Online Appendix to "Developing Novel Drugs"

Joshua Krieger[32] Danielle Li[33] Dimitris Papanikolaou[34]

May 2, 2018

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

A 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 market imperfections in drug development.

A.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.[35]

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

[35] 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.

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.

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).

A.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, Grabowski, and Hansen (2016) estimate that the direct cost of developing a single approved drug is

over $1.4 billion and has been increasing over time.[36] This total cost of development is spread unevenly across 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

[36]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`.

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.[37] 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).

A.3 A simple model of financial frictions and drug development choice

Time is discrete and there is an infinite number of periods. Consider a firm that discounts the future at a rate β, with a cashflow stream θX_t that has decided to invest I in developing a new drug. The firm has a choice between developing a novel (N) or a me-too (M) drug. For simplicity, assume that either drug, if it is developed at time t will only generate excess profits at $t+1$, after which competition eliminates profits. If the firm decides to develop a me-too drug, then development is always successful and the incremental drug yields profits equal to X_t. Novel drugs generate a cashflow equal to λX_t, where $\lambda \geq 1$, but development is only successful with probability π. Developing either drug costs I. Each period, the firm chooses the type of drug it will develop.

Each period the firm realizes profits from newly developed novel n_t or me-too m_t drugs, along with profits from other sources (previously developed drugs whose patents have not expired). In period t, firm profits are:

$$\pi_t = (\theta + \lambda n_t + m_t) X_t. \tag{7}$$

Financial frictions take the form of convex costs of external finance. Specifically, if the firm needs to raise external funds e, then it needs to pay a cost $c(e)$ that is increasing and (weakly) convex and is paid by the firm owners. Specifically, we assume $c(e) = e$ if $e \leq 0$ and $c(e) > e$ and is strictly convex if $e > 0$. That is, if profits exceed the cost

[37] 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.

of investment I, in which case $e < 0$, then owners obtain a dividend equal to $-e$. By contrast, if the firm needs to raise an amount equal to e, then the cost to firm owners is $c(e)$.

Given the form of the financing friction, we can write the firm's flow profits, net of investment and financing costs, as

$$\pi_t = u\left(\left(\theta + \lambda n_t + m_t\right) X_t - I_t\right), \tag{8}$$

where $u(z) = -c(-z)$ is an increasing and (weakly) concave function. Equation (8) illustrates the main intuition from models with (smooth) financing frictions, for instance (Froot et al., 1993): convex costs of external finance are on some level isomorphic to risk aversion. To see this, consider the firm's decision to develop a novel, versus a me-too, drug at time t.

The firm will develop a novel drug at time t if

$$E_t\, u\Big(\left(\theta + \lambda \tilde{z}_t\right) X_{t+1} - I\Big) \geq u\Big(\left(\theta + 1\right) X_{t+1} - I\Big), \tag{9}$$

where $\tilde{z}$ is a random variable that takes the value 1 with probability π. Approximating around $\tilde{z} = \bar{\pi} \equiv 1/\lambda$, we get

$$\begin{aligned} u\Big(\left(\theta + \lambda \tilde{z}_t\right) X_{t+1} - I\Big) \approx u\Big(\left(\theta + 1\right) X_{t+1} - I\Big) + u'\Big(\left(\theta + 1\right) X_{t+1} - I\Big)\left(\tilde{z} - \bar{\pi}\right) \\ + \frac{1}{2} u''\Big(\left(\theta + 1\right) X_{t+1} - I\Big)\left(\tilde{z} - \bar{\pi}\right)^2 \end{aligned}$$

Replacing the term inside the expectation in (9), we see that the firm will develop a novel drug as long as

$$\left(\pi - \bar{\pi}\right) \geq -\frac{1}{2} \frac{u''\Big(\left(\theta + 1\right) X_{t+1} - I\Big)}{u'\Big(\left(\theta + 1\right) X_{t+1} - I\Big)} E\left(\tilde{z} - \bar{\pi}\right)^2 \tag{10}$$

Examining Equation (10), we see that the firm will develop a novel drug as long as the expected benefit of doing so $\hat{\pi} - \pi$ (which is positive as long as $\lambda\pi > 1$) is sufficiently high to compensate the firm for the risk of failure. The firm acts as if it were risk averse because it internalizes the fact that if the drug is unsuccessful, its internal funds may be insufficient to cover its development costs next period, which would imply that it would need to raise costly external funds.

Importantly, the decision to develop a novel or me-too drugs depends on expectations about *future* profitability. If the firm anticipates higher future profits, then the likelihood that it will need to raise (costly) external funds tomorrow decreases, which increases the incentive to undertake risk today.

It is worth emphasizing that these predictions are more general than this particular model. For instance, consider a different version in which more novel drugs are more expensive to finance using external funds—that is, the costs of external finance $c(e)$ depend on the type of drug being developed. One reason why this might be the case, based on the intuition put forth by Myers and Majluf (1984), is that the degree of information 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.

B Data Construction

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

B.1 Drug Development Histories

Our drug development data primarily comes from the Cortellis Investigational Drugs and Clinical Trials databases.[38] 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

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

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 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.

B.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..

B.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.[39] 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.[40]

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.

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

[40] 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.

B.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 3.2), and the latter helps us describe the correlation between our novelty measure and private value to drug developers (see Section 2.3).

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 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 correllations 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).

B.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.

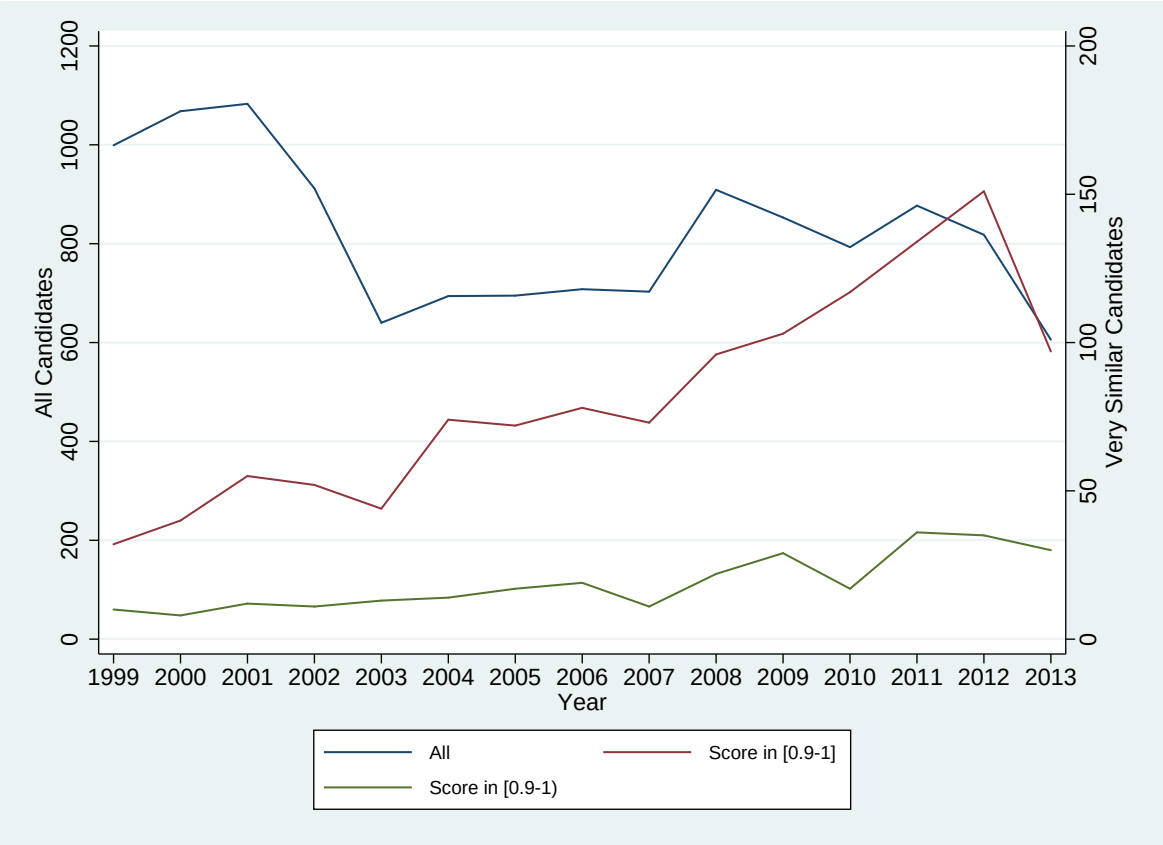
B.6 Public Firms

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.

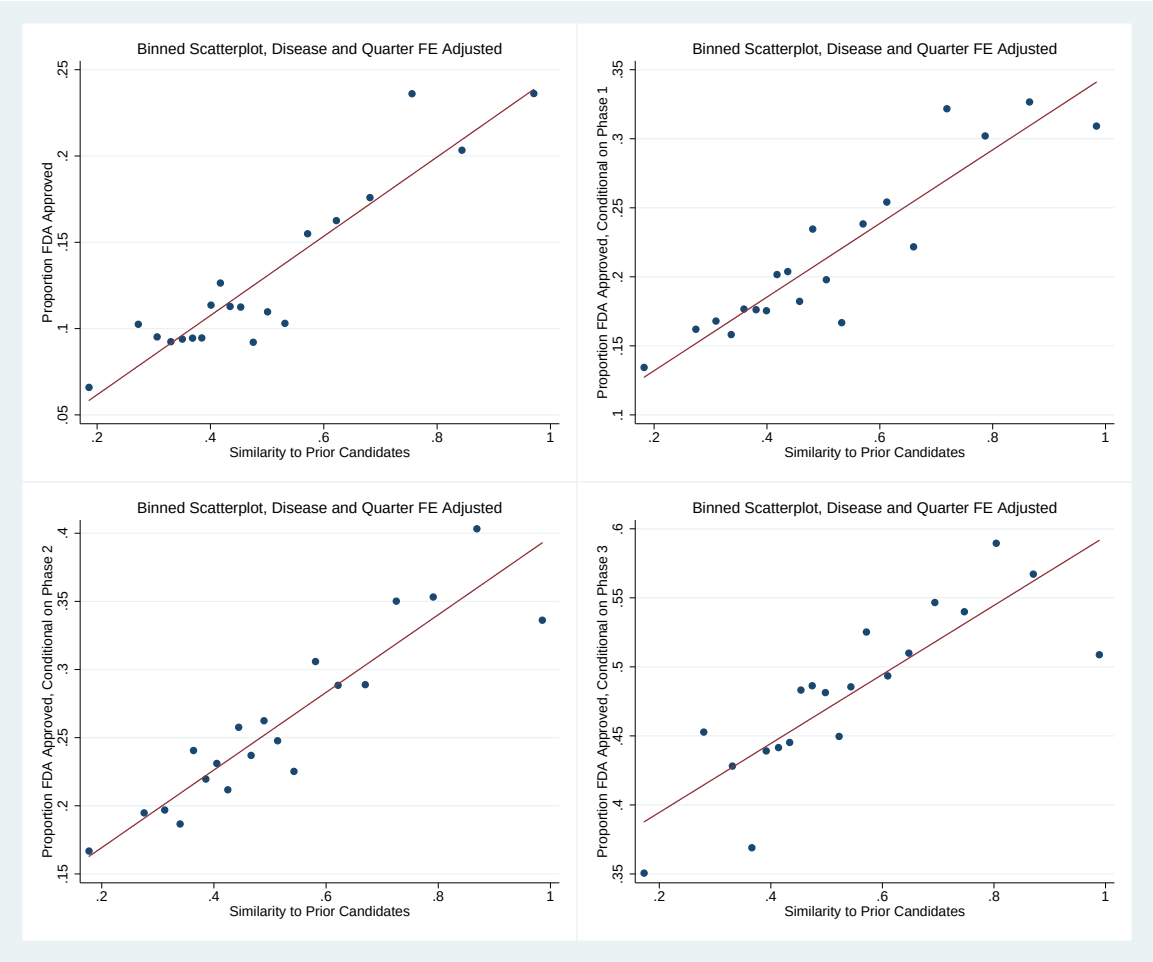
Appendix Tables and Figures

Figure A.1: # of Drug Candidates over Time



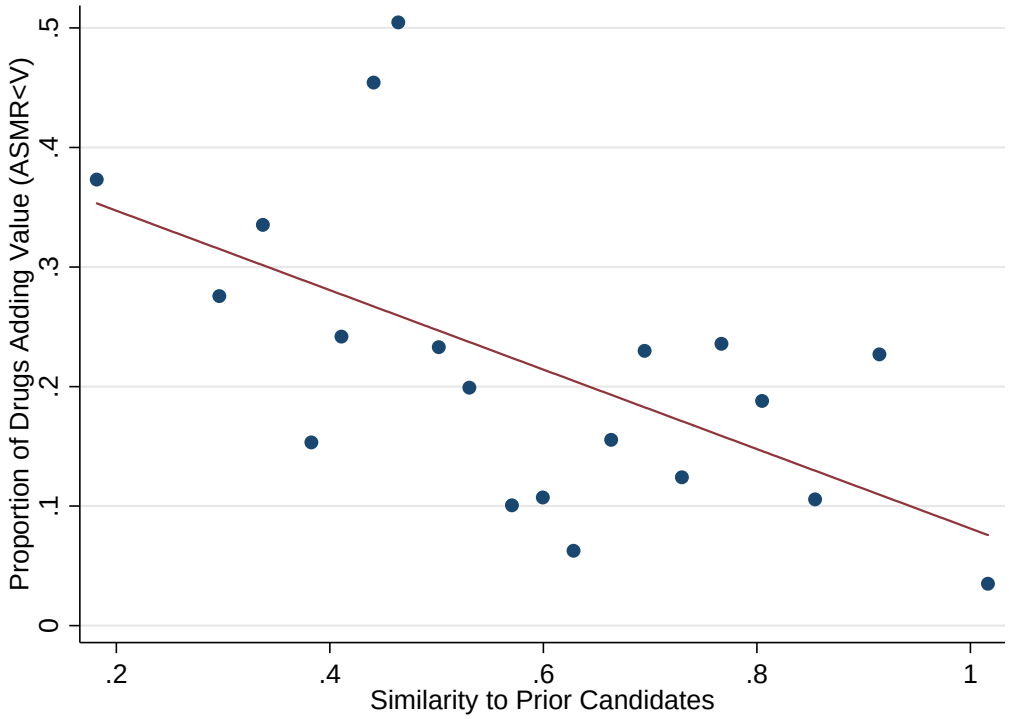
Notes: This figure plots the number of new drug candidates for which we have data on molecular structure over time. The blue line all drug candidates. The red line represents drugs with similarity scores greater than 0.9, which indicates over 90% overlapping chemical structures. The green line plots 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



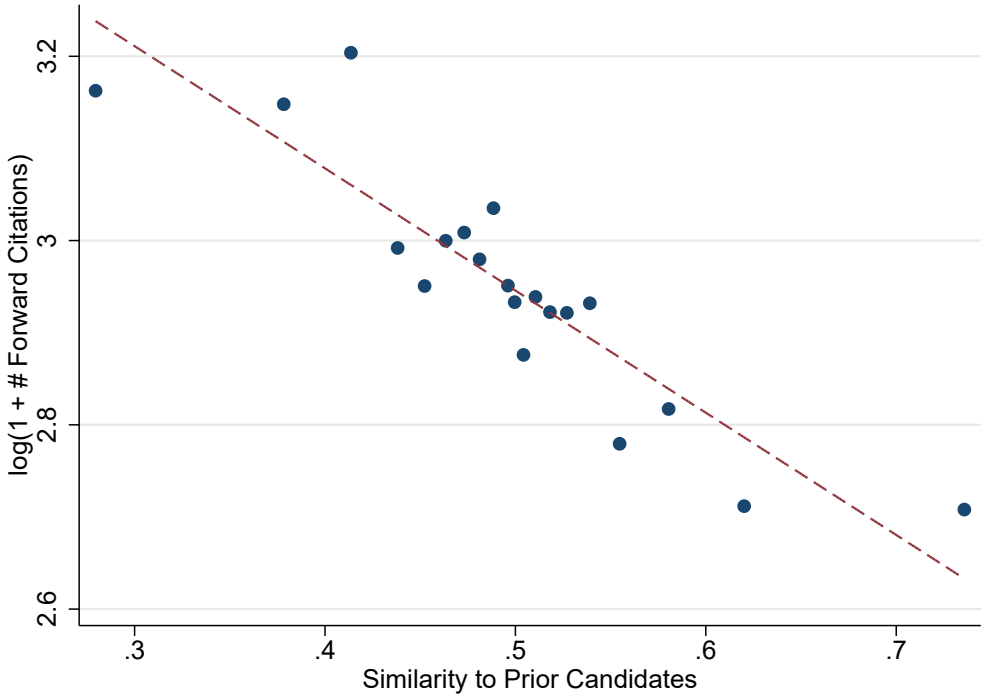
Notes: Figure A.2 presents binned scatterplots of drug-level similarity against whether a drug is FDA approved. Each dot represents the proportion of candidates that FDA approved, among all candidates within a given similarity score bin, conditional on disease (ICD9) and quarter of development fixed 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



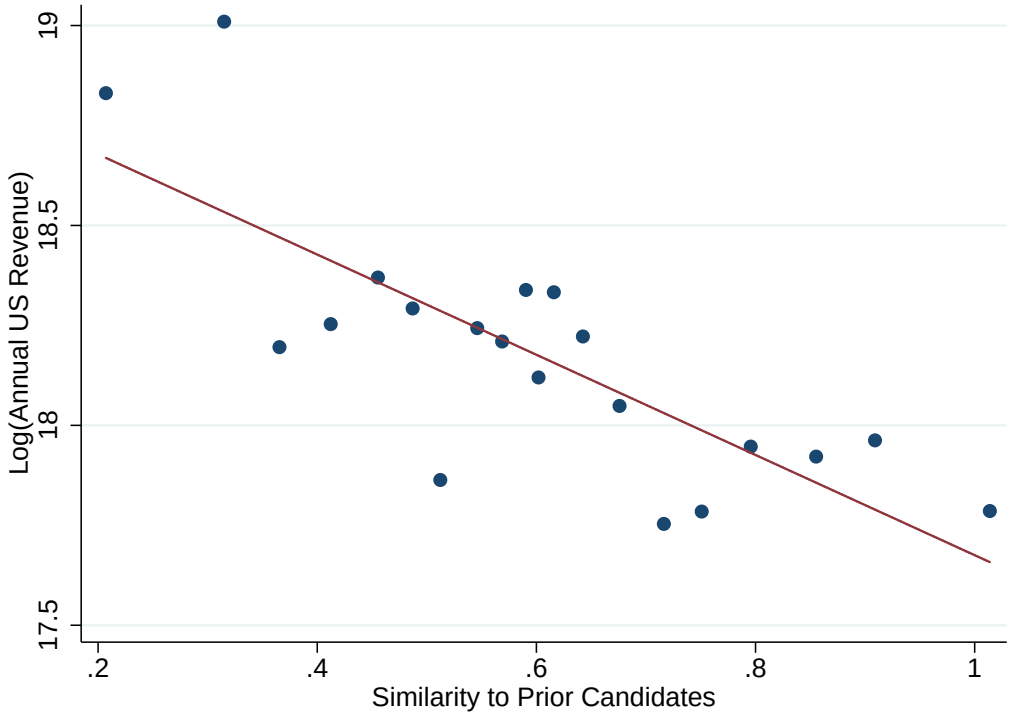
Notes: Figure A.3 presents a binned scatterplot of drug-level similarity against drug added benefits. A drug's added benefit is derived from the from the French Haute Autorité de Santé (HAS) health system's clinical added benefits scores (Amélioration du Service Medical Rendu, or ASMR), which range from 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



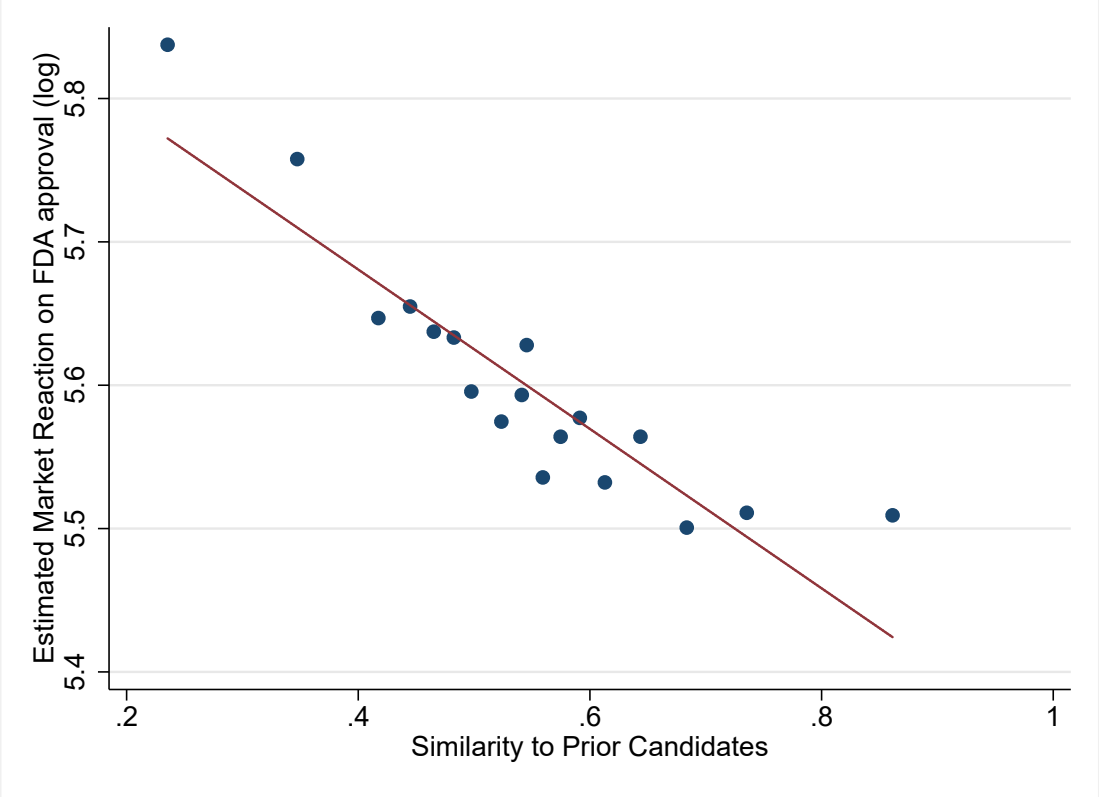
Notes: Figure A.4 presents a binned scatterplot of drug-level similarity against the logarithm of one plus the number of forward citations the patent receives. Each dot represents mean log value, among all candidates within a given similarity score bin, conditional on disease (ICD9); patent issue year; 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



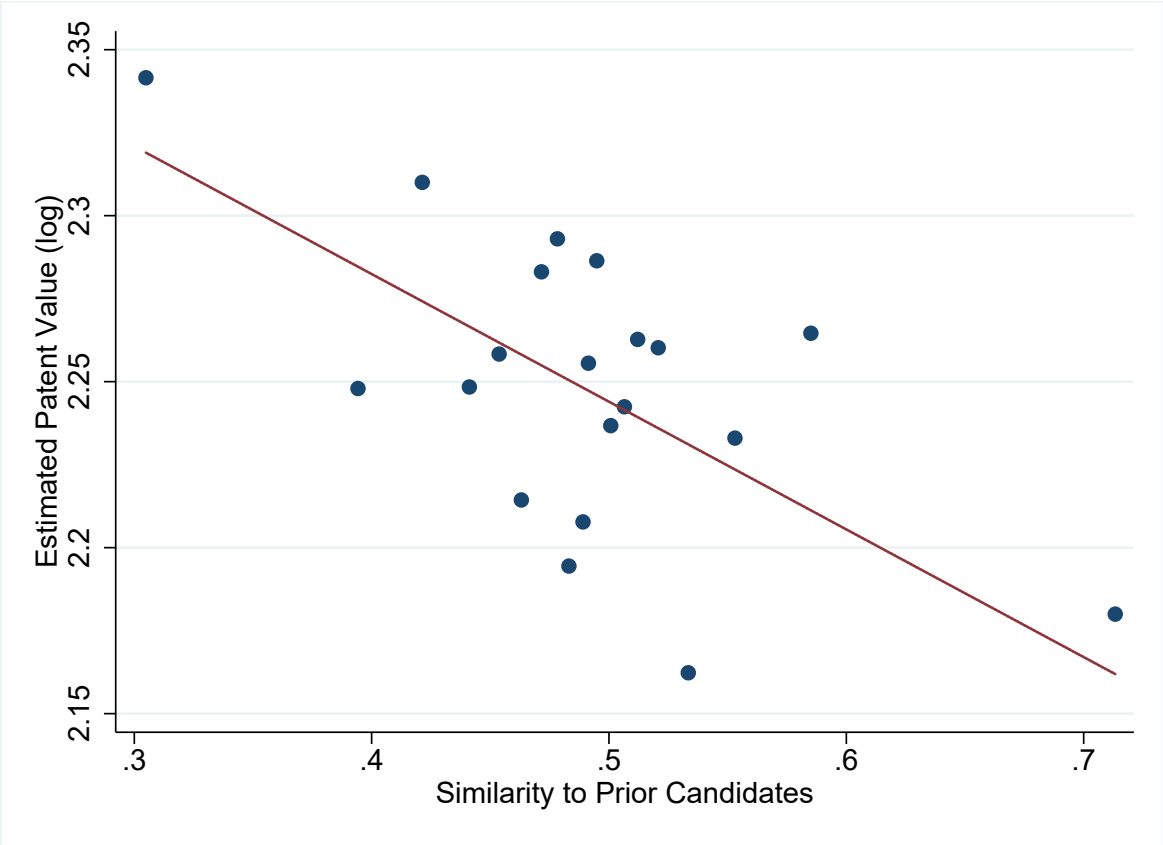
Notes: Figure A.5 presents a binned scatterplot of drug-level similarity against revenue conditional on approval. The plot corresponds to the regression in Column (4) of Table A.8, which includes controls for drug indication, drug age, and firm dummies.

Figure A.6: Drug Similarity and Stock Market reaction on FDA Approval



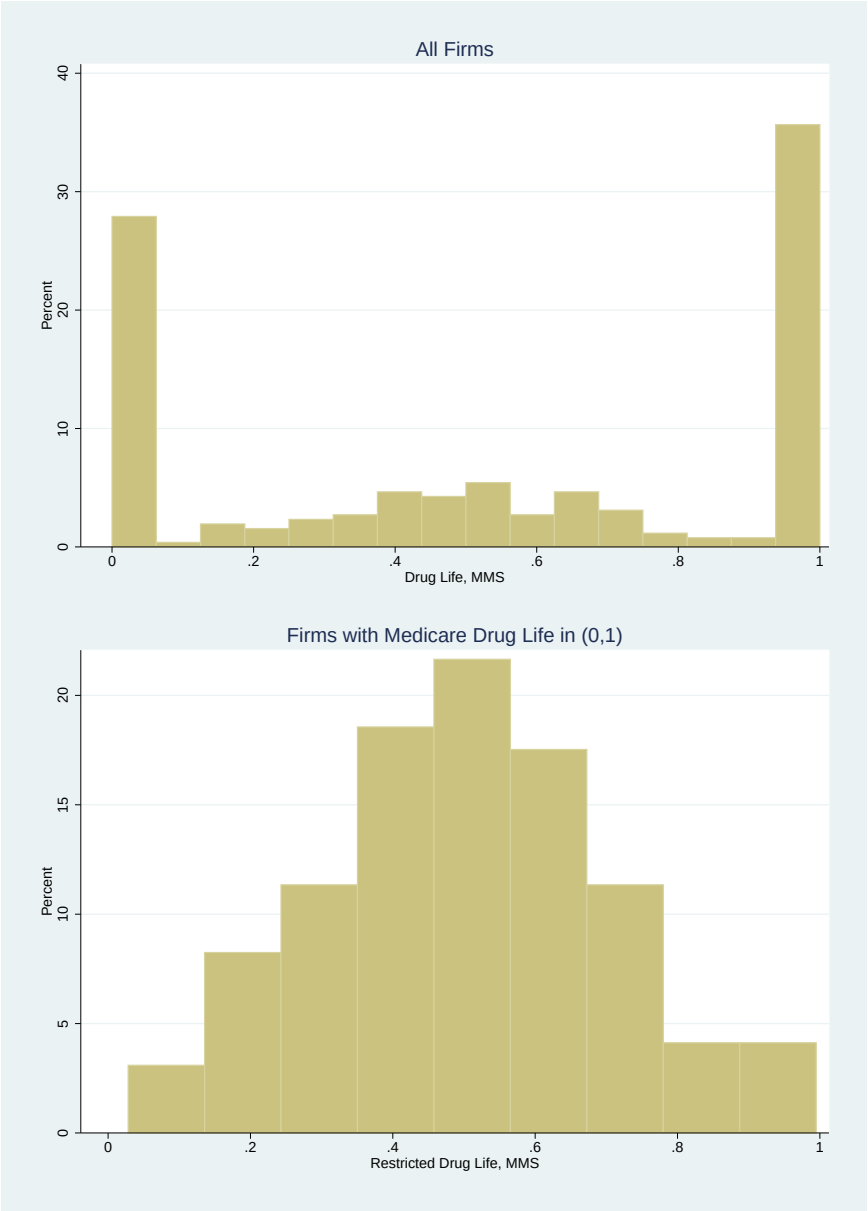
Notes: Figure A.6 presents a binned scatterplot of drug-level similarity against the logarithm of the estimated dollar reaction on the (first) approval of the drug by the FDA. The dollar reaction to the FDA approval is estimated following the methodology of Kogan et al. (2017) and uses a 5-day window 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



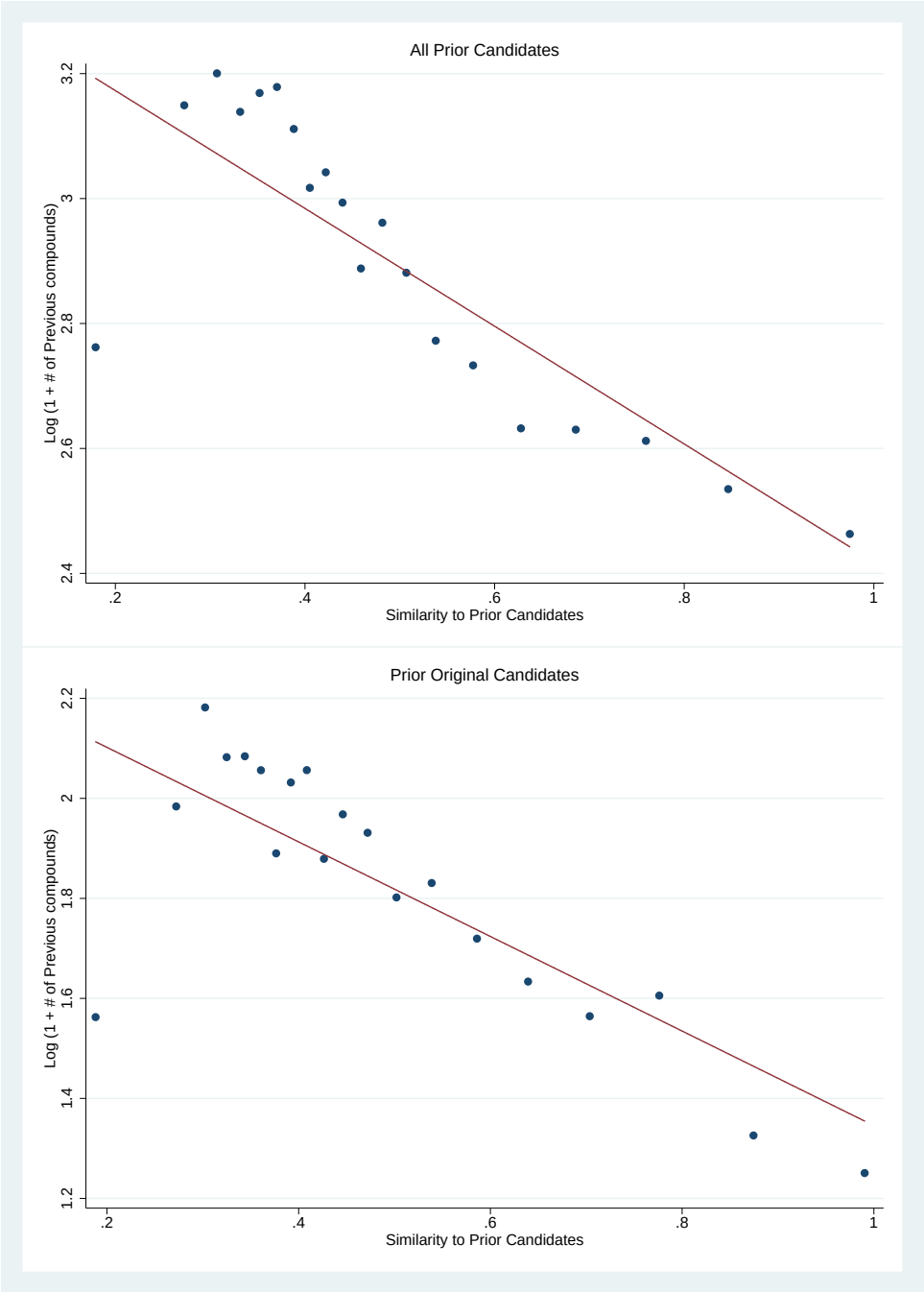
Notes: Figure A.7 presents a binned scatterplot of drug-level similarity against the logarithm of the Kogan et al. (2017) estimated patent values. Each dot represents mean log value, among all candidates within a given similarity score bin, conditional on disease (ICD9); patent issue year; and year of 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



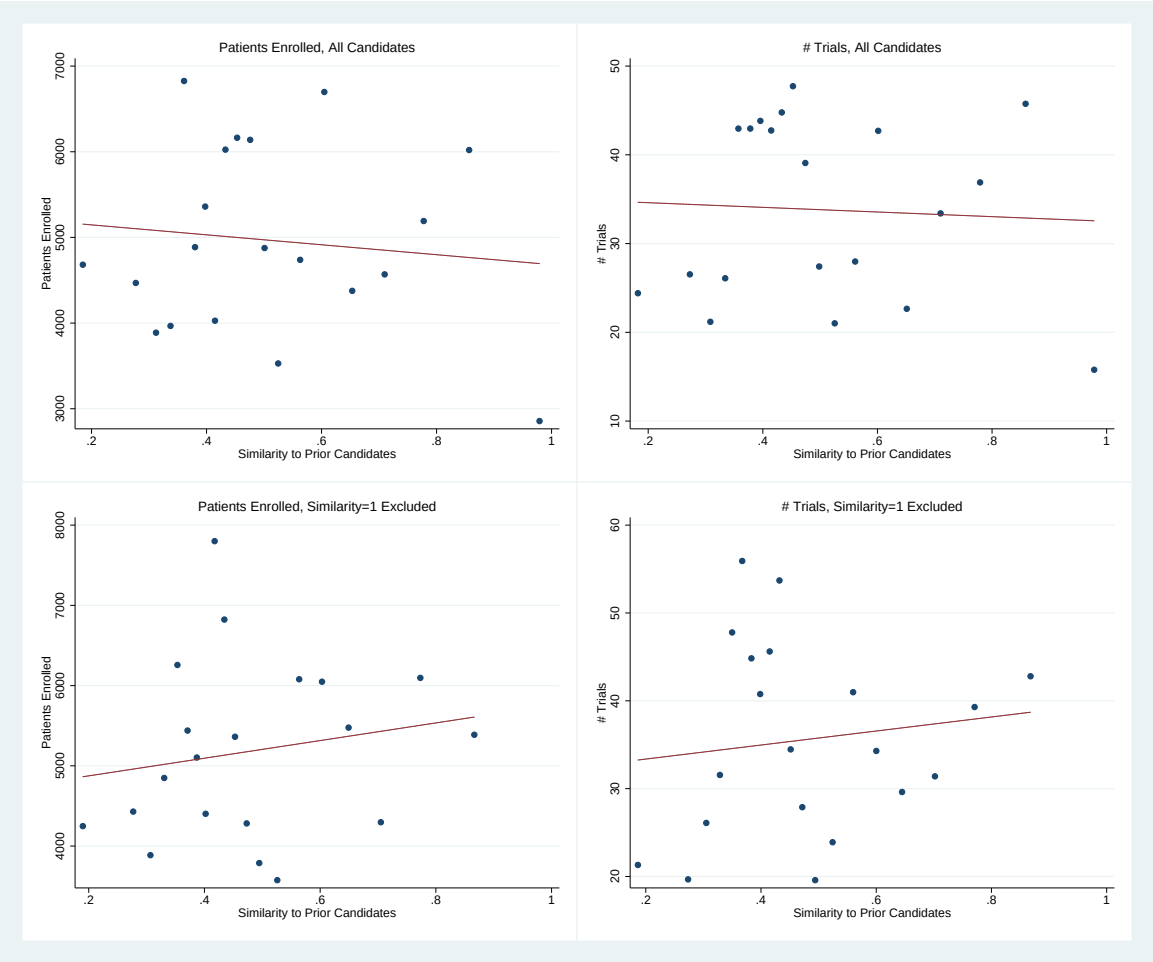
Notes: Figure A.8 plots the distribution of Medicare Drug Life in 2003. Each observation is a firm in our main analysis sample.

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.

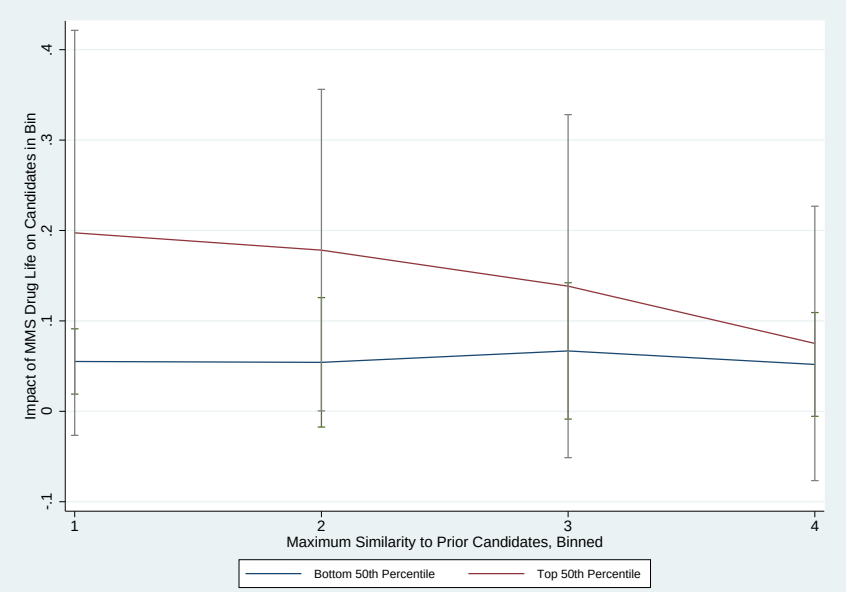
Figure A.10: Proxies for Development Costs, by Drug Similarity



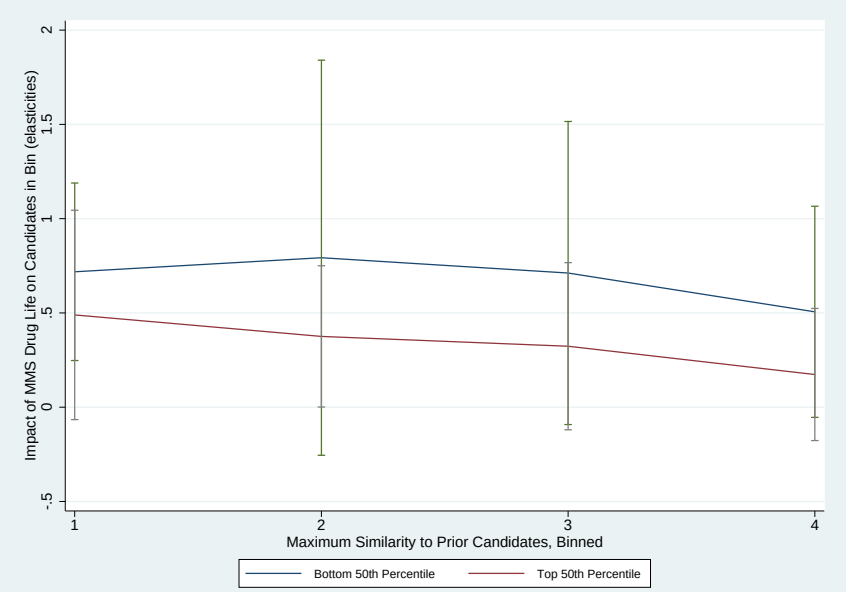
Notes: Figure A.10 presents a binned scatterplot of drug-level similarity against proxies for the direct cost of drug development. Each dot represents the mean number of patients enrolled (or number of trials conducted), among all candidates within a given similarity score bin, conditional on disease (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

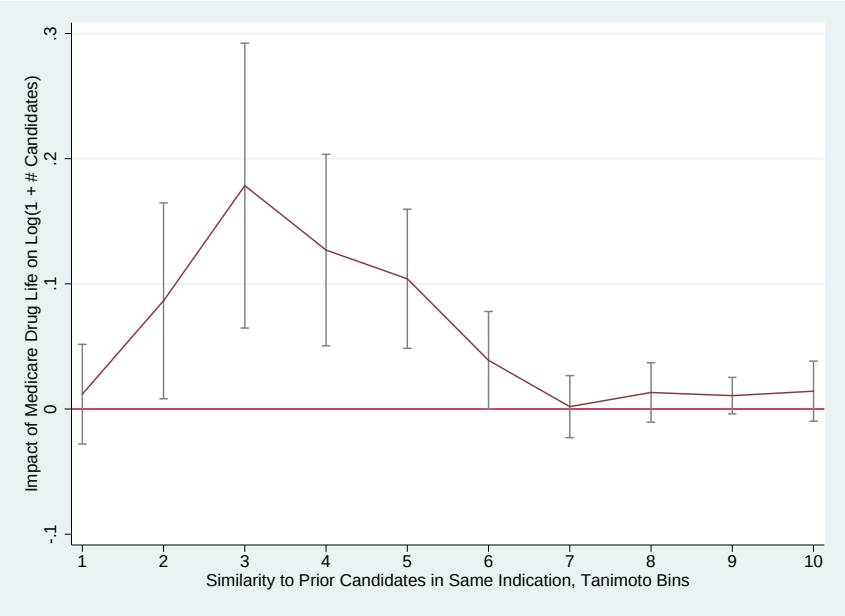


B. Elasticities



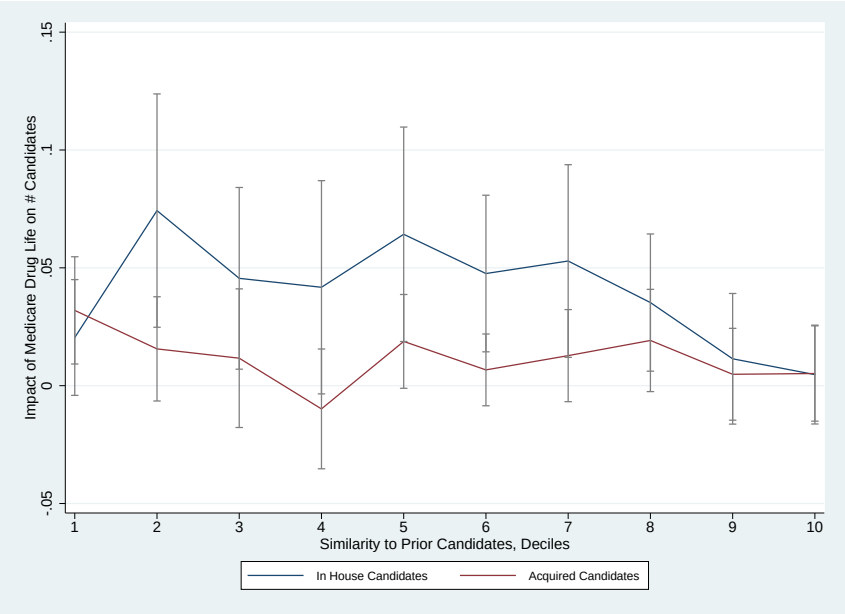
Notes: Figure A.12 plots the estimated coefficients on Post × Medicare Drug Life$_{f,2003}$ from our main regression specification defined by Equation (3) across firm size groups (defined by total revenue generated by approved drugs prior to 2003). The outcome variable is number of drug candidates across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

Figure A.12: Impact of Additional Resources on Novelty, within Indication



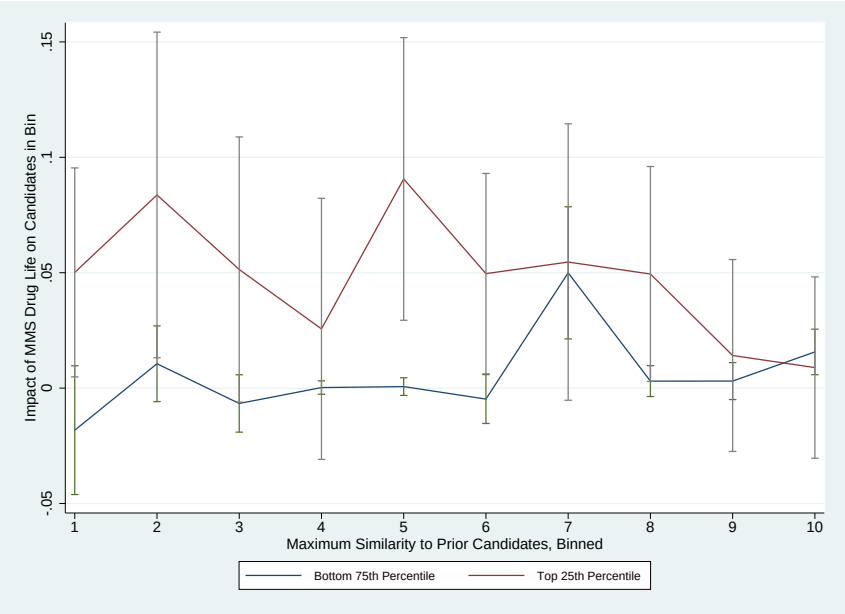
Notes: Figure A.12 plots the estimated coefficients on Post $\times$ Medicare Drug Life$_{f,2003}$ from our main regression specification defined by Equation (3). This figure is analogous to the bottom panel of Figure 5 of the main text, except that similarity is measured with respect to other drugs in the same indication (disease).

Figure A.13: Original vs. Acquired



Notes: Figure A.13 plots the estimated coefficients on Post $\times$ 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 $\times$ 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)
Maximum Similarity	-580.394	-2.621
	(509.676)	(3.550)
R^2	0.194	0.170
Qtr of Development FEs	Yes	Yes
ICD-9 FEs	Yes	Yes
Observations	8801	10546

(b) Candidates with Similarity Score < 1

	Patients Enrolled	# of Trials
	(1)	(2)
Maximum Similarity	1099.570*	7.982*
	(633.287)	(4.418)
R^2	0.201	0.175
Qtr of Development FEs	Yes	Yes
ICD-9 FEs	Yes	Yes
Observations	8280	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). Panel B excludes candidates with similarity scores of exactly 1, which may include extended release formulations that require fewer additional 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 A.10. $^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$.

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
R^2	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 $\times$ 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 $\times$ 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.15: Firm Experience

	Log(1 + New Candidates), by Experience		
	(1) All	(2) Top 25	(3) Bottom 75
Post 2003 X Medicare Drug Life	0.263***	0.260**	0.053
	(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 × 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 × Overall Drug Life$_{f,2003}$ and Post × 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 × 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 bottome three quartiles of firm experience. 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.17: Impact of Resources on # New Candidates, Similarity within Indication

	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	0.089**	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 $\times$ Medicare Drug $\text{Life}_{f,2003}$. Robust standard errors in parentheses, clustered around company identifiers. $^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$.

Table A.18: New Biologics

	Log(1 + New Biologics)		
	(1)	(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: New Targets

	Log(1 + New Target drugs)			
	(1) New Target-Actions	(2) Coarse Target (6-levels)	(3) Coarser Target (5-levels)	(4) Novel Target Score
Post 2003 X Medicare Drug Life	0.039*	0.021*	0.016**	0.024*
	(0.021)	(0.011)	(0.007)	(0.013)
R^2	0.237	0.123	0.097	0.156
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 A.19 reports the main specification coefficient for Post × Medicare Drug Life$_{f,2003}$. All new drugs, including both small molecules and biologic drugs are included in the dependent variable counts. The dependent variable in Column 1 is the log of one plus the number of drugs that the focal firm developed (in the given quarter) for new molecular target-actions. We define drugs with "new" target-actions as drugs that were the first drug candidate (chronologically across all firms) developed to treat any condition via the given target-action. The dependent variables in Columns 2 and 3 use coarser definitions of targets, based on the Cortellis target tree ontology. The "coarse" definition of targets in Column 2 counts the log of one plus the number of new drugs that were the first entrant to a target group six levels deep into the Cortellis target tree, while the "coarser" outcome in Column 3 is the same but for target groups five levels into the Cortellis ontology. Column 4 defines new target drugs as those in the top 10% of a "target novelty" score. This score is based off target tree position and entry order for targets associated with a given drug. 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. Robust standard errors in parentheses, clustered around company identifiers.

Table A.20: 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.20 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.21: 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.21 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 $\times$ Overall Drug Life$_{f,2003}$ and Post $\times$ Firm MMS$_{f,2003}$ both included as additional independent variables, but not reported in the table. 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.22: 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 Medicare Drug Life	0.167***	0.111***	0.079**	0.084**	0.065**
	(0.059)	(0.040)	(0.035)	(0.035)	(0.028)
Post 2003 X Overall Drug Life	-0.138**	-0.104**	-0.060*	-0.054	-0.062**
	(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.22 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.23:

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.23 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.24: 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.24 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.25: 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.25 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.26: 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.26 shows that our results are robust to excluding firms with extreme values of Medicare exposure of 0 or 1. $^{*}p < 0.10$,** $p < 0.05$,*** $p < 0.01$.