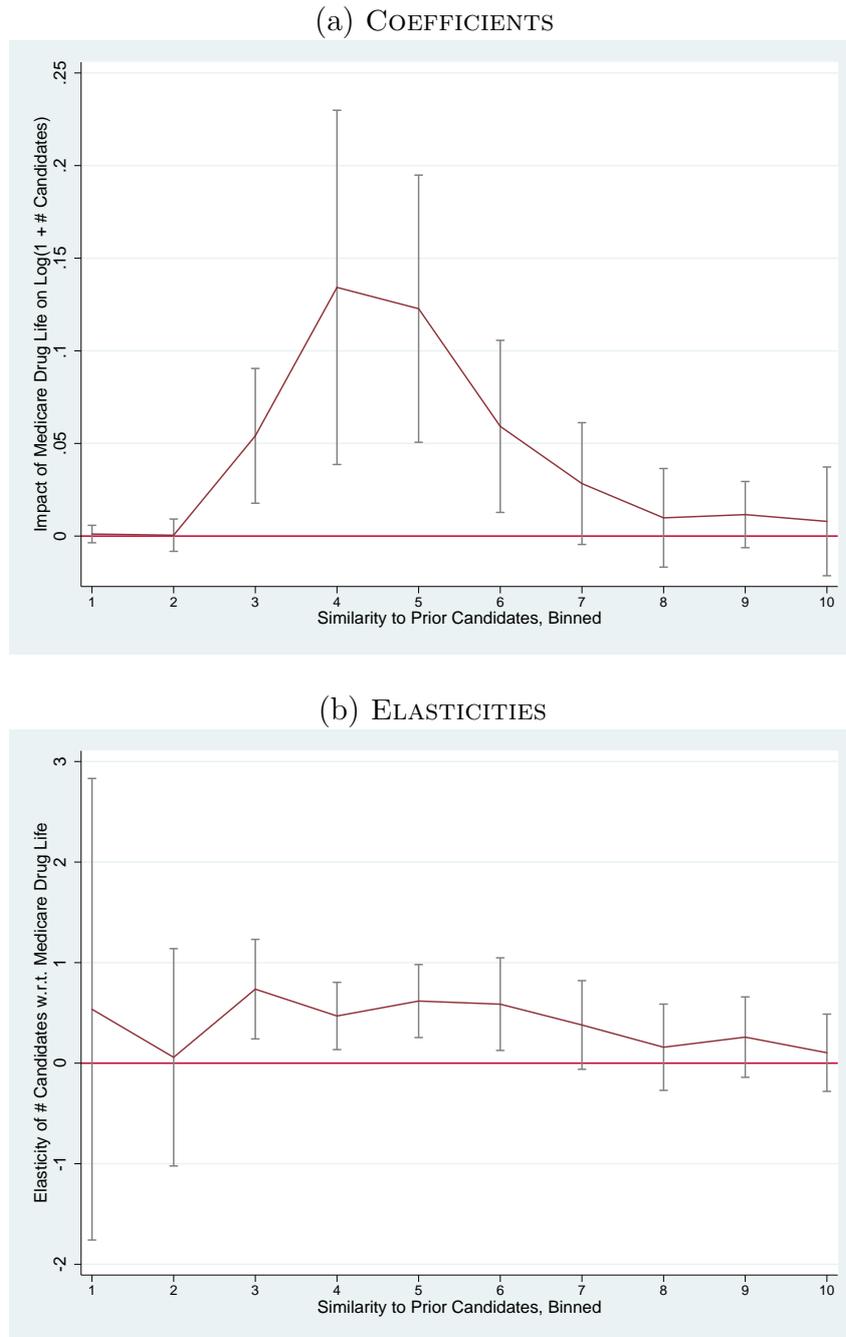
NOTES: Figure 5 plots the estimated coefficients on Post $\times$ Medicare Drug Life$_{f,2003}$ from our main regression specification defined by (6). Each point represents a different outcome variable: the number of new drug candidates in a given bin of similarity. Bins are specified by absolute similarity scores: Bin 1, for example, counts the impact of our treatment on the number of drugs with similarity score between 0 and 0.1, while Bin 10 is the impact on drugs with similarity between 0.9 and 1.0. The bottom figure reports the estimated elasticities for drugs in each novelty bin. We note that for a regression of the form $log(1 + y) = bx + e$, the elasticity is given by $b \times \frac{1+y}{xy}$. We evaluate these elasticities at the corresponding means of x and y.

Figure 6: Event Studies: # of New Candidates
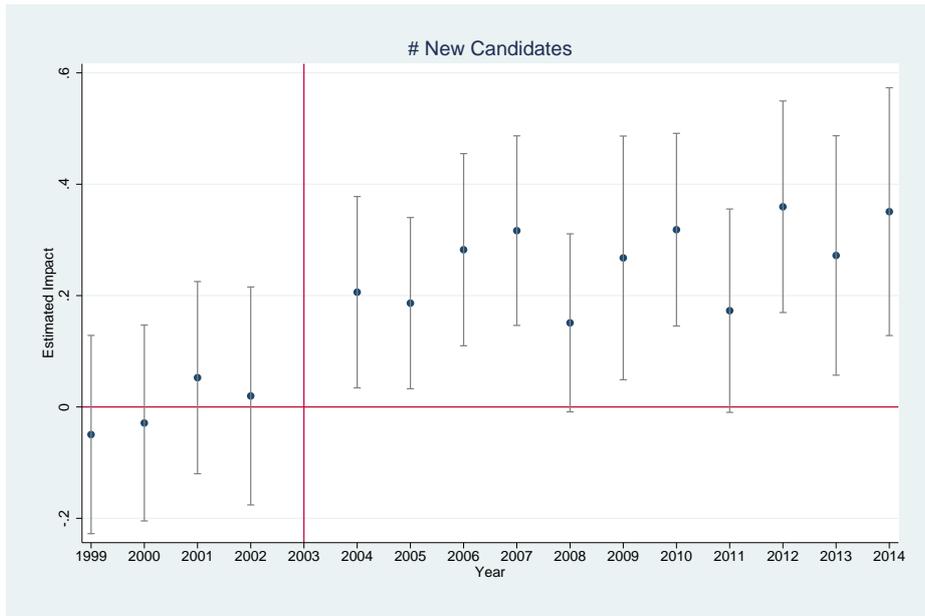
NOTES: Figure 6 reports the accompanying event study associated with Column 6 of Table 4. Each dot represents the coefficient on Medicare Drug Life$_{f,2003}$ interacted with an indicator variable for that given year. 2003 is the omitted year, and 90th percentile confidence intervals are reported.

Figure 7: Event Studies: # of New Candidates, by Quartiles of Similarity
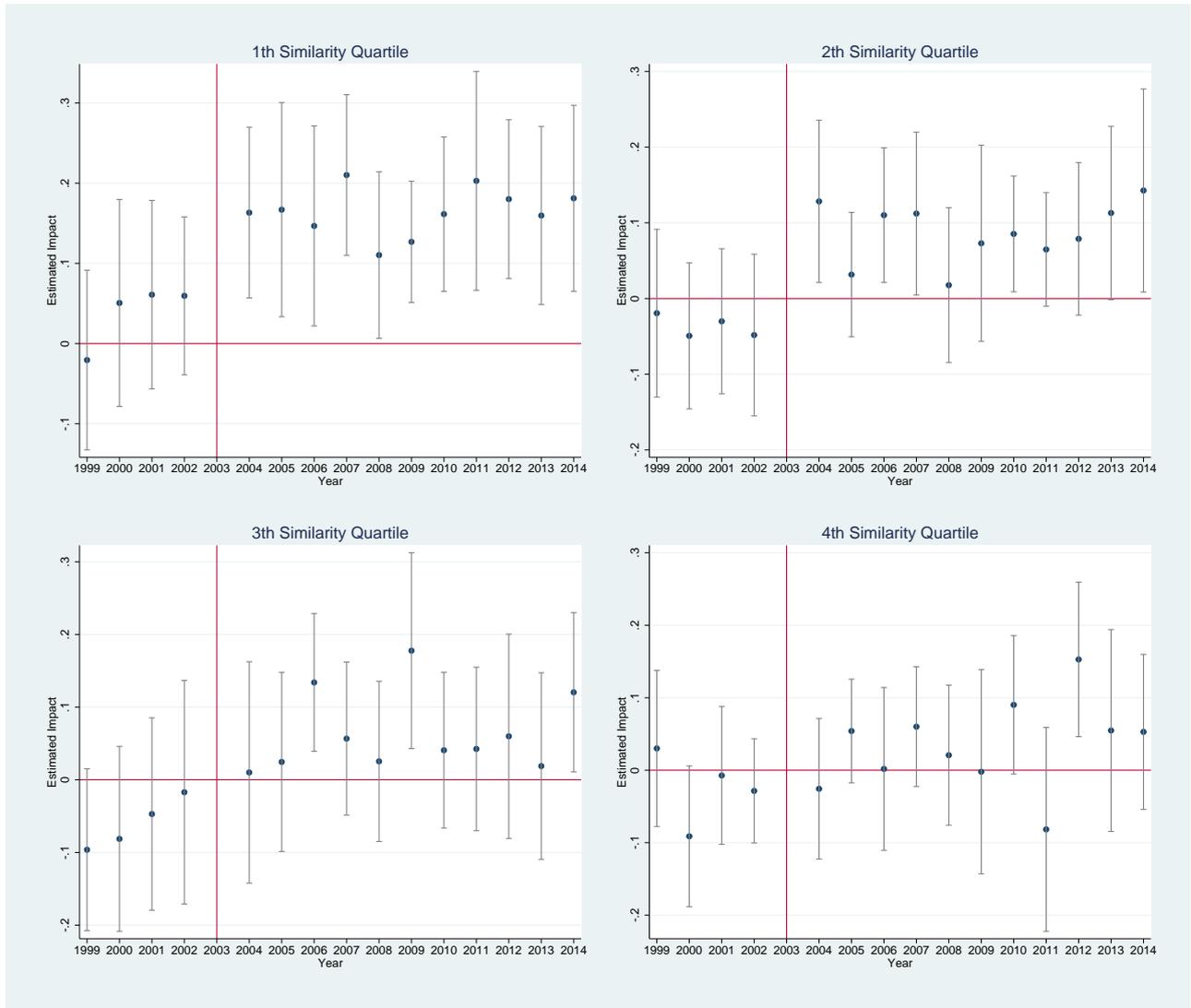
NOTES: Figure 7 reports event studies for number of novel drugs. In the interest of space, our outcome variables are the number of new candidates in different quartiles of similarity (rather than deciles as reported in Figure 5. Each dot represents the coefficient on Medicare Drug Life$_{f,2003}$ interacted with an indicator variable for that given year. 2003 is the omitted year, and 90th percentile confidence intervals are reported. The upper left panel presents the event study associated with the (log of one plus the) number of drugs in the bottom quartile of similarity (the most novel drugs); the upper right presents the 25th-50th percentile quartile; bottom left the 50th to 75th percentile and, finally, the bottom right figure presents the impact on the most similar drugs.

Table 1: DRUG CANDIDATES SUMMARY STATISTICS

	All Drug Candidates 1999-2014	All Drug Candidates, Sample Firms 1999-2014
Compound Characteristics		
# Compounds	12,191	6,374
# US Phase 1 or above	3,043	1,894
# US Phase 2 or above	2,251	1,443
# US Phase 3 or above	988	756
# FDA Approved	392	356
Maximum Similarity to Prior Compounds	0.53	0.50
% between 0 and 0.1	0.20	0.06
% between 0.1 and 0.2	0.66	0.31
% between 0.2 and 0.3	6.60	6.48
% between 0.3 and 0.4	29.70	34.77
% between 0.4 and 0.5	21.97	23.25
% between 0.5 and 0.6	10.57	10.06
% between 0.5 and 0.6	7.65	7.08
% between 0.7 and 0.8	6.20	5.65
% between 0.8 and 0.9	5.96	4.88
% between 0.9 and 1.0	10.48	7.47
Coverage Characteristics		
# Target-Actions	2,211	1,448
# Disease Categories	430	363

NOTES: Table 1 reports characteristics of our full sample of drug candidates versus the sample of candidates associated with firms for which we are able to compute Medicare exposure in 2003. See Section 5.1 for details about phases of drug approval in the United States. See Section 1.2 for details about how similarity is defined.

Table 2: Drug Novelty and Risk, Social and Private Values—Summary Table

	Risk	Measures of Social Value		Measures of Private Value		
	Likelihood of FDA Approval	Drug Effectiveness (ASMR < V)	Patent Citations	Revenue	Stock Reaction to FDA Approval	Patent Value
	(1)	(2)	(3)	(4)	(5)	(6)
Maximum Similarity	0.272***	-0.332**	-1.470***	-0.641***	-0.556***	-0.383**
	(0.024)	(0.099)	(0.141)	(0.281)	(0.064)	(0.128)
Observations	19,127	385	3,448	11,230	399	1,643
Appendix Table/Column	A.5.(2)	A.6.(4)	A.7.(4)	A.8.(4)	A.9.(3)	A.10.(4)

NOTES: Table 2 summarizes the relation between drug novelty and drug characteristics—specifically: risk (defined as the likelihood of FDA approval); proxies for social value (measured either using the AMR score, or the number of citations to related patents); and estimates of private value (measured either by drug revenues, the stock market reaction following a drug's FDA approval, or via the Kogan et al. (2017) measure of value for the associated patents). The last row indicates the Appendix Tables referenced in this summary table (along with the relevant columns). For brevity, we report the coefficients on novelty (along with standard errors) using the most conservative specification, which, whenever possible, control for disease (indication); drug age (drug launch or patent issue year); company; Please see the notes to the relevant Appendix Tables for more details. $^*p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 3: FIRM–QUARTER SUMMARY STATISTICS

	Mean	p10	p25	p50	p75	p90
Firm-Quarter Output						
# New Drug Candidates	0.55	0	0	0	0	2
. . . own	0.36	0	0	0	0	1
. . . acquired	0.19	0	0	0	0	1
Average Max Similiarity Score	0.53	0.31	0.37	0.48	0.66	0.85
Firm Characterisitics (2003)						
Medicare Drug Life	0.54	0	0	0.54	1	1
Firm MMS	0.35	0.12	0.20	0.32	0.49	0.65
Overall Drug Life	0.57	0	0	0.60	1	1

NOTES: Table 3 reports characteristics of our firm–quarter sample. A drug is considered a firm's own if it is assigned to that firm on the first date it enters development (as recorded in Cortellis); it is considered acquired if, on that date, it becomes associated with our focal firm even though it had previously been associated with another firm. Similarity is defined as the maximum similarity score, compared to all candidates that had previously entered development. We also compute distributions separately for prior candidates within the same indication or the same firm. Medicare drug life is the proportion of a firm's approved drugs in 2003 that had greater than 5 years of exclusivity left, weighted by the drug's Medicare Market Share (MMS). Firm MMS is the average MMS across that firm's approved drugs as of 2003. Overall drug life is the unweighted proportion of a firm's approved drugs in 2003 that had greater than 5 years of exclusivity left. Number of high patent life drugs is the total number of such drugs.

Table 4: IMPACT OF RESOURCES ON # NEW CANDIDATES

	# New Candidates			Log(1 + New Candidates)		
	(1)	(2)	(3)	(4)	(5)	(6)
Post 2003 X Medicare Drug Life	0.211**	0.860**	0.847**	0.057**	0.268***	0.263***
	(0.084)	(0.363)	(0.365)	(0.027)	(0.096)	(0.096)
Post 2003 X Overall Drug Life		-0.707*	-0.694*		-0.229**	-0.225**
		(0.366)	(0.368)		(0.098)	(0.098)
Post 2003 X Firm MMS			-0.153			-0.049
			(0.140)			(0.044)
R^2	0.556	0.556	0.557	0.594	0.595	0.595
Company FEs	Yes	Yes	Yes	Yes	Yes	Yes
Qtr of Development FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	16442	16442	16442	16442	16442	16442

NOTES: Table 4 examines the impact of additional resources on the number of new drug candidates. The dependent variable is the count of new drug candidates entering development (models 1-3), or the log of one plus the number of new drug candidates entering development. All models include a full set of company and quarter indicator variables to control for firm and calendar time fixed effects. Models 3 and 6 correspond to our main regression specification in defined by (6), with Post × Overall Drug Life$_{f,2003}$ and Post × Firm MMS$_{f,2003}$ both included as independent variables. Robust standard errors in parentheses, clustered around company identifiers. $^*p < 0.10,$$^{**}p < 0.05,$$^{***}p < 0.01$.

Table 5: PROPORTION OF NEW DRUGS ACROSS MMS QUARTILES

	Log(1+ New Candidates), by MMS Quartile			
	(1)	(2)	(3)	(4)
Post 2003 X Medicare Drug Life	0.085**	0.084**	0.110**	0.115**
	(0.041)	(0.042)	(0.043)	(0.046)
R^2	0.337	0.343	0.366	0.358
Company FEs	Yes	Yes	Yes	Yes
Qtr of Development FEs	Yes	Yes	Yes	Yes
Overall Drug Life/Firm MMS	Yes	Yes	Yes	Yes
Observations	16442	16442	16442	16442

NOTES: Table 5 examines whether firms developing more drugs in response to cashflow shocks do so in areas that experience a greater increase in demand (proxied by MMS Share). The table reports the main specification coefficient for Post $\times$ Medicare Drug Life$_{f,2003}$. The dependent variable in each column corresponds to each quartile of the Medicare market share (MMS) distribution. All models include a full set of company and quarter indicator variables, with Post $\times$ Overall Drug Life$_{f,2003}$ and Post $\times$ Firm MMS$_{f,2003}$ both included as additional independent variables, but not reported in the table. Robust standard errors in parentheses, clustered around company identifiers. $^*p < 0.10,^{**}p < 0.05,^{***}p < 0.01$.

Table 6: PORTFOLIO EXPANSION (CANDIDATES NEW TO FIRM)

	New Indications		New Targets	
	(1)	(2)	(3)	(4)
	Log(1+ #)	Δ HHI	Log(1+ #)	Δ HHI
Post 2003 X Medicare Drug Life	0.160**	-0.013*	0.101*	-0.020***
	(0.069)	(0.008)	(0.060)	(0.007)
R^2	0.260	0.029	0.440	0.025
Company FEs	Yes	Yes	Yes	Yes
Qtr of Development FEs	Yes	Yes	Yes	Yes
Overall Drug Life/Firm MMS	Yes	Yes	Yes	Yes
Observations	16442	12220	16442	12220

NOTES: Table 6 examines whether firms choose to diversify their drug portfolio, by pursuing candidates that are sufficiently different that their existing portfolio. We report the main specification coefficient for Post × Medicare Drug Life$_{f,2003}$. All models include a full set of company and quarter indicator variables, with Post × Overall Drug Life$_{f,2003}$ and Post × Firm MMS$_{f,2003}$ both included as additional independent variables, but not reported in the table. The first model reports the main effect of the Medicare Part D shock on the number of new (to the firm) indications entered. The second model reports how the introduction of Part D impacted the change in firm project concentration, as measured by a Herfindahl-Hirschman index of projects by therapeutic indication. The dependent variables in the third and fourth models are number of new drug targets, and the change in project concentration across drug targets, respectively. Robust standard errors in parentheses, clustered around company identifiers. $^*p < 0.10,^{**} p < 0.05,^{***} p < 0.01$.

Table 7: EFFECT OF CASHLOWS ON NUMBER OF DRUG CANDIDATES, BY FIRM SIZE

	Log(1 + New Candidates), by Size		
	(1)	(2)	(3)
	All	Top 50	Bottom 50
Post 2003 X Medicare	0.263***	0.299	0.192*
Drug Life	(0.096)	(0.214)	(0.100)
R^2	0.595	0.641	0.209
Company FEs	Yes	Yes	Yes
Qtr of Development FEs	Yes	Yes	Yes
Overall Drug Life/Firm MMS	Yes	Yes	Yes
Observations	16442	5950	6207

NOTES: Table 7 examines how the treatment effect of cashflows on number of developed drugs varies by firm size. We measure firm size as the sum of revenue generated by approved drugs prior to 2003. We split firms into equal sized groups based on their size as of 2003; the number of observations differs due to firm exit. We report the main specification coefficient for Post × Medicare Drug Life$_{f,2003}$. We estimate the model with the full set of company and quarter indicator variables, including Post × Overall Drug Life$_{f,2003}$ and Post × Firm MMS$_{f,2003}$, separately across groups. All control variables are allowed to vary across specifications, but are not reported in the table. Robust standard errors in parentheses, clustered around company identifiers. $^*p < 0.10,^{**}p < 0.05,^{***}p < 0.01$.

Table 8: Impact on R&D and Profits

A. Full Sample

	(1) Log(RD)	(2) Log(Profits)	(3) Log(Debt)	(4) Leverage
Post 2003 X Medicare Drug Life	0.975*	1.046*	0.967	0.108
	(0.573)	(0.564)	(1.118)	(0.108)
R^2	0.934	0.930	0.800	0.463
Company FEs	Yes	Yes	Yes	Yes
Year of Development FEs	Yes	Yes	Yes	Yes
Observations	1774	1572	1657	1925

B. Excluding the Financial Crisis (pre-2008)

	(1) Log(RD)	(2) Log(Profits)	(3) Log(Debt)	(4) Leverage
Post 2003 X Medicare Drug Life	1.098*	1.189**	1.281	0.103
	(0.559)	(0.513)	(1.005)	(0.087)
R^2	0.949	0.937	0.800	0.582
Company FEs	Yes	Yes	Yes	Yes
Year of Development FEs	Yes	Yes	Yes	Yes
Observations	1110	978	1067	1218

NOTES: Table 8 examines the response of firm-level research spending, operating cashflow, and debt to our main treatment variable, Post $\times$ Medicare Drug Life$_{f,2003}$. The dependent variable is either the logarithm of R&D spending; the logarithm of operating cashflows (Compustat: ib + dp); the logarithm of long-term debt (Compustat: dltt); and the logarithm of leverage (Compustat: dltt scaled by at). Panel A examines the full sample (years 1999–2013), while panel B only includes the years 1999–2008 in order to exclude the effects of the financial crisis. All specifications include a full set of company and quarter indicator variables, with Post $\times$ Overall Drug Life$_{f,2003}$ and Post $\times$ Firm MMS$_{f,2003}$ both included as additional independent variables, but not reported in the table. Robust standard errors in parentheses, clustered around company identifiers.*$p < 0.10$,** $p < 0.05$,*** $p < 0.01$.

Online Appendix to "Developing Novel Drugs"

Joshua Krieger[19] Danielle Li[20]Dimitris Papanikolaou[21]

November 21, 2017

---

[19]Harvard Business School, jkrieger@hbs.edu
[20]MIT Sloan and NBER, d_li@mit.edu
[21]Kellogg School of Management and NBER, d-papanikolaou@kellogg.northwestern.edu

5 A Primer on Drug Development

Here, we provide a brief description of the process of drug development by pharmaceutical firms, while also emphasizing the potential role of financial constraints.

5.1 Development Process

The drug development process is typically divided into five stages: discovery and pre-clinical research, and Phase 1, 2, and 3 of human clinical trials. From start to end, this process may take anywhere from 5 to 15 years. In the first stage of this process, discovery, researchers identify biological mechanisms that impact diseases and symptoms. For example, they may want to develop a drug that inhibits the functioning of a particular target, such as an enzyme or the gene that encodes it, This becomes the biological "target" of the drug. Having identified a potential target, developers then screen potential compounds looking for structures that have some desired action on this target. At some point during this first stage of development, firms will apply for patents on promising candidates.[22]

Having identified a set of promising compounds, researchers focus next on testing its pharmacokinectic and pharmacodynamic properties: how the body impacts the drug (its absorption, bioavailability, etc.) and how the drug impacts the body (drug actions, toxicity, etc.), respectively. If a drug performs well in animal models, firms may choose to file an Investigational New Drug (IND) application with the FDA to begin human clinical trials. Clinical trials have three phases. Phase 1 clinical trials mainly test for toxicity and help set dosage levels, using a few dozen healthy patients. Phase 2 trials involve hundreds of patients with the conditions of interest, and are typically randomized. Phase 3 trials are randomized controlled trials on a focused subset of patients likely to show the greatest response to the drug. These trials often include thousands of patients and involve tracking outcomes over long periods to assess both safety and efficacy. At the end of Phase 3, firms may submit a New Drug Application (NDA) to the FDA that includes the results of all trials and preclinical testing. After a formal review process, the FDA decides whether or not to approve the drug.

Throughout the development process, firms make many decisions about what types of compounds to invest in. These decisions are important for the ultimate novelty of drugs that are brought to market. For instance, firms may choose to develop drug candidates that act on known targets through known channels, or they can attempt to develop drugs that differ in either their mode of action.

---

[22]Firms typically apply for broad patents that would cover a collection of similar compounds, rather than a single compound itself. This set of claims is described by a "Markush structure," which is a generalized molecular structure used to indicate a collection of similar compounds.

One aspect of drug pipeline decisions that has attracted a lot of attention is the issue of "me–too" innovation. The idea behind "me–too" or "copycat" drugs is that firms prefer to modify existing drugs or create similar compounds in order to avoid the costs and uncertainty of more novel drug development. Developing such drugs has the benefit of providing doctors with a menu of valuable alternatives if a patient is not responding or having an adverse reaction to a specific drug. For example, Berndt, Cockburn, and Grépin (2006) find that drugs that gained supplemental approvals for new dosages, formulations and indications account for a large portion of drug utilization and economic benefits. A common critique of these type of drugs, however, is that they yield only marginal clinical improvements while increasing drug costs and diverting resources from the development of truly innovative therapies. For example, Joseph Ross, a professor of medicine and public health at Yale University School of Medicine, describes me-too drugs as those that "may have some unique niche in the market, but they are fairly redundant with other therapies that are already available" (New York Times, 2015). It is also worth noting that two similar drugs that are both brought to market may have been developed in parallel ("racing") rather than through a scenario in which one drug imitated the other in order to capture a piece of the same, or similar, pie (DiMasi and Chakravarthy, 2016).

5.2   Development Costs and Financing

Drug development is financed through a number of different mechanisms, both public and private. First, an important input into drug development is the scientific knowledge that enables researchers to identify biological targets, and which enables the development of tools and techniques used in drug discovery. This type of "basic" research is usually funded by the government, most often through the National Institutes of Health. Translational research, in which insights from basic research are advanced toward medical applications and commercialization, may also involve public funding. For example, early stage biotechnology firms working on a proof-of-concept for a new type of drug may receive capital from the government's Small Business Innovation Research (SBIR) program, as well as from private foundations and venture capital. In general, however, the direct public funding of private–sector drug development is limited.

The direct cost of drug discovery to firms themselves is substantial: DiMasi et al. (2016) estimate that the direct cost of developing a single approved drug is over $1.4 billion and has been increasing over time.[23] This total cost of development is spread unevenly across

---

[23]This estimate is subject to some debate. See for instance, `http://www.nytimes.com/2014/11/19/upshot/calculating-the-real-costs-of-developing-a-new-drug.html`.

the stages of drug development, with discovery and preclinical costs accounting for one third and clinical costs accounting for the remaining two thirds. Phase 3 trials, in particular, can be extremely costly and involve multiple thousands of patients over several years. Because of this escalating cost structure, investments in drug development are essentially staged, with firms putting in smaller amounts of money in early stages and making greater capital commitments only if the drug shows promise.

One possible reason why firms, especially smaller ones, may not choose to invest in novel drugs is because these drugs may be more costly to develop. In general, assessing the costs of development is difficult because we do not have access to internal investment data and, furthermore, a large part of R&D spending is on scientific staff, who then work on multiple projects. A noisy proxy for development costs, however, are the number of patients enrolled in clinical trials and the number of trials associated with drugs: because trials are so expensive, recruiting patients and running trials constitutes a substantial proportion of a drug's development cost. In Table A.1 and Figure A.10, we consider how the number of patients and number of trials associated to a compound vary by its chemical novelty. In general, we find no consistent relationship between these proxies of development cost and drug novelty. The left hand side panels of Figure A.10, for instance, plot bin scatters of the relationship between drug novelty and number of patients or trials for our full set of drug candidates. We find no relationship between novelty and the number of enrolled patients across all trials. We find a weakly negative relationship between similarity and the total number of trials; however, these appear to be driven by the set of drug candidates with similarity scores of exactly 1, which may include extended release formulations that should require fewer additional trials. Restricting to the set of candidates with similarity strictly less than 1, we find, if anything, that more similar drugs are more expensive. Further, to the extent that novel drugs are less likely to survive to later stages, our evidence suggests that their initial expected cost is likely to be weakly lower.

Accessing external finance for such costly and uncertain projects can be particularly challenging. In general, the pecking order theory (Myers and Majluf, 1984) predicts that using internal funds is the cheapest form of financing, followed by debt and then equity. By now, a broadly accepted view in corporate finance is that information asymmetries and moral hazard frictions make it particularly costly for both public and private firms to raise external equity finance. For several reasons, these frictions may be particularly salient for innovative firms (see, e.g. Kerr and Nanda, 2015; Hall and Lerner, 2010, for a review of the literature).

Financing drug development with debt is also difficult because few pharmaceutical firms have assets that can be reliably used as collateral. Patents for drug candidates, for instance,

are taken out early in the development process, making their use as collateral something of a Catch 22—in order to know whether the patent is valuable as collateral, a bank would have to lend the firm the money to put it through testing, which is what the firm wanted the loan for in the first place.[24] Consistent with this view, firms in the pharmaceutical industry have indeed lower leverage ratios than comparable firms in other industries (see Appendix Table A.2 for more details).

6 Data Construction

Here, we describe the construction of the data in more detail.

6.1 Drug Development Histories

Our drug development data primarily comes from the Cortellis Investigational Drugs and Clinical Trials databases.[25] For drugs in the Cortellis data, we have information on characteristics, as well as associated companies and clinical trials. Most notably, Cortellis uses information from patents, regulatory filings, press releases, public press and company materials (e.g., pipeline "tables" and company website) to derive key dates for each drug's development history by company, therapuetic indication and country. For example, Cortellis might list an earliest "discovery" date based on the scientific publication or patent that describes a drug candidate's use for a particular disease, followed by dates corresponding to the start of clinical trials of each phase, and finally an approval or market launch date.

In our various analyses, we distinguish between a drug-indication's earliest development date with any company, its first development milestone with a non-originating company that acquired the drug, and the drug candidate's entry dates into phase I/II/III clinical trials. We calculate our primary drug novelty measures by taking the maximuma new drug candidate's chemical structure similarity (at the time of earliest entry) to all prior drug candidates that ever reached phase I clinical trials. While we also tested alternative definitions of novelty that compare new drugs to all prior developed drug candidates of any stage, we prefer to compare to the phase I drugs because doing so reduces the likelihood of comparing a new

---

[24]A growing set of papers have shown, however, that pharmaceutical patents are sometimes pledged as collateral by public firms, although this phenomenon is small compared to the use of patents in electronics or medical devices (Mann, 2016). Hochberg, Serrano, and Ziedonis (2016) conduct a similar analysis examining the use of debt in venture financing; their study includes some medical devices firms but few if any biopharmaceutical firms.

[25]At the time of our data access agreement, Cortellis was owned by Thomson Reuters. In October 2016, Thomson Reuters sold Cortellis to Clarivate Analytics.

drug candidate to another compound that was developed independently and simultaneously, but by chance was disclosed (or captured by Cortellis) at a slightly earlier date.

6.2 Chemical Similarity Scores

Section (1.2) in the paper provides a basic summary of our method for calculating drug similarity scores. This section provides more details on the mechanics of gathering pairwise similarity scores, and then calculating our novelty measures. The starting point for these scores is information on the drug candidate's chemical structure. Cortellis contains information about the chemical structure of small molecule drugs, when that information is available. Chemical structure information is not available for vaccines and biologic drugs, which involve more complex mixtures of substances generated through biotechnology. Often, the chemical structure is also not available for drugs that never progress out of very early stage drug development stages. Roughly 36% of Cortellis drug entries contain information on drug structure. This percentage is higher for small molecule drugs (53%), and for small molecule drugs that reach clinical trials (70%). When the chemical structure is known, Cortellis provides standardized chemical identifiers such as the simplified molecular-input line-entry system (SMILES). SMILES codes represent chemical structures as ASCII strings, with components of the string identifying atoms, bonds, branching, order and shape of a compound. These SMILES strings serve as the inputs to our similarity calculations.

In practice, calculating Tanimoto distance requires an algorithm that can convert a chemical identifier like a SMILES string into its component fragments and compare to other compounds. This process is both complex and computationally intensive. We used features of ChemMine Tools (publicly available at `http://chemmine.ucr.edu/`) a system developed by chemical informatics researchers at the University of California, Riverside (Backman, Cao, and Girke, 2011) in order to process and calculate pairwise Tanimoto scores. We used the R package version of ChemMine (ChemmineR) to batch submit similarity calculation requests for the unique SMILES codes represented in our drug development data from Cortellis. For data management purposes, we only kept pairwise similarity score results for pairs of compounds that had a Tanimoto distance greater than or equal to 0.1.

After generating all the pairwise similarity score data, we merge in the key development dates (e.g., earliest, phase I/II/III) for each drug, and calculate our novelty measures by drug candidate, as of the drug candidate's earliest development date, and based on the maximum similarity score to all previously developed drugs, all drugs that previously reached phase I, all drugs that previously reached phase I etc..

6.3 Drug Patents

In order to build our firm-level measure of drug patent life, we start by gathering patent expiration and market exclusivity information for drugs that had been approved prior to the passage of Medicare Part D in 2003. To maximize our drug patent life coverage, we combine multiple data sources. As a starting point, we use information from the Federal Register on the key patents for approved drugs, along with the patents' expiration dates and market exclusivity extensions. Extensions are usually the result of FDA rules that grant additional exclusivity after marketing approval for new chemical entities, pediatric drugs, antibiotics, and orphan drugs.[26] When we could not match an approved drug to the Federal Register data, we used the patent expiration dates of the drugs' affliated "Orange Book" patents listed by the FDA.[27]

After identifying exclusivity periods for approved drugs, we use drug names to merge this information into our Cortellis drug data. We first match on exact names, then use a "fuzzy" match technique to identify potential additional matches and reviewed that set manually. Once merged to Cortellis entries, we can aggregate remaining exclusivity into a firm-level measure of drug patent life as of 2003.

6.4 Matching Drugs to MEPS

An important data step for our analyses is matching our drug development history and novelty data with the Medical Expenditure Panel Survey (MEPS). The MEPS program is run by the Agency for Healthcare Research and Quality at the U.S. Department of Health & Human Services, and tracks data on health services use and cost for a large nationally representative sample of households. For 2003, the year congress approved Medicare Part D, the MEPS consolidated data file includes 11,929 household identifiers.

Our matching process (described below) serves two purposes: 1) to estimate drug-specific Medicare market share ("elderly share"), and 2) to estimate relative drug revenues. We aggregate the former up to the firm-level to calculate one of the two components of our main "treatment" variable (Medicare drug Life, see Section 3.1), and the latter helps us describe the correlation between our novelty measure and private value to drug developers (see Section ??).

To match our drug development and novelty data to the MEPS data, we use all the drug names affiliated with Cortellis drug identifiers, and merge them with drug names represented

---

[26]We thank Duncan Gilchrist for sharing this Federal Registrar data.

[27]The Orange Book covers all FDA approved drugs; however, a key limitation of Orange Book patents is that they are designated by the producing firm and are subject to patent challenges.

in MEPs. After finding all the perfect name matches, we manually inspect any potential matches using a "fuzzy" name matching algorithm. Matching drug names from the MEPS prescription data to Cortellis can also be challenging due to inconsistencies in the naming of drugs. For example, a common antibiotic prescription may be listed as "Zithromax ," "Zithromax Z-Pak," or "Zithromax 250 Z-PAK."

If a drug is not matched in the 2003 MEPS data, we attempt to match it to observations in the 2002 survey; 2001 if that is also not available, and so forth. For drugs we are unable to match, we infer the drug's MMS using information on MMS for the other drugs in MEPS that share the same therapeutic indications. Therapeutic level MMS is computed in MEPS by taking the average share of revenues coming from elderly patients for all approved drugs in a particular ICD9 class in the year 2003. For example, if a drug is used to treat two different conditions, we assign that drug the average of the Medicare shares associated with each of these conditions, weighted by the relative importance of the conditions. The weights assigned to ICD9s are the share of total revenue in the 2003 MEPS data that come from drugs associated with that ICD9.

For drug revenue, we use all the years in our MEPS data (1996–2012) and adjust dollar expenditures to 2015 dollars using the Consumer Price Index for All Urban Consumers (CPI-U). After matching to the Cortellis drug development data, we then estimate the correlations between our drug novelty measure and annual drug revenue, controlling for sales year, the drug's approval year, and therapeutic area (see Section 2.3).

6.5 Matching Drugs to Companies

One of the challenges in studying drug development pipelines is assigning drug candidates to their developer firms in a given point in time. The reason for this issue is that multiple firms may be connected with a single drug development project. Firms may team up to develop a drug through joint ventures, financing partnerships, or web of licensing and subsidiary arrangements. Ideally, one would assign ownership weights for a given drug (e.g., Firm A owns 30% and Firm B owns 70%). But due to complicated licensing and royalty arrangements, the outside analyst cannot easily infer such weights.

As a result, we are left with two distinct options: a) allow a single drug candidate to count as as a (full or equal weighted) member of multiple firms' portfolios, or b) determine which company is likely the central company in the development alliance, and assign that firm as the sole "lead" developer. We use the former method—allowing multiple firms to get credit for a single drug candidate or approved therapy. But when possible, we limit the set of

assigned companies to those that were most recently "active" with the drug in the Cortellis data.

6.6 Public Firms

A number of our analyses require data on public firms in our drug development data. To identify public companies in the Cortellis drug development data, we started by running all Cortellis company names through Bureau Van Dijk's Orbis software, which matches strings to company identifiers (including ticker and cusip CUSIP identifiers for publicly traded firms). To ensure that the Orbis process did not miss any notable public firms, we checked the match against historical lists of public pharmaceutical firms (e.g., Nasdaq and Standard & Poor's pharmaceutical indices) to make sure we had positively matched major firms. In total, we match over 600 tickers to Cortellis company identifiers. When we limit to publicly traded firms in our main analysis sample of 17,775 small molecule drugs, we are left with 140 public firms. While this may seem like a small number given that we have over 3,585 distinct company identifiers linked to drugs in the sample, we also see that these 140 public firms are responsible for more than half of the drug development activity in the sample. After linking to public company identifiers (tickers and CUSIPS), we are able to download daily stock data from The Center for Research in Security Prices (CRSP), as well as historical profits and R&D spending from Compustat. Out of these firms, approximately 71 are in the United States and are publicly traded at some point (appear in CRSP). When estimating the market reaction to an FDA approval, we further restrict the set to firms that were publicly traded at the time of the drug's first approval, we have 462 first-time approvals from 35 unique firms.

References

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

Almeida, H., M. Campello, B. Laranjeira, and S. Weisbenner (2011). Corporate Debt Maturity and the Real Effects of the 2007 Credit Crisis. *Critical Finance Review 1*, 3–58.

Azoulay, P., J. Graff-Zivin, D. Li, and B. Sampat (2016). Public r&d investments and private-sector patenting: Evidence from nih funding rules.

Backman, T. W. H., Y. Cao, and T. Girke (2011, 05). Chemmine tools: an online service for analyzing and clustering small molecules. *Nucleic Acids Research 39*(suppl_2), W486–W491.

Baldi, P. and R. Nasr (2010, 07). When is chemical similarity significant? the statistical distribution of chemical similarity scores and its extreme values. *Journal of Chemical Information and Modeling 50*(7), 1205–1222.

Benmelech, E., N. K. Bergman, and A. Seru (2011, June). Financing labor. Working Paper 17144, National Bureau of Economic Research.

Benmelech, E., C. Frydman, and D. Papanikolaou (2017, March). Financial frictions and employment during the great depression. Working Paper 23216, National Bureau of Economic Research.

Bernardo, A. E. and I. Welch (2001). On the evolution of overconfidence and entrepreneurs. *Journal of Economics & Management Strategy 10*(3), 301–330.

Berndt, E. R., I. M. Cockburn, and K. A. Grépin (2006). The impact of incremental innovation in biopharmaceuticals. *PharmacoEconomics 24*(2), 69–86.

Berndt, E. R., D. Nass, M. Kleinrock, and M. Aitken (2015). Decline in economic returns from new drugs raises questions about sustaining innovations. *Health Affairs 34*(2), 245–252.

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

Bloom, N., C. I. Jones, J. V. Reenen, and M. Webb (2017). Are ideas getting harder to find? working paper, Stanford University.

Blume-Kohout, M. E. and N. Sood (2013). Market size and innovation: Effects of medicare part d on pharmaceutical research and development. *Journal of Public Economics 97*, 327 – 336.

Bond, S., D. Harhoff, and J. van Reenen (2005). Investment, r&d and financial constraints in britain and germany. *Annals of Economics and Statistics* (79-80), 433–460.

Brown, J. R., S. M. Fazzari, and B. C. Petersen (2009). Financing innovation and growth: Cash flow, external equity, and the 1990s r&d boom. *The Journal of Finance 64*(1), 151–185.

Budish, E., B. N. Roin, and H. Williams (2015, July). Do firms underinvest in long-term research? evidence from cancer clinical trials. *American Economic Review 105*(7), 2044–85.

Chodorow-Reich, G. (2014). The employment effects of credit market disruptions: Firm-level evidence from the 2008-09 financial crisis. *Quarterly Journal of Economics 129*(1), 1–59. Lead article.

DeBondt, W. F. M. and R. H. Thaler (1995). *Handbooks in Operations Research and Management Science*, Volume 9, Chapter Financial Decision-Making in Markets and Firms: A Behavioral Perspective. Elsevier Press.

DiMasi, J. and R. Chakravarthy (2016). Competitive development in pharmacologic classes: Market entry and the timing of development. *Clinical Pharmacology and Therapeutics 100*(6), 754–760.

DiMasi, J. and C. Paquette (2004). The economics of follow-on drug research and development: Trends in entry rates and the timing of development. *PharmacoEconomics 22*(2), 1–14.

DiMasi, J. A. and L. B. Faden (2011, 01). Competitiveness in follow-on drug r&d: a race or imitation? *Nat Rev Drug Discov 10*(1), 23–27.

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

Dranove, D., C. Garthwaite, and M. Hermosilla (2014). Breakthrough or me-too? the impact of medicare part d on biotech innovation.

Duggan, M. and F. Scott Morton (2010, March). The effect of medicare part d on pharmaceutical prices and utilization. *American Economic Review 100*(1), 590–607.

Duygan-Bump, B., A. Levkov, and J. Montoriol-Garriga (2015). Financing constraints and unemployment: Evidence from the Great Recession. *Journal of Monetary Economics 75*(C), 89–105.

Farre-Mensa, J. and A. Ljungqvist (2016). Do measures of financial constraints measure financial constraints? *The Review of Financial Studies 29*(2), 271–308.

Friedman, J. (2009). The incidence of the medicare prescription drug benefit: Using asset prices to assess its impact on drug makers.

Froot, K. A., D. S. Scharfstein, and J. C. Stein (1993, December). Risk Management: Coordinating Corporate Investment and Financing Policies. *Journal of Finance 48*(5), 1629–58.

Frydman, C., E. Hilt, and L. Y. Zhou (2015). Economic effects of runs on early "shadow banks": Trust companies and the impact of the panic of 1907. *Journal of Political Economy 123*(4), 902 – 940.

Galasso, A. and T. S. Simcoe (2011). Ceo overconfidence and innovation. *Management Science 57*(8), 1469–1484.

Goel, A. M. and A. V. Thakor (2008). Overconfidence, ceo selection, and corporate governance. *The Journal of Finance 63*(6), 2737–2784.

Hall, B. H., A. B. Jaffe, and M. Trajtenberg (2005). Market value and patent citations. *The RAND Journal of Economics 36*(1), pp. 16–38.

Hall, B. H. and J. Lerner (2010). Chapter 14 - the financing of r&d and innovation. In B. H. Hall and N. Rosenberg (Eds.), *Handbook of The Economics of Innovation, Vol. 1*, Volume 1 of *Handbook of the Economics of Innovation*, pp. 609 – 639. North-Holland.

Harhoff, D., F. Narin, F. M. Scherer, and K. Vopel (1999). Citation frequency and the value of patented inventions. *The Review of Economics and Statistics 81*(3), 511–515.

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

Hochberg, Y. V., C. J. Serrano, and R. H. Ziedonis (2016). Patent collateral, investor commitment, and the market for venture lending.

Howell, S. T. (2017, April). Financing innovation: Evidence from r&d grants. *American Economic Review 107*(4), 1136–64.

Johnson, M. A. and G. M. Maggiora (1990). *Concepts and applications of molecular similarity.* Wiley.

Jones, B. F. (2010). Age and great invention. *The Review of Economics and Statistics 92*(1), 1–14.

Kaplan, S. N. and L. Zingales (1997). Do investment-cash flow sensitivities provide useful measures of financing constraints? *The Quarterly Journal of Economics 112*(1), 169–215.

Kerr, W. R. and R. Nanda (2015). Financing innovation. *Annual Review of Financial Economics 7*(1), 445–462.

Kogan, L., D. Papanikolaou, A. Seru, and N. Stoffman (2017). Technological innovation, resource allocation, and growth*. *The Quarterly Journal of Economics 132*(2), 665–712.

Kyle, M. and H. L. Williams (2017, January). Is american health care uniquely inefficient? evidence from prescription drugs. Working Paper 23068, National Bureau of Economic Research.

Lanthier, M., K. L. Miller, C. Nardinelli, and J. Woodcock (2013). An improved approach to measuring drug innovation finds steady rates of first-in-class pharmaceuticals, 1987–2011. *Health Affairs 32*(8), 1433–1439.

Lichtenberg, F. (2013). The effect of pharmaceutical innovation on longevity: Patient level evidence from the 1996–2002 medical expenditure panel survey and linked mortality public-use files. *Forum for Health Economics and Policy 16*(1), 1–33.

Lichtenberg, F. R. and S. X. Sun (2007). The impact of medicare part d on prescription drug use by the elderly. *Health Affairs 26*(6), 1735–1744.

Light, D. W. and J. R. Lexchin (2012). Pharmaceutical research and development: what do we get for all that money? *BMJ 345*.

Lin, H. and D. Paravisini (2013). The effect of financing constraints on risk. *Review of Finance 17*(1), 229–259.

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

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

Malmandier, U. and G. Tate (2005). Ceo overconfidence and corporate investment. *The Journal of Finance 60*(6), 2661–2700.

Malmandier, U., G. Tate, and J. Yan (2011). Overconfidence and early-life experiences: The effect of managerial traits on corporate financial policies. *The Journal of Finance 66*(5), 1687–1733.

Mann, W. (2016). Creditor rights and innovation: Evidence from patent collateral.

Metrick, A. and S. Nicholson (2009). Do financial constraints have a real effect on pharmaceutical and biotech investment?

Moser, P., J. Ohmstedt, and P. Rhode (2011). Patents, citations, and inventive output - evidence from hybrid corn.

Myers, S. C. and N. S. Majluf (1984). Corporate financing and investment decisions when firms have information that investors do not have. *Journal of Financial Economics 13*(2), 187–221.

Naci, H., A. W. Carter, and E. Mossialos (2015). Why the drug development pipeline is not delivering better medicines. *BMJ 351*.

Nanda, R. and T. Nicholas (2014). Did bank distress stifle innovation during the great depression? *Journal of Financial Economics 114*(2), 273 – 292.

OHagan, S., N. Swainston, J. Handl, and D. B. Kell (2015). A 'rule of 0.5'for the metabolite-likeness of approved pharmaceutical drugs. *Metabolomics 11*(2), 323–339.

Pye, C. R., M. J. Bertin, R. S. Lokey, W. H. Gerwick, and R. G. Linington (2017). Retrospective analysis of natural products provides insights for future discovery trends. *Proceedings of the National Academy of Sciences 114*(22), 5601–5606.

Ralf Otto, Alberto Santagostino, and Ulf Schrader (2014, December). Rapid growth in biopharma: Challenges and opportunities. *McKinsey & Company*.

Smith, C. W. and R. M. Stulz (1985). The determinants of firms' hedging policies. *The Journal of Financial and Quantitative Analysis 20*(4), 391–405.

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

Appendix Tables and Figures

A Appendix Materials

Figure A.1: # OF DRUG CANDIDATES OVER TIME
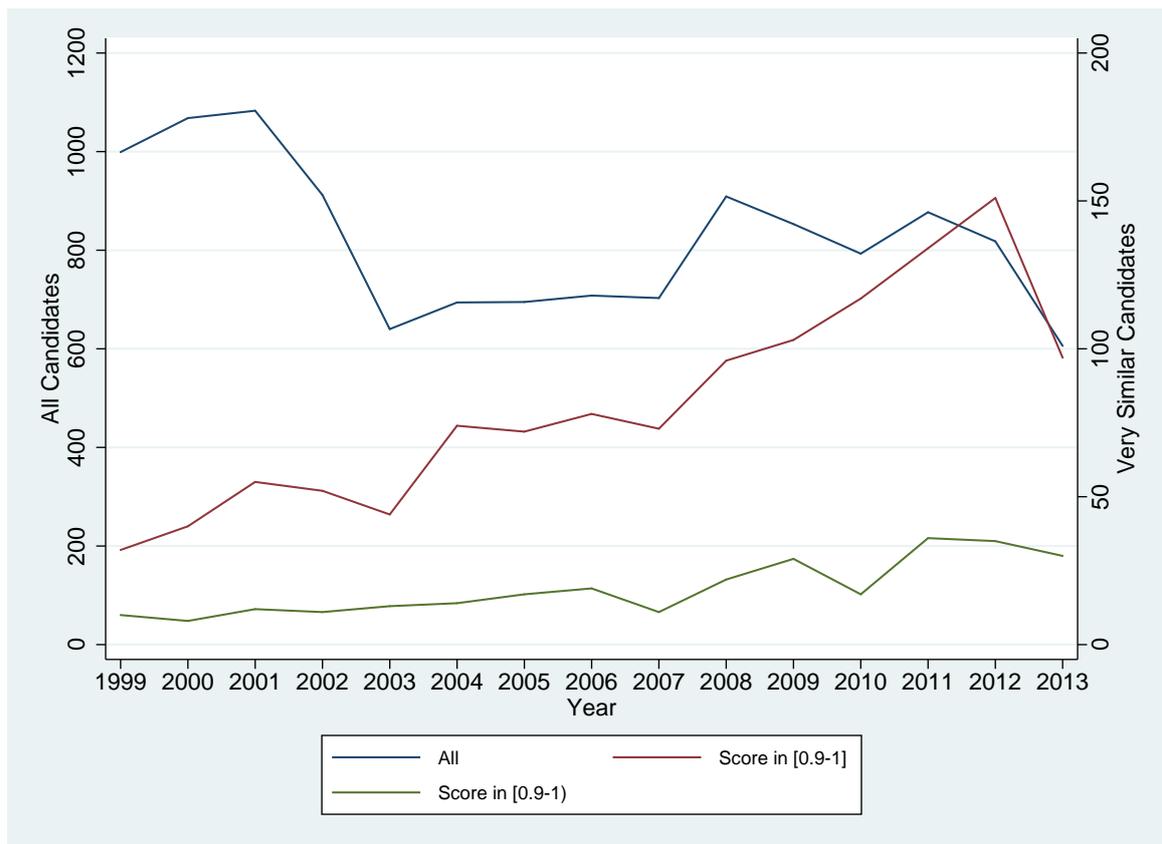
NOTES: This figure plots the number of new drug candidates for which we have data on molecular structure over time. The blue line all drug candidates. The red line represents drugs with similarity scores greater than 0.9, which indicates over 90% overlapping chemical structures. The green line plots the same pattern, excluding drugs with similarity equal to one; this is to avoid counting combination therapies which may use the same molecule in conjunction with another molecule.

Figure A.2: PROPORTION FDA APPROVED, BY DRUG SIMILARITY
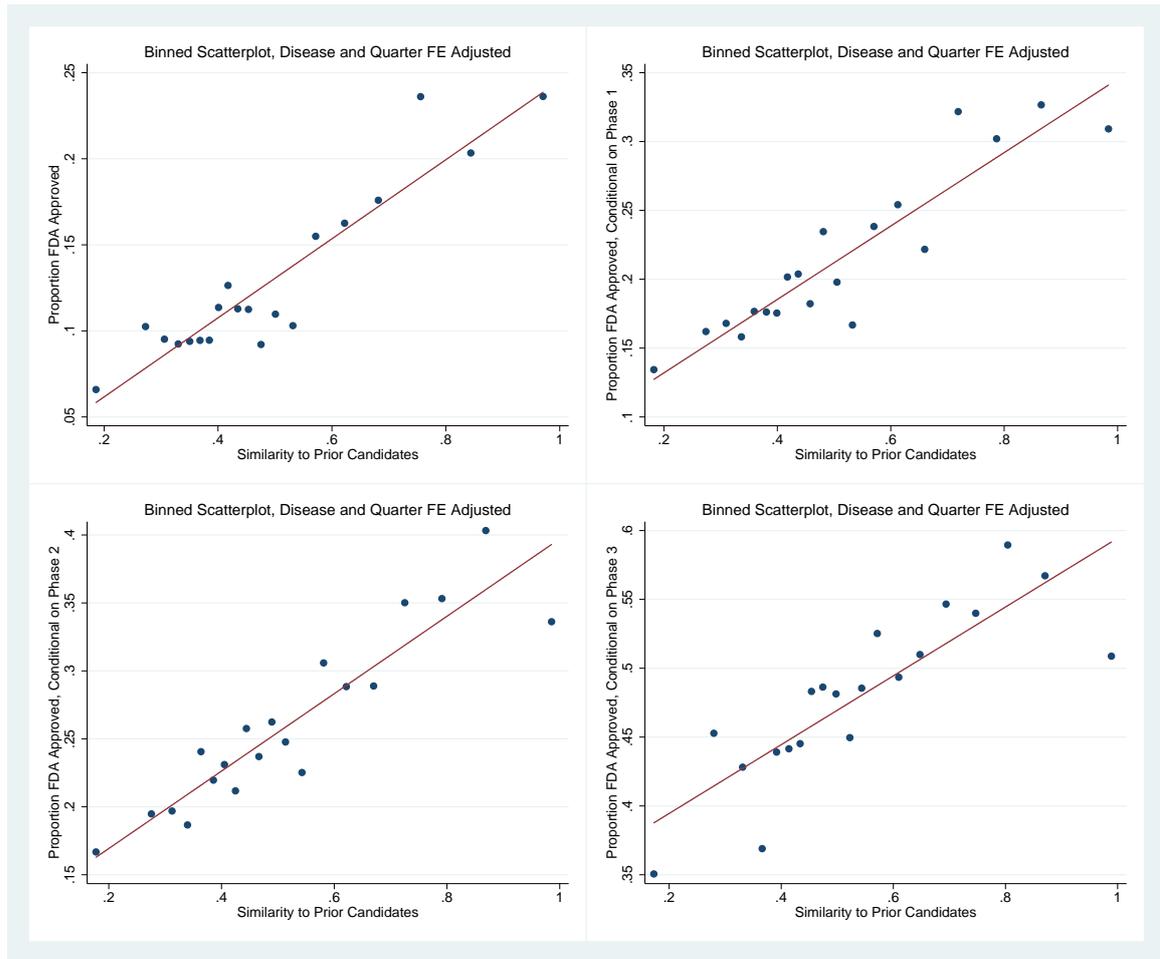
NOTES: Figure A.2 presents binned scatterplots of drug-level similarity against whether a drug is FDA approved. Each dot represents the proportion of candidates that FDA approved, among all candidates within a given similarity score bin, conditional on disease (ICD9) and quarter of development fixed effects. The top left panel examines all drug candidates; the top right represents only candidates that have made it into Phase 1 testing; the bottom left examines approval outcomes conditional on making it into Phase 2; the final figure examines outcomes conditional on Phase 3.

Figure A.3: Drug Similarity and Drug Effectiveness
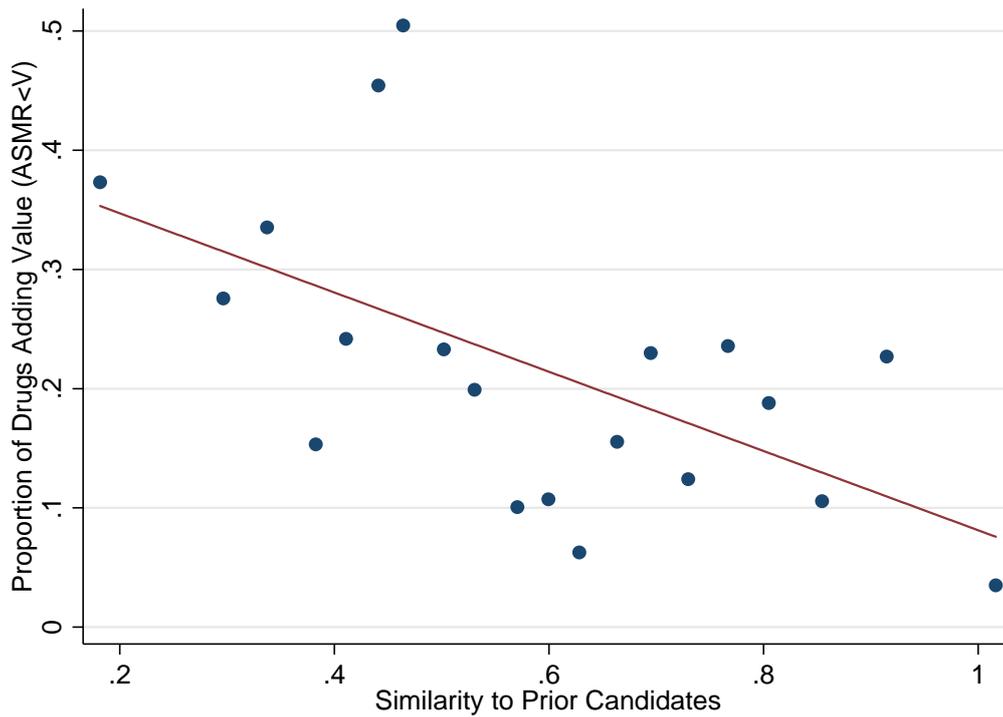
NOTES: Figure A.3 presents a binned scatterplot of drug-level similarity against drug added benefits. A drug's added benefit is derived from the from the French Haute Autorité de Santé (HAS) health system's clinical added benefits scores (Amélioration du Service Medical Rendu, or ASMR), which range from one to five (I to V), with V indicating no value added. In the plot above, the y-axis values represent the proportion of drugs in each similarity bin that had ASMR values less than V, after normalizing by disease area (ICD9) and the year of each drug's first regulatory approval year.

Figure A.4: DRUG SIMILARITY AND PATENT CITATIONS
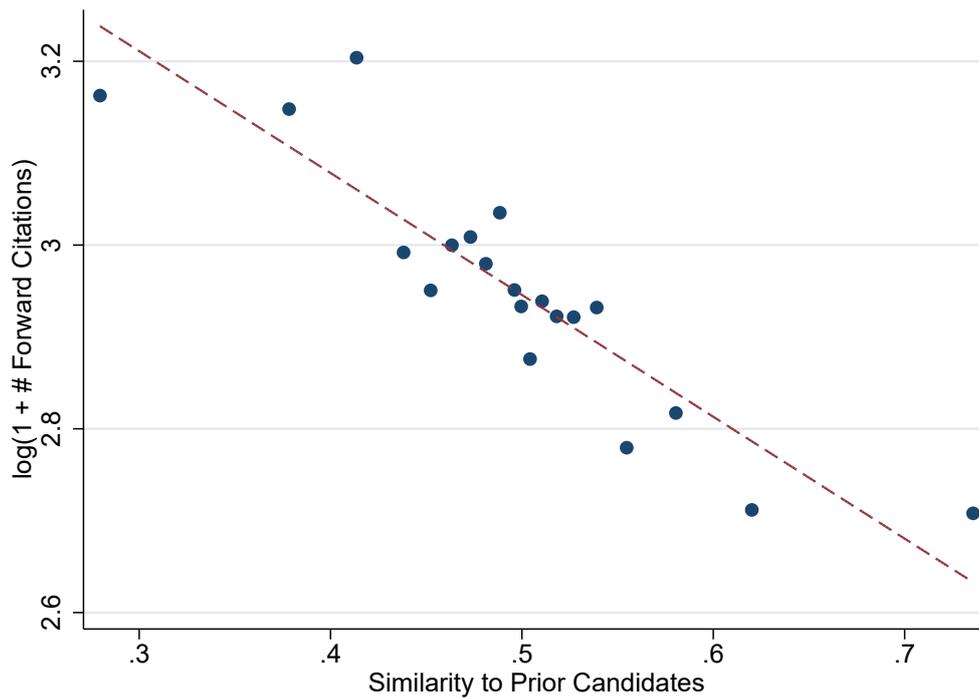
NOTES: Figure A.4 presents a binned scatterplot of drug-level similarity against the logarithm of one plus the number of forward citations the patent receives. Each dot represents mean log value, among all candidates within a given similarity score bin, conditional on disease (ICD9); patent issue year; company (assignee code), and year of development fixed effects. This specification corresponds to Column (4) of Table A.7. Please see Table A.7 for additional specifications.

Figure A.5: REVENUE, BY DRUG SIMILARITY
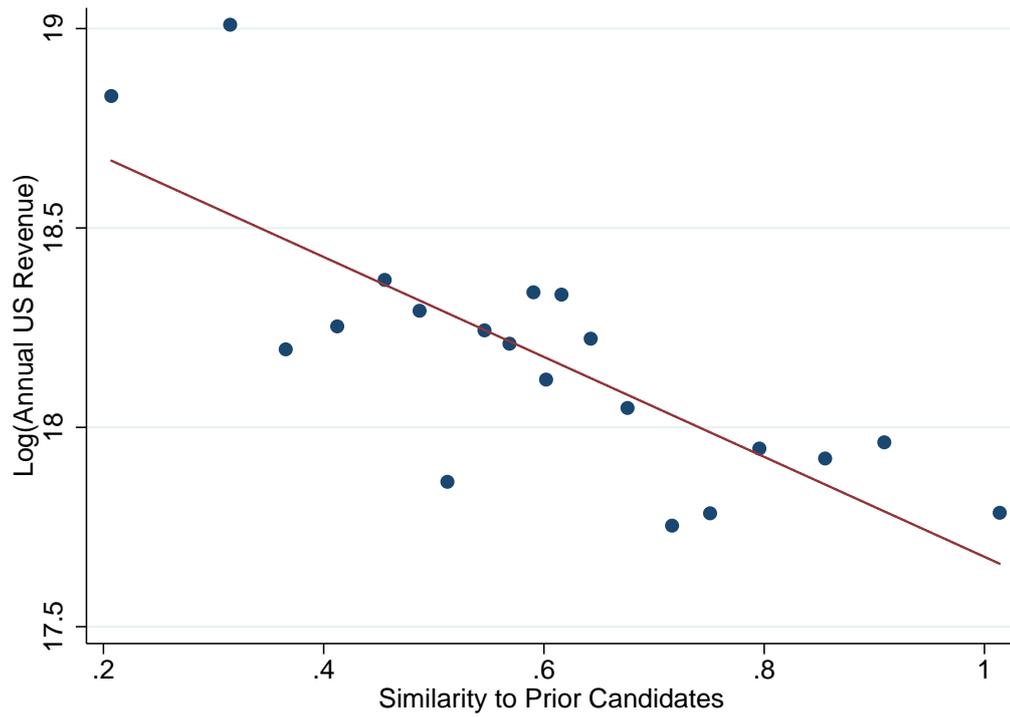
NOTES: Figure A.5 presents a binned scatterplot of drug-level similarity against revenue conditional on approval. The plot corresponds to the regression in Column (4) of Table A.8, which includes controls for drug indication, drug age, and firm dummies.

Figure A.6: DRUG SIMILARITY AND STOCK MARKET REACTION ON FDA APPROVAL
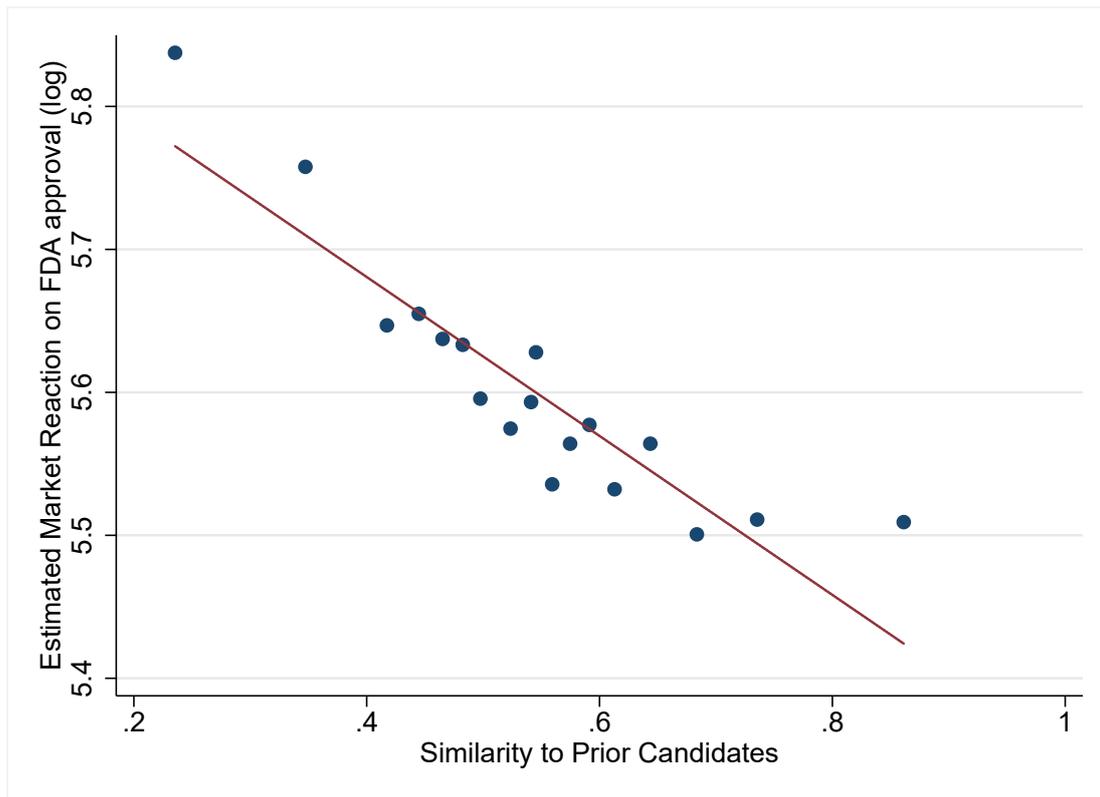
NOTES: Figure A.6 presents a binned scatterplot of drug-level similarity against the logarithm of the estimated dollar reaction on the (first) approval of the drug by the FDA. The dollar reaction to the FDA approval is estimated following the methodology of Kogan et al. (2017) and uses a 5-day window following the FDA approval. Each dot represents mean log value, among all candidates within a given similarity score bin, conditional on disease (ICD9); patent issue year; and year of development fixed effects, along with controls for the (log) firm's stock market capitalization prior to the patent issue. This specification corresponds to Column (4) of Table A.9.

Figure A.7: Drug Similarity and Market Value of Patents
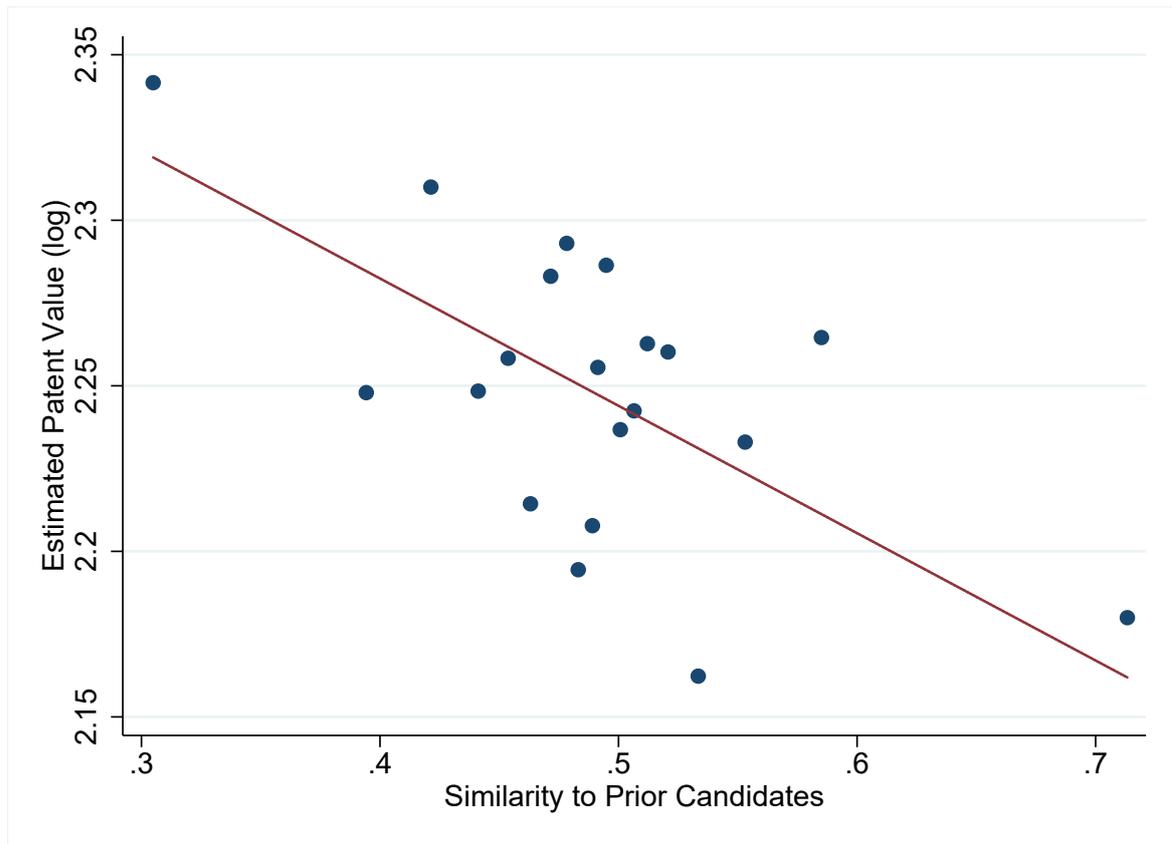
NOTES: Figure A.7 presents a binned scatterplot of drug-level similarity against the logarithm of the Kogan et al. (2017) estimated patent values. Each dot represents mean log value, among all candidates within a given similarity score bin, conditional on disease (ICD9); patent issue year; and year of development fixed effects, along with controls for the (log) firm's stock market capitalization prior to the patent issue. This specification corresponds to Column (4) of Table A.10. Please see Table A.10 for additional specifications.

Figure A.8: Distribution of Medicare Drug Life in 2003
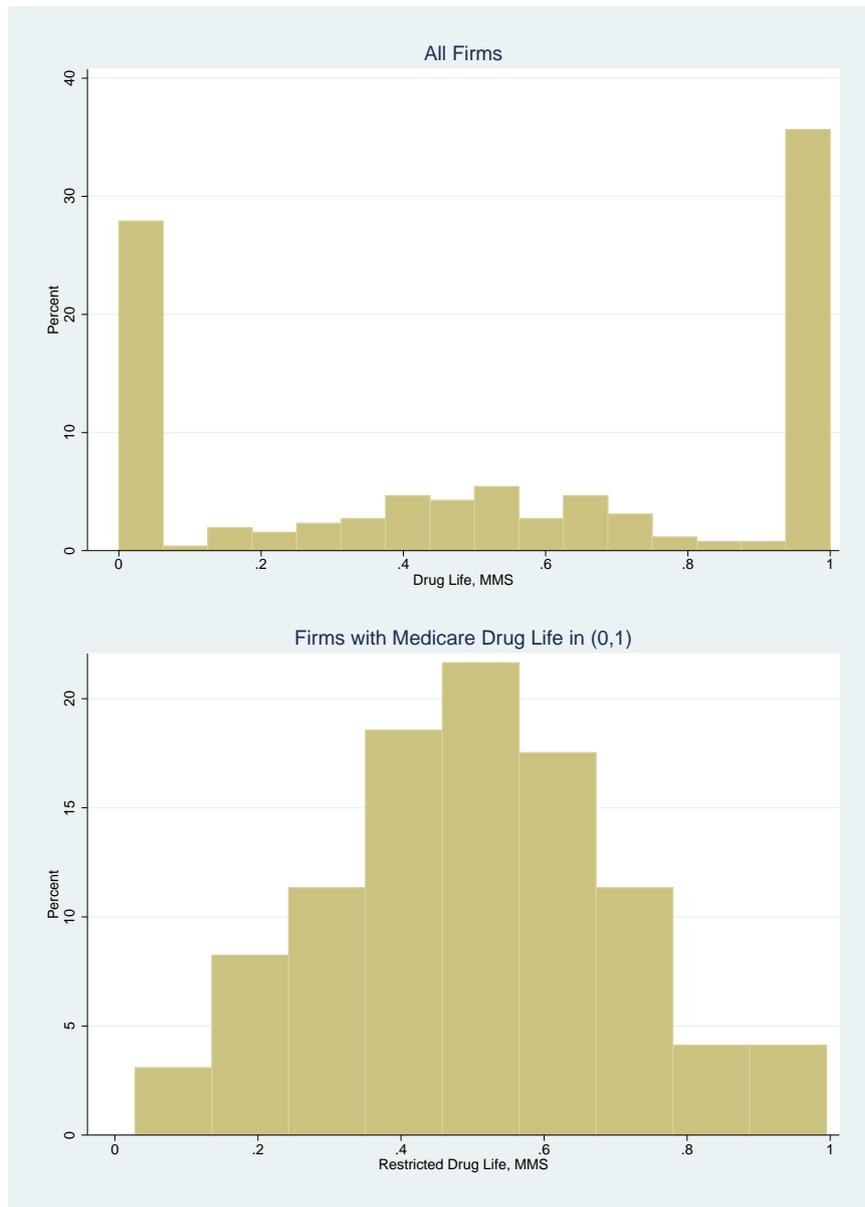
NOTES: Figure A.8 plots the distribution of Medicare Drug Life in 2003. Each observation is a firm in our main analysis sample.

Figure A.9: FIRM EXPERIENCE, BY DRUG SIMILARITY
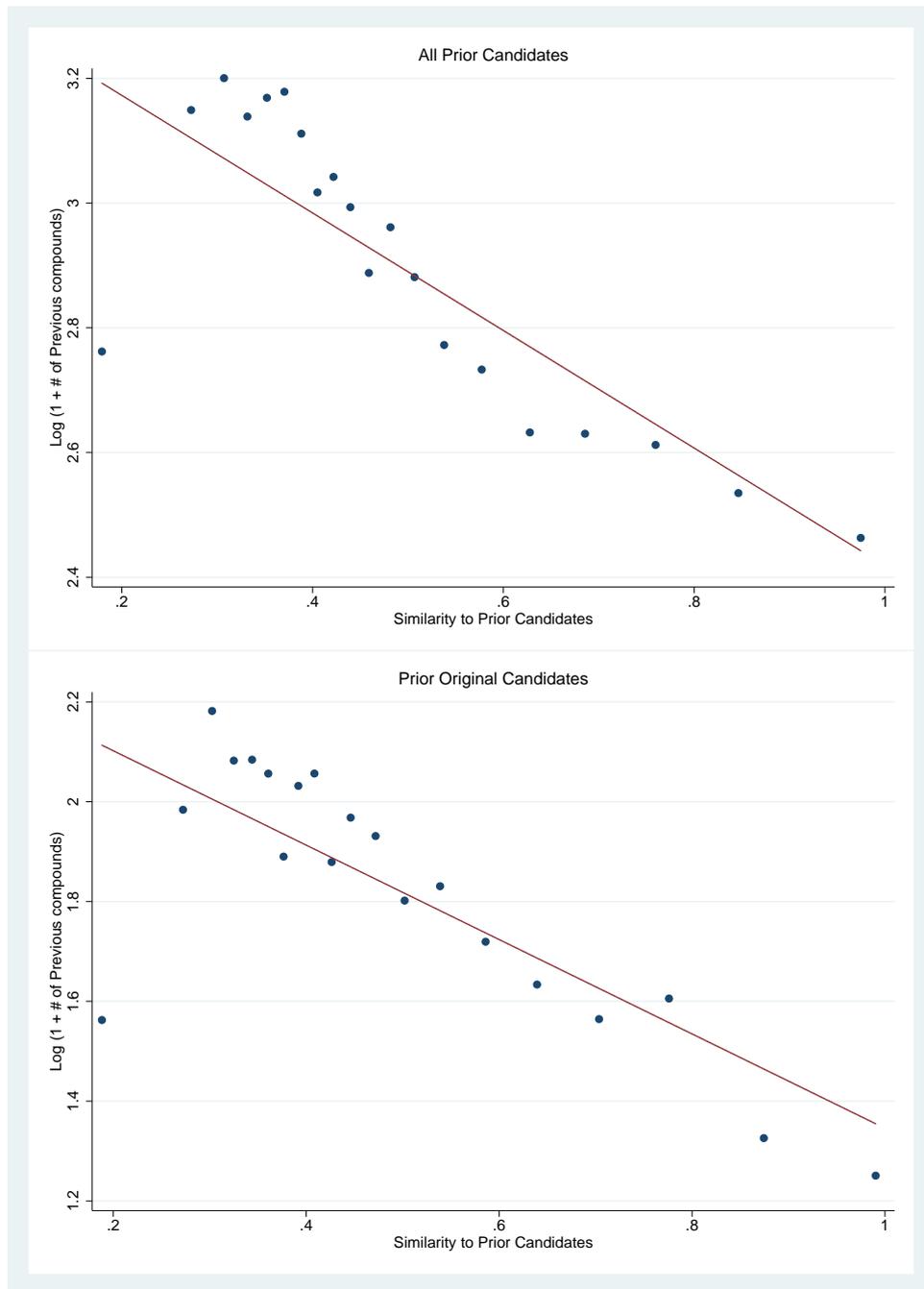
NOTES: Figure A.9 presents a binned scatterplot of drug-level similarity against measures of firm experience. Each dot represents the mean log of past firm experience, among all candidates within a given similarity score bin, conditional on disease (ICD9) and quarter of development fixed effects. In the top panel, past firm experience is defined as one plus the total number of compounds developed by this firm prior to a the drug candidate in question. In the bottom panel, we count experience using only past compounds for which the given firm had ownership at the time the compound first enters development.

Figure A.10: PROXIES FOR DEVELOPMENT COSTS, BY DRUG SIMILARITY
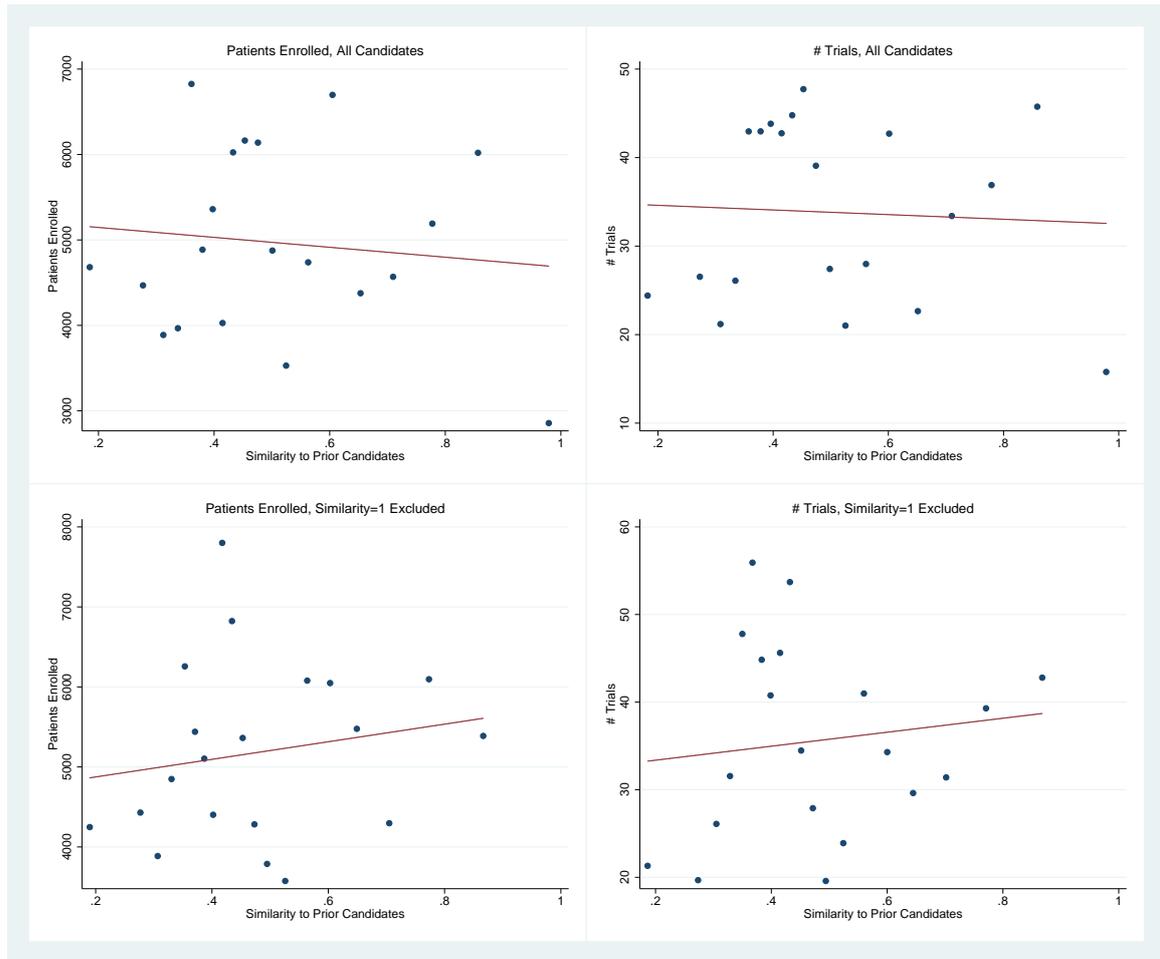
NOTES: Figure A.10 presents a binned scatterplot of drug-level similarity against proxies for the direct cost of drug development. Each dot represents the mean number of patients enrolled (or number of trials conducted), among all candidates within a given similarity score bin, conditional on disease (ICD9) and quarter of development fixed effects. In the bottom two panels, we exclude drug candidates with a similarity score of 1 to restrict to candidates that likely did not rely on results of trials conducted for an identical past drug.

Figure A.11: IMPACT OF ADDITIONAL RESOURCES ON NOVELTY, WITHIN INDICATION
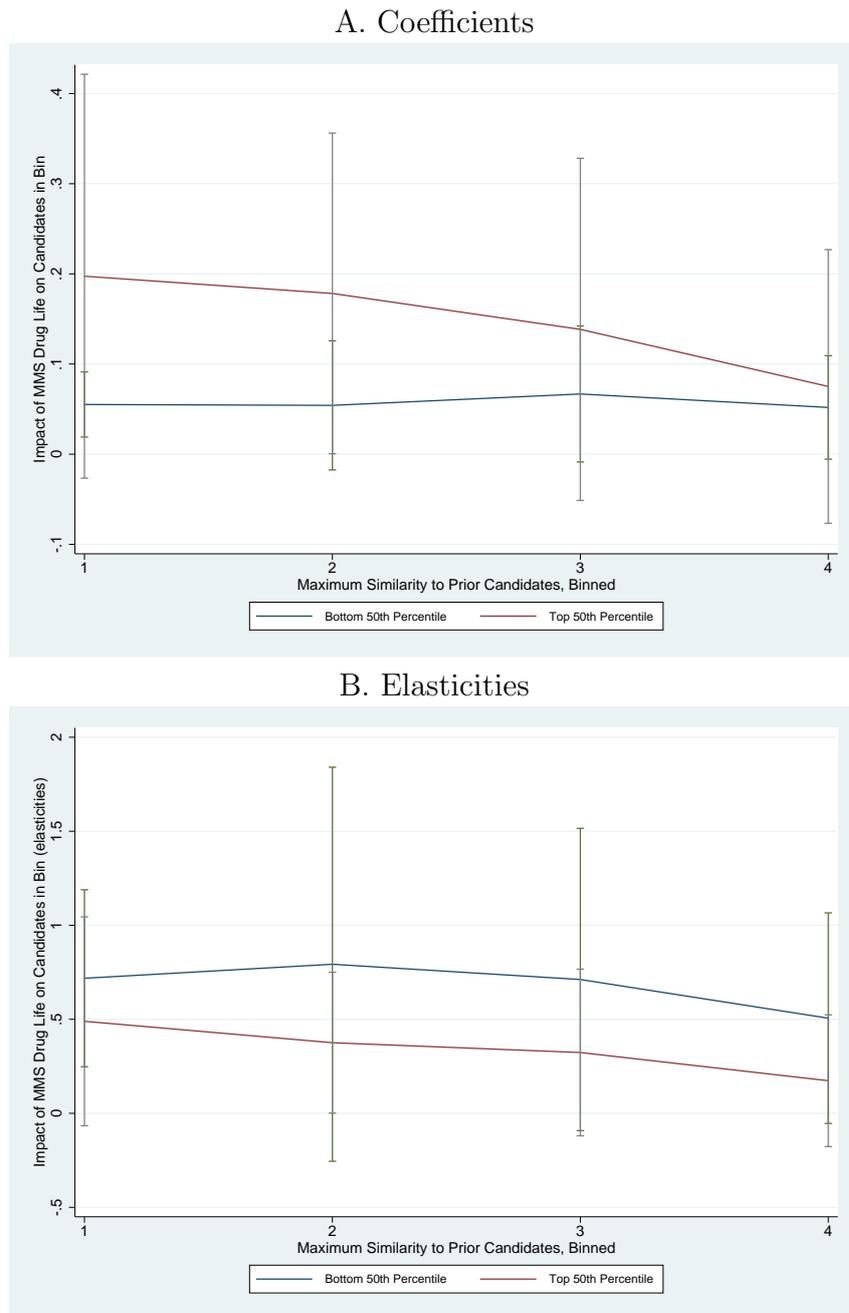
A. Coefficients

B. Elasticities

NOTES: Figure A.12 plots the estimated coefficients on Post × Medicare Drug Life$_{f,2003}$ from our main regression specification defined by Equation (3) across firm size groups (defined by total revenue generated by approved drugs prior to 2003). The outcome variable is number of drug candidates across novelty bins.

Figure A.12: IMPACT OF ADDITIONAL RESOURCES ON NOVELTY, WITHIN INDICATION
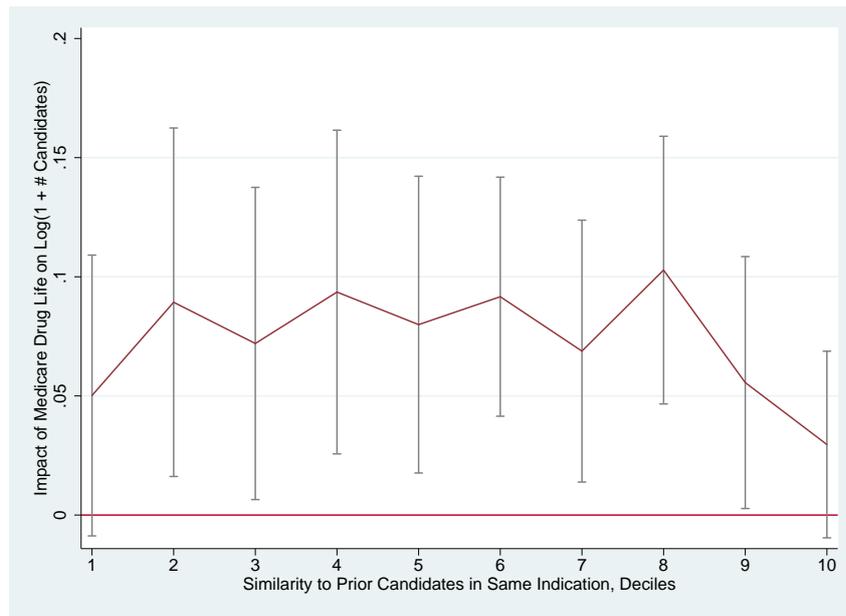
NOTES: Figure A.12 plots the estimated coefficients on Post × Medicare Drug Life$_{f,2003}$ from our main regression specification defined by Equation (3). This figure is analogous to the bottom panel of Figure 5 of the main text, except that similarity is measured with respect to other drugs in the same indication (disease). Bin 1 in this case is the impact of the policy on the number of drugs that fall into the bottom 10th percentile of similarity in our sample.

Figure A.13: ORIGINAL VS. ACQUIRED
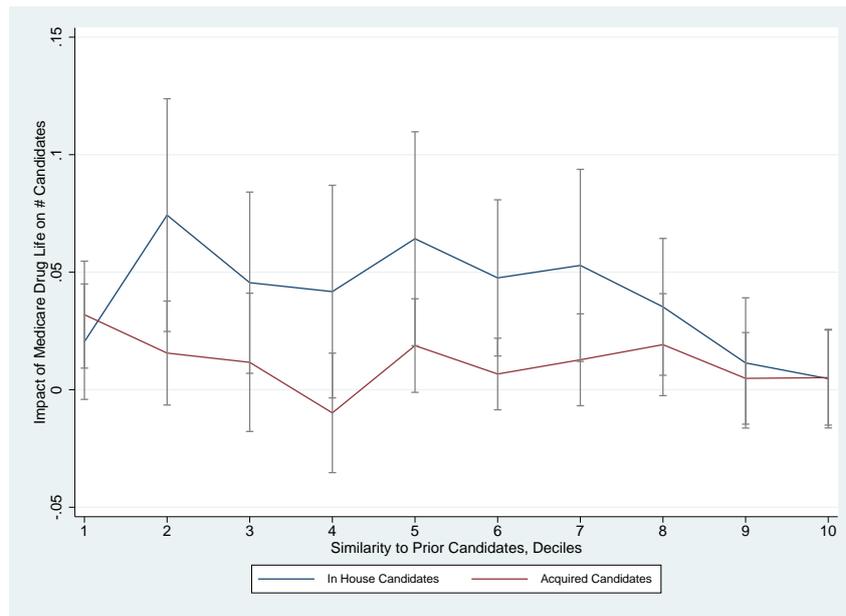
NOTES: Figure A.13 plots the estimated coefficients on Post × Medicare Drug Life$_{f,2003}$ from our main regression specification defined by Equation (3), with the sample split based on firm experience in drug development. Each point represents a different outcome variable: the number of new drug candidates in a given bin of similarity. The blue line (above) represents the coefficients corresponding firms. The red line (below) displays the coefficients for drugs that the developer acquired. Both sets of coefficients include 95% confidence intervals around the point estimates.

Figure A.14: EXPERIENCED VS. INEXPERIENCED FIRMS
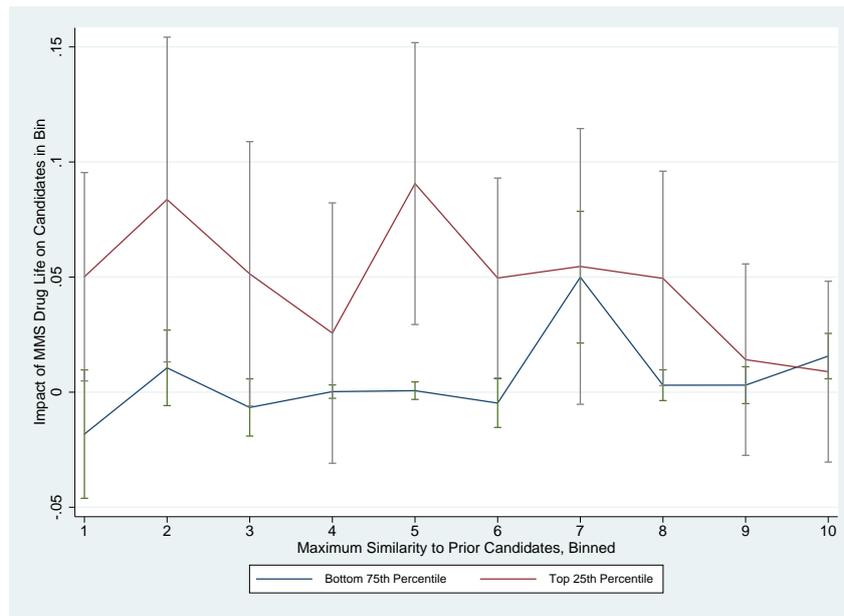
NOTES: Figure A.14 plots the estimated coefficients on Post × Medicare Drug Life$_{f,2003}$ from our main regression specification defined by Equation (3), with the sample split based . Each point represents a different outcome variable: the number of new drug candidates in a given bin of similarity. The red line (above) represents the coefficients corresponding to firms in the top 25th percentile of experience (as proxied by one plus the number of new drug candidates the firm had previously developed) while the blue line (below) displays the coefficients for firms in the bottom 75th percentile of firm experience. Both sets of coefficients include 95% confidence intervals around the point estimates.

Table A.1: Drug Novelty and Development Costs

(a) All Candidates

	Patients Enrolled		# of Trials	
	(1)	(2)	(3)	(4)
Maximum Similarity	-1690.088***	-580.394	-14.413***	-2.621
	(556.994)	(509.676)	(4.019)	(3.550)
R^2	0.117	0.194	0.088	0.170
Qtr of Development FEs	Yes	Yes	Yes	Yes
ICD-9 FEs		Yes		Yes
Observations	8887	8801	10624	10546

(b) Candidates with Similarity Score < 1

	Patients Enrolled		# of Trials	
	(1)	(2)	(3)	(4)
Maximum Similarity	83.851	1099.570*	0.109	7.982*
	(687.068)	(633.287)	(4.727)	(4.418)
R^2	0.122	0.201	0.090	0.175
Qtr of Development FEs	Yes	Yes	Yes	Yes
ICD-9 FEs		Yes		Yes
Observations	8359	8280	9975	9903

NOTES: Table A.1 examines the relationship between drug level similarity (maximum similarity to any prior developed drug candidate) and the the cost of drug development (as proxied for by the number of patients and number of clinical trials). Observations are at the drug level-ICD9 and results are reported with standard errors clustered by ICD9. The accompanying binned scatterplots of results are shown in Figure **??**. $^*p < 0.10,^{**} p < 0.05,^{***} p < 0.01$.

Table A.2: PHARMACEUTICAL FIRMS AND DEBT FINANCE

	A. Compustat North America			B. Compustat Global		
	(1)	(2)	(3)	(1)	(2)	(3)
Pharmaceutical	-0.0330**	-0.0709***	-0.0808***	-0.00775	-0.0287***	-0.0324***
	(-2.60)	(-4.75)	(-5.24)	(-0.96)	(-3.40)	(-3.72)
Size, log		0.0188***	0.0239***		0.00650***	0.00658***
		(32.09)	(40.39)		(37.41)	(35.70)
Profitability			-0.0384***			-0.0270***
			(-12.46)			(-13.46)
Mean leverage ratio	0.174	0.174	0.174	0.118	0.118	0.118
N	261,158	261,158	249,845	533,580	533,577	493,448
R^2	0.008	0.058	0.086	0.003	0.024	0.022

NOTES: Table A.2 compares leverage ratios of the pharmaceutical firms in our sample and compares them to the broader Compustat universe. Standard errors are clustered by firm. Firm size is book assets (Compustat: at); profitability is income before extraordinary items (Compustat: ib) plus depreciation (Compustat: dp) over book assets. Panel A presents results for firms in Compustat North America; Panel B for Compustat Global. All specifications include time fixed effects. We report t-statistics in parentheses, with standard errors clustered by firm. $^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$.

Table A.3: DRIVERS OF PAIRWISE DRUG SIMILARITY

	Drug Candidate Pairwise Similarity			
	Mean = 0.106			
	(1)	(2)	(3)	(4)
Share Target-Action	0.167***	0.122***		
Mean: 0.022	(6.24e-05)	(0.00838)		
Share Indication			0.0102***	0.0285***
Mean: 0.149			(8.51e-06)	(0.00200)
N	955,921,961	955,921,961	955,921,961	955,921,961
R²	0.025	0.265	0.002	0.075
Target-Action FEs		X		
Indication FEs				X

NOTES: Table A.3 examines the relationship between indicator variables for sharing the same target-action or the same indication (ICD9) on the pairwise similarity of two drug candidates, call them drug A and drug B. Because single drug can be associated with multiple target-actions and indications, each observation is a drugA-actionA-indicationA-drugB-actionB-indicationB pair. We include such a pair for every pair of drugs in our data. $^*p < 0.10,^{**} p < 0.05,^{***} p < 0.01$.

Table A.4: PROPORTION FIRST IN TARGET, BY DRUG SIMILARITY

	First in Narrow Target		First in Broad Target	
	Mean: 0.194		Mean: 0.068	
	(1)	(2)	(3)	(4)
Similarity Measure	-0.210***	-0.175***	-0.144***	-0.141***
	(0.0148)	(0.0153)	(0.00858)	(0.00921)
N	15,160	15,160	15,160	15,160
R^2	0.052	0.129	0.044	0.076
Quarter of Development FEs	X	X	X	X
Disease FEs		X		X

NOTES: Table A.4 examines the relationship between drug level similarity (maximum similarity to any prior developed drug candidate) and a drug's likelihood of being the first in its target, defined narrowly (target and action) and broadly (coarse target family). Observations are at the drug level and results are reported with robust standard errors. The accompanying binned scatterplot of results is shown in Figure 2. $^*p < 0.10,^{**}p < 0.05,^{***}p < 0.01$.

Table A.5: Proportion FDA Approved, by Drug Similarity

	All		Phase 1		Phase 2		Phase 3	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Maximum Similarity	0.272***	0.230***	0.303***	0.267***	0.321***	0.285***	0.288***	0.250***
	(0.024)	(0.025)	(0.024)	(0.027)	(0.025)	(0.028)	(0.026)	(0.030)
R^2	0.108	0.181	0.123	0.188	0.121	0.187	0.122	0.192
Qtr of Development FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
ICD-9 FEs		Yes		Yes		Yes		Yes
Observations	19191	19127	11476	11400	9508	9431	5158	5069

NOTES: Table A.5 examines the relationship between drug level similarity (maximum similarity to any prior developed drug candidate that ever reached Phase 1 clinical trials) and a drug's likelihood of reaching FDA approval. Observations are at the drug-ICD9 level and results are reported with standard errors clustered at the ICD-9 level. The analysis sample changes by column, including all drugs (Columns 1 and 2), drugs that reach Phase 1 (Columns 3 and 4), drugs that reach Phase 2 (Columns 5 and 6), and drugs that reach Phase 3 (Columns 7 and 8). The accompanying binned scatterplot of results is shown in Figure A.2.
$^*p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01.$

Table A.6: DRUG NOVELTY AND DRUG EFFECTIVENESS

	(1) Any Value Added	(2) Any Value Added	(3) High Importance	(4) High Importance	(5) ASMR Value	(6) ASMR Value	(7) ASMR Value
	ASMR<V	ASMR<V	ASMR<IV	ASMR<IV			Ordered Logit
Maximum Similarity	-0.270**	-0.332**	-0.126**	-0.129*	0.491**	0.459*	2.436**
	(0.069)	(0.099)	(0.043)	(0.061)	(0.143)	(0.178)	(0.734)
Controls							
Disease Area (ICD9)		Y		Y		Y	Y
Drug Launch Year		Y		Y		Y	Y
Nb. of Drugs	385	385	385	385	369	369	369

NOTES: Table A.6 examines the relationship between drug level similarity (maximum similarity to any prior drug candidate that had reached phase 1 clinical trials) and the French Haute Autorité de Santé (HAS) health system's measure of clinical added benefits (Amélioration du Service Medical Rendu, or ASMR). The ASMR scores range from I (major value added) to V (no value added). The analysis sample includes approved small molecule drugs that recieved ASMR scores and that we were able to match to drugs in the Cortellis database. Controls include broad disease area (ICD9 codes grouped into 20 more general categories), drug launch year and company identifiers. Standard errors are clustered by broad disease area. The accompanying binned scatterplot of results is shown in Figure A.3. $^*p < 0.10,^{**} p < 0.05,^{***} p < 0.01$.

Table A.7: Patent citations and Drug Similarity

	(1)	(2)	(3)	(4)
Maximum Similarity	-0.392***	-0.459***	-0.566***	-1.470***
	(0.103)	(0.114)	(0.132)	(0.141)
N	3539	3479	3449	3448
R^2	0.421	0.527	0.773	0.811
Controls				
Patent Issue Year	Y	Y	Y	Y
ICD-9 FEs		Y	Y	Y
Firm FEs			Y	Y
Drug Cohort FEs				Y

NOTES: Table A.7 examines the relationship between drug level similarity (maximum similarity to any prior developed drug candidate) and the logarithm of one plus the number of forward citations. The matching between drugs and patents is from Cortellis. We restrict attention to patents filed prior to the FDA approval. Observations are at the drug-disease(ICD9)-patent level. We report standard errors in parentheses clustered by indication (ICD9). Controls include: 1) the year the patent is granted; 2) the ICD9 disease area treated by the drug; 3) company (PERMCO) fixed effects; 4) the year the drug is developed. The accompanying binned scatterplot of results is shown in Figure A.4. $^*p < 0.10,^{**} p < 0.05,^{***} p < 0.01$.

Table A.8: REVENUE, BY DRUG SIMILARITY

| | Log(Annual US Revenue) | | | |
	(1)	(2)	(3)	(4)
Maximum Similarity	-1.449***	-1.307***	-1.253***	-0.641**
	(0.250)	(0.275)	(0.281)	(0.281)
R^2	0.091	0.272	0.292	0.573
Year FEs	Yes	Yes	Yes	Yes
ICD-9 FEs		Yes	Yes	Yes
Drug Cohort FEs			Yes	Yes
Firm FEs				Yes
Observations	11256	11243	11243	11230

NOTES: Table A.8 examines the relationship between drug level similarity (maximum similarity to any prior developed drug candidate that ever reached Phase 1 clinical trials) and a drug's revenue conditional on approval. Drug revenue data is derived by matching approved drugs to the Medicare Expenditure Panel Survey. To control for differences in when and how often drug revenue is observed for various drugs, drug revenue is calculated as the fixed effect associated with a drug, holding constant year fixed effects: drug revenue is thus measured relative to other drugs observed in that year, averaged over years. Observations are at the drug-ICD9 level and results are reported with standard errors clustered at the ICD-9 level. The accompanying binned scatterplot of results is shown in Figure A.5. $^*p < 0.10,^{**}p < 0.05,^{***}p < 0.01$.

Table A.9: Market reaction to FDA approval, by Drug Similarity

	(1)	(2)	(3)
Maximum Similarity	-1.321**	-0.519***	-0.556***
	(0.576)	(0.069)	(0.064)
N	462	451	399
R^2	0.065	0.980	0.988
Controls			
Approval Year	Y	Y	Y
Firm Size (Market Capitalization)		Y	Y
Company		Y	Y
Firm Volatility		Y	Y
ICD-9 FEs			Y

NOTES: Table A.9 examines the relationship between drug level similarity (maximum similarity to any prior developed drug candidate) and the logarithm of the estimated dollar reaction on the (first) approval of the drug by the FDA. The dollar reaction to the FDA approval is estimated following the methodology of Kogan et al. (2017) and uses a 5-day window following the FDA approval. Observations are at the drug level. We report standard errors in parentheses clustered by firm. Controls include: 1) the year the drug is approved; 2) the firm's market capitalization on the day prior to the first approval by the FDA, to ensure that we are not simply capturing differences in firm size; 3) the ICD9 disease area treated by the drug; 4) company fixed effects; and 5) the firm's stock market volatility, since the measurement error adjustment results in a non-linear transformation of the firm's stock return. The accompanying binned scatterplot of results is shown in Figure A.6. $^*p < 0.10,^{**}p < 0.05,^{***}p < 0.01$.

Table A.10: PATENT MARKET VALUE, BY DRUG SIMILARITY

	(1)	(2)	(3)	(4)
Maximum Similarity	-0.641***	-0.527**	-0.227**	-0.383***
	(0.244)	(0.265)	(0.092)	(0.128)
N	1785	1740	1644	1643
R^2	0.268	0.446	0.958	0.961
Controls				
Patent Issue Year	Y	Y	Y	Y
Disease Area (ICD9)		Y	Y	Y
Firm Market Capitalization			Y	Y
Company			Y	Y
Firm Volatility			Y	Y
Drug Development Year				Y

NOTES: Table A.10 examines the relationship between drug level similarity (maximum similarity to any prior developed drug candidate) and the logarithm of the estimated patent value, where the latter is based on Kogan et al. (2017). The matching between drugs and patents is from Cortellis. We restrict attention to patents filed prior to the FDA approval. Observations are at the drug-disease(ICD9)-patent level. We report standard errors in parentheses clustered by firm. Controls include: 1) the year the patent is granted; 2) the ICD9 disease area treated by the drug; 3) the firm's market capitalization on the day prior to the patent grant, to ensure that we are not simply capturing differences in firm size; 4) company (PERMCO) fixed effects; 5) the firm's stock market volatility, since the measurement error adjustment results in a non-linear transformation of the firm's stock return; and 6) the year the drug is developed. The accompanying binned scatterplot of results is shown in Figure A.7. $^*p < 0.10,$$^{**}p < 0.05,$$^{***}p < 0.01.$

Table A.11: FIRM EXPERIENCE, BY DRUG SIMILARITY

	Log(1 + All Prior Candidates)		Log(1 + Prior Original Candidates)	
	(1)	(2)	(3)	(4)
Maximum Similarity	-0.764**	-0.751***	-0.906***	-0.837***
	(0.315)	(0.291)	(0.204)	(0.198)
R^2	0.030	0.078	0.069	0.124
Company FEs	Yes	Yes	Yes	Yes
Qtr of Development FEs	Yes	Yes	Yes	Yes
ICD-9 FEs		Yes		Yes
Observations	28521	28486	21220	21182

NOTES: Table A.11 examines the relationship between drug level similarity (maximum similarity to any prior developed drug candidate) and the experience of the firm (as measured by the log of past compounds). Observations are at the drug-icd9-firm level and results are reported with standard errors clustered by firm. The accompanying binned scatterplot of results is shown in Figure A.9. $^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$.

Table A.12: In-House vs. Acquired Drug Candidates

	(1) All	(2) In House	(3) Acquired
Post 2003 X Medicare Drug Life	0.263***	0.223**	0.094*
	(0.096)	(0.086)	(0.049)
R^2	0.595	0.593	0.321
Company FEs	Yes	Yes	Yes
Qtr of Development FEs	Yes	Yes	Yes
Overall Drug Life/Firm MMS	Yes	Yes	Yes
Observations	16442	16442	16442

NOTES: Table A.12 reports the main specification coefficient for Post $\times$ Medicare Drug Life$_{f,2003}$. Model 1 repeats the result from our main regression specification (Column 6 of table 4). Model 2 limits the dependent variable to the number of new drug candidates that originated within the focal firm (in-house), while Model 3 includes only drug candidates that the focal firm acquired (originated at another firm) All models include a full set of company and quarter indicator variables, with Post$\times$Overall Drug Life$_{f,2003}$ and Post$\times$Firm MMS$_{f,2003}$ both included as additional independent variables, but not reported in the table. Robust standard errors in parentheses, clustered around company identifiers. $^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$.

Table A.13: Impact of Resources on # Original New Candidates, by Similarity Decile

(A) In House Candidates

	Log(1 + New In House Candidates), by Similarity Decile									
	1	2	3	4	5	6	7	8	9	10
Post 2003 X Medicare Drug Life	0.020	0.074**	0.046*	0.042	0.064**	0.048**	0.053**	0.035**	0.011	0.005
	(0.015)	(0.030)	(0.023)	(0.027)	(0.028)	(0.020)	(0.025)	(0.018)	(0.017)	(0.013)
R^2	0.169	0.273	0.272	0.302	0.310	0.238	0.218	0.187	0.172	0.104
Company FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Qtr of Development FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Overall Drug Life/Firm MMS	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	16442	16442	16442	16442	16442	16442	16442	16442	16442	16442

(A) Acquired Candidates

	Log(1 + New Acquired Candidates), by Similarity Decile									
	1	2	3	4	5	6	7	8	9	10
Post 2003 X Medicare Drug Life	0.032**	0.016	0.012	-0.010	0.019	0.007	0.013	0.019	0.005	0.005
	(0.014)	(0.013)	(0.018)	(0.015)	(0.012)	(0.009)	(0.012)	(0.013)	(0.012)	(0.012)
R^2	0.069	0.084	0.081	0.079	0.079	0.066	0.056	0.083	0.085	0.076
Company FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Qtr of Development FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Overall Drug Life/Firm MMS	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	16442	16442	16442	16442	16442	16442	16442	16442	16442	16442

Notes: Table A.13 reports the main specification coefficient for Post × Medicare Drug $Life_{f,2003}$. In Panel A, the dependent variable is limited to new drug candidates that were originally developed in the focal firm, and varies by new drug candidates' deciles of maximum similarity compared to all prior drug candidates that reached phase I trials. In Panel B, dependent variable includes only newly acquired drug candidates that originated at other firms. All models include a full set of company and quarter indicator variables, with Post × Overall Drug $Life_{f,2003}$ and Post × Firm $MMS_{f,2003}$ both included as additional independent variables. Robust standard errors in parentheses, clustered around company identifiers. $^*p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

Table A.14: IMPACT OF RESOURCES ON # NEW CANDIDATES, BY NOVELTY

(A) ABSOLUTE SIMILARITY BINS

	Log(1 + New Candidates), by Similarity Bin									
	1	2	3	4	5	6	7	8	9	10
Post 2003 X Medicare Drug Life	0.001	0.000	0.054**	0.134**	0.123***	0.059**	0.028	0.010	0.012	0.008
	(0.003)	(0.005)	(0.022)	(0.058)	(0.044)	(0.028)	(0.020)	(0.016)	(0.011)	(0.018)
R^2	0.023	0.034	0.188	0.506	0.395	0.231	0.163	0.128	0.111	0.118
Company FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Qtr of Development FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Overall Drug Life/Firm MMS	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	16442	16442	16442	16442	16442	16442	16442	16442	16442	16442

(B) DECILES OF SIMILARITY

	Log(1 + New Candidates), by Similarity Decile									
	1	2	3	4	5	6	7	8	9	10
Post 2003 X Medicare Drug Life	0.049**	0.084**	0.053*	0.029	0.083***	0.051**	0.064**	0.052**	0.017	0.009
	(0.023)	(0.035)	(0.029)	(0.028)	(0.031)	(0.022)	(0.029)	(0.024)	(0.020)	(0.019)
R^2	0.176	0.280	0.283	0.314	0.324	0.247	0.223	0.210	0.201	0.141
Company FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Qtr of Development FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Overall Drug Life/Firm MMS	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	16442	16442	16442	16442	16442	16442	16442	16442	16442	16442

NOTES: Table A.14 reports the main specification coefficient for Post × Medicare Drug Life$_{f,2003}$. In Panel A, the dependent variable varies by new drug candidates' absolute maximum similarity compared to all prior drug candidates that reached phase I trials (e.g. bin 6 represents all drugs with maximum similarity scores in the range 0.5-0.6). In Panel B, the dependent variable is split into bins that represent new drugs' deciles of maximum similarity score. All models include a full set of company and quarter indicator variables, with Post × Overall Drug Life$_{f,2003}$ and Post × Firm MMS$_{f,2003}$ both included as additional independent variables. Robust standard errors in parentheses, clustered around company identifiers. $^*p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.15: FIRM EXPERIENCE

	Log(1 + New Candidates), by Experience		
	(1)	(2)	(3)
	All	Top 25	Bottom 75
Post 2003 X Medicare	0.263***	0.260**	0.053
Drug Life	(0.096)	(0.118)	(0.040)
R^2	0.595	0.578	0.049
Company FEs	Yes	Yes	Yes
Qtr of Development FEs	Yes	Yes	Yes
Overall Drug Life/Firm MMS	Yes	Yes	Yes
Observations	16442	11122	4040

NOTES: Table A.15 reports the main specification coefficient for Post $\times$ Medicare Drug Life$_{f,2003}$. Column (1) repeats the result from our main regression specification (Column (6) of Table 4). Column (2) limits the sample to firms in the top 25% of the experience distribution (as proxied by number of drug candidates previously developed), while Column (3) includes firms in the bottom 75th percentile in terms of experience. All models include a full set of company and quarter indicator variables, with Post $\times$ Overall Drug Life$_{f,2003}$ and Post $\times$ Firm MMS$_{f,2003}$ both included as additional independent variables, but not reported in the table. Robust standard errors in parentheses, clustered around company identifiers. $^*p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

Table A.16: Impact of Resources on # New Candidates, by Firm Experience

(A) Experienced Firms (top 25th percentile)

	Log(1 + New Candidates), by Similarity Decile									
	1	2	3	4	5	6	7	8	9	10
Post 2003 X Medicare Drug Life	0.050*	0.084*	0.051	0.026	0.091**	0.050*	0.055	0.049*	0.014	0.009
	(0.027)	(0.043)	(0.035)	(0.034)	(0.037)	(0.026)	(0.036)	(0.028)	(0.025)	(0.024)
R^2	0.171	0.276	0.274	0.308	0.317	0.241	0.220	0.202	0.195	0.133
Company FEs										
Qtr of Development FEs										
Overall Drug Life/Firm MMS										
Observations	11122	11122	11122	11122	11122	11122	11122	11122	11122	11122

(A) Less Experienced Firms (bottom 75th percentile)

	Log(1 + New Candidates), by Similarity Decile									
	1	2	3	4	5	6	7	8	9	10
Post 2003 X Medicare Drug Life	-0.018	0.011	-0.007	0.000	0.001	-0.005	0.050***	0.003	0.003	0.016**
	(0.017)	(0.010)	(0.008)	(0.002)	(0.002)	(0.006)	(0.017)	(0.004)	(0.005)	(0.006)
R^2	0.045	0.039	0.032	0.030	0.028	0.043	0.034	0.039	0.033	0.054
Company FEs										
Qtr of Development FEs										
Overall Drug Life/Firm MMS										
Observations	4040	4040	4040	4040	4040	4040	4040	4040	4040	4040

NOTES: Table A.16 reports the main specification coefficient for Post $\times$ Medicare Drug Life$_{f,2003}$. In Panel A, the sample includes only firms in the top 25th percentile of experience (number of drugs developed by 2003). The sample Panel B includes only the remaining firms in the bottom three quartiles of firm experience. All models include a full set of company and quarter indicator variables, with Post $\times$ Overall Drug Life$_{f,2003}$ and Post $\times$ Firm MMS$_{f,2003}$ both included as additional independent variables. Robust standard errors in parentheses, clustered around company identifiers. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.17: Impact of Resources on # New Candidates, Similarity within Indication

	1	2	3	4	5	6	7	8	9	10
					Log(1 + New Candidates), by Similarity Decile					
Post 2003 X Medicare Drug Life	0.050	0.089**	0.072*	0.094**	0.080**	0.092***	0.069**	0.103***	0.056*	0.030
	(0.036)	(0.044)	(0.040)	(0.041)	(0.038)	(0.030)	(0.033)	(0.034)	(0.032)	(0.024)
R^2	0.186	0.234	0.293	0.317	0.348	0.365	0.333	0.300	0.251	0.209
Company FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Qtr of Development FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Overall Drug Life/Firm MMS	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	16442	16442	16442	16442	16442	16442	16442	16442	16442	16442

NOTES: Table A.17 shows that our results are robust to alternative definitions of novelty: we compute drug similarities relative to all prior drug candidates that reached phase I trials and were developed for the same disease area as the focal drug. We report the main specification coefficient for Post×Medicare Drug Life$_{f,2003}$. Robust standard errors in parentheses, clustered around company identifiers. $^*p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.18: NEW BIOLOGICS

	(1)	Log(1 + New Biologics) (2)	(3)
	All	Past Exp.	No Past Exp.
Post 2003 X Medicare Drug Life	0.045	0.352**	0.007
	(0.048)	(0.152)	(0.012)
R^2	0.366	0.306	0.083
Company FEs	Yes	Yes	Yes
Qtr of Development FEs	Yes	Yes	Yes
Overall Drug Life/Firm MMS	Yes	Yes	Yes
Observations	16442	825	15609

NOTES: Table A.18 reports the main specification coefficient for Post $\times$ Medicare Drug Life$_{f,2003}$ but focuses on the development of biologics. The dependent variable is the log of one plus the number of new biologics introduced into development per company-quarter. New biologic drugs are identified through the Cortellis Investigational Drugs drug development histories. All models include a full set of company and quarter indicator variables, with Post$\times$Overall Drug Life$_{f,2003}$ and Post$\times$Firm MMS$_{f,2003}$ both included as additional independent variables, but not reported in the table. Column 1 includes all firms, while Columns 2 and 3 separate firms by whether or not they had developed biologic drugs prior to 2004. Robust standard errors in parentheses, clustered around company identifiers. $^*p < 0.10,^{**}p < 0.05,^{***}p < 0.01$.

Table A.19: IMPACT OF RESOURCES ON # NEW CANDIDATES, COMPANY TIME TRENDS

	Log(1 + New Candidates), by Similarity				
	All	Q 1	Q 2	Q 3	Q 4
Post 2003 X Medicare Drug Life	0.174*	0.116**	0.095*	0.074	0.010
	(0.099)	(0.057)	(0.049)	(0.050)	(0.042)
R^2	0.644	0.471	0.527	0.432	0.339
Company FEs	Yes	Yes	Yes	Yes	Yes
Qtr of Development FEs	Yes	Yes	Yes	Yes	Yes
Patent Life/Firm MMS X Post	Yes	Yes	Yes	Yes	Yes
Company-Qtr Trends	Yes	Yes	Yes	Yes	Yes
Observations	16442	16442	16442	16442	16442

NOTES: Table A.19 shows that our results are not driven by company-specific trends. The table reports the main specification coefficient for Post $\times$ Medicare Drug Life$_{f,2003}$. The outcome variable in the first models includes all new drug candidates, while the other four models limit the dependent variable to the count of new drug candidates that fall into the given similarity quartile. All models include a full set of company and quarter indicator variables, with Post $\times$ Overall Drug Life$_{f,2003}$ and Post $\times$ Firm MMS$_{f,2003}$ both included as additional independent variables, but not reported in the table. Additionally, these models include company-quarter indicator variables to capture any firm-specific time trends. Robust standard errors in parentheses, clustered around company identifiers. $^*p < 0.10,^{**}p < 0.05,^{***}p < 0.01$.

Table A.20: IMPACT OF RESOURCES ON # NEW CANDIDATES, POISSON QUASI MAXIMUM LIKELIHOOD

	# New Candidates, by Similarity				
	All	Q 1	Q 2	Q 3	Q 4
Post 2003 X Medicare Drug Life	0.790**	0.830	0.962**	0.693	0.631
	(0.389)	(0.593)	(0.445)	(0.514)	(0.577)
Post 2003 X Overall Drug Life	-0.397	-0.592	-0.312	0.208	-0.607
	(0.429)	(0.614)	(0.513)	(0.547)	(0.659)
Post 2003 X Firm MMS	-0.495	-0.147	-0.592	-0.125	-0.622
	(0.354)	(0.477)	(0.462)	(0.428)	(0.591)
R^2					
Company FEs	Yes	Yes	Yes	Yes	Yes
Qtr of Development FEs	Yes	Yes	Yes	Yes	Yes
Observations	15611	11136	10354	12319	12861

NOTES: Table A.20 reports the coefficients corresponding to those in our main specification, but obtained from a Poisson quasi-maximum likelihood estimation regression. The outcome variable in the first models includes all new drug candiates, while the other four models limit the dependent variable to the count of new drug candidates that fall into the given similarity quartile. All models include a full set of company and quarter indicator variables, with Post × Overall Drug Life$_{f,2003}$ and Post × Firm MMS$_{f,2003}$ both included as additional independent variables, but not reported in the table. One can interpret the coefficient from the first column (0.790) as a one unit change in Medicare drug life leading to a 79% increase in all new drug candidates. This coefficient translates into an elasticity of 0.43. QML (robust) standard errors in parentheses, clustered around company identifiers. $^*p < 0.10,^{**}p < 0.05,^{***}p < 0.01$.

Table A.21: IMPACT OF RESOURCES ON # NEW CANDIDATES, BINARY TREATMENT

	Log(1 + New Candidates), by Similarity				
	All	Q 1	Q 2	Q 3	Q 4
Post 2003 X Above Median	0.167***	0.111***	0.079**	0.084**	0.065**
Medicare Drug Life	(0.059)	(0.040)	(0.035)	(0.035)	(0.028)
Post 2003 X Overall Drug	-0.138**	-0.104**	-0.060*	-0.054	-0.062**
Life	(0.063)	(0.041)	(0.035)	(0.036)	(0.030)
Post 2003 X Firm MMS	-0.048	-0.014	-0.019	-0.014	-0.012
	(0.042)	(0.022)	(0.018)	(0.020)	(0.020)
R^2	0.596	0.397	0.480	0.386	0.301
Company FEs	Yes	Yes	Yes	Yes	Yes
Qtr of Development FEs	Yes	Yes	Yes	Yes	Yes
Observations	16442	16442	16442	16442	16442

NOTES: Table A.21 shows our results are robust to a less parametric definition of the treatment variable, given that treatment might not be linear in medicare drug life because many of our firms have a Medicare exposure of 0 or 1. We define a binary treatment depending on whether our treatment variable is above or below the median. $^*p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

Table A.22:

Impact of Resources on # New Candidates, Alternative Definitions of Remaining Exclusivity

(A) 7 Year Threshold for Remaining Drug Life

	Log(1 + New Candidates), by Similarity				
	All	Q 1	Q 2	Q 3	Q 4
Post 2003 X Medicare Drug Life	0.236**	0.106**	0.093**	0.118**	0.028*
	(0.098)	(0.054)	(0.046)	(0.052)	(0.037)
Post 2003 X Overall Drug Life	-0.214**	-0.101*	-0.075*	-0.090*	-0.030*
	(0.098)	(0.053)	(0.047)	(0.052)	(0.037)
Post 2003 X Firm MMS	-0.056*	-0.020*	-0.022*	-0.015*	-0.016*
	(0.042)	(0.022)	(0.017)	(0.020)	(0.019)
R^2	0.595	0.394	0.479	0.385	0.300
Company FEs	Yes	Yes	Yes	Yes	Yes
Qtr of Development FEs	Yes	Yes	Yes	Yes	Yes
Observations	16442	16442	16442	16442	16442

(A) 10 Year Threshold for Remaining Drug Life

	Log(1 + New Candidates), by Similarity				
	All	Q 1	Q 2	Q 3	Q 4
Post 2003 X Medicare Drug Life	0.249**	0.107*	0.111**	0.129**	0.048
	(0.103)	(0.056)	(0.048)	(0.059)	(0.040)
Post 2003 X Overall Drug Life	-0.218**	-0.110**	-0.092*	-0.103*	-0.039
	(0.105)	(0.055)	(0.049)	(0.061)	(0.041)
Post 2003 X Firm MMS	-0.052	-0.021	-0.020	-0.013	-0.014
	(0.043)	(0.022)	(0.016)	(0.020)	(0.020)
R^2	0.595	0.394	0.479	0.385	0.300
Company FEs	Yes	Yes	Yes	Yes	Yes
Qtr of Development FEs	Yes	Yes	Yes	Yes	Yes
Observations	16442	16442	16442	16442	16442

NOTES: Table A.22 shows that our results are robust to different definitions of the threshold for having long remaining patent life. $^*p < 0.10,$ $^{**}p < 0.05,$ $^{***}p < 0.01.$

Table A.23: IMPACT OF RESOURCES ON # NEW CANDIDATES, ANY DEVELOPMENT

	Any New Candidates, by Similarity				
	All	Q 1	Q 2	Q 3	Q 4
Post 2003 X Medicare Drug Life	0.187**	0.130***	0.113**	0.108**	0.068*
	(0.078)	(0.048)	(0.048)	(0.053)	(0.037)
Post 2003 X Overall Drug Life	-0.166**	-0.123**	-0.091*	-0.070*	-0.063*
	(0.078)	(0.049)	(0.048)	(0.055)	(0.039)
Post 2003 X Firm MMS	-0.046*	-0.015*	-0.018*	-0.010*	-0.011*
	(0.040)	(0.023)	(0.018)	(0.023)	(0.023)
R^2	0.400	0.313	0.387	0.306	0.250
Company FEs	Yes	Yes	Yes	Yes	Yes
Qtr of Development FEs	Yes	Yes	Yes	Yes	Yes
Observations	16442	16442	16442	16442	16442

NOTES: Table A.23 shows that our results are robust to considering a binary dependent variable and are not driven purely by the intensive margin. $^*p < 0.10,$ $^{**}p < 0.05,$ $^{***}p < 0.01.$

Table A.24: IMPACT OF RESOURCES ON # NEW CANDIDATES, TOTAL PATENT LIFE CONTROLS

	Log(1 + New Candidates), by Similarity				
	All	Q 1	Q 2	Q 3	Q 4
Post 2003 X Medicare Drug Life	0.180***	0.119***	0.105***	0.111***	0.038**
	(0.035)	(0.021)	(0.019)	(0.019)	(0.019)
Post 2003 X Log(1 + Total Patent Life)	-0.085***	-0.066***	-0.051***	-0.048***	-0.021**
	(0.014)	(0.010)	(0.010)	(0.009)	(0.008)
Post 2003 X Firm MMS	-0.036	-0.004	-0.011	-0.006	-0.010
	(0.039)	(0.020)	(0.016)	(0.019)	(0.020)
R^2	0.604	0.417	0.490	0.396	0.302
Company FEs	Yes	Yes	Yes	Yes	Yes
Qtr of Development FEs	Yes	Yes	Yes	Yes	Yes
Observations	16442	16442	16442	16442	16442

NOTES: Table A.24 shows that our results are robust to alternative specifications that control for the overall length of remaining patents. Specifically, we control for the total patent life instead of proportion of drugs on patent – this controls for the differential effect of part D by scale of firm more directly than controlling for the proportion of drugs with patent life remaining. $^*p < 0.10,^{**} p < 0.05,^{***} p < 0.01$.

Table A.25: IMPACT OF RESOURCES ON # NEW CANDIDATES, EXTREME TREATMENT VALUES EXCLUDED

	Log(1 + New Candidates), by Similarity				
	All	Q 1	Q 2	Q 3	Q 4
Post 2003 X Medicare Drug Life	0.303**	0.130*	0.098*	0.143**	0.110*
	(0.141)	(0.077)	(0.063)	(0.068)	(0.056)
Post 2003 X Overall Drug Life	0.111*	0.043*	0.035*	0.134*	0.077*
	(0.166)	(0.085)	(0.084)	(0.080)	(0.067)
Post 2003 X Firm MMS	-0.179*	-0.143*	-0.061*	-0.089*	0.048*
	(0.167)	(0.088)	(0.088)	(0.082)	(0.076)
R^2	0.621	0.406	0.478	0.400	0.322
Company FEs	Yes	Yes	Yes	Yes	Yes
Qtr of Development FEs	Yes	Yes	Yes	Yes	Yes
Observations	6208	6208	6208	6208	6208

NOTES: Table A.25 shows that our results are robust to excluding firms with extreme values of Medicare exposure of 0 or 1. $^*p < 0.10,^{**}p < 0.05,^{***}p < 0.01$.