Missing Markets for Innovation: Evidence from New Uses for Existing Drugs*

Eric Budish Maya Durvasula Benjamin Roin Heidi Williams

November 2026

Abstract

We investigate missing markets for innovation: contexts where the intellectual property rights necessary to incentivize investment in research and development are either unavailable or unenforceable. A simple model characterizes the conditions under which these gaps arise and illustrates why the canonical "Nordhaus parameters" emphasized in the literature—patent length and patent breadth—cannot restore incentives. We identify an empirical setting—research into new uses for existing drugs—with sharp variation in the enforceability of intellectual property rights on otherwise comparable inventions. Using newly constructed data tracing research investment, commercialization, and intellectual property protection, we show that when intellectual property rights become unenforceable, commercialization essentially ceases, even as new scientific leads continue to arrive. We use variation in the duration of enforceable intellectual property protection across drugs to quantify the scale of the missing market for new uses and estimate that the social value of missing investment is on the order of trillions of dollars. This same variation allows us to provide the first direct estimate of the elasticity of research investment with respect to the duration of intellectual property protection. We discuss the implications of our findings for the design of innovation policy.

*_Budish_: University of Chicago & NBER (eric.budish@chicagobooth.edu); _Durvasula_: University of Chicago & U.S. Courts of Appeals (maya.durvasula@gmail.com); _Roin_: Independent (bnroin@gmail.com); _Williams_: Dartmouth College & NBER (heidi.lie.williams@dartmouth.edu). Budish and Durvasula are co-lead and corresponding authors. We are grateful to Pierre Azoulay (discussant), Daron Acemoglu, Oren Bar-Gill, Daniel Fetter, Amy Finkelstein, Peter Ganong, Matthew Gentzkow, Edward Glaeser, Daniel Hemel, Scott Hemphill, David Kamin, Louis Kaplow, Mark Lemley, Paul Milgrom, Abhishek Nagaraj, Matt Notowidigdo, Lisa Larrimore Ouellette, Ariel Pakes, Al Roth, Elizabeth Sepper, Carolyn Stein, Chad Syverson, Melissa Wasserman, Brian Wright, and seminar participants at AALS (Law & Economics), BFI Health, Chicago Booth, Harvard Law & Economics, MIT Sloan, NBER Assessing the U.S. Medical Innovation System, NBER Health, NBER Market Design, NBER Productivity Lunch, NBER Summer Institute (Law & Economics), Northwestern Kellogg, NYU Law, NYU Law & Economics, Penn Carey Law, Stanford, Texas Law, the Radcliffe Institute, TNIT, UC-Berkeley Haas, UCLA, Utah WBEC, and the Harvard/MIT IO seminar for comments; to Yali Friedman, Ben Berger, Amitabh Chandra, and Craig Garthwaite for data; and to Joey Anderson, Ryan Broll, Jeremy Brown, AnneMarie Bryson, Henry Manley, Tamri Matiashvili, Gideon Moore, Lia Petrose, Maya Roy, Mahnum Shahzad, Ralph Skinner, Liyang Sun, and Zahra Thabet for excellent research assistance. Research reported in this publication was supported by the National Institute on Aging (NIA) and the NIH Common Fund, Office of the NIH Director, through Grant U01-AG046708 to the National Bureau of Economic Research (NBER); the content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or NBER. This research was also funded in part by NIA grant P01AG005842, the Alfred P. Sloan Foundation, NSF Grant Number 1151497, the Polsky Center for Entrepreneurship and Innovation at the University of Chicago Booth School of Business, the Hultquist Faculty Research Endowment, the Knight Hennessy Scholars Program at Stanford University, the John M. Olin Program in Law and Economics at Stanford Law School, the B.F. Haley and E.S. Shaw Fellowship for Economics via the Stanford Institute for Economic Policy Research, the Sloan Foundation Graduate Fellowship on Innovation and Productivity via the NBER, and the National Science Foundation through grant DGE-1656518 (Durvasula) to Stanford University. While this paper was written, Williams was a visiting scholar and is currently working on contract as a research advisor to the Congressional Budget Office (CBO); the views expressed in this paper should not be attributed to CBO.

1 Introduction

Most inventions are far more costly to discover and develop than to imitate and reproduce. Economists have long recognized that, absent some mechanism that incentivizes firms to engage in these costly processes of research and development, competitive markets will under-provide socially valuable innovation (Nelson, 1959; Arrow, 1962; Nordhaus, 1969). Policies based on this idea go back at least half a millennium to 15th century Italy and are enshrined in the U.S. Constitution.[1]

In this paper, we investigate the idea that, for large classes of potentially valuable inventions, intellectual property rights that exist "on paper" are impossible to enforce in practice. When this is the case, a key mechanism used to induce private investment in research and development in fact provides no incentive for innovation. We provide evidence consistent with the idea that there are many scientifically feasible, socially valuable inventions that are missing as a result.[2]

Consider three hypothetical inventions: a rigorously tested regimen of nutritional supplements that reduces the risk of developing Alzheimer's disease for adults or that improves cognitive development for babies; a manufacturing process that would substantially increase the precision of pharmaceutical manufacturing and reduce the risk of large-scale medicine recalls; and a motorcycle helmet that outperforms existing alternatives on metrics of safety and comfort, available for sale on an online platform.[3] For each, an inventor could file for some form of intellectual property protection—likely, a patent from an entity like the U.S. Patent and Trademark Office.[4] Yet any resulting grants of intellectual property would offer little protection from imitation: in each case, inventors would find it nearly impossible to enforce those rights in practice.

Patents and other forms of intellectual property rights are not self-enforcing. An inventor must be able both to identify instances of infringement and to take effective legal action to stop it. But when it is difficult to determine, say, that someone is combining widely-available supplements in a way that infringes on a patent or that a bottle of pills was manufactured using a protected method,

[1] Article I, Section 8, Clause 8 ("[t]o promote the Progress of Science and useful Arts, by securing for a limited Time to Authors and Inventors the exclusive Right to their respective Writings and Discoveries").

[2] Note that we use the term "invention" in the same manner as the literature on economic growth, to refer to any instance of knowledge production, whether in the form of physical products, sets of instructions, or pieces of information (Arrow, 1962; Romer, 1994; Jones, 2005).

[3] We choose, for the sake of illustration, three types of socially valuable inventions for which the enforceability problem of interest to this paper can arise: medical inventions, business process inventions, and product inventions. For a discussion of the value of rigorous evaluation of health interventions, see Mozaffarian et al. (2011); for a discussion of the difficulty of establishing such an evidence base, see Oster (2024). For a discussion of the social costs of low-quality methods of pharmaceutical manufacturing, see Price (2014); for an analysis of the consequences of low pharmaceutical manufacturing quality—which can trigger contamination events and large-scale recalls—see Yurukoglu et al. (2017). For an example of the consequences of easy imitation of critical protective products like motorcycle helmets, see https://www.congress.gov/117/meeting/house/114092/documents/HMKP-117-JU00-20210929-SD217.pdf.

None of these products are amenable to protection with trade secrecy, as each must be publicly disclosed to end users and regulators as a condition of sale. In Section 2, we develop a framework that identifies instances in which trade secrecy is likely to be an effective incentive for innovation. Mansfield (1986) and Moser (2005) provide empirical evidence on the practical importance of secrecy.

[4] It may surprise some readers that all three kinds of inventions described in the text are eligible for patent protection. This reflects, in part, the distinction this paper draws between the availability of patent protection in principle and the enforceability of patent protection in practice. Section 2 provides a primer on the standards for patentability in the United States.

inventors will be unable to enforce their patents simply because they cannot observe infringement. And even when an inventor can observe, for instance, that dozens of small, overseas sellers are offering products that infringe her rights, a patent is of little use if there is no viable party against whom she can file suit.

This paper argues that these legal and practical gaps in intellectual property protection are widespread and shows that their social costs are significant. We develop a simple theoretical framework that uses the characteristics of inventions to identify settings in which gaps in intellectual property protection yield missing private incentives for investment. We identify one such setting—research into new therapeutic uses for existing drugs—where the enforceability of intellectual property rights on otherwise comparable inventions changes sharply over time and hence allows us to test the link between intellectual property rights and private incentives central to this paper. When intellectual property rights become unenforceable, we show that research and development essentially cease.

We begin by formalizing the heterogeneity that generates these differences in enforceability across inventions. We introduce a model in the style of Nordhaus (1969), in which intellectual property rights confer some period of time under monopoly. To the standard framework, we add three complications. First, some inventions can be profitably commercialized while remaining secret. In these cases, firms can earn an economic return on their investments in research and development without formal grants of intellectual property—so long as they can maintain secrecy. Second, only some inventions are eligible for formal intellectual property protection. By law, others are not. Third, of those inventions that are eligible for intellectual property rights, only some can be protected by rights that are enforceable in practice.

We say that a *missing market* arises when formal intellectual property rights are necessary to incentivize research and development, but such rights are either unavailable or unenforceable. Note that this characterization rules out a standard policy solution to the problem of inadequate incentives for innovation: canonical models linking intellectual property rights to incentives embed the idea that sufficiently strong property rights—that is, longer or broader periods of market exclusivity—can make any invention profitable to develop (e.g., Nordhaus, 1969; Gilbert and Shapiro, 1990; Klemperer, 1990). The key implication is that when gaps in incentives arise because intellectual property rights are completely unavailable or unenforceable, there are no "Nordhaus parameters" that can induce private investment.

This model allows us to identify large classes of potential inventions, across sectors of the economy, where missing markets are likely to arise. Our empirical analysis focuses, however, on a specific context—investments into new uses for existing drugs—where sharp variation in the enforceability of intellectual property rights over time allows us to test for evidence of a missing market.

In most missing markets identified by our model, intellectual property rights are unenforceable from the moment that they are granted. New uses for existing drugs provide an unusual setting in which this enforceability instead changes over time.

To fix ideas, consider a drug that has been developed and approved to treat some disease A. By the time it receives regulatory approval, it has already cleared two demanding hurdles: it has been shown to be sufficiently safe for use in humans and effective in treating at least one disease. Scientists often describe chemicals that satisfy these two criteria as therapeutically "promiscuous," reflecting the empirical regularity that chemicals that can successfully treat one disease without causing serious harm are more likely to prove useful in treating others (Wermuth, 2006). While some of these additional uses are anticipated at the outset, others are identified only years or decades later as advances in basic science or clinical experience reveal new opportunities.

Even if potential new uses are identified with substantial delay, firms can file for intellectual property rights that, say, protect the use of a given drug as a treatment for some disease B. In the United States, the U.S. Patent and Trademark Office routinely grants "method of use" patents on new uses, and the Food and Drug Administration issues "regulatory exclusivities" that provide a similar protection. But whether a firm can expect to earn a return on the investment associated with the lengthy clinical trials required to establish a drug's effectiveness in treating B depends entirely on when new scientific leads arise in a drug's lifecycle.

First, suppose that the firm that developed, patented, and sought regulatory approval of the drug for use A identifies potential new use B relatively early in the drug's life, while the drug remains protected from generic competition. If the firm files for formal intellectual property protection and successfully demonstrates that the drug is effective in treating B, then it can market the drug for both A and B and set a price accordingly. So long as the drug faces no generic competition—what we refer to as the drug's period of *exclusivity*—the firm can earn economic profits from any and all uses. During this period, intellectual property rights on new uses are thus fully enforceable.

Suppose instead that the firm identifies potential use B after the drug's exclusivity period has ended. On paper, the same formal intellectual property rights remain available. But in practice, the firm can no longer capture the value of the information that its drug is effective in treating B: any patient or physician can make use of the information about B freely—by using or prescribing the generic version—without compensating the firm that undertook the expensive clinical trial. Because a generic manufacturer may lawfully sell the same drug for A, the sale of the generic version does not reveal whether a patent on use B has been infringed. Even if the innovator could detect unauthorized use, enforcement would be practically impossible: the parties on whom liability would most clearly fall are the patients using the drug to treat their own illnesses, hardly an attractive litigation target. Formal intellectual property rights on B may then remain in force even as their practical enforceability drops essentially to zero.

Intellectual property rights on new uses thus switch from enforceable to unenforceable when the drug becomes available in generic form. This timing—a consequence of the regulatory structure for pharmaceuticals in the United States—provides our research design.

Mapping this variation to our framework yields three predictions. First, firms will have the strongest incentives to invest in new uses for existing drugs early in the drug's lifecycle. Second, as exclusivity expiration approaches for the underlying drug, the number of years in which invest-

ments in new uses will yield profits declines—and thus, incentives should fall. Third, exclusivity expiration creates a missing market for new uses, and private research and development should cease.

Testing these predictions is the focus of our empirical work. Doing so requires data on three objects: a sample of drugs whose initial approval can be observed; measures of subsequent investment in, and commercialization of, new uses for those drugs; and, most importantly, measures of the point in each drug's lifecycle at which intellectual property rights on new uses cease to be practically enforceable. No existing dataset contains these objects together. To facilitate our analysis, we construct new drug-level panel data by linking administrative, legal, and scientific records that were created for different purposes, use different identifiers, and describe drugs, uses, and intellectual property rights at different levels of aggregation. In doing so, we constructed the NBER Orange Book Dataset by digitizing, harmonizing, and validating historical FDA records of patents and regulatory exclusivities associated with approved drugs.[5] The resulting data allow us to trace, drug by drug, the arrival of scientific information, research investment, commercialization, and enforceable intellectual property protection.

Using these data, we show that investments in new uses for existing drugs closely track our predictions. At peak, ten years prior to exclusivity expiration, seven percent of drugs in our sample are re-approved for at least one new use. As expiration approaches, the likelihood of re-approval declines sharply; following exclusivity expiration, it falls to approximately zero. Of the roughly 500 new uses approved for innovator drugs in our data, 97 percent are approved no later than the calendar year in which exclusivity expires (and we hand checked and explain each exception). We show, however, that potential scientific "leads"—newly documented links between drugs and diseases in the scientific literature—continue to arise at roughly the same rate before and after exclusivity expiration. Before expiration, these leads predict subsequent commercialization; after, they do not translate into new use approvals.

Out next exercise leverages variation in the duration of the exclusivity period across drugs. Firms in the pharmaceutical sector have strong incentives to file key patents on new drugs early in the research and development process, often many years before drugs are approved for sale. Sizable differences in the time required to complete clinical trials and obtain regulatory approval mean that certain drugs reach the market with many more years of effective protection remaining than others. We use this variation to show that drugs with longer periods of enforceable protection are ultimately approved for more new uses—and also show, using a set of proxies, that drugs with longer periods of protection do not have systematically greater measured scientific potential. Thus, longer periods of enforceable protection affect the total amount of commercialization, rather than simply shifting when commercialization occurs. This variation also allows us to provide, to our knowledge, the first direct estimate of the elasticity of research investment with respect to the duration of intellectual property protection.

[5] A separate paper, Durvasula et al. (2023), serves as a "user's guide" to the U.S. Food and Drug Administration's "Orange Book" dataset, which offers a unique linkage between products (drugs) and patents, which is key in enabling analyses of the kind discussed in this paper.

We then compare new uses with related products that do not face the same change in enforceability at generic entry. Specifically, we consider *fixed-dose combinations*, which bundle multiple chemicals into a single form and are regulated as distinct new drugs. Because generic versions of the individual components are not perfect substitutes for the bundled product, intellectual property rights on these combinations may remain enforceable after exclusivity on the underlying drugs expires. We show that firms substantially increase their investment in fixed-dose combinations at the time of exclusivity expiration: the probability of a combination approval more than doubles relative to the preceding year and remains elevated in the years that follow. That commercialization shifts toward related products involving the same molecules, precisely when intellectual property rights on new uses become practically unenforceable, points more specifically to enforceability as the mechanism.

Taken together, this body of evidence provides support for the idea that the loss of enforceable intellectual property protection creates a missing market for new uses. Scientific opportunities continue to arise after exclusivity expiration, but commercialization largely ceases; drugs with longer periods of enforceable protection are ultimately approved for more new uses; and, when exclusivity expires, commercialization shifts toward related products for which intellectual property rights remain enforceable. Across these three exercises, then, commercialization tracks the availability of enforceable intellectual property protection rather than the underlying flow of scientific opportunities.

These empirical exercises allow us not only to establish the existence of a missing market but also to quantify the social value of missing research investments. Our estimated elasticity implies that if intellectual property rights on new uses were enforceable, there would be between 0.1 and 0.25 additional new uses per drug. As roughly 2,000 drugs that satisfy our sample criteria have been approved since the modern FDA was created in 1962, this implies between 200 and 500 missing new uses. To offer a sense of magnitudes, we draw on existing estimates of the social value of new drugs and adopt an intentionally conservative figure, valuing each missing use at $0.5 billion per year. The value of missing research and development in this market, then, is on the order of $100–250 billion per year. At a conservative social discount rate, the present value of this missing innovation is on the order of $2.5 to $6 trillion.[6]

Although our back-of-the-envelope estimates of the social cost of missing new uses are rough, they are consistent with evidence offered by the scientific literature. Gelijns et al. (1998), for example, document that within a sample of "blockbuster" drugs sold in 1993, nearly 90 percent had important uses beyond those for which they were originally approved. More recently, Every Cure, a nonprofit initiative that systematically searches for new therapeutic uses for existing drugs, reported that its medical team identified more than 9,000 promising repurposing opportunities in 2025 alone (EveryCure, 2025).

The idea that there are insufficient private incentives to develop new uses for existing drugs has

[6]Prior work has estimated comparably large magnitudes for the value of missing medical innovation. See, for example, Murphy and Topel (2006), Hall and Jones (2007), and Budish et al. (2015a).

been a long-standing concern for legal scholars, scientists, and policymakers (Gelijns et al., 1998; Mossinghoff, 1999; Eisenberg, 2005; US Institute of Medicine, 2008; Rai, 2012; Roin, 2014; Sachs et al., 2017; Conti et al., 2020; EveryCure, 2025). This paper is, to the best of our knowledge, the first to formalize the source of missing incentives in this setting theoretically and to document underinvestment empirically. We confirm that the scale of underinvestment is extraordinarily large. As advances in health data infrastructure and screening technology, including those that leverage emerging artificial intelligence technologies (EveryCure, 2025), reduce the cost of discovering potential new uses, the problem may only be growing (Collins, 2011a,b, 2012; Collins and Varmus, 2015).

Evidence of a quantitatively significant missing market for new uses provides, as well, clear empirical support for the paper's simple model of intellectual property and innovation incentives. More specifically, our framework embeds and formalizes two ideas that have appeared in various forms and under many names across much of the legal and economic scholarship on incentives for innovation—that formal intellectual property rights are not necessary to incentivize the development of certain inventions (e.g., Mansfield et al., 1981; Teece, 1986; Mansfield, 1986; Levin et al., 1987; Cohen et al., 2000; Moser, 2005), and that formal intellectual property rights cannot incentivize the development of others (e.g., Nelson, 1959; Arrow, 1962; Teece, 1986; Lemley and Shapiro, 2005; Eisenberg, 2005; Abramowicz, 2007; Kapczynski and Syed, 2012; Rai, 2012; Budish et al., 2015a). In theoretical and empirical studies that aim to inform the design of optimal patent policy, however, every invention is typically assumed patentable (and patent-protected), and every patent perfectly enforceable (e.g., Machlup, 1958; Nordhaus, 1969; Scherer, 1972; Nordhaus, 1972; Klemperer, 1990; Gilbert and Shapiro, 1990; Scotchmer, 1991; Budish et al., 2015a).[7]

The theoretical predictions generated by this model hinge on a causal link between intellectual property protection and incentives for innovation. Although decades of theory and centuries of policymaking have assumed such a relationship—that stronger intellectual property rights induce greater investment—there is essentially no direct empirical support, owing to the difficulty of identifying meaningful variation in the protection provided to otherwise comparable inventions.[8] Even in the pharmaceutical sector, often held up as the "poster child" for the patent system (Boldrin and Levine, 2008), there is no direct empirical evidence that longer or stronger patent rights affect the level or composition of research and development. Budish et al. (2015a) document evidence consistent with the idea that longer patent terms induce additional research investments, but their setting does not allow them to measure patent term directly or to rule out corporate short-termism as an alternative explanation.

Our simple model implies an alternative research design, which allows us to test for direct evidence that intellectual property rights affect the level of research and development by leveraging

[7] Since at least Pakes (1986), economists have grappled with other consequences of meaningful heterogeneity in the use and utility of patent rights. We formalize and test the form of heterogeneity conjectured in Kapczynski and Syed (2012)—that there are "domain[s] of innovation that patents, whatever their scope, cannot adequately address."

[8] A small number of papers—notably, Sakakibara and Branstetter (2001) and Lerner (2002)—have leveraged changes in national patent laws to test for such a relationship, but have found no effect, likely because country-level changes have relatively small impacts on global markets for innovation.

variation in the enforceability of intellectual property rights. We find that increases in the duration of enforceable intellectual property protection cause increases in the number of new uses brought to market and—with the caveat that this estimate is specific to our empirical context—estimate the elasticity of research and development with respect to patent term. In doing so, we contribute more broadly to the empirical literature on the determinants of innovation (for surveys, see Williams, 2017 and Bryan and Williams (2021)) and, more specifically, to the strands of this literature that study how property rights affect incentives for investment (e.g., Moser, 2005; Budish et al., 2015a; Hodgson, 2024; Dix and Lensman, 2024) and how gaps in incentives distort research and development specifically in the health context (e.g., Kremer and Snyder, 2015; Budish et al., 2015a; Krieger et al., 2022; Michelman and Msall, 2024; Dix and Lensman, 2024).

This paper proceeds as follows. Section 2 introduces our conceptual framework. Section 3 provides context on the problem of new uses. Section 4 describes our data and measurement. Section 5 presents three sets of evidence that, together, establish the existence of a missing market for new uses for existing drugs. In Section 6, we present evidence on the magnitude of missing investment—in terms of missing "uses" and dollars of social value. Section 7 discusses implications for innovation policy and concludes.

2 Characterizing Missing Markets for Innovation

This section defines missing markets for innovation and develops a simple classification procedure for identifying potential missing markets in practice. Section 2.1 extends a standard Nordhaus (1969)–style model by adding three parameters that determine whether an innovator can capture any economic return from investment: whether the invention can remain secret, whether it is eligible for formal intellectual property protection, and whether any resulting intellectual property right can be enforced in practice. These three modifications allow us to distinguish between settings in which private incentives for investment are *weak* and settings in which incentives are *non-existent* because neither longer nor broader intellectual property rights—the canonical "Nordhaus parameters" of innovation policy—are effective. Section 2.2 provides institutional and legal detail on each parameter. Section 2.3 summarizes the classification exercise graphically.

2.1 Defining Missing Markets

2.1.1 Preliminaries

Our starting point is a Nordhaus (1969)–style framework, as adapted by Budish et al. (2015a). A representative firm must decide whether to attempt to commercialize a potential invention, the characteristics of which it regards as exogenous.[9] Commercialization refers to the costly and risky process of attempting to bring a potential invention to market—e.g., running clinical trials to assess a drug's safety and efficacy, prototyping and iterating on product design, etc.

[9]We use the term "invention" to refer to any instance of knowledge production—including in the form of physical products, sets of instructions, and pieces of information (Arrow, 1962; Romer, 1994; Jones, 2005).

A potential invention is characterized by the following sets of parameters, the first six of which are common to prior work.

- **Timing parameters**: The firm must decide in year t_{invent} whether to attempt to commercialize the potential invention. Commercialization involves a deterministic delay t_{comm}. We normalize $t_{\mathsf{invent}} = 0$.
- **Cost of commercialization**: If the firm decides to attempt to commercialize the potential invention it incurs a positive cost $c > 0$, measured in net present value terms at t_{invent}.
- **Likelihood of successful commercialization**: Commercialization may or may not be successful. We denote the probability of success by p_{comm}. We restrict consideration to potential inventions for which p_{comm} is greater than zero.
- **Obsolescence risk**: An invention may become obsolete as new, substitutable alternatives are brought to market. We capture this possibility with an exogenous risk of obsolescence $1-\gamma$ in each year following t_{invent}.
- **Discount factor**: The firm discounts future profits by a factor δ.
- **Monopoly profits and social value**: A successfully commercialized invention that is not obsolete and produced by the firm as a monopolist yields profits π per year. The invention also yields social value v^{monop} when priced by the monopolist and v when priced by the social planner, where $v > v^{\mathsf{monop}}$.

To this standard framework, we add three sets of parameters.

- **Secrecy**: Some inventions can be profitably commercialized while remaining secret. That is, the firm can bring the product to market in a way that does not reveal its underlying invention to potential competitors. Let $p_{\mathsf{secret}} \in [0,1]$ denote the probability that the invention remains secret upon successful commercialization, if the firm chooses secrecy.[10] After an invention is commercialized the firm faces further risk of losing secrecy. We capture this by assuming the firm faces an exogenous risk $1-\eta$ of losing secrecy each year following successful commercialization.
- **Eligibility**: Not all inventions are eligible for protection with grants of intellectual property rights. The probability that a commercialized invention satisfies these eligibility requirements is given by $p_{\mathsf{eligible}} \in [0,1]$. The associated protection lasts for t_{patent} years.
- **Enforceability**: Grants of intellectual property rights are not self-enforcing. To enforce these rights, the firm must seek out instances of imitation and successfully bring suit against each infringer. The probability that the firm can enforce its intellectual property rights is $p_{\mathsf{enforce}} \in [0,1]$.

[10] One could, instead, interpret $1 - p_{\mathsf{secret}}$ as the fraction of monopoly profits lost to imitation in the absence of a formal intellectual property grant.

The firm can seek formal intellectual property protection or rely on secrecy. We treat these as mutually exclusive alternatives: seeking patent protection requires disclosure, whereas a firm that relies on secrecy attempts to commercialize without revealing the invention. In Section 2.2, we discuss factors that affect the value of each parameter.

2.1.2 Monopoly Duration

For each potential invention, there is some period in which the firm will, in expectation, earn monopoly profits. The firm's expected time under monopoly depends on the mechanism used to deter imitation: secrecy or formal intellectual property rights.

If the firm seeks to protect the invention through secrecy, the expected time under monopoly is given by

$$T_{\text{secret}} = p_{\text{comm}} \cdot p_{\text{secret}} \cdot \sum_{t=t_{\text{comm}}}^{\infty} (\delta\gamma)^t \eta^{t-t_{\text{comm}}}. \tag{1}$$

In words, time under secrecy T_{secret} is determined by (i) the likelihood of successful commercialization p_{comm}, (ii) the likelihood that secrecy effectively deters imitation p_{secret}, and (iii) the number of years in which secrecy is maintained, adjusted for discounting (by δ), the risk of obsolescence (by γ), and the risk of losing secrecy (by η).

If, instead, the firm seeks formal intellectual property protection (e.g., a patent), the expected discounted years of monopoly returns are given by

$$T_{\text{patent}} = p_{\text{comm}} \cdot p_{\text{eligible}} \cdot p_{\text{enforce}} \cdot \sum_{t=t_{\text{comm}}}^{t_{\text{patent}}-1} (\delta\gamma)^t. \tag{2}$$

We take the sum to be zero when $t_{\text{patent}} \leq t_{\text{comm}}$.

In words, time under intellectual property protection T_{patent} is determined by (i) the likelihood of successful commercialization p_{comm}, (ii) the likelihood that the invention is eligible for intellectual property protection p_{eligible}, times the likelihood that the intellectual property rights can be enforced p_{enforce}, and (iii) the expected duration of effective intellectual property protection $t_{\text{patent}} - t_{\text{comm}}$, adjusting for discounting and obsolescence risk.

2.1.3 Incentives to Invest

A profit-maximizing firm will attempt to commercialize a potential invention if and only if the expected profits at least cover the costs of commercialization. Define T^* as the firm's expected time under monopoly under whichever of trade secrecy and patent protection yields more time in expectation:

$$T^* = \max(T_{\text{secret}}, T_{\text{patent}}). \tag{3}$$

The firm will pursue the invention if and only if

$$T^* \cdot \pi - c \geq 0. \tag{4}$$

If $T_{\text{secret}} > T_{\text{patent}}$ then the firm chooses trade secrecy, and T^* is given by (1). Otherwise, the firm chooses intellectual property protection, and T^* is given by (2).

2.1.4 Missing Markets

Definition 1 (Missing Market)**.** *A missing market arises when an invention's expected time under monopoly T^* is zero. That is, both $T_{\text{secret}} = 0$ and $T_{\text{patent}} = 0$.*

Assume that the invention is plausibly successful $p_{\text{comm}} > 0$, not immediately obsolete $\gamma > 0$, and that the firm is not completely impatient $\delta > 0$. Under what more substantive conditions is $T^* = 0$, and hence there is a missing market? There are two cases.

The first case where $T^* = 0$ is if $p_{\text{secret}} = 0$ and $t_{\text{patent}} \leq t_{\text{comm}}$. In words, trade secrecy is ineffective and the invention's patent term is too short relative to the time it takes to bring the invention to market. This is the case studied in Budish et al. (2015a), which provides evidence that firms underinvest in projects that are completed over long time horizons. Notice that incentives can be restored by extending the patent term t_{patent}, or by starting the patent clock at time t_{comm} instead of time t_{invent}. In this sense, this first case reflects a misalignment of patent timing parameters.

The second case, our focus in this paper, arises if $p_{\text{secret}} = 0$ and either $p_{\text{eligible}} = 0$ or $p_{\text{enforce}} = 0$. That is, trade secrecy is ineffective and the invention is either not eligible for intellectual property protection or the available intellectual property rights are unenforceable. In this case, extending the duration or increasing the breadth of rights that remain unavailable or unenforceable does not create an incentive to commercialize the invention.

Proposition 2 (No Nordhaus Parameters)**.** *If $p_{\text{secret}} = 0$ and either $p_{\text{eligible}} = 0$ or $p_{\text{enforce}} = 0$, then neither longer ($t'_{\text{patent}} > t_{\text{patent}}$) nor broader ($\pi' > \pi$) patent rights make commercialization profitable.*

Theoretical studies of innovation and intellectual property rights often characterize optimal policy in terms of two "Nordhaus parameters"—t_{patent} (capturing patent duration) and π (capturing patent breadth).[11] Implicit in these analyses is the idea that careful calibration of these dimensions of intellectual property rights can, in fact, strike an optimal balance between incentives for innovation and deadweight loss. But when intellectual property rights are unavailable or unenforceable, changes to these Nordhaus parameters t_{patent} and π have *zero* impact on the firm's investment decision in (4).

2.2 Identifying Missing Markets

In Section 2.1, we introduce three parameters—$p_{\text{secret}}, p_{\text{eligible}}, p_{\text{enforce}}$—that affect both whether an invention is commercially viable and whether the classic "Nordhaus parameters" of innovation

[11] See Gilbert and Shapiro (1990) and Klemperer (1990) for models of patent breadth. Gilbert and Shapiro (1990) define patent breadth as the profit parameter π—that is, as the flow rate of profit available to the firm so long as the patent is in force. Klemperer (1990) defines patent breadth in a spatial horizontal differentiation model, where greater breadth gives the firm greater market power and hence greater profits in equilibrium.

policy can be used to calibrate incentives for investment. This section provides institutional and legal context on each parameter and, in doing so, helps the reader understand the practical conditions under which missing markets are likely to arise.

2.2.1 p_{secret} : Can the invention be profitably commercialized while remaining secret?

For many inventions, both private and social value come entirely from public disclosure. Consider inventions that are, in effect, "recipes" (Jones, 2023) or sets of directions—such as regimens of diet and exercise, instructions for the use of an existing drug in the treatment of a new disease (Section 3), or protocols for the use of checklists to reduce the risk of error (Gawande, 2009; Kapczynski and Syed, 2012). Each invention consists entirely of the knowledge that some series of steps will yield a particular outcome. Under some circumstances, firms can implement restrictions that limit how widely this information will diffuse—for example by binding recipients contractually with non-disclosure and non-compete agreements (e.g., Singh and Marx, 2011). Absent such restraints, secrecy is not viable.[12] For other types of inventions, secrecy is impractical. If an invention must be placed on display or disclosed to a potential customer to facilitate sale, trade secrecy provides no meaningful innovation incentive (Arrow, 1962).[13] When there are no circumstances under which the invention can be commercialized while remaining secret, p_{secret} takes a value of zero.

Existing work has identified some settings in which p_{secret} is clearly positive (e.g., Mansfield, 1986; Levin et al., 1987; Cohen et al., 2000; Moser, 2005, 2007).[14] Manufacturing processes used entirely within a firm—including proprietary recipes, like the formula for Coca-Cola (Carr and Choi, 2023)—can be held as secrets for as long as the firm can limit disclosure.[15] Other inventions may be amenable to secrecy simply because they are complex or otherwise difficult to reverse engineer. Vaccines are a classic example: even when the chemical makeup of a vaccine is known, the technical know-how required to produce a product that is stable and safe can bar widespread replication (Price et al., 2020; Gault, 2021). Secrecy can also result from deliberate engineering choices. The "fuzzy logic" used to modulate the cook time and temperature of the award-winning Zojirushi rice cooker, for example, has been held as a trade secret since its introduction in the 1980s; the key piece of technology is embedded within the product in a way that ensures disassembly will destroy the relevant component (Santo, 2023).

[12] State and federal laws in the United States provide formal trade secret protection. This protection is both expansive and fragile. Legal protection lasts until secrecy is destroyed, which may mean that it is indefinite. But any instance of disclosure, reverse-engineering, or independent discovery is secrecy-destroying. So long as the owner takes "reasonable efforts" to maintain secrecy, these laws provide recourse, but they cannot restore an already-disclosed secret.

[13] This observation in Arrow (1962) is often described as Arrow's Information Paradox.

[14] Moser (2005) documents that 19th century inventors in countries without patent laws disproportionately concentrated in industries where secrecy was thought to be effective. Moser (2007) tests this idea by leveraging a breakthrough—the publication of the periodic table in 1869—that made it easier for competitors to reverse engineer chemical innovations. When the cost of reverse engineering dropped—in our framework, when p_{secret} fell close to zero—inventors' propensity to patent increased. Mansfield (1986) provides survey evidence consistent with these findings.

[15] For details on Coca-Cola's efforts to preserve secrecy over nearly every aspect of its product, see Carr and Choi (2023).

2.2.2 p_{eligible} : Is the invention eligible for formal intellectual property protection?

Government-granted intellectual property rights take many forms—including patents, copyrights, and "quasi-patents" administered by government agencies (see Section 4). Whether any invention is eligible for such a grant depends, first, on its subject matter. In the United States, certain inventions—by statute and by judge-made doctrine—are excluded from particular forms of protection (e.g., most forms of DNA cannot be protected by patents), and some inventions are ineligible for all forms of intellectual property rights (e.g., most databases, which cannot be protected by copyrights or patents).[16]

In our model, p_{eligible} takes a value of one for categories of inventions where patent eligibility is clear, and p_{eligible} takes a value of zero when an invention falls squarely within a category excluded from protection by law. For many categories of inventions, recent shifts in legal standards have introduced considerable uncertainty in both the tests used to determine eligibility and in the validity of previously issued patents (e.g., Hoyt, 2022). For such inventions, then, p_{eligible} lies between zero and one.

2.2.3 p_{enforce} : Can the firm enforce an intellectual property grant in practice?

An invention may be eligible for formal intellectual property protection but protected only by rights that are practically unenforceable. For our purposes, enforceability requires two things. First, the firm must be able to observe infringement. Second, the firm must be able to obtain a useful remedy against a party whose conduct erodes the innovator's returns. If either condition fails, the formal grant does not allow the firm to exclude imitation or charge for use of the invention. In the limiting case, $p_{\text{enforce}} = 0$.

For some inventions, infringement is easy to observe. Unauthorized copies of luxury handbags (Hemphill and Suk, 2008) or consumer electronics (Robbins, 2018), for example, are visible in the marketplace. Advances in reverse-image search, digital watermarking, and algorithmic plagiarism detection have substantially improved this visibility (Eckhause, 2022).

In other contexts, infringement may be difficult or impossible to observe. Manufacturing processes, for example, may be copied and used inside a competitor's facility, thus producing no evidence of imitation (Price, 2014). Firms can sometimes respond by designing transactions or technologies that make use of the invention easier to detect. Teece (1986), for example, describes sellers of petroleum-refining methods who required licensees to use a specialized catalyst supplied only by the patent holder—in effect, requiring the use of an observable complement to an otherwise easily imitated process. Regulators and other intermediaries can sometimes serve a similar function (Kapczynski and Syed, 2012).

Even when infringement can be observed, enforcement may be impossible because the legally

[16] At present, utility patents may not be granted on inventions that claim abstract ideas, natural laws, or physical phenomena. Judicial decisions in recent years—in particular, *Mayo v. Prometheus* (2012), *Association for Molecular Pathology v. Myriad Genetics* (2013), and *Alice Corp. v. CLS Bank* (2014)—have clarified these restrictions and, in doing so, substantially narrowed the set of patentable potential inventions.

responsible infringer is not a practical defendant. The difficulty arises, in part, from legal rules that specify how certain valuable patents are infringed. In the United States, patents on methods—specific ways of using a product or carrying out a set of instructions—are directly infringed only by the individual or entity performing each step of the method. Consider a firm that holds a patent on a method for using orthodontic retainers to straighten teeth. Nothing about selling a retainer performs the patented method. Direct infringement occurs only when a patient uses the copycat retainer to straighten her teeth. Enforcing the patent would therefore require the innovator to sue the retainer-wearing patients themselves. For largely intuitive reasons, this is rarely an attractive strategy: such suits are likely to engender backlash and, even if successful, individual consumers are unlikely to have sufficient resources to support meaningful damages awards.

A related limitation may arise when competitors can reproduce nearly all of an invention's characteristics without technically infringing on the innovator's intellectual property rights. Patent claims define the boundaries of a legal right, and competitors may be able to reverse engineer a protected product while "designing around" these boundaries. When this is the case, intellectual property rights might be practically unenforceable because they fail to capture the relevant imitation.

2.3 Summary Figure

Figure 1 provides a classification tree to identify missing markets for innovation, in the sense of Definition 1. Some potential inventions are amenable to trade secrecy and hence protectable that way. Some potential inventions are both eligible for intellectual property rights and these rights are practically enforceable, so the inventions are protectable that way. A missing market arises when trade secrecy is unavailable and intellectual property rights are either unavailable or unenforceable.

3 The New Uses Problem

3.1 Background

Modern medicine relies on a relatively small set of chemical compounds—roughly 2,000 in total—that have been shown to be safe for use in humans and effective for at least one therapeutic indication. Scientists often describe compounds that meet both of these demanding criteria as "promiscuous," by which they mean that if a drug can successfully treat one disease without causing serious harm, it is likely to be useful in treating others as well (Wermuth, 2006).

In some cases, multiple therapeutic uses are anticipated at the outset of drug development—for example, when related conditions (e.g., depression and anxiety disorders) are expected to respond similarly to treatment. Often, however, additional uses are identified only with substantial delay. New opportunities may emerge as advances in basic science improve understanding of how different diseases are related, or as clinical experience with an approved drug reveals previously unrecognized benefits in new patient populations.[17] Technological change can accelerate the identification of such

[17] Cancer provides a useful example. Traditionally, cancers have been classified by the organ in which they first ap-

opportunities by enabling new scientific insights and uncovering patterns in accumulated clinical data (EveryCure, 2025). Even so, recognition of potential new uses often occurs many years—or decades—after a drug's initial approval. As Gelijns et al. (1998) observe, "it took half a century for the cardiovascular benefits of aspirin, the most widely used drug in the world, to be recognized, and nearly 40 more years before it was widely used for cardiovascular indications."

Several other high-profile examples make clear how wide-ranging new uses may be. Thalidomide, initially developed as a treatment for morning sickness (and famously kept off the U.S. market before its global recall in the early 1960s), was approved in 1998 as a treatment for leprosy and again in 2006 as a treatment for multiple myeloma.[18] Minoxidil, initially approved in 1979 as an antihypertensive, was reapproved in 1988 for male pattern baldness and in 1991 for female pattern hair loss.[19] Recently, the biologic drug omalizumab (marketed under the brand-name Xolair), initially approved in 2003 to treat allergic asthma and related conditions, received a new approval as the first medicine shown to reduce the frequency and severity of allergic reactions (Wood et al., 2024).[20] Notably, each of these new use approvals preceded generic (or, in the case of Xolair, biosimilar) competition.

Prominent researchers have argued that systematic efforts to develop new uses for existing drugs could "convert cancer into a treatable chronic disease" (Telleria, 2012), unlock treatments for many or all of the 8,000 rare diseases (Muthyala, 2011), and yield breakthroughs in diseases that have long resisted therapeutic intervention, such as Alzheimer's (Corbett et al., 2012; Butcher, 2013). Recent efforts to leverage artificial intelligence to identify potential new uses are premised on the same idea (EveryCure, 2025).

Despite this scientific promise, investment in the systematic development of new uses for existing drugs remains limited (Eisenberg, 2005; Rai, 2012; Butcher, 2013; Pahud et al., 2014; Roin, 2014; Sachs et al., 2017; Conti et al., 2020; EveryCure, 2025). Although some new indications receive regulatory approval, historical accounts and descriptive evidence suggest that many potential uses identified through advances in basic science and clinical observation are never explored in rigorous clinical trials.[21] Roin (2014) illustrates the breadth of this problem, collecting case stud-

pear (e.g., "breast cancer" or "lung cancer"). Recent work shows that this approach can be misleading: tumors in the same organ can differ substantially, while tumors arising in different organs can share therapeutically important molecular features. These changes in the scientific understanding of cancer have reshaped both treatment decisions and drug development, such that physicians and scientists have increasingly moved away from organ-based approaches and toward approaches that emphasize shared characteristics across tumors. This has created concrete opportunities to develop and test existing drugs for uses that were not well defined at the time of initial approval (Garraway and Lander, 2013; Hoadley et al., 2014, 2018).

[18] NDA # 020785, NDA # 021430.

[19] NDA # 019501, NDA # 020834, NDA #021812.

[20] BLA #103976.

[21] Drugs may be prescribed "off label" even where rigorous clinical trial evidence does not support their use for a particular indication. Historically, many off label uses have later been validated by rigorous clinical trials—for example, the off label use of an existing heart medication was later unequivocally shown to be safe and effective in the treatment of a childhood vascular condition (Léauté-Labrèze et al., 2015). However, there are also many historical examples of widespread off label use that was later shown to be, on net, harmful—for example, the widespread use of hormone therapy to prevent heart disease in postmenopausal women, which was ultimately shown to increase the risk of stroke, heart attack, and blood clots (Manson et al., 2003). Thus, while off label use may be useful in generating hypotheses or collecting early signals about potential new uses, it is not considered a perfect substitute for investment in large-scale

ies involving possible treatments for "cancer (including cancer prevention), Alzheimer's disease, depression, diabetes, stroke, tuberculosis, malaria, multi-drug resistant bacteria."

The case of metformin provides an especially instructive example. In 1995, the U.S. Food and Drug Administration (FDA) approved Glucophage (metformin hydrochloride) as a treatment for diabetes.[22] Generic versions of the drug entered the market in 2002. In the years that followed, oncologists began to collect tentative evidence that metformin may also be effective in treating or preventing cancer. Though these researchers proceeded with various "preclinical" studies, commercial interest in funding a clinical trial that would evaluate metformin as a treatment for human cancer was nonexistent (Gale and Gerlin, 2012). As one oncologist told journalists (Gale and Gerlin, 2012), "*When they realized the results wouldn't be available until they lost their patent, they pulled out*." Decades later, whether metformin has clinically meaningful effects in cancer treatment or prevention remains unresolved (O'Connor et al., 2024).[23]

3.2 A Missing Market for New Uses

In this section, we show that the conditions that give rise to a missing market are satisfied in the case of new uses for existing drugs. To illustrate, we walk through each step of the stylized classification exercise introduced in Section 2.

p_{secret} : Can the invention be profitably commercialized while remaining secret? *No.*

It is difficult to imagine how the information that a drug has a new therapeutic use could be profitably commercialized while remaining secret. In the United States, federal law requires drugs to carry labels specifying each FDA-approved therapeutic indication.[24] This legal requirement aside, secrecy is impractical. A new use has economic value only if it is disclosed: physicians must know when to prescribe the drug, pharmacists must know when to dispense it, patients must know how to take it, and insurers must know how to process reimbursements. Thus, $p_{\text{secret}} = 0$.

p_{eligible} : Is the invention eligible for formal intellectual property protection? *Yes.*

In the United States, new uses for existing drugs are eligible for two distinct forms of intellectual property protection. First, they may be protected by method-of-use patents. Even if a drug has

clinical trials—the margin of investment that is our focus.

[22] NDA #020357, approved on 03 March 1995.

[23] Since at least Gale and Gerlin (2012) reported on the difficulties of garnering commercial support for clinical trials, an academic literature has explored whether metformin may have anti-cancer effects. Much of this work consists of observational studies, secondary analyses, and relatively small, exploratory randomized trials conducted in diabetic populations (where estimates of the effects of metformin are confounded). Though well suited to generating hypotheses and characterizing potential biological mechanisms, these studies have not been designed to provide definitive evidence on metformin's efficacy for any specific cancer indications. Accordingly, meta-analyses have catalogued problems with scale, design, and publication bias (Lord and Harris, 2023). A recent, comprehensive meta-analysis in the *Journal of the National Cancer Institute* concludes as follows: "Metformin may be associated with a decreased risk of many cancer types, but high heterogeneity and risk of publication bias limit confidence in these results. Additional studies in populations without diabetes are needed to better understand the utility of metformin in cancer prevention" (O'Connor et al., 2024).

[24] *See* 21 U.S.C. § 352(f); 21 C.F.R. § 201.57(c).

already been commercialized (and its active ingredient is already protected by a patent), firms may seek patent protection for methods of using a drug to treat a new disease or condition.[25]

Second, new uses are eligible for regulatory exclusivities administered by the FDA. The FDA grants a three-year period of "new clinical investigation exclusivity" (NCI) to incentivize firms to develop new indications, dosage regimens, patient populations, and formulations of previously approved drugs. During this three-year exclusivity period, the FDA will not approve applications from generic competitors for the specific use protected by NCI exclusivity. The limits of this restriction are worth emphasizing: when NCI exclusivity is in force, the FDA may still approve generic competitors for *other* indications. That is, a brand-name firm with NCI exclusivity protecting one approved use may still face generic competition for other, unprotected uses. As new uses may be protected by both patents and regulatory exclusivities, $p_{\text{eligible}} = 1$.

p_{enforce} : **Can the available intellectual property protection be enforced in practice?** ***Yes, then no.***

Patents on new uses for existing drugs are fully enforceable so long as the drug itself faces no generic competition.

The Hatch-Waxman Act of 1984 created the regulatory system for generic drugs in the United States. Its key provisions included a change to clinical trial requirements for generic firms: since 1984, generic firms have not been required to conduct independent, costly clinical trials as a condition of approval. Instead, sponsors must satisfy two requirements. First, the generic firm must prove that its product is therapeutically equivalent to the original brand-name drug. Second, for whatever scope of approval it seeks—intended use(s), product form, patient population, etc.—the applicant must demonstrate that any relevant patents associated with the brand-name drug are invalid, not infringed, or expired. The list of relevant patents is provided by the sponsor of the brand-name drug and reported in a regulatory document maintained by the FDA, colloquially referred to as the "Orange Book."[26] While challenges to listed patents are in progress, the FDA will not approve a competitor. Until the competitor is approved, patents associated with the brand-name drug, including patents on its uses, are thus perfectly enforceable (and enforced, in effect, by the FDA as an intermediary).[27]

Once a generic version of the drug is approved for an unprotected use, enforceability of intellectual property rights on new uses collapses. To enforce intellectual property rights after generic entry, the brand-name firm must solve two problems. First, it must identify instances in which generic products approved by the FDA only for some unprotected use *A* are, instead, used for a protected use *B*. Second, it must put an end to infringement, typically by filing and winning a lawsuit.

[25] In re Marshall, 578 F.2d 301, 304 (C.C.P.A. 1978). Federal courts have consistently upheld patent grants on new uses for existing drugs. E.g., Vanda Pharmaceuticals Inc. v. West-Ward Pharmaceuticals International Ltd., 887 F.3d 1117, 1134–36 (Fed. Cir. 2018).

[26] We provide details on this patent list in Section 4.4.

[27] If a generic firm markets its product without FDA approval, it faces civil liability, criminal charges, and regulatory penalties. The FDA can immediately seek injunctions to halt sales, issue warning letters, seize offending products, and recommend criminal prosecution.

Neither is practical.

On the first problem: the brand-name firm can neither observe when generic drugs are prescribed or purchased, nor can it observe what indication the generic is used for. Although insurers often collect data that link prescriptions to diagnoses, that information is generally not available to the brand-name firm (see Roin, 2014).

On the second: even if the brand-name firm could observe infringement of a patent on a new use, the nature of liability in this setting makes enforcement challenging. Generic firms typically rely on the Hatch-Waxman Act's "skinny labeling" provisions, which allow them to sell generic versions of drugs that have patent-protected uses, so long as the protected uses are "carved out" of the generic label. This means that a generic firm selling a drug approved for an off-patent use A and a patented use B can obtain approval for a "skinny" generic label that lists only A. Of course, physicians remain free to prescribe the generic for B—technically doing so "off label"—and courts have consistently declined to hold the generic firm liable when they do, absent active encouragement of the patented use.[28]

Patent liability is, instead, likely to fall on those parties who actually "practice" the method of using the drug in violation of a patent—hospitals or clinicians who administer drugs in inpatient settings, and patients who take the drug as instructed at home. In doing so, each technically risks a lawsuit. But even setting aside the impracticability of filing expensive lawsuits against entities with limited capacity to pay damages, Eisenberg (2005) notes that it is rarely a winning strategy for firms to sue their own (often sick) customers en masse.[29]

Thus, before generic entry, p_{enforce} takes a value of one: the brand-name firm can fully enforce its portfolio of intellectual property rights. After generic entry, p_{enforce} falls to zero. It is generic entry, then, that creates a missing market in the sense of Definition 1.

We refer to the period during which the innovator remains protected from generic competition for the underlying drug as the drug's *exclusivity period*. That protection may come from patents, FDA-administered regulatory exclusivities, or both. Once it ends, rights covering a new use may continue to exist without providing effective protection against use of a generic version for that indication.

3.3 Empirical Signatures

The conditions described in Section 3.2 generate testable empirical predictions. Consider a drug that is approved by the FDA at time $t = 0$. For some period of time, the innovator firm is the

[28] *Warner-Lambert v. Apotex Corp.*, 316 F.3d 1348 (Fed. Cir. 2003). This is an evolving area of doctrine—though all doctrinal changes in this area post-date our analysis and thus do not complicate the interpretation of our empirical results. Between 2020 and 2026, the Court of Appeals for the Federal Circuit developed a theory of liability wherein a generic manufacturer could be found liable for patent infringement if it, in effect, hinted at the existence of a patent-protected use, regardless of whether the drug's label and marketing materials included any references. See, for example, *GlaxoSmithKline LLC v. Teva Pharmaceuticals USA, Inc.*, 976 F.3d 1347 (Fed. Cir. 2020). For a discussion of this theory of liability and the associated case law, see Sherkow and Gugliuzza (2026). In June 2026, the Supreme Court rejected this broad theory of liability and left substantially intact the conventional understanding of the Hatch-Waxman Act's skinny-labeling provisions. *Hikma Pharmaceuticals USA Inc. v. Amarin Pharma*, Inc., 608 U.S. ___ (2026).

[29] For extended discussions, see Eisenberg (2005) and Roin (2014).

sole seller of the drug. During this exclusivity period, intellectual property rights on new uses are fully enforceable in practice. Since each new use that is commercialized yields monopoly profits until exclusivity expires, incentives for investing in new uses are strongest at $t = 0$. As exclusivity expiration approaches, the length of time during which any newly identified use can be protected from competition drops: a new use identified at year $t + 1$ has one fewer year of enforceability than a use identified at year t.

This framework, thus, suggests three empirical signatures of a missing market for new uses:

1. Firms will have the strongest incentives to invest in developing and commercializing new uses early in the drug's lifecycle, when the potential duration of enforceable monopoly rights is longest.

2. As exclusivity expiration approaches, incentives to invest in new uses decline monotonically towards zero, as each newly identified use has fewer remaining years of effective protection.

3. When exclusivity ends, p_{enforce} drops to zero. Commercialization of new uses should, thus, fall to zero.

The remainder of this paper constructs the data and develops the measurement needed to test these predictions at scale.

4 Data and Measurement

Testing for the existence of a missing market for new uses requires data on three objects: a sample of drugs whose initial approval can be observed; measures of subsequent investment in, and commercialization of, new uses for those drugs; and, most importantly, measures of the point in each drug's life cycle at which intellectual property rights on new uses cease to be practically enforceable. No existing dataset contains these objects together. We therefore construct new drug-level panel data by linking administrative and scientific records that were created for different purposes, use different identifiers, and describe drugs, uses, and intellectual property rights at different levels of aggregation.

The third object presents a particular conceptual challenge. As Section 3 explains, intellectual property rights on new uses become practically unenforceable once the innovator loses what we refer to as its *exclusivity*: protection from generic competition for the underlying drug. Measuring the end of exclusivity requires identifying both when generic competition becomes possible and the portfolio of patents and regulatory exclusivities that may delay it. We therefore construct new data infrastructure linking each FDA-approved drug to its generic competitors and to the overlapping rights protecting it.

As part of this project, we created the NBER Orange Book Dataset by digitizing, harmonizing, and validating the FDA's Orange Book records of firm-reported patents and regulatory exclusivities associated with approved small-molecule drugs. In Durvasula et al. (2023), we describe the data

construction, error correction, and validation—completed in conjunction with a team of technical experts—required to turn these administrative records into a research-ready dataset. Here, we link those records to FDA approval and generic-entry histories to measure exclusivity for each innovator drug, and to new-use approvals, clinical trials, and scientific publications to trace research and commercialization over the drug's life cycle.

The remainder of this section describes our data construction. Sections 4.1 and 4.2 detail how we construct a sample of drugs and how we measure approvals for new uses. Section 4.3 introduces complementary measures of information about, and investment in, new uses. Section 4.4 describes our measures of exclusivity. Section 4.5 summarizes our data construction and presents basic summary statistics. For each step of our data construction, Appendix A provides additional details.

4.1 Measuring Drug Approvals

We construct a census of new drugs approved by FDA between 1985 and 2014 using an administrative database, Drugs@FDA (U.S. Food and Drug Administration, 1985-2014).[30] We classify a drug as an *innovator* and include it in our primary sample if it satisfies two criteria:

1. It was approved under a New Drug Application (NDA).

2. It represents the first approval of a unique combination of chemical components.

The first restriction limits our analysis to small-molecule drugs, which are relatively simple, stable chemical compounds synthesized through chemical processes. It excludes biologic drugs—large, complex molecules derived from living organisms—which are subject to distinct legal and regulatory frameworks in the United States.[31]

The second restriction ensures that we focus on research incentives associated with each drug's first approval.[32] Many drugs are approved under multiple brand names over time, usually after generic entry; these are often cases in which brand name products can capture some substantial premium over generic alternatives (Bronnenberg et al., 2015). Acetaminophen, for example, is

[30] Our sample begins in 1985, one year after the 1984 Hatch-Waxman Act introduced the modern system of pharmaceutical patent protection and generic drug regulation. The Hatch-Waxman Act increased terms of patent protection for certain new drugs and introduced a new regulatory pathway for generic entry. From a practical perspective, it led to the creation of the Orange Book database that is central to our measurement of exclusivity, introduced in Section 4.4. We select 2014 as an end point for two reasons. First, a key table in the Drugs@FDA database that we use to identify re-approvals of drugs for new indications is unavailable in later years (from the middle of 2016 on), and we have been unable to obtain analogous records via Freedom of Information Act requests. Second, for drugs approved in later cohorts, measures of exclusivity may be censored (e.g., if a firm is still applying for new patents or regulatory exclusivities).

[31] Biologic drugs include gene therapies, monoclonal antibodies, and vaccines. A different set of statutes govern federal research incentives for biologics, and the FDA applies a different set of standards for generic ("biosimilar") approval. One important distinction, from the perspective of our empirical analysis, is that our data construction relies on the existence of a mapping between small-molecule drug patents and approved drug products, the Orange Book. No analogous regulatory document exists for biologic drugs. Frakes and Wasserman (2025) hand-review drug patent filings to construct what is, to our knowledge, the most complete listing of biologics patents.

[32] This restriction is similar to, but not the same as, the regulatory designation of "new molecular entity" (NME). NMEs are a strict subset of our sample. See Appendix A.1.2 for details.

available both as a generic and under the brand names "Tylenol" and "Panadol." Ibuprofen, too, is available both as "Advil" and "Motrin." We keep only the first of these brand name approvals.

4.2 Measuring New Use Approvals

Our primary measure of successful commercialization of new uses is FDA approval of an existing innovator drug for a new or modified indication. We refer to these events as new use approvals. We identify these approvals using FDA administrative records that classify supplemental approvals by type, and collect those tagged as approvals for new or modified indications. We also construct a broader measure that follows the same drug across regulatory applications and sponsors: for each innovator drug, we collect new use approvals associated with any new drug application with the same active-moiety combination, even if the approval is not linked to the original application or sponsor.

FDA's supplemental-approval codes identify that an approval concerns a new or modified indication, but they do not provide a consistent, machine-readable description of the substantive use that was approved. We therefore reconstruct the indication associated with every initial approval and re-approval event, building on Berger et al. (2021).[33] We collect indication text from FDA labels and approval packets and supplement these records with dated company press releases, scientific articles, and other contemporaneous public sources. Whenever possible, we use the FDA's full-text drug-labeling service, FDALabel (U.S. Food and Drug Administration, 2005-2016b); because its coverage begins primarily in 2005, the supplementary sources are necessary to recover indications over our full sample period. We then assign each approval or re-approval one or more codes from the Tenth Revision of the International Classification of Diseases (ICD-10), yielding a standardized drug-by-indication approval history (World Health Organization, 2019).

4.3 Measuring Scientific Opportunity and Research Investment

An important feature of our data is that they allow us to distinguish the arrival of scientific opportunities from investment in developing those opportunities and from successful commercialization. FDA approval of a new use captures the final stage. We use clinical trials to measure entry into costly clinical development and scientific publications to trace the arrival of potential scientific leads about new uses. Constructing these measures requires linking records from disparate sources to common drug and disease identifiers, often using information reported only in unstructured text.

Clinical Trials To obtain coverage extending back to the 1980s, we combine clinical trial records from four sources: NDA Pipeline (F-D-C Reports, 1982-2001), Pharmaprojects (Citeline, 1995-2010), Cortellis (Clarivate, 1995-2016), and ClinicalTrials.gov (National Library of Medicine, 2010-2016). The sources differ in the time period covered, reporting conventions, drug-name formatting

[33] We are grateful to Ben Berger, Amitabh Chandra, and Craig Garthwaite for sharing a crosswalk between drugs and indications, which covers a subset of our period of interest. We build on Berger et al. (2021) to produce a crosswalk between drugs and uses that covers our full sample period.

conventions, and formatting. To create a harmonized dataset, we develop and validate linkage procedures for each database by standardizing drug names by hand, and mapping records to the active moiety identifiers introduced in Section 4.1. This yields a drug-by-disease-by-year panel of clinical trial activity. Appendix A.3 describes these procedures in detail.

Scientific Publications We collect records of scientific publications from the National Library of Medicine's PubMed database (National Library of Medicine, 1966-2022). PubMed indexes the near-universe of biomedical publications and assigns standardized Medical Subject Headings (MeSH) to the chemicals and diseases discussed in each paper. To facilitate our analysis, we construct a crosswalk from PubMed's MeSH chemical terms to the active moieties in our sample by hand.

We use these records to construct two measures. First, for each publication, we record whether the authors acknowledge support from a public funding source. Second, we use disease descriptors in the MeSH vocabulary to trace the arrival of newly documented associations between drugs and diseases. For each innovator drug, we identify every disease descriptor that appears alongside the drug in the PubMed-indexed literature and record the first year in which each drug–disease pairing appears. We refer to the first observed appearance of a previously unseen pairing as a *new disease association*, which we interpret as a coarse proxy for a genuine lead linking drug i to disease j. Appendix A.2 describes the construction of these data in detail.

4.4 Measuring Exclusivity Expiration

The framework introduced in Section 2 offers a simple prediction: when intellectual property rights become unenforceable, private firms have no incentive to pursue further research and development. For new uses, this loss of enforceability occurs when substitutable generic competitors enter the market. Empirically, then, we are interested in measuring trends in investment and commercialization in the periods around expected exclusivity expiration.[34]

The most natural empirical proxy for a firm's expectation of the time when it will lose exclusivity is the date on which generic entry actually occurs. This should reflect market participants' rational expectations. However, the difficulty in using the date of generic approval as a proxy for expected exclusivity expiration is that more than half of innovator drugs in our sample (548 / 990, 55 percent) do not experience generic entry during our sample period. Similar trends in generic entry have been well-documented and thus are not peculiar to our data. Although a series of legal reforms in the 1980s—including the 1984 Hatch-Waxman Act—are credited with increasing generic entry for small-molecule drugs, FDA (2025) lists more than 350 ever-approved small-molecule drugs with no active patents (or regulatory exclusivities) for which there is no associated generic version.[35]

[34] Durvasula et al. (2023) offers an extended discussion of both the legal and empirical challenges associated with measuring market exclusivity.

[35] Scott Morton (1999) documents similar patterns and analyzes the determinants of these generic entry decisions. For a recent analysis of challenges that limit generic entry in the "patent afterlife," see Hemel and Ouellette (2023a). There are fewer small-molecule drugs without generic entry in FDA (2025) than in our data because FDA (2025) covers an

An alternative proxy might be the date on which the patent or patents on the drug expire.[36] There are three challenges with constructing such a measure. First, new drugs approved for sale in the United States are, often, protected both by patents and by non-patent "regulatory exclusivities." The FDA will deny approval to a competing generic product so long as certain exclusivities are in force, and these exclusivities may run sequentially or concurrently with patents.[37] Second, new drugs are, increasingly, protected by many patents, not all of which will block generic competition.[38] The median innovator drug approved in 1990 was protected by a single patent grant. In 2014, the median drug had five associated patents. These patent portfolios can confer periods of nominal exclusivity that far exceed that provided by a single patent. Third, and relatedly, would-be generic competitors are incentivized—as part of the regulatory design of the 1984 Hatch-Waxman Act—to "challenge" patents on brand name drugs by asserting that any patents in force are invalid, unenforceable, or not infringed by the generic.[39] Successful challenges allow generic entry before the expiration of all patents. Any potential measure based on the expiration of patents, then, must account for the likelihood that each patent and regulatory exclusivity will successfully block generic entry.

With these practical limitations in mind, our primary measure of exclusivity is constructed as follows. We use the actual date of generic approval if generic entry occurs before the expiration of all of the drug's patents and exclusivities. Otherwise, we assume that firms expect all patents and regulatory exclusivities to bind and use the latest expiration date.[40] Formally, for each drug, we define the year that exclusivity expires as

$$\text{exclusivity} = \min\{\text{generic approval}, \text{expiration of all patents/exclusivities}\}.$$

In effect, our definition selects an upper bound on the firm's expected period of exclusivity. We define an additional measure of exclusivity, *minimum patent exclusivity*, that takes the earlier of generic approval and the expiration date of the first-filed patent. This second measure functions, in

additional eleven years of generic drug approvals.

[36] As Section 4.5 discusses, given our substantive interest in the relationship between research investments and periods of market exclusivity provided by intellectual property rights, we restrict consideration to those drugs for which we observe at least one form of intellectual property. Section 4.1 of Durvasula et al. (2023) investigates the very small number of drugs for which there is no linked patent or regulatory exclusivity in the Orange Book.

[37] Regulatory exclusivities are often called quasi-patents, as they also allow a firm to exclude competitors from the market for some fixed period of time. They differ from patents in two key ways. First, regulatory exclusivities in the pharmaceutical sector are awarded by the FDA on the basis of its own assessments of a product (e.g., the FDA itself determines whether a drug is a "new chemical entity" and thus eligible for a five-year exclusivity grant). Second, exclusivities are enforced by agency: so long as exclusivities are in force, the FDA will limit the actions that may be taken by generic competitors.

[38] For more details on "patent portfolios" in pharmaceutical markets, see Hemphill and Sampat (2012); Durvasula and Ouellette (2026).

[39] 21 U.S.C. § 355(j)(5)(B)(iv). For instance, the first (successful) generic challenger is potentially eligible for a bounty in the form of exclusive rights to the generic market for a 180-day period.

[40] We restrict consideration to the four regulatory exclusivities that we determined, in Durvasula et al. (2023), are "generic blocking" in the sense that the FDA will not approve a generic form of the active moiety while these exclusivities are in place: new chemical entity exclusivity (NCE), orphan drug exclusivity (ODE), pediatric exclusivity (PED—only instances of PED that extend a form of exclusivity that is generic blocking), and Generating Antibiotic Incentives Now exclusivity (GAIN). Technically, there are sixteen types of FDA-granted regulatory exclusivities. Table 5 in Durvasula et al. (2023) provides a list.

effect, as a lower bound. We return to its interpretation in Section 5.2, where it is used in one of our key empirical tests.

To construct our primary exclusivity measure for each innovator drug in our sample, we collect records of generic competitor approval and records of all patents and exclusivities. Specifically, we identify the first approval of a generic competitor for each drug in our sample in Drugs@FDA, as the first approval of the same combination of active moieties via the FDA's generic approval pathway.[41] We also collect records of patents and regulatory exclusivities in the Orange Book, as digitized in the NBER Orange Book Dataset, for each of our sample years (U.S. Food and Drug Administration, 1984-2016a). Appendix Figure A1 plots the distribution of this exclusivity measure for our sample of innovator drugs. The median drug in our sample receives 13.37 years of exclusivity (mean, 12.90 years) and receives 8.37 years of minimum patent exclusivity (mean, 9.35 years).

4.5 Sample Construction and Summary Statistics

There are 1,102 drugs that meet our criteria for innovator drugs. Because our empirical design requires measuring the period during which legal protection from generic competition remains in force, we further restrict consideration to drugs with at least one patent or regulatory exclusivity.[42] This yields a final sample of 990 drugs. Table 1 presents summary statistics.

5 Empirical Evidence

This section presents three sets of evidence that, together, indicate that the loss of enforceable intellectual property protection creates a missing market for new uses. We begin in Section 5.1 by establishing that each of the three empirical signatures of a missing market (Section 3.3) is borne out in our data: the likelihood that a new use is approved is high shortly after initial approval, declines as exclusivity expiration approaches, and falls to approximately zero after. We also show that new

[41] This definition of generic approval introduces an important simplification. Generics are typically approved at the level of a particular product (e.g., gemcitabine in a one-gram injectable form) and may not be approved for every dosage or form that is available for the brand name product. In some cases, these limited approvals reflect the existence of patents and exclusivities that protect some forms of a drug but not others. The full details of the Gemzar case study are reported in Durvasula et al. (2023). In other cases, limited entry is part of a negotiated settlement. A second feature of this definition is neither an assumption nor simplification, but a piece of institutional context worth noting: we focus on the timing of generic approval, rather than generic entry. Hemphill and Sampat (2012) establish that these two dates are typically close in time. In a small number of cases, however, there may be lengthy delays between generic approval and entry—if, for example, there is some agreement between the parties to delay generic entry.

[42] We also exclude a small number of drugs for which, in our data, generic competitors appear to have been approved prior to the approval of the innovator drug itself, which yields a negative exclusivity duration: NDA #19008 (bretylium); NDA #19931 (sulfacetamide); NDA #21794 (dapsone); NDA #22041 (hydroxocobalamin); NDA #22556 (carbinoxamine); NDA #50606 (vancomycin); NDA #50640 (oxacillin); NDA #50763 (mitomycin); NDA #204031 (acetaminophen; oxycodone). For each, we review FDA records and drug approval histories to understand why generic entry appears to predate NDA approval. In each case, although our data construction method properly captures the first NDA approved by the modern FDA—that is, since the 1962 passage of the Kefauver–Harris Amendment to the Federal Food, Drug, and Cosmetic Act created the modern regulatory system—our data do not capture earlier approvals under the pre-1962 regime. Generic versions of these drugs were approved before the NDAs that we flagged as innovators. There are a small number of other conceptual difficulties that arise when drugs in our sample are grandfathered into new regulatory schemes introduced in the 1960s and 1980s. We offer an extended discussion in Durvasula et al. (2023).

scientific leads arise at roughly the same rate before and after exclusivity expiration but translate into new use approvals only before expiration. Section 5.2 shows that drugs with longer exclusivity periods have more commercialized new uses and uses this relationship to estimate the elasticity of research investment with respect to the duration of effective intellectual property protection. Section 5.3 provides additional evidence on the mechanism underlying the new facts in Sections 5.1 and 5.2: at exclusivity expiration, commercialization shifts toward related products for which intellectual property rights remain at least partially enforceable. Section 5.4 summarizes our empirical evidence.

5.1 Timing of Approvals

Figures 2 and 3 plot trends in the probability that an innovator drug is approved for a new use in each year.

Figure 2 plots this probability of re-approval relative to the year of initial FDA approval. Consistent with the predictions in Section 3.3, the probability that a new use is approved rises rapidly after initial FDA approval, remains elevated early in the drug's lifecycle, and declines as exclusivity expiration approaches. At peak, between two and five years after initial approval, approximately five percent of drugs are re-approved for a new use in each year. Within fifteen years of initial approval, the probability that a drug is reapproved for a new use is approximately zero. This timing roughly coincides with exclusivity expiration—the point at which the innovator loses protection from generic competition for the underlying drug and rights covering new uses become practically unenforceable—for the typical drug in our sample.[43]

To examine the relationship between commercialization and exclusivity directly, Figure 3 plots the same outcome relative to the year of exclusivity expiration, where $t = 0$ is the calendar year in which exclusivity expires. The probability that an innovator drug is approved for a new use peaks ten years prior to exclusivity expiration, when approximately seven percent receive approval for at least one new use. It declines sharply as expiration approaches and remains close to zero in the years that follow. Appendix B shows that our results are robust to several alternative cuts of the data, including dividing each drug's exclusivity period into equal-length intervals and examining separately cohorts of drugs that receive the same number of years of exclusivity.

Of the 511 new uses approved for innovator drugs in our data, 494—96.7 percent—are approved no later than the calendar year in which exclusivity expires. Only 17 are approved in $t \geq 1$, the calendar years following the expiration year. That is, only three percent of new use approvals associated with the nearly 1,000 innovator drugs in our sample occurred after intellectual property rights became unenforceable. We hand checked each of these 17 cases and found that nearly all reflected development and regulatory review already underway before expiration, occurred before any generic version of the drug had been approved, or involved narrow changes that FDA records suggest were required for the manufacturer to continue selling the drug safely.[44]

[43] Recall that although drugs in our sample receive quite different periods of exclusivity, the mean and median are approximately thirteen years.

[44] Appendix B.1.4 reviews the 17 approvals that occurred in $t \geq 1$. The only two approvals that occurred more than five years after expiration involve unusual safety and public-health contexts.

Because Figure 3 uses all available drug-years, the set of observations contributing to the figure changes in the earliest and latest relative years. Figure 4 reports five- and ten-year balanced panels that hold the sample of drugs fixed around exclusivity expiration. The pattern in Figure 3 is robust to both modifications; in both panels, the probability of a new use approval declines as expiration approaches and remains near zero in the subsequent calendar years.

Our primary outcome follows approvals linked to the innovator drug's original application (a "new drug application") to the FDA. Appendix Figure A5 considers a broader outcome that includes, instead, every new use approved under any application to the FDA, submitted by any sponsor, containing the same combination of active moieties. Trends in calendar time and event time are essentially unchanged with this modification. That is, the decline in commercialization that we document in Figures 2 and 3 does not mask, for example, a shift toward new uses pursued by other sponsors or through separate FDA applications once exclusivity expires.[45]

A natural concern is that these sharp declines in commercialization reflect a decline in the underlying flow of scientific ideas, rather than the loss of enforceable intellectual property protection. Put differently, perhaps fewer new uses are commercialized after exclusivity expires simply because fewer promising new uses remain to be discovered.

Our first observation is that such an explanation requires a somewhat astonishing coincidence—that commercially relevant scientific opportunities also decline sharply at the same point that exclusivity expires, a date that varies substantially across drugs in our sample. Second, our data allow us to speak directly to this concern. Specifically, we use scientific publication data to construct a coarse proxy for the arrival of potential scientific "leads": newly documented relationships between drugs in our sample and diseases. For each publication indexed in PubMed, the National Library of Medicine assigns standardized Medical Subject Headings (MeSH) identifying, among other things, the chemicals and diseases discussed in the article. We link these chemical headings to the innovator drugs in our sample and record the first year in which each drug–disease pair appears together in the scientific literature. For example, if an article were the first PubMed-indexed publication in our data to associate acetaminophen with breast cancer, we would count one new acetaminophen–breast-cancer association in that year. We refer to these first appearances as *new disease associations*.[46] The key advantage of this measure is that it primarily reflects research investment by academics, physicians, and publicly funded researchers, whose incentives are not directly tied to the innovator's ability to capture an economic return from the sale of a new use for an existing drug. If new disease associations continue to appear in the scientific literature after exclusivity expires even as commercialization declines, then, such a pattern weighs against the idea that the decline simply reflects an exhaustion of scientific opportunities.

Figure 5 plots the number of new disease associations appearing in each year relative to exclu-

[45] Appendix Figure A9 provides complementary evidence from clinical development. Using our clinical trial records, we identify the first year in which each drug is studied for a previously unobserved indication. The arrival of newly studied indications declines as exclusivity expiration approaches and remains low thereafter.

[46] Although the full texts of scientific publications are not available at the scale required to classify each drug–disease relationship by its substantive content, the standardized structure of the MeSH vocabulary allows us to construct these links for our full sample of innovator drugs. Appendix A.2 and Appendix B.2 provide details on data construction.

sivity expiration. There is no discernible break in the flow of new disease associations at exclusivity expiration, in sharp contrast to the collapse in approvals of new uses.[47]

Panel A of Table 2 makes this comparison explicit. The mean number of new disease associations changes little across the five years before and after exclusivity expiration, from 9.92 to 9.77 associations per drug-year. Across the same windows, however, the probability that an innovator drug is approved for a new use falls by more than 80 percent, from 3.7 percent to 0.6 percent. Thus, our proxy for the arrival of new scientific information is essentially unchanged across exclusivity expiration even as commercialization of new uses falls to near zero.

Of course, our measure of drug-disease associations is broad and captures newly documented relationships between drugs and diseases, not only potential therapeutic uses. This breadth leaves open the question of whether these associations are informative about scientific opportunities for new uses. We take up this question in Panel B of Table 2, which asks whether recent drug-disease associations *predict* subsequent approvals of new uses before and after exclusivity expiration. Before expiration, ten additional drug–disease associations observed over the preceding five years are associated with a 0.39 percentage point increase in the probability of a new-use approval—approximately 8 percent relative to the relevant sample mean. After expiration, ten additional associations are associated with just a 0.03 percentage point increase in the probability of a new-use approval. The difference between these relationships is 0.36 percentage points and statistically significant ($p = 0.003$). Put differently, newly arriving scientific information predicts the successful commercialization of new uses only while intellectual property rights remain enforceable.

5.2 Quantity of Approvals

Section 5.1 shows that each of the three empirical signatures of a missing market is borne out in our data. Commercialization is positive after initial approval, declines as the remaining period of enforceable protection shortens, and falls to approximately zero once protection ends. New scientific leads continue to arise even as commercialization disappears. These facts provide strong evidence of a missing market. We next ask whether shorter protection also reduces the total quantity of commercialization, rather than simply causing firms to bring otherwise identical projects to market earlier. Our framework generates a further prediction that distinguishes these possibilities: drugs with longer periods of enforceable intellectual property protection should ultimately yield more commercialized new uses.

[47] Appendix Figure A7 plots the cumulative number of newly documented disease associations. Appendix Figure A8 provides complementary evidence using instead data on the aggregate number of publications associated with each innovator drug in our sample. Specifically, Figure A8 plots the total number of publications associated with each innovator drug, separately for publications with and without recorded public funding. Over the twenty-year window around exclusivity expiration, total publication volume declines by roughly one-half. Publications acknowledging public funding, however, remain approximately constant, with roughly ten publicly funded publications per drug-year; the aggregate decline is driven entirely by the set of publications without public support.

5.2.1 Investigating the effect of drug-level exclusivity on the quantity of new uses

In principle, patents provide uniform protection to all inventions: a twenty-year term, which begins to elapse from the moment of initial filing. In practice, however, features of patent law and regulatory law in the pharmaceutical industry mean that some drugs receive much more *effective* protection from a patent grant than do others.

Patent law encourages pharmaceutical firms to file for patent protection very early in the research and development process—in the notation of Section 2, at t_{invent} rather than at t_{comm}. Delays in filing introduce at least two substantial risks: first, that another inventor will file first; and second, that any information about the invention released by the firm itself—for example, the clinical trial protocols and results that firms are required by law to disclose—will become "prior art" that defeats patentability.[48] As a result, core patents—most often, compound or molecule patents that provide the most certain protection from generic competition (Lemley and Shapiro, 2005; Hemphill and Sampat, 2011; Durvasula et al., 2023)—are typically filed prior to the initiation of clinical trials (Mossinghoff, 1999; Patrick, 2005; Thomas, 2007). Because patent terms run during development, drugs that take longer to proceed through clinical trials and regulatory approval enter the market with fewer years of protection remaining from these early-filed patents.

The time required to complete satisfactory clinical trials varies considerably across drug-disease pairs, due to differences in disease biology, drug characteristics, and regulatory requirements.[49] For some conditions, clinically meaningful endpoints can be observed relatively quickly, allowing for short trials. For other diseases—for example, diseases that progress slowly or that involve chronic symptoms—however, establishing efficacy can require much longer periods of study. Variation in development time can thus generate substantial variation in the amount of patent protection remaining at commercialization (Roin, 2013; Budish et al., 2015a). Budish et al. (2015a) leverage this same institutional feature to show that fixed patent terms distort cancer research toward projects with shorter commercialization lags. Their empirical analysis uses five-year survival to proxy for differences in commercialization time and combines this evidence with additional tests designed to isolate the role of commercialization lags. Our setting allows us to observe variation in the duration of effective protection directly. Starting at initial approval, we observe drug by drug how much protection remains and ask whether drugs with longer remaining protection generate more subsequent new uses.

We measure this variation using the measure of *minimum patent exclusivity* defined in Section 4.4. Minimum patent exclusivity captures the period between initial FDA approval and the earlier of two events: generic approval and the expiration of the first-filed patent and all generic-blocking regulatory exclusivities. This measure has two advantages for our test. First, first-filed patents gen-

[48] Patents are granted only on those inventions that are novel, non-obvious, and useful. Publicly available information about an invention renders a claim *non-novel*. Disclosures about other, related aspects of development can render a claim *obvious*.

[49] FDA regulations require "adequate and well-controlled" studies designed to distinguish treatment effects from other influences, including spontaneous changes in the progression of disease.

erally protect drugs' molecules[50] and thus—together with generic-blocking exclusivities—provide the strongest protection against generic competition, as they block competitors from making and selling the chemical itself. Second, the measure excludes protection provided by later-filed patents, which firms may obtain in response to the drug's realized or anticipated commercial value. Minimum patent exclusivity therefore maps more directly to the variation generated by differences in commercialization lags than does a measure that incorporates the full patent portfolio.

Consistent with the expected heterogeneity in clinical trial duration across drug–disease pairs, our measure of minimum patent exclusivity varies substantially across innovator drugs. The average innovator drug receives 9.4 years of minimum patent exclusivity, while those at the 10th and 90th percentiles of this measure receive approximately five and fifteen years, respectively.[51] Throughout this section, we also report estimates using the broader measure of exclusivity used in Section 5.1 but, because this measure includes each drug's full patent portfolio, its mapping to the variation generated by commercialization lags is less precise.

Interpreting these estimates as the effect of longer enforceable protection requires assuming that variation in a drug's minimum patent exclusivity at approval is not systematically related to its scientific potential for new uses. In Appendix B.4, we examine this assumption using two proxies for scientific potential derived from the biological characteristics of drugs in our sample. First, we identify the number of human proteins that each drug is known to act on in producing its therapeutic effects using data from DrugBank. Second, we identify the number of diseases linked, through human-genetic evidence, to such proteins using data from the Open Targets Platform. Intuitively, drugs that act on more proteins, or on proteins implicated in more diseases, likely have greater potential for additional therapeutic uses.[52]

Appendix B.4 first asks whether drugs that score higher on these measures are, in fact, approved for more new uses. This relationship is positive for both measures, although it is estimated more precisely for the human-genetic measure. We next ask whether drugs with longer periods of minimum patent exclusivity also score higher on either measure of scientific potential. Across both proxies and both measures of exclusivity, the estimated relationships are small and statistically indistinguishable from zero. That is, we find no evidence that drugs with longer periods of effective

[50] In our data, the first-filed patent generally protects the drug's molecule. These grants are often described as "compound patents" (Kapczynski et al., 2012). We reviewed the text of the first-filed patent for a random 20 percent sample of our data ($N = 198$). For 159 drugs ($\sim 80\%$), the patent text described a claimed chemical structure.

[51] Appendix Figure A2 plots the distribution of minimum patent exclusivity. In 18 cases (two percent), this measure exceeds the (statutory) standard twenty-year patent term. In nearly all cases, this arises because the earliest patent was filed after the drug's initial approval. While it is generally true, as we note above, that firms have strong incentives to file key patents before approval, these cases are the rare exception. On inspection, many of these later-filed patents appear to be method-of-use patents on drugs for which there is no Orange Book patent that protects the active ingredient. As Footnote 50 notes, the vast majority of first-filed patents in our data are so-called compound patents that protect the active ingredient.

[52] These measures are especially useful for our purposes because they do not mechanically assign greater scientific potential to drugs that have already attracted research investment. A drug's protein targets are largely determined by its chemical structure. Our measure captures, in effect, the number of distinct pieces of human biology on which the molecule can act, in a mechanistic sense. Our human-genetic measure is constructed using data from studies in which researchers scanned patient genomes in order to map naturally occurring genetic variation across large populations and asked which genes and proteins are associated with particular diseases.

protection have systematically greater measured scientific potential for additional uses.

Figure 6 plots the relationship between exclusivity duration and the number of new uses realized over each drug's lifecycle. Panel A uses minimum patent exclusivity, while Panel B uses the broader exclusivity measure. In both panels, we residualize exclusivity duration and new use approvals with respect to initial-approval-year fixed effects and group drugs into twenty equal-sized bins. Under either measure, drugs with longer periods of enforceable protection have more approved new uses.

Table 3 formalizes this relationship. For innovator drug i, we estimate

$$y_i = \beta \cdot \text{exclusivity}_i + \mu_{a(i)} + \varepsilon_i \tag{5}$$

where y_i captures the number of approvals of innovator drug i for a new or modified indication and $\mu_{a(i)}$ is a fixed effect for the drug's year of initial approval.[53] The coefficient of interest, β, captures the change in the number of commercialized new uses associated with an additional year of drug-level exclusivity—and therefore with an additional year during which rights covering new uses remain practically enforceable. Columns (1) and (2) report results using minimum patent exclusivity; Columns (3) and (4) instead use the broader measure of exclusivity. The relationship between this measure of exclusivity and the quantity of realized new uses is positive and significant in all specifications, with and without the approval year fixed effect μ. In what follows, we focus on the interpretation of the estimate in Column (2) in Table 3, as the specification with an approval year fixed effect accounts for the concern that earlier-approved drugs have had more years in which to realize additional new uses.

In Column (2), an additional year of minimum patent exclusivity is associated with 0.023 additional approvals of new uses, with a standard error of 0.008. Column (4) yields a similar estimate of 0.020 using the broader measure of exclusivity, with a standard error of 0.006. Thus, an additional year of enforceable protection is associated with approximately two additional commercialized uses for every one hundred innovator drugs.

Because these estimates capture averages over drugs' full periods of enforceable intellectual property protection, they may obscure discrete changes in commercialization at the time of exclusivity expiration. Appendix B.3 presents a complementary test that focuses on this margin. To fix ideas for this test, consider two drugs that are ten years out from initial approval ($t = 10$). Suppose that one drug receives ten years of exclusivity ($T = 10$), while the other receives eleven ($T = 11$). If the duration of enforceable intellectual property protection affects the level of commercialization, we should expect to see a difference in the probability that the still-protected drug is reapproved for a new use relative to the drug whose exclusivity has just expired in $t = 11$. We calculate this difference for each cohort of drugs with observed exclusivity duration T, comparing the probability that it is reapproved for a new use in lifecycle year $t = T + 1$ with the corresponding probability

[53] Baseline therapeutic class need not describe the opportunity set for later uses, which may involve different disease areas. Our baseline specification includes initial approval-year fixed effects, and Appendix B.4 examines available measures of scientific opportunity. Neither choice eliminates the possibility that other drug characteristics are associated with both protection duration and new-use commercialization.

for drugs that receive exactly $T+1$ years of exclusivity. The estimated difference is positive but, given that approvals for new uses are rare and the sample sizes in these cohort-level comparisons are small, imprecisely estimated. To increase the power of this test, we broaden the comparison group and compare drugs with T years of exclusivity to all drugs that remain protected in lifecycle year $t = T+1$. In this broader comparison, drugs that remain protected are approximately 1.9 percentage points more likely to receive a new use approval in year $t = T+1$, using either measure of exclusivity.

Note that nearly three-quarters of innovator drugs in our sample ($N = 725$, 73 percent) have zero re-approvals for new uses. In Panel B of Table 3, we report estimates from a two-part model that estimates the extensive and intensive margins separately. On the extensive margin, the dependent variable is an indicator for whether the drug is reapproved for *any* new use, and we estimate a logit model. On the intensive margin, we restrict the sample to drugs with at least one new use approval and estimate the relationship between exclusivity duration and the number of approvals using ordinary least squares.

The extensive margin estimates are positive and significantly different from zero. In the specifications with initial approval year fixed effects, an additional year of minimum patent exclusivity increases the odds that a drug is approved for any new use by approximately six percent. The corresponding estimate using our broader measure of exclusivity is approximately four percent.[54] Although the intensive margin restriction leaves us underpowered, estimated coefficients from each specification are positive and comparable in magnitude to those reported in Panel A; none, however, are significantly different from zero.

In addition to providing support for the existence of missing incentives for investment in new uses, this empirical exercise speaks directly to a long-standing open question: do longer periods of intellectual property protection cause additional investments in research and development? That more new uses are commercialized when drugs receive longer periods of enforceable protection provides, to our knowledge, the first direct evidence that the answer to this question is yes.

5.2.2 Interpreting these estimates as an elasticity

This empirical exercise also yields an estimate of the elasticity of commercialization with respect to the duration of enforceable intellectual property protection. This parameter has long been difficult to estimate because the largely uniform structure of patent protection provides little variation in the duration of protection across otherwise comparable inventions (Nordhaus, 1969; Budish et al., 2016).[55] Our setting allows us to make progress by using variation not in the nominal patent term, but in the period during which intellectual property rights are enforceable in practice. Though this

[54] We exponentiate the logit coefficients and subtract one to obtain the proportional change in odds. In columns (2) and (4), the coefficients 0.059 and 0.042 imply changes of approximately 6.1 and 4.3 percent, respectively.

[55] A survey article by Budish et al. (2016) writes: "A key parameter needed to inform optimal patent policy design is the elasticity of research investments with respect to the patent term. Estimating this elasticity is conceptually difficult because it requires constructing a counterfactual in which we can infer that some scientifically feasible inventions would have been brought to market under an alternative patent policy design. Despite a near half-century of research effort, we have essentially no credible empirical evidence on this elasticity."

elasticity is necessarily specific to our empirical context, it is directly relevant to any efforts to alter the structure and availability of intellectual property rights. In Section 6, we use the corresponding level estimates to estimate the number of uses that would have been developed had intellectual property rights on new uses been enforceable for a longer period of time, to offer a rough quantification of the social costs of inadequate incentives for investment.

To express our estimates as elasticities, we evaluate them at the sample means. Let $\bar{y} = 0.52$ denote the average number of new use approvals per innovator drug and denote by $\bar{x}$ the average years of exclusivity (9.35 for minimum patent exclusivity and 12.90 for our broader exclusivity measure). The point elasticity at the mean is given by $\beta \cdot \frac{\bar{x}}{\bar{y}}$. Using the estimate for minimum patent exclusivity yields $0.023 \times {}^{9.35}/_{0.52} \approx 0.41$. Using the estimate associated with our broader exclusivity measure yields $0.020 \times {}^{12.9}/_{0.52} \approx 0.50$. That is, a one-percent increase in the duration of enforceable protection is associated with an approximately 0.4 to 0.5 percent increase in commercialized new uses, evaluated at the sample means. Equivalently, an additional year of minimum patent exclusivity increases the expected number of new-use approvals by approximately 4.4 percent; the corresponding estimate using exclusivity is approximately 3.8 percent.

Budish et al. (2015a) and Abrams (2009) offer two points of comparison, although neither estimate captures precisely the same object. Budish et al. (2015a) estimate that a one-year increase in commercialization lag reduces private investment in cancer research by approximately 7 to 23 percent. In their setting, a longer commercialization lag both delays the time until an innovator can capture an economic return on her investment and reduces the patent term remaining at commercialization. Their estimates then capture both the effect of corporate short-termism and the fixed patent term. Our setting instead allows us to use variation in the period of effective protection remaining after an innovator drug reaches the market to measure the number of new uses that are ultimately commercialized. Abrams (2009) uses changes in the length of patent terms induced by the adoption of the 1995 Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS) to estimate the responsiveness of patent filings to changes in patent term length. In that context, a one-year increase in patent term is estimated to generate a 66 percent increase in patent filings. While the setting of Abrams (2009) yields policy-induced variation in patent duration, it focuses on the production of *patents*, which in turn captures incentives to seek patent protection, rather than the actual discovery and commercialization of *inventions*. These differences in design and outcome make it difficult to compare the estimates quantitatively.

5.3 Composition of approvals

Sections 5.1 and 5.2 show that commercialization of new uses tracks the enforceability of intellectual property rights. We next ask whether the decline in commercialization at generic entry is specific to inventions whose intellectual property protection becomes practically unenforceable at that point. To do so, we compare new uses with a related class of inventions involving the same molecules for which intellectual property protection may remain enforceable after generic entry.

To fix ideas, suppose an innovator drug A loses exclusivity. Once generic versions of A are

available, physicians and patients can capture the value of the new use using its generic equivalent. But the firm might instead develop a new product that combines A with another drug, B. Generic A is not equivalent to $A+B$, and, as we discuss below, even generic A plus generic B may not be a perfect substitute for $A+B$. Generic entry for A, then, need not eliminate the firm's ability to protect and earn returns from $A+B$.

We test this idea by considering the case of *fixed-dose combinations*—drugs dispensed as single "dosage forms" (e.g., a single tablet) that contain two or more chemicals that are regulated as a single, distinct drug. The widely used HIV treatment and preventive Truvada, for instance, is a single tablet that contains two distinct drugs, emtricitabine and tenofovir disoproxil fumarate, neither of which on its own offers the same therapeutic benefits.[56] fixed-dose combinations function like new small-molecule drugs and are eligible for the same types of intellectual property rights, including patents on molecules.[57] When the fixed-dose combination contains at least one never-approved chemical, the entire combination is eligible for a five-year FDA-administered exclusivity that bars all generic entry. Even if the components of a fixed-dose combination are available in generic form, the combination form is often preferred by patients and providers, as it can simplify complex treatment regimens and increase prescription adherence.[58]

We examine whether firms shift investments toward fixed-dose combinations as exclusivity expiration approaches. Specifically, we restrict attention to innovator drugs composed of a single active moiety and track combination products containing the same moiety. We count cases in which the combination receives initial FDA approval or an existing combination receives approval for a new or modified indication. We exclude supplemental approvals limited to new strengths, dosage forms, formulations, or patient populations.

Figure 7 plots the probability of a fixed-dose-combination approval in each year relative to exclusivity expiration for the underlying innovator drug. In the calendar year of expiration, the probability that a new combination is approved—or that an existing combination receives approval for a new use—more than doubles relative to the preceding year. The probability of a combination approval remains elevated in the years that follow, while approvals of new uses for the underlying

[56] Truvada (emtricitabine; tenofovir disoproxil fumarate), NDA#021752, was approved by the FDA on August 2, 2004. A generic version, manufactured by Teva Pharmaceuticals, was approved on June 8, 2017. Emtricitabine was approved on July 2, 2003, and a generic approved on April 17, 2017. Tenofovir disoproxil fumarate was first approved on October 26, 2001, and a generic was approved on January 17, 2012.

[57] A recent paper by Dix and Lensman (2024) studies a distinct incentive problem associated with investment in combinations of drugs. Motivated by features of markets for cancer drugs, they develop a model in which two existing drugs A and B may be complementary, in the sense that their therapeutic value when used together exceeds the therapeutic value of each taken in isolation; that is, $v(A+B) > v(A)+v(B)$, where $v(.)$ denotes therapeutic value. The authors document evidence consistent with the idea that firms underinvest in tests of combinations, especially when components are owned by different firms. Relevant here, the combinations of interest to Dix and Lensman (2024) differ in form, function, and regulation relative to the fixed-dose combinations we consider in this section. Combination cancer therapies are dispensed as separate components of a single regimen, almost always in inpatient settings; each component is regulated as a distinct drug. Fixed-dose combinations, by contrast, are dispensed as single products, themselves regulated as new drugs.

[58] Combination drugs have many advantages for patients, chief among which is simplicity. The development of the first single-tablet regimen for HIV, Atripla (Efavirenz/emtricitabine/tenofovir), was lauded as a significant improvement for patients, many of whom had previously been taking up to 36 distinct pills per day. See Pollack (2006). Analyses of claims data suggest that treatment adherence is higher in single-tablet regimens than in multi-tablet regimens. See Silverman (2018).

innovator drug fall to approximately zero.

This fact has two additional implications worth emphasizing. First, that the number of approvals of fixed-dose combinations *increases* at the time of exclusivity expiration further counsels against the idea that scientific opportunities coincidentally drop off at this point in time. Second, the timing of the change is consistent with the idea that our measured expiration date captures an economically relevant change in incentives.[59]

5.4 Summary

Sections 5.1 through 5.3 provide complementary evidence on the existence and source of a missing market for new uses. Together, our results address three concerns with this interpretation: that the decline in commercialization may reflect a shift in the timing of investment rather than missing investment; that the relationship between exclusivity and commercialization may arise for reasons other than the loss of enforceability; and that the patterns we document may instead reflect a decline in underlying scientific opportunities. We consider each concern in turn.

Evidence of Missing Investment—Viewed in isolation, the trends in Figure 3, although consistent with the empirical signatures of a missing market predicted by our framework, raise a natural concern. The timing of exclusivity expiration is, as we note, not an unexpected discontinuity; it is known to all market participants well in advance. If firms know both when exclusivity will expire and which new uses are worth pursuing, they may strategically sequence investment so that those uses reach the market while intellectual property rights remain enforceable. Put differently, the sharp decline in commercialization around exclusivity expiration could reflect a shift in the timing of investment without indicating that any potential new uses are, in fact, missing.

Sections 5.1 and 5.2 provide evidence to the contrary. In Section 5.1, we show that new scientific leads continue to arise at roughly the same rate after exclusivity expiration, even as they largely cease to translate into new use approvals. Even if firms can bring forward investment in *known* new uses, they cannot similarly accelerate the development of opportunities that have not yet been identified. The continued arrival of new leads after expiration thus provides prima facie evidence that some potential new uses are, in fact, missing. Section 5.2 addresses the timing concern more directly. Drugs with longer periods of enforceable protection remaining at initial approval are ultimately approved for more new uses over their lifecycles, indicating that longer protection affects the total amount of commercialization rather than simply when commercialization occurs.

Evidence of Enforceability as the Mechanism—A second concern is that the decline in commercialization may reflect weak commercial incentives for new uses more generally, rather than the loss of enforceability in particular. All three exercises suggest otherwise. In Section 5.1, we document that, although the duration of exclusivity varies substantially across drugs in our sample,

[59]The pattern we are documenting is consistent with practices that are often labeled as "product hopping" (Carrier and Shadowen, 2016), wherein firms respond to the loss of exclusivity on one product by directing innovation toward products that are still protected by enforceable patents. This practice can impose real social costs. We do not undertake a complete welfare analysis of innovation decisions around exclusivity expiration, but note that any such analysis must take care to account for both incentives for innovation and the deadweight loss associated with delayed generic entry.

commercialization closely tracks this period of enforceable intellectual property protection. In Section 5.2, we show that drugs with longer periods of enforceable protection are ultimately approved for more new uses. And in Section 5.3, we establish that, when exclusivity expires, commercialization shifts toward related products for which intellectual property rights remain at least partially enforceable. Across all three exercises, then, commercialization tracks the availability of enforceable intellectual property protection.

Evidence against Depletion of "Scientific Opportunities" as the Cause—A third concern is that the patterns we document may reflect the exhaustion of scientific opportunities, rather than the loss of commercial incentives. All three exercises suggest otherwise. As we discuss in Section 5.1, it would be a remarkable coincidence for the exhaustion of scientific opportunities to precisely track the expiration of enforceable intellectual property rights. More directly, Section 5.1 establishes that potential scientific leads arrive at the same rate before and after exclusivity expiration. Section 5.2 likewise finds no evidence that drugs with longer periods of enforceable intellectual property protection have systematically greater scientific potential. The increase in fixed-dose-combination approvals at exclusivity expiration in Section 5.3 provides further evidence that commercially relevant opportunities involving the same molecules have not disappeared. Taken together, these findings point to a loss of incentives to invest in developing scientific opportunities, rather than a disappearance of the opportunities themselves.

6 Quantification

Section 5 collects evidence of a missing market for new uses. This section uses the estimated relationship between the duration of enforceable intellectual property protection and the level of new use commercialization in Section 5.2 to develop a rough quantification of the social value of the missing innovation that results.

We proceed in three steps. First, we use the level estimates reported in Section 5.2 to estimate how many additional new uses would have been realized had innovator drugs received longer periods of enforceable intellectual property protection. Second, we use estimates from the existing literature to generate a rough estimate of the social value of a missing indication for an already-approved drug. Third, we multiply the estimates from the first two steps and take a present value. Given the considerable uncertainty in each stage of this exercise, our objective in this section is not to arrive at a point estimate that captures the precise extent of missing research and development but, rather, to offer a sense of the magnitudes associated with the results in Section 5 and, in turn, the missing markets problem more broadly.

We emphasize that the calculations reported in Section 6.1 are not intended to represent policy proposals. They are designed to measure the potential quantity of missing research and development using the variation available in our data. Lengthening intellectual property protection for innovator drugs would also cause deadweight loss for the usual reasons associated with patent protection; for a classic treatment, see Nordhaus (1969). We discuss potential policy responses to the missing

markets problem in the concluding section.

6.1 How Many New Uses are Missing?

To estimate the number of missing new uses, ideally we would ask how many additional new uses would have been commercialized had intellectual property rights on new uses remained enforceable after exclusivity for the underlying drug expired. We consider two benchmarks using the variation available in our data. First, how many additional new uses would have been developed if protection for new uses remained enforceable for an additional five years? Five-year periods are common in pharmaceutical regulation: new chemical entities have received five years of market exclusivity since the 1984 Hatch-Waxman Act, and certain new antibiotics receive an additional five years under reforms enacted in 2012 (Durvasula et al., 2023). Second, how many additional new uses would have been developed if the innovator drug received a total effective patent term of twenty years—that is, if its patent clock started at commercialization rather than initial filing?

To operationalize these thought experiments using the variation available in our data, we use the relationship between new use commercialization and the duration of exclusivity for the underlying drug estimated in Section 5.2.

For the five-year benchmark, we ask what our estimates imply if each drug had five additional years of enforceable protection. The estimate in Column (2) of Table 3, Panel A implies 0.12 additional new uses per innovator drug; the estimate in Column (4) implies 0.10. Since 1962, the FDA has approved approximately 2,000 innovator drugs that satisfy our sample criteria. Scaling these estimates to that population implies roughly 200 to 240 additional new uses.

For the full patent-term benchmark, we ask instead what our estimates imply if each drug received twenty years of effective protection beginning at initial approval. Because drugs generally enter the market with less than twenty years of effective intellectual property protection remaining, we calculate the difference between twenty years and each innovator drug's minimum patent exclusivity; we set the difference to zero for drugs that already receive more than twenty years of enforceable protection. To the average drug, we add 10.72 years. Applying the estimates from Columns (2) and (4) yields 0.25 and 0.21 additional new uses per drug, respectively. Scaled to approximately 2,000 innovator drugs, these estimates imply roughly 420 to 500 additional new uses.

6.2 How Valuable are New Uses?

Though previous studies have concluded that new drugs, on average, generate large health benefits, there is enormous heterogeneity in the social value of particular drug approvals. Philipson and Jena (2006) estimate, for example, that the social value of survival gains from breakthrough HIV/AIDS therapies in the late 1980s was $1.4 trillion over a twenty-year period in the United States alone; Yin et al. (2012) estimate that early tyrosine kinase inhibitors, which transformed the treatment of leukemia and other cancers, produced roughly $6 billion per year immediately following initial approval. Other studies suggest that many "me-too" drugs—products that provide little clinical

differentiation from existing therapies—contribute little in the way of social value (Dranove et al., 2014). This heterogeneity in therapeutic value is compounded by an additional complication: new drugs that offer limited incremental clinical benefit to current patients may nonetheless represent real scientific advances that enable important follow-on innovation.[60]

Our goal with this back-of-the-envelope calculation, then, is modest: to implement a transparent calculation that offers a sense of the magnitude of missing discoveries. Rather than attempting to aggregate estimates from available case studies, we follow a standard approach in studies of cost-effectiveness, which values new therapeutics in terms of quality-adjusted life years (QALYs) that capture improvements in both well-being and life expectancy in standardized units.

Based on a survey of existing literature, we select values for the following: (i) the incremental QALYs generated by a new drug relative to the standard of care; (ii) the social value of one QALY, per patient, per year; and (iii) the patient population treated. Throughout, where the literature admits a range of plausible values, we err on the side of conservative assumptions. These quantities, scaled by a factor that reflects the lower average value of new uses compared to initial approvals, give our estimate of social value as $V = Q \times P \times N \times \theta$, where Q denotes incremental QALYs per patient, P is the value per QALY, N the number of patients, and θ a scaling factor for new uses.

Shafrin et al. (2023) estimate that the average drug approved by the FDA between 2011 and 2021 added 1.04 QALYs per patient relative to the standard of care, based on published cost-effectiveness analyses available for roughly half of all approvals.[61] For our back-of-the-envelope calculation, we round the average estimate to $Q = 1$. In the United States, cost-effectiveness analyses typically value one QALY at $P = \$100,000$.[62] For patient population, we assume that $N = 10,000$, a conservative estimate given industry reports suggest that specialty drugs are used to treat 10,000 to 50,000 patients per year, with an average market size of 40,000. (IQVIA Institute for Human Data Science, 2021). Under these assumptions, the annual social value of an initial approval is roughly \$1 billion ($=1\times$ 100,000 $\times$ 10,000).

We next scale these values to reflect existing evidence that new uses are, on average, less valuable than initial approvals. Vokinger et al. (2023) find that roughly one-third of new uses have high therapeutic value compared to half of initial indications, based on a sample of drug approvals from the United States and European Union between 2011 and 2020. In oncology, Michaeli et al. (2022) find that first indications provide nearly twice as many QALYs per patient, around 0.99 QALYs, relative to second or third indications, which average approximately 0.5–0.6 QALYs.[63]

[60] Trusheim et al. (2010) write, "[i]t is difficult if not impossible to quantify reliably, objectively and unambiguously the extent to which new biopharmaceuticals embody significant innovation and address unmet medical needs."

[61] The authors note substantial heterogeneity: many drugs provided less than 0.5 QALYs, while a small number provided multiple QALYs; drugs at the 25th percentile provided 0.08 QALYs, drugs at the 75th percentile provided 1.18 QALYs, and a small number of drugs provided more than 3. Note that the Shafrin et al. (2023) sample does not include the restriction that new drugs are "innovators" (the first approvals of a unique combination of active moieties) and hence is more expansive.

[62] Cutler (2004) also values life-years at \$100,000, though Hemel and Ouellette (2023b) note that this estimate is "extraordinarily low relative to the values that federal agencies use in other contexts."

[63] As complementary evidence, we assess one measure of value—market size—in our own data, using data drawn from the Medical Expenditure Panel Survey on prescriptions written for diseases associated with initial approvals and new uses. In roughly one quarter of cases, new uses are approved for indications in different disease areas than the initial

Motivated by this evidence, and recognizing the considerable heterogeneity across indications, we set $\theta = 0.5$. We thus assume that new uses generate half the value of initial approvals on average. Applying this factor implies that the average social value of a new use is \$0.5 billion.

6.3 Valuing the Missing Market for New Uses

To quantify the gross social value of missing investment in new uses, we begin by calculating the annual social value of missing new uses. Using estimates from Section 6.1 and Section 6.2, then, we multiply \$0.5 billion per new use per year (Section 6.2) by the 200–500 new uses that we estimate could have been developed with alternative exclusivity periods (Section 6.1). This yields an estimate of \$100–250 billion in gross social value per year.

To translate these annual estimates into a measure of total social value, we compute the present value of the missing market for new uses with a simple constant growth calculation that divides the annual estimated missing social value by the difference between the social discount rate r and the long-run population growth rate g. With standard, conservative values for r and g—$r = 0.05$ and $g = 0.01$—the total social value of the missing market for new uses is on the order of \$2.5 to \$6.3 trillion. This estimate is comparable to estimates of the value of missing medical innovation in existing work (see, for example, Murphy and Topel (2006), Hall and Jones (2007) and Budish et al. (2015a)).

Although these present-value magnitudes are large, the implied expansion in the stock of approved uses is relatively modest. There are approximately 2,000 existing innovator drugs, each with an average of roughly 1.5 approved uses, implying approximately 3,000 approved drug-disease combinations in the current market. Against this benchmark, the 200–500 missing new uses that we estimate corresponds to an expansion of the set of approved uses by roughly 7–17 percent. Assuming once more that new uses generate roughly half the social value of initial uses, this implies an increase in aggregate social value on the order of 3–8 percent relative to the value generated by currently approved uses.

A second way to benchmark the magnitude is to ask how small the annual value per missing use would need to be for the missing market still to be worth \$1 trillion in present value. A \$1 trillion present value corresponds to an annual flow of \$40 billion. Across 200–500 missing uses, this requires each missing use to generate only \$80–200 million per year in social value—well below the \$0.5 billion baseline used above.

One final complication might be instructive. Thus far, we have abstracted from obsolescence—the risk, represented in Section 2 as $1-\gamma$, that a use loses value over time as substitutable therapies reach the market. A use that obsolesces at an annual rate h has a present value given by the annual flow divided by r minus g plus h. Although we do not estimate h, evidence from existing literature suggests it is modest. Of interest is how quickly the social value of an innovator drug erodes

approval, where disease areas are defined broadly. Although the exercise is necessarily rough, we find that, on average, the market size for a new use is roughly 70 percent that of each drug's initial indication. This estimate is consistent with the findings of a similar exercise in Berger et al. (2021), which examines the effects of re-approvals for new uses on market size.

as superior alternatives arrive—what Lichtenberg (2002) terms the obsolescence rate of medical innovation and characterizes, following Lichtenberg and Philipson (2002), as creative destruction. Lichtenberg (2002) reports that roughly twenty percent of drugs approved between 1950 and 1993 had exited the market by 1999, implying an obsolescence rate on the order of one percent per year.

We can use this evidence to gauge the sensitivity of our calculation. At $h = 0$, the present value is \$2.5 to \$6.3 trillion; at $h = 0.01$, roughly Lichtenberg's benchmark, it is \$2.0 to \$5.0 trillion; and at $h = 0.05$, it is \$1.1 to \$2.8 trillion. Even if each missing use were worth only half of our baseline value and obsolesced at 5 percent per year, the present value would remain roughly \$0.6 to \$1.4 trillion. The qualitative conclusion thus holds: even under substantially more conservative assumptions than those used in our baseline calculation, the value of missing investment appears to be very large.

7 Conclusion and Implications for Innovation Policy

In this paper, we investigate the idea that there are large classes of potential inventions for which private incentives to invest in research and development are missing. We develop a theoretical model that clarifies conditions under which these incentive gaps arise—when inventions are at risk of immediate imitation in the absence of intellectual property protection, but intellectual property rights are either unavailable or unenforceable—and predicts that *missing markets* will arise across sectors of the economy. In doing so, we formalize a set of ideas that have long been central to theoretical studies of innovation and intellectual property—that stronger and broader intellectual property rights induce additional investment in research and development, and that not all inventions are amenable to protection with any existing form of intellectual property. Our empirical work offers, to our knowledge, the first direct tests of both sets of ideas. We identify a setting, investment in new uses for existing drugs, in which there is sharp variation in the enforceability of intellectual property rights on otherwise comparable inventions over time. In this setting, intellectual property rights shift from perfectly enforceable to wholly unenforceable at a measurable point in time. We show that when enforceability drops to zero, so too does private investment. Plausibly exogenous variation in these periods of enforceability allows us to provide causal evidence that longer periods of (enforceable) intellectual property protection yield higher levels of research and development.

We use this elasticity of research and development with respect to intellectual property duration to estimate the number of additional new uses that would have been commercialized under longer periods of enforceable intellectual property protection for new uses, using exclusivity for the underlying drug as a proxy. That is, we quantify, roughly, the size and value of a missing market for new uses. We estimate that, since 1962, between 200 and 500 new uses would have been commercialized but for these gaps in intellectual property protection. Estimates from the scientific literature on the therapeutic potential of existing drugs suggest that this range is, if anything, conservative (Gelijns et al., 1998; Wermuth, 2006). Using measures of the value of a new use drawn from the literature, we estimate that the forgone social value associated with this *particular* missing market is on the

order of $2 to $6 trillion in present value terms. Although these estimates are necessarily rough, they suggest that the social costs of the missing markets problem, in general, may be very large.

The potential scale of the missing markets problem makes more salient the question of solutions. Is it possible to "fix" a missing market? Our theoretical model rules out one standard tool, patent policy, typically deployed to correct inadequate incentives for innovation. Necessarily, neither longer nor broader periods of formal intellectual property protection—what we term "Nordhaus parameters"—affect incentives to develop potential inventions in missing markets.

Our model does, however, offer a structured way to identify alternatives. Consider a stylized version of (4) that captures the conditions under which a firm receiving a grant of intellectual property will commercialize an invention—if and only if

$$p_{\text{eligible}} \cdot p_{\text{enforce}} \cdot (\text{timing parameters})\pi - c \geq 0.$$

To create incentives for innovation, policymakers could intervene on the two parameters that directly generate the missing market, p_{eligible} and p_{enforce}, the profitability of commercialization π, or the cost of commercialization c. We discuss each briefly in turn.

p_{eligible}—Eligibility rules for intellectual property rights are not immutable; a series of Supreme Court decisions imposed novel restrictions on patentability in the early 2000s, and policymakers have proposed new legislation that would expand the scope of patentable subject matter.[64] But any change to the expansiveness of intellectual property rights would affect all inventions, across sectors, and requires careful consideration of Nordhaus (1969)-style tradeoffs between creation of incentives for innovation and the deadweight loss of monopoly pricing.

p_{enforce}—Changes to enforceability we view as likely context specific. In the new uses context, Roin (2014) sketches how electronic medical records, combined with indication-specific prescriptions and insurance reimbursement, might enable a firm that discovers and patents a new use B to enforce its intellectual property even if the drug is available generically for its initial use A. Our empirical results suggest that the innovation benefits of such enforceability might be very large, in the trillions of dollars. The potential cost we see, in addition to the standard Nordhaus (1969) deadweight loss of monopoly pricing, is if firms are able to obtain patents for uses B that are already known to doctors and utilized off-label. In principle, the novelty and non-obviousness provisions of the patent code should prevent already-known new uses from being patented, but these provisions are not perfect.

π—An alternative to intellectual property is for a funder, such as a government or a foundation, to incentivize research and development with a financial reward for successful innovation. This is sometimes called a "pull mechanism" (see, for example, Barder et al., 2005 and Snyder et al., 2020). For example, in the new uses context, a funder might announce a prize π for a Phase III clinical trial that shows that a generic drug reduces the five-year mortality associated with stage IV breast cancer by 10 percent. The key design challenge is setting the proper reward size. If the reward is too low,

[64] See, for example, S. 1546 – Patent Eligibility Restoration Act of 2025.

few, if any, firms will participate. And if the reward is too high, the funder will overpay relative to what was required to induce entry. Optimal reward design then requires balancing the benefits of greater entry associated with higher rewards against the costs of overpayment. We formalize this tradeoff with two stylized models in Appendix C. In a setting with complete information, the optimal prize simply covers the firm's risk-adjusted costs, $\pi^* = \frac{c}{p_{\text{comm}}}$. With incomplete information about the probability of success, under some stylized assumptions, the optimal prize is $\pi^* = \sqrt{cV}$, where V is the social value of the realized invention (e.g., the social value of reducing five-year mortality for stage IV breast cancer by 10 percent).[65]

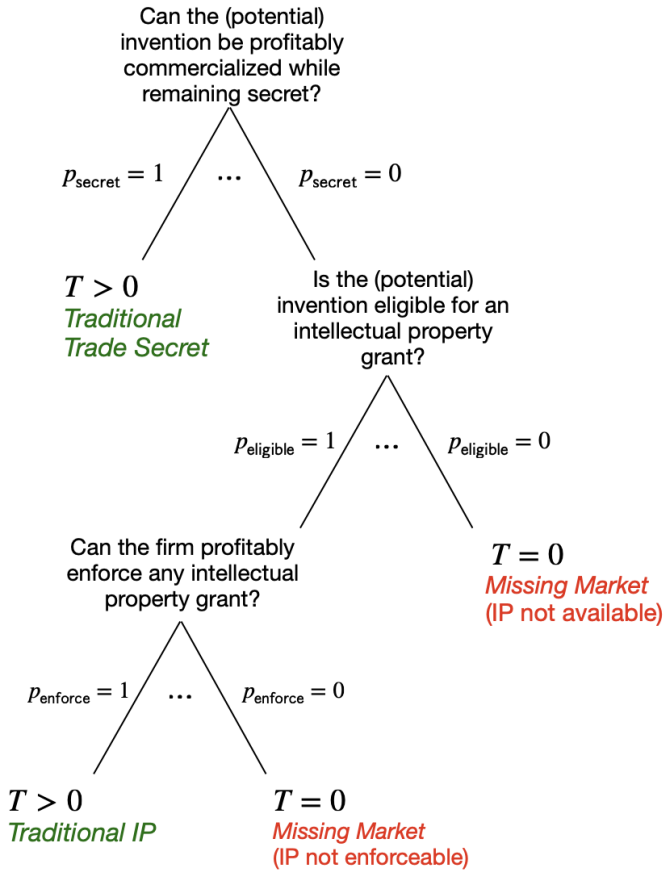
c—Finally, and most simply, a funder could directly pay the cost of the missing research and development. This is sometimes called "push" financing (see, for example, Ahuja et al., 2021 and Castillo et al., 2021). For example, a funder could directly pay for a clinical trial that tests a specific off-patent drug (e.g., metformin as discussed in Section 3) as a treatment for Stage IV breast cancer. The core challenges are informational: can the funder properly determine c, identify the appropriate recipient, and monitor effort? In settings where missing inventions are thought to be "technologically close" (Kremer et al., 2020) direct funding may be especially attractive.

Each of these potential policy interventions has important tradeoffs, and a more complete analysis of how to fix missing markets is left for future research. Still, one conclusion is clear: the social returns to improved innovation incentives may be extraordinarily large. In the single market we study, the potential gains are measured in trillions of dollars. Extending such mechanisms across the broader landscape of missing markets could enable progress on some of the thorniest challenges of our time.

[65] We are grateful to Chris Snyder for discussions that motivated this analysis. For details on ongoing efforts to design such a mechanism to incentivize investment in new uses for existing drugs, see https://marketshaping.uchicago.edu/winner/repurposing-generic-drugs/.

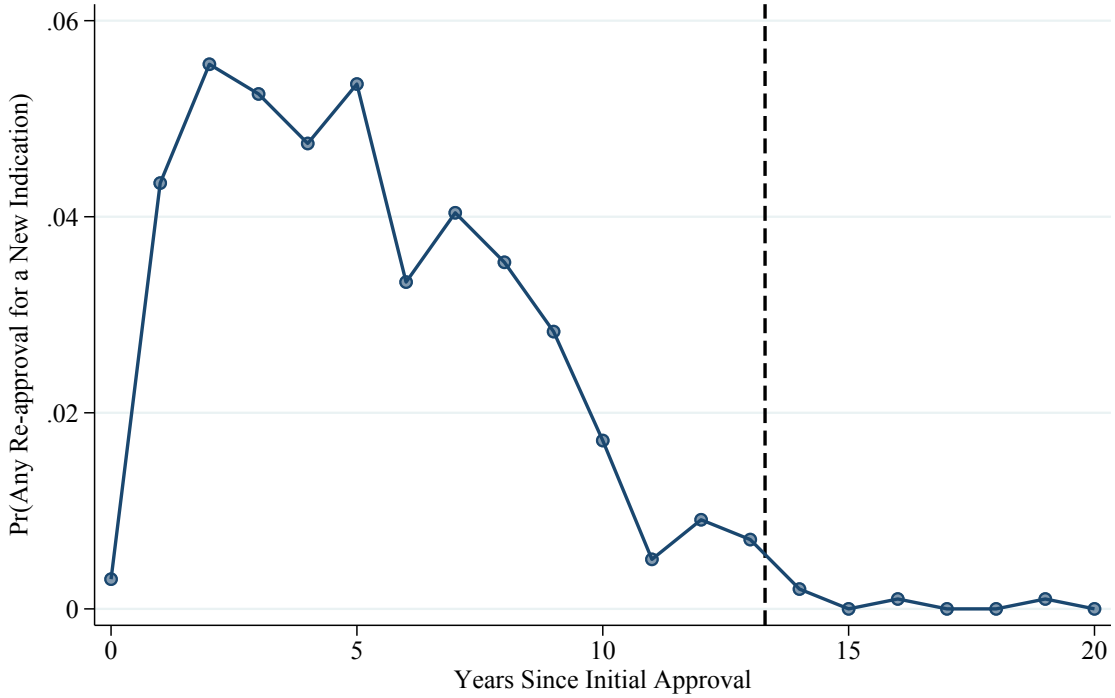
Figures and Tables

Figure 1: Missing Markets for Innovation



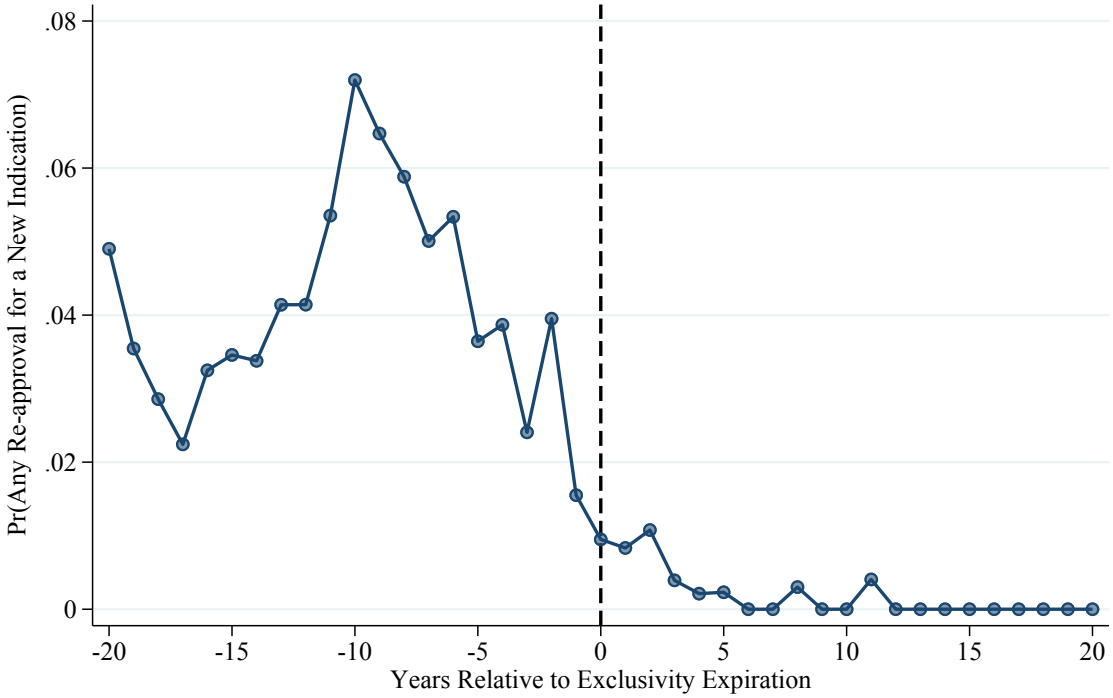
Notes: Figure summarizes the framework introduced in Section 2.1. The figure classifies potential innovations according to whether profitable commercialization is possible while the innovation remains secret, whether the innovation is eligible for formal intellectual property protection, and whether those rights are practically enforceable. $p_{\text{secret}}, p_{\text{eligible}}$, and p_{enforce} denote the corresponding probabilities. $T > 0$ denotes a positive period during which the innovator can earn returns from commercialization. $T = 0$ denotes a missing market in which the available innovation-policy instrument does not provide a period of effective protection. For simplicity, the figure depicts the case where if $p_{\text{secret}} = 1$ the firm chooses secrecy, even though in the model the firm could choose the intellectual property path even when $p_{\text{secret}} = 1$ if it is the case that $T_{\text{patent}} > T_{\text{secret}}$.

Figure 2: New Use Approvals Relative to Initial FDA Approval



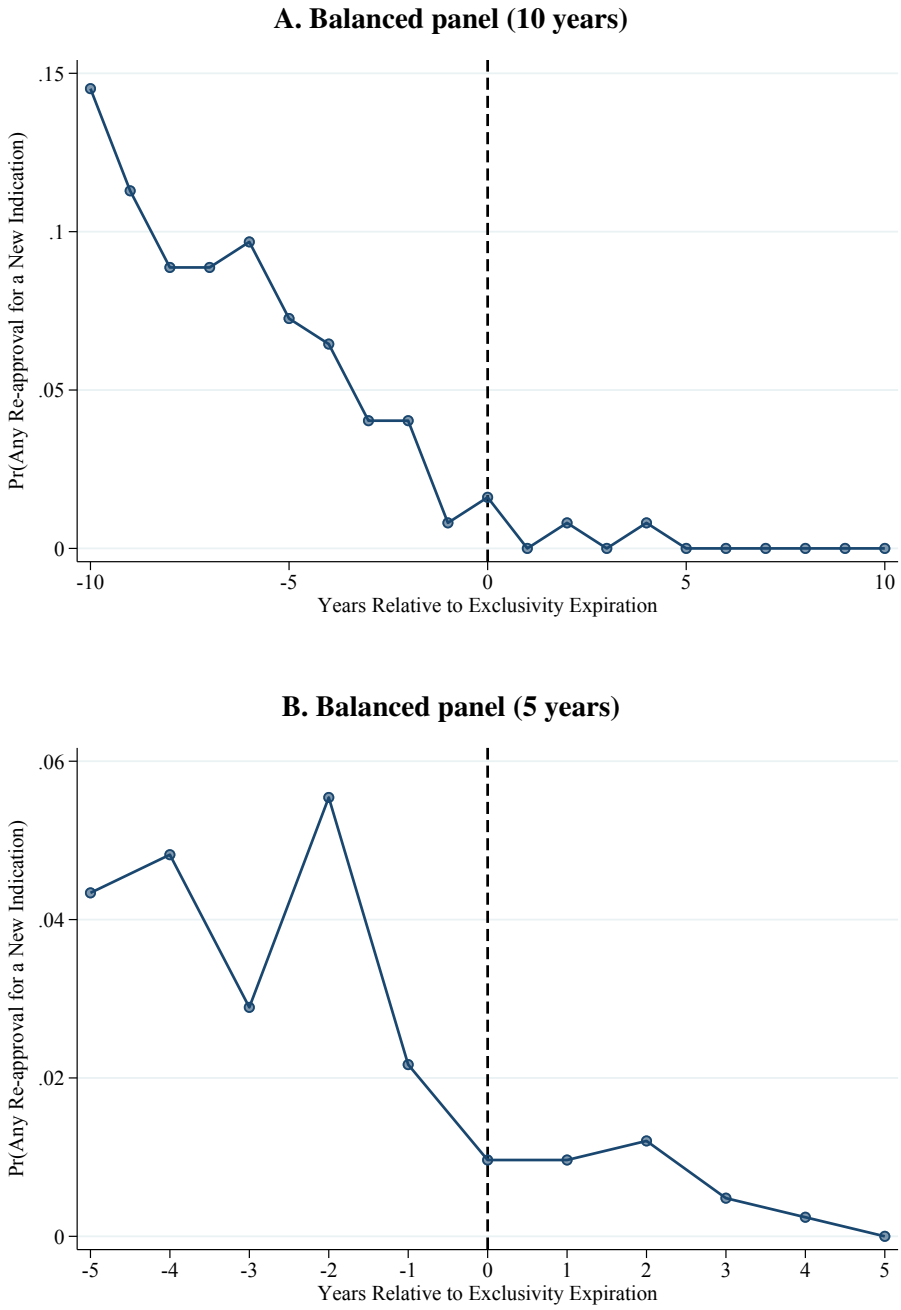
Notes: Figure plots the probability of re-approval for a new use for the sample of innovator drugs defined in Section 4.5 ($N = 990$) relative to the drug's initial approval by the FDA. Event year zero is the calendar year of initial approval. Drug-years in which a drug receives no approval for a new use are assigned a value of zero. A vertical line plots the median duration of exclusivity in our sample of innovator drugs ($t = 13.3$ years).

Figure 3: New Use Approvals Relative to Exclusivity Expiration



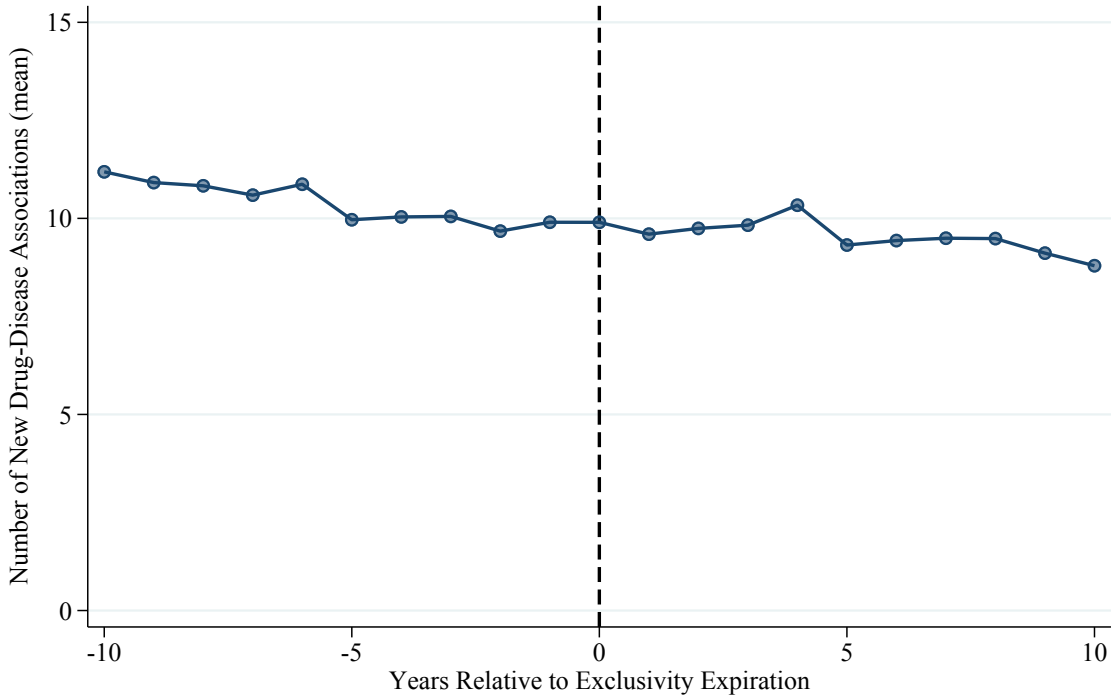
Notes: Figure plots the probability of re-approval for a new use for the sample of innovator drugs defined in Section 4.5 ($N = 990$), relative to exclusivity expiration. We plot the probability of re-approval for a new indication for the twenty years preceding and following exclusivity expiration for all drugs in our sample.

Figure 4: New Use Approvals—Balanced Panels



Notes: Figure plots the probability that an innovator drug receives approval for at least one new or modified indication in each calendar year relative to exclusivity expiration, holding the sample of drugs fixed across the displayed event window. Panel A is a ten-year balanced panel ($N = 124$), and Panel B is a five-year balanced panel ($N = 415$). Exclusivity expiration is defined as the earlier of two events: expiration of all patents and regulatory exclusivities linked to the innovator drug in the FDA's Orange Book, or FDA approval of a generic drug with an identical set of active moieties.

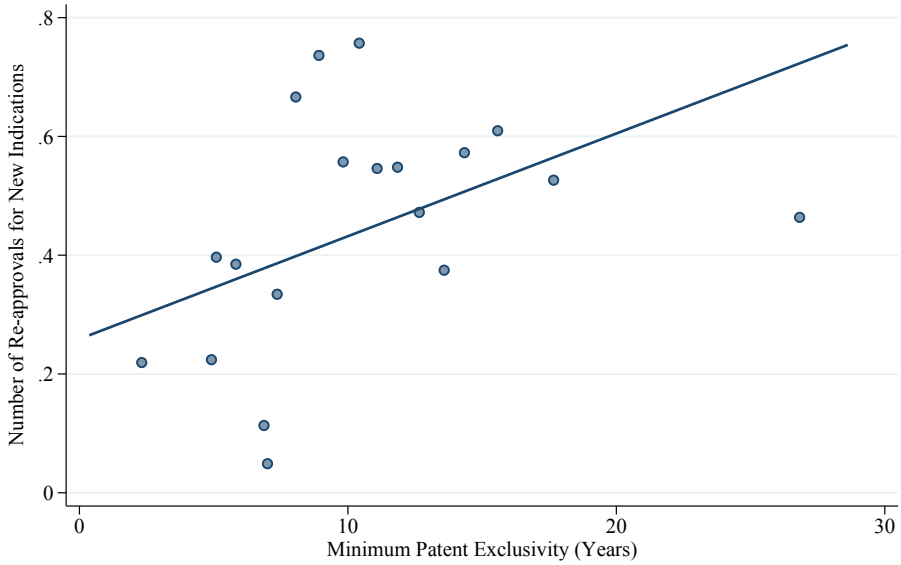
Figure 5: New Disease Associations Relative to Exclusivity Expiration



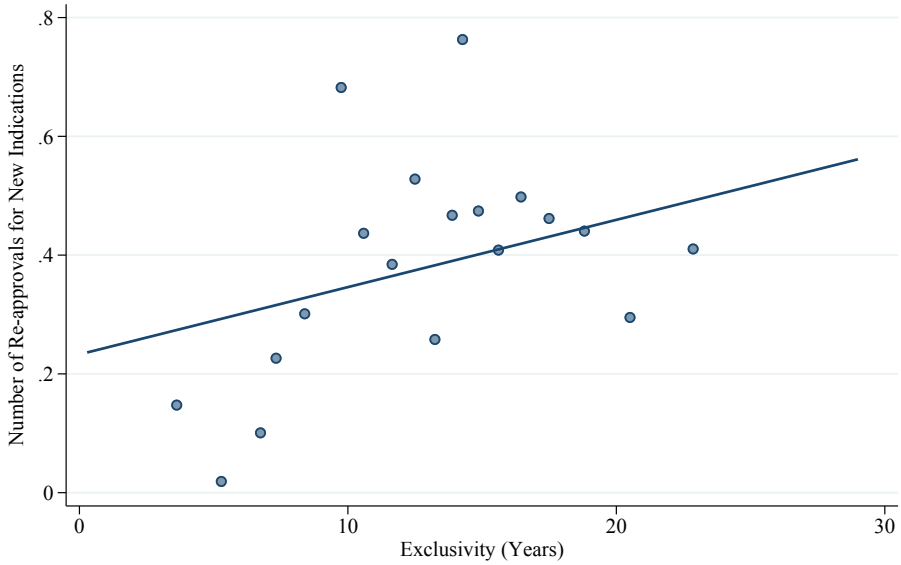
Notes: Figure plots the mean number of new associations between innovator drugs and diseases appearing in PubMed-indexed publications in each calendar year relative to exclusivity expiration. For each drug, a new disease association is the first calendar year in which a MeSH disease descriptor appears alongside the drug in an indexed publication. Each drug–disease pair is counted once. Drug-years without a new disease association are assigned a value of zero. Associations whose first appearance coincides with the introduction of the disease descriptor into the MeSH vocabulary are excluded. The vertical line marks exclusivity expiration; $t = 0$ is the calendar year in which exclusivity expires. Appendix Figure A6 plots estimates using a balanced panel; Appendix Figure A7 plots the cumulative number of new disease associations.

Figure 6: Quantity of New Uses by Exclusivity Duration

A. Minimum Patent Exclusivity

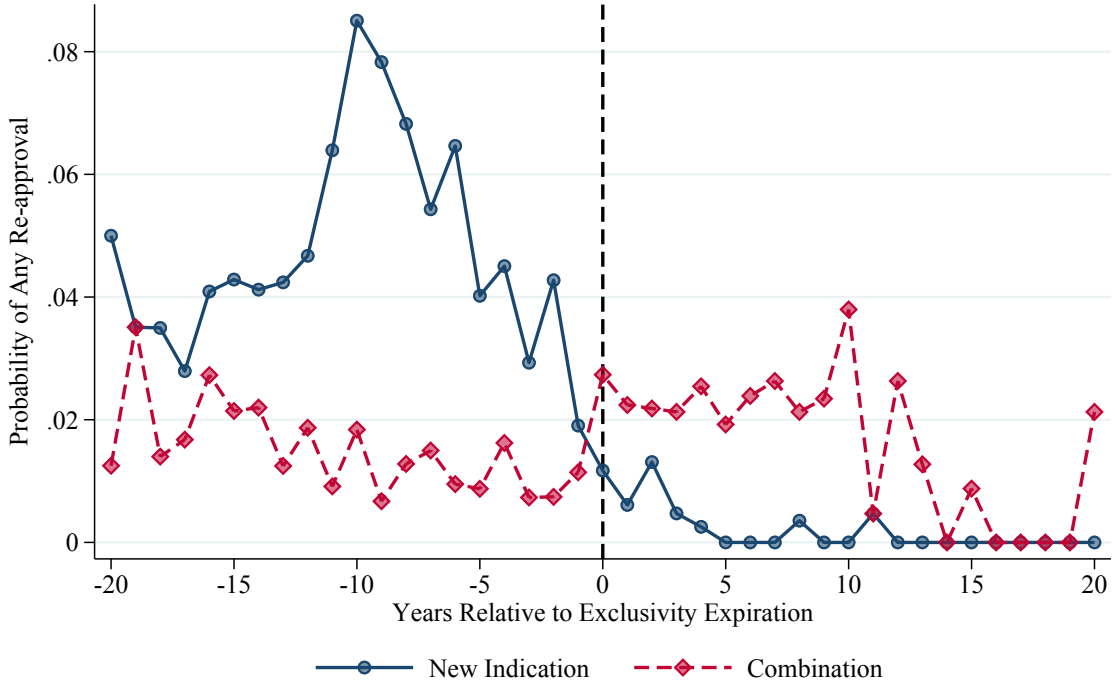


B. Exclusivity



Notes: Figure plots the relationship between exclusivity duration and the number of approvals for new uses over an innovator drug's lifecycle for the sample of innovator drugs defined in Section 4.5 ($N = 990$). Panel A uses minimum patent exclusivity; Panel B uses exclusivity. Minimum patent exclusivity is the period between initial approval and the earlier of generic approval or the date by which the first-filed patent and all generic-blocking regulatory exclusivities have expired. Exclusivity is the period between initial approval and the earlier of generic approval or expiration of all listed patents and generic-blocking regulatory exclusivities. In each panel, exclusivity duration and the number of new use approvals are residualized with respect to initial approval year fixed effects. We split our sample of drugs into twenty equally-sized bins based on the residualized exclusivity measure; for each bin, we plot the mean residualized outcome against the mean residualized exclusivity duration. We add sample means to both axes to facilitate interpretation.

Figure 7: Fixed-Dose Combination Approvals Relative to Exclusivity Expiration



Notes: Figure plots the probability that a fixed-dose-combination product containing the active moiety of an innovator drug receives FDA approval in each calendar year relative to exclusivity expiration for the underlying innovator drug. The sample is restricted to innovator drugs containing a single active moiety ($N = 792$). The outcome equals one if, in a given drug-year, a fixed-dose combination containing that active moiety receives either initial FDA approval or approval for a new or modified indication, and zero otherwise. Approvals limited to a new strength, dosage form, formulation, or patient population of an existing combination are excluded. A combination approval is associated with each eligible component active moiety contained in the product. The vertical line marks exclusivity expiration; $t = 0$ is the calendar year in which exclusivity expires. Exclusivity expiration is defined as the earlier of generic approval or expiration of all patents and generic-blocking regulatory exclusivities linked to the innovator drug in the FDA's Orange Book.

Table 1: Summary Statistics

	Mean	Median	SD	Min.	Max.	# obs.
Years to Generic Entry	11.27	11.47	4.52	0.21	28.19	442
exclusivity (years)	12.90	13.37	5.48	0.18	42.46	990
Minimum Patent Exclusivity (years)	9.35	8.37	4.55	0.18	35.32	990
Number of Re-Approvals (total)	14.87	11	13.43	0	87	990
Number of New Use Approvals	0.52	0	1.18	0	10	990

Notes: Table shows summary statistics for the sample of innovator drugs defined in Section 4.5 ($N = 990$). Observations are innovator drugs, defined as the first FDA-approved application containing a unique combination of active moieties.

Table 2: New Drug–Disease Associations Relative to Exclusivity Expiration

	Before expiration $t = -5$ to -1	After expiration $t = 1$ to 5	Difference Post − Pre
Panel A. *Scientific information and commercialization*			
New drug-disease associations per drug-year	9.92	9.77	–0.15
			(0.37)
Probability of any new use approval (percent)	3.7	0.6	–3.1
			(0.50)
Panel B. *Relationship between scientific information and commercialization*			
Change in approval probability per 10 recent	0.39	0.03	–0.36
associations (percentage point)	(0.11)	(0.04)	(0.12)
Drug-years	2,590	1,814	4,404

Notes: Table compares the arrival of new scientific information and the commercialization of new uses in the five calendar years before and after exclusivity expiration. The calendar year of expiration ($t = 0$) is excluded from all comparisons. A new drug–disease association is the first appearance of a previously unobserved pairing of an innovator drug and a disease descriptor in PubMed-indexed publications. For example, if a PubMed-indexed article is the first article in our data in which acetaminophen and breast cancer appear together, we count one new acetaminophen–breast-cancer association in that year. This measure is intentionally broad, and co-occurrence does not imply that the article identifies the drug as a potential treatment for the disease. Panel A reports the mean number of new drug–disease associations per drug-year and the probability that an innovator drug receives FDA approval for at least one new or modified indication. The final column reports the post-expiration mean minus the pre-expiration mean. Panel B reports estimates from a pooled linear probability model in which the dependent variable is an indicator for any approval of a new use in a drug-year. The explanatory variable is the number of new drug–disease associations observed over the preceding five years, divided by ten, interacted with an indicator for the post-expiration period. The specification includes relative-year fixed effects. Columns report the implied pre-expiration slope, post-expiration slope, and their difference. Coefficients are expressed in percentage points per ten associations. Standard errors, clustered at the innovator-drug level, are reported in parentheses. The difference in slopes in Panel B is significantly different from zero ($p = 0.003$).

Table 3: Exclusivity Duration and New Use Approvals

Panel A. *Re-approvals for new/modified indication*

	(1)	(2)	(3)	(4)
Minimum Patent Exclusivity	0.037	0.023		
(years)	(0.008)	(0.008)		
exclusivity	—	—	0.013	0.020
(years)			(0.006)	(0.006)
Approval Year FEs	no	yes	no	yes
Observations	990	990	990	990

Panel B: *Two-part model*

Extensive Margin	*Any re-approvals for new/modified indication*			
Minimum Patent Exclusivity	0.083	0.059	—	—
(years)	(0.015)	(0.018)		
Exclusivity	—	—	0.026	0.042
(years)			(0.012)	(0.014)
Approval Year FEs	no	yes	no	yes
Observations	990	990	990	990
Intensive Margin	*No. of re-approvals for new/modified indication*			
Minimum Patent Exclusivity	0.017	0.011	—	—
(years)	(0.021)	(0.022)		
Exclusivity	—	—	0.015	0.020
(years)			(0.017)	(0.018)
Approval Year FEs	no	yes	no	yes
Observations	256	256	256	256

Notes: Table reports estimates of the relationship between exclusivity duration and approvals of new uses for the sample of 990 innovator drugs defined in Section 4.5. Observations are innovator drugs. Columns (1) and (2) use minimum patent exclusivity; Columns (3) and (4) use exclusivity. Minimum patent exclusivity and exclusivity are defined in the note to Figure 6. Columns (2) and (4) include fixed effects for the calendar year of initial FDA approval. Panel A reports ordinary-least-squares estimates in which the dependent variable is the total number of new use approvals over the drug's lifecycle. The first section of Panel B reports logit estimates in which the dependent variable is an indicator for whether the drug is ever approved for a new use. Reported coefficients are log-odds coefficients. The second section of Panel B restricts the sample to drugs approved for at least one new use and reports ordinary-least-squares estimates of the number of approvals. We report heteroskedasticity-robust standard errors in parentheses.

References

Abramowicz, Michael, "The danger of underdeveloped patent prospects," *Cornell Law Review*, 2007, *92* (6), 1065–1121.

Abrams, David, "Did TRIPS spur innovation? An analysis of patent duration and incentives to innovate," *University of Pennsylvania Law Review*, 2009, *157* (6), 1613–1647.

Ahuja, Amrita, Susan Athey, Arthur Baker, Eric Budish, Juan Camilo Castillo, Rachel Glennerster, Scott Duke Kominers, Michael Kremer, Jean Lee, Canice Prendergast et al., "Preparing for a pandemic: Accelerating vaccine availability," in "AEA Papers and Proceedings," Vol. 111 American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203 2021, pp. 331–335.

Arrow, Kenneth, "Economic Welfare and the Allocation of Resources to Invention," in Universities-National Bureau Committee for Economic Research and the Committee on Economic Growth of the Social Science Research Councils, eds., *The Rate and Direction of Inventive Activity: Economic and Social Factors*, Princeton, NJ: Princeton University Press, 1962, pp. 609–626.

Barder, Owen, Michael Kremer, and Ruth Levine, "Answering concerns about Making Markets for Vaccines," *Geneva: Commission on Intellectual Property Rights, Innovation and Public Health (CIPIH). Available: http:// www. who. int/intellectualproperty/submissions/BarderSubmission. pdf. Accessed*, 2005, *15*.

Berger, Benjamin, Amitabh Chandra, and Craig Garthwaite, "Regulatory approval and expanded market size," Technical Report, National Bureau of Economic Research 2021.

Boldrin, Michele and David K Levine, *Against intellectual monopoly*, Vol. 62, Cambridge University Press Cambridge, 2008.

Bronnenberg, Bart J, Jean-Pierre Dubé, Matthew Gentzkow, and Jesse M Shapiro, "Do pharmacists buy Bayer? Informed shoppers and the brand premium," *The Quarterly Journal of Economics*, 2015, *130* (4), 1669–1726.

Bryan, Kevin A and Heidi L Williams, "Innovation: market failures and public policies," in "Handbook of industrial organization," Vol. 5, Elsevier, 2021, pp. 281–388.

Budish, Eric, Benjamin N Roin, and Heidi L Williams, "Do firms underinvest in long-term research? Evidence from cancer clinical trials," *American Economic Review*, 2015, *105* (7), 2044–2085.

_ , _ , and _ , "Patents and research investments: Assessing the empirical evidence," *American Economic Association Papers and Proceedings*, 2016, *106* (5), 183–187.

_ , **Peter Cramton, and John Shim**, "The high-frequency trading arms race: Frequent batch auctions as a market design response," *The Quarterly Journal of Economics*, 2015, *130* (4), 1547–1621.

Butcher, Nancy, "Old Drugs and New Tricks: Repurposing Drugs to Treat Psychiatric Disorders," *IMS Magazine*, 2013, *1*, 21–21.

Carr, Austin and Ann Choi, "The Plot to Steal the Secret Coke Can Liner Formula," May 2023. Accessed: March 10, 2025.

Carrier, Michael A and Steve D Shadowen, "Product hopping: A new framework," *Notre Dame L. Rev.*, 2016, *92*, 167.

Castillo, Juan Camilo, Amrita Ahuja, Susan Athey, Arthur Baker, Eric Budish, Tasneem Chipty, Rachel Glennerster, Scott Duke Kominers, Michael Kremer, Greg Larson et al., "Market design to accelerate COVID-19 vaccine supply," *Science*, 2021, *371* (6534), 1107–1109.

Citeline, "Pharmaprojects," Database 1995-2010. Proprietary data; version current through May 2011.

Clarivate, "Cortellis Clinical Trials Intelligence," Database 1995-2016. Proprietary data; accessed 2016.

Cohen, Wesley, Richard Nelson, and John Walsh, "Protecting their intellectual assets: Appropriability conditions and why US manufacturing firms patent (or not)," 2000. Published: National Bureau of Economic Research (NBER) working paper 7552.

Collins, Francis, "Testimony before the US House Subcommittee on Labor - HHS - Education Appropriations," 2012. Published: Department of Health and Human Services, National Institutes of Health.

Collins, Francis S, "Mining for therapeutic gold," *Nature Reviews Drug Discovery*, 2011, *10* (6), 397.

_ , "Reengineering translational science: the time is right," *Science translational medicine*, 2011, *3* (90), 90cm17–90cm17. Publisher: American Association for the Advancement of Science.

Collins, Francis S. and Harold Varmus, "A New Initiative on Precision Medicine," *New England Journal of Medicine*, 2015, *372* (9), 793–795.

Conti, Rena M, Susan Athey, Richard G Frank, and Jonathan Gruber, "Generic drug repurposing for public health and national security: COVID-19 and beyond," *Health Affairs Forefront*, 2020.

Corbett, Anne, James Pickett, Alistair Burns, Jonathan Corcoran, Stephen B Dunnett, Paul Edison, Jim J Hagan, Clive Holmes, Emma Jones, Cornelius Katona et al., "Drug repositioning for Alzheimer's disease," *Nature reviews Drug discovery*, 2012, *11* (11), 833–846.

Cutler, David, *Your Money or Your Life: Strong Medicine for America's Health Care System*, Oxford University Press, 2004.

Dix, Rebekah and Todd Lensman, "Combining Complements: Theory and Evidence from Cancer Treatment Innovation," 2024.

Dranove, David, Craig Garthwaite, and Manuel Hermosilla, "Pharmaceutical profits and the social value of innovation," Technical Report, National Bureau of Economic Research 2014.

Durvasula, Maya and Lisa Larrimore Ouellette, "Beyond the Pharmaceutical Patent Arms Race," *Yale Journal on Regulation*, 2026.

_ , C. Scott Hemphill, Lisa Larrimore Ouellette, Bhaven Sampat, and Heidi L Williams, "The NBER Orange Book dataset: A user's guide," *Research Policy*, 2023, *52* (7), 104791.

Eisenberg, Rebecca, "The problem of new uses," *Yale Journal of Health Policy, Law, & Ethics*, 2005, *717* (5), 717–739.

EveryCure, "EveryCure: Unlocking the Full Potential of Existing Drugs," 2025.

F-D-C Reports, "NDA Pipeline," Database 1982-2001. Proprietary data; accessed April 2007.

FDA, "List of Patent and Exclusivity for Drugs without Approved Generic," 2025. Accessed: 2025-03-14.

Frakes, Michael D. and Melissa F. Wasserman, "Strategic Patenting: Evidence from the Biopharmaceutical Industry," 2025.

Gale, Jason and Andrea Gerlin, "Five-cent diabetes pill from 1958 may be new cancer drug," *Bloomberg Business*, 2012, *26 September.*

Garraway, Levi A. and Eric S. Lander, "Lessons from the Cancer Genome," *Cell*, 2013, *153* (1), 17–37.

Gault, Matthew, "Stanford Scientists Reverse Engineer Moderna Vaccine, Post Code on Github," *VICE*, mar 2021. 9:35 a.m. (PDT).

Gawande, Atul, "The checklist manifesto: How to get things right," 2009.

Gelijns, Annetine, Nathan Rosenberg, and Alan Moskowitz, "Capturing the unexpected benefits of medical research," *New England Journal of Medicine*, 1998, *339* (10), 693–697.

Gilbert, Richard and Carl Shapiro, "Optimal patent length and breadth," *RAND Journal of Economics*, 1990, *21* (1), 106–112.

Hall, Robert E and Charles I Jones, "The value of life and the rise in health spending," *The Quarterly Journal of Economics*, 2007, *122* (1), 39–72.

Hemel, Daniel J and Lisa Larrimore Ouellette, "The generic drug trilemma," *Entrepreneurship and Innovation Policy and the Economy*, 2023, *2* (1), 41–77.

_ and _, "Valuing medical innovation," *Stan. L. Rev.*, 2023, *75*, 517.

Hemphill, C. Scott and Bhaven Sampat, "Evergreening, patent challenges, and effective market life in pharmaceuticals," *Journal of Health Economics*, 2011, *31* (2), 327–339.

_ and _, "Evergreening, Patent Challenges, and Effective Market Life in Pharmaceuticals," *Journal of Health Economics*, 2012, *31*, 327–339.

_ and Jeannie Suk, "The law, culture, and economics of fashion," *Stan. L. Rev.*, 2008, *61*, 1147.

Hoadley, Katherine A. et al., "Multiplatform Analysis of 12 Cancer Types Reveals Molecular Classification within and across Tissues of Origin," *Cell*, 2014, *158* (4), 929–944.

_ and _, "Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer," *Cell*, 2018, *173* (2), 291–304.

Hodgson, Charles, "Information externalities, free riding, and optimal exploration in the uk oil industry," Technical Report, National Bureau of Economic Research 2024.

Hoyt, A Sasha, "The impact of uncertainty regarding patent eligible subject matter for investment in US medical diagnostic technologies," *Wash. & Lee L. Rev.*, 2022, *79*, 397.

IQVIA Institute for Human Data Science, "The Use of Medicines in the U.S.," 2021. Accessed: 2025-03-19.

Jones, Charles I, "Growth and ideas," in "Handbook of economic growth," Vol. 1, Elsevier, 2005, pp. 1063–1111.

_, "Recipes and economic growth: A combinatorial march down an exponential tail," *Journal of Political Economy*, 2023, *131* (8), 1994–2031.

Kapczynski, Amy and Talha Syed, "The continuum of excludability and the limits of patents," *Yale LJ*, 2012, *122*, 1900.

_, Chan Park, and Bhaven Sampat, "Polymorphs and prodrugs and salts (oh my!): An empirical analysis of 'secondary' pharmaceutical patents," *PLoS One*, 2012, *7* (12), e49470.

Klemperer, Paul, "How broad should the scope of patent protection be?," *RAND Journal of Economics*, 1990, *21* (1), 113–130.

Kremer, Michael and Christopher M Snyder, "Preventives versus treatments," *The Quarterly Journal of Economics*, 2015, *130* (3), 1167–1239.

_ , Jonathan Levin, and Christopher M Snyder, "Advance market commitments: insights from theory and experience," in "AEA Papers and Proceedings," Vol. 110 American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203 2020, pp. 269–273.

Krieger, Joshua, Danielle Li, and Dimitris Papanikolaou, "Missing novelty in drug development," *The Review of Financial Studies*, 2022, *35* (2), 636–679.

Léauté-Labrèze, Christine, Peter Hoeger, Josiane Mazereeuw-Hautier, Laurent Guibaud, Eva Baselga, Giedre Posiunas, Robert J. Phillips, Horacio Caceres, Juan Carlos Lopez Gutierrez, Rafael Ballona et al., "A Randomized, Controlled Trial of Oral Propranolol in Infantile Hemangioma," *New England Journal of Medicine*, 2015, *372* (8), 735–746.

Lemley, Mark A and Carl Shapiro, "Probabilistic patents," *Journal of Economic Perspectives*, 2005, *19* (2), 75–98.

Lerner, Josh, "Patent protection and innovation over 150 years," 2002. Published: National Bureau of Economic Research (NBER) working paper 8977.

Levin, Richard, Alvin Klevorick, Richard Nelson, Sidney Winter, Richard Gilbert, and Zvi Griliches, "Appropriating the returns from industrial research and development," *Brookings Papers on Economic Activity*, 1987, *1987* (3), 783–831.

Lichtenberg, Frank R., "Sources of U.S. Longevity Increase, 1960–1997," NBER Working Paper 8755, National Bureau of Economic Research, Cambridge, MA February 2002.

_ and Tomas J. Philipson, "The Dual Effects of Intellectual Property Regulations: Within- and Between-Patent Competition in the U.S. Pharmaceuticals Industry," *Journal of Law and Economics*, 2002, *45* (S2), 643–672.

Lord, Simon R and Adrian L Harris, "Is it still worth pursuing the repurposing of metformin as a cancer therapeutic?," *British journal of cancer*, 2023, *128* (6), 958–966.

Machlup, Fritz, "An economic review of the patent system," 1958. Published: Study of the Subcommittee on Patents, Trademarks, and Copyrights of the Committee on the Judiciary, United States Senate, 85th Congress, Second Session.

Mansfield, Edwin, "Patents and innovation: An empirical study," *Management Science*, 1986, *32* (2), 173–181.

_ , **Mark Schwartz, and Samuel Wagner**, "Imitation costs and patents: An empirical study," *The Economic Journal*, 1981, *91* (364), 907–918.

Manson, JoAnn E., Judith Hsia, Karen C. Johnson, Jacques E. Rossouw, A. Richey Assaf, Norman L. Lasser, Maurizio Trevisan, Howard R. Black, Susan R. Heckbert, Robert Detrano et al., "Estrogen plus Progestin and the Risk of Coronary Heart Disease," *New England Journal of Medicine*, 2003, *349* (6), 523–534.

Michaeli, Daniel Tobias, Mackenzie Mills, and Panos Kanavos, "Value and price of multi-indication cancer drugs in the USA, Germany, France, England, Canada, Australia, and Scotland," *Applied health economics and health policy*, 2022, *20* (5), 757–768.

Michelman, Valerie and Lucy Msall, "Sex, Drugs, and R&D: Missing Innovation from Regulating Female Enrollment in Clinical Trials," Technical Report, University of Chicago Working paper 2024.

Morton, Fiona M Scott, "Entry decisions in the generic pharmaceutical industry," *The Rand journal of economics*, 1999, pp. 421–440.

Moser, Petra, "How do patent laws influence innovation? Evidence from nineteenth-century world fairs," *American Economic Review*, 2005, *95* (4), 1214–1236.

_ , "Why don't inventors patent?," 2007.

Mossinghoff, Gerald, "Overview of the Hatch-Waxman Act and its impact on the drug development process," *Food & Drug Law Journal*, 1999, *54* (2), 187–194.

Mozaffarian, Dariush, Tao Hao, Eric B Rimm, Walter C Willett, and Frank B Hu, "Changes in diet and lifestyle and long-term weight gain in women and men," *New England journal of medicine*, 2011, *364* (25), 2392–2404.

Murphy, Kevin and Robert Topel, "The value of health and longevity," *Journal of Political Economy*, 2006, *114* (5), 871–904.

Muthyala, Ramaiah, "Orphan/rare drug discovery through drug repositioning," *Drug Discovery Today: Therapeutic Strategies*, 2011, *8* (3-4), 71–76.

National Library of Medicine, "PubMed," Database 1966-2022. Accessed February 2026.

_ , "ClinicalTrials.gov," Database 2010-2016. Accessed July 28, 2022.

Nelson, Richard R, "The simple economics of basic scientific research," *Journal of political economy*, 1959, *67* (3), 297–306.

Nordhaus, William, *Invention, Growth, and Welfare: A Theoretical Treatment of Technological Change*, Cambridge, MA: MIT Press, 1969.

_ , "The optimum life of a patent: Reply," *American Economic Review*, 1972, *62* (3), 428–431.

Oster, Emily, "Why Is Nutrition So Stressful?," ParentData Podcast Dec. 2024. Updated Jan. 3, 2025.

O'Connor, Lauren, Maeve Bailey-Whyte, Manami Bhattacharya, Gisela Butera, Kaitlyn N Lewis Hardell, Andrew B Seidenberg, Philip E Castle, and Holli A Loomans-Kropp, "Association of metformin use and cancer incidence: a systematic review and meta-analysis," *JNCI: Journal of the National Cancer Institute*, 2024, *116* (4), 518–529.

Pahud, Dominique, Lesa Mitchell, John Wilbanks, Melissa Stevens, Scott Weir, John McBride, Lili Portilla, Ed Pezalla, Michael A Nameth, Bernadette O'Donoghue, and others, "A New Market Access Path for Repurposed Drugs," *Available at SSRN 2437502*, 2014.

Pakes, Ariel, "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks," *Econometrica*, 1986, *54* (4), 755–784.

Patrick, Graham, *An Introduction to Medicinal Chemistry*, 2 ed., Oxford University Press Oxford, 2005.

Philipson, Tomas and Anupam Jena, "Who benefits from new medical technologies? Estimates of consumer and producer surpluses for HIV/AIDS drugs," *Forum for Health Economics & Policy*, 2006, *9* (2).

Pollack, Andrew, "Debate Over AIDS Drug's Effect on H.I.V. Infection Rate," July 2006. Accessed: March 10, 2025.

Price, W Nicholson, Arti K Rai, and Timo Minssen, "Knowledge transfer for large-scale vaccine manufacturing," *Science*, 2020, *369* (6506), 912–914.

Price, W Nicholson II, "Making do in making drugs: innovation policy and pharmaceutical manufacturing," *BCL Rev.*, 2014, *55*, 491.

Rai, Arti, "Use patents, carve-outs, and incentives - A new battle in the drug-patent wars," *New England Journal of Medicine*, 2012, *367* (6), 491–493.

Robbins, Christopher, "Fake Luxury Goods: Inside the Secret World of Counterfeit Handbags," August 2018. Accessed: March 10, 2025.

Roin, Benjamin N, "The case for tailoring patent awards based on time-to-market," *UCLA L. Rev.*, 2013, *61*, 672.

_ , "Solving the problem of new uses by creating incentives for private industry to repurpose off-patent drugs," 2014. Published: mimeo, MIT Sloan School of Management.

Romer, Paul M, "The origins of endogenous growth," *Journal of Economic perspectives*, 1994, *8* (1), 3–22.

Sachs, Rachel E, Paul B Ginsburg, and Dana P Goldman, "Encouraging new uses for old drugs," *Jama*, 2017, *318* (24), 2421–2422.

Sakakibara, Mariko and Lee Branstetter, "Do stronger patents induce more innovation? Evidence from the 1998 Japanese patent law reforms," *RAND Journal of Economics*, 2001, *32* (1), 77–100.

Santo, Brian, "The Consumer Electronics Hall of Fame: Zojirushi Micom Electric Rice Cooker/Warmer," September 2023. Accessed: March 10, 2025.

Scherer, F. M., "Nordhaus' theory of optimal patent life: A geometric reinterpretation," *American Economic Review*, 1972, *62* (3), 422–427.

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

Shafrin, Jason, Sabiha Quddus, Moises Marin, and Dennis Scanlon, "A decade of health innovation: the impact of new medicines on patient health and the implications for NICE's size of benefit multiplier," *Value in Health*, 2023, *26* (10), 1435–1439.

Sherkow, Jacob S and Paul R Gugliuzza, "Infringement by Drug Label," *Stanford Law Review*, 2026, *78.*

Silverman, Ed, "Single Tablet Regimen: How Gilead Scientists Developed a Pill That Changed the Course of HIV Prevention," November 2018. Accessed: March 10, 2025.

Singh, Jasjit and Matt Marx, "The Geographic Scope of Knowledge Spillovers: Spatial Proximity, Political Borders and Non-Compete Enforcement," 2011.

Snyder, Christopher M, Kendall Hoyt, Dimitrios Gouglas, Thomas Johnston, and James Robinson, "Designing Pull Funding For A COVID-19 Vaccine: Article proposes a "pull" program that incentivizes late-stage development and manufacturing of COVID-19 vaccines by awarding advance purchase commitments to bidding firms.," *Health Affairs*, 2020, *39* (9), 1633–1642.

Teece, David J, "Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy," *Research policy*, 1986, *15* (6), 285–305.

Telleria, Carlos M, "Drug repurposing for cancer therapy," *Journal of cancer science & therapy*, 2012, *4* (7), ix.

Thomas, Gareth, *Medicinal Chemistry: An Introduction*, Wiley-Interscience, 2007.

Trusheim, Mark, Murray Aitken, and Ernst Berndt, "Characterizing markets for biopharmaceutical innovations: Do biologics differ from small molecules?," *Forum for Health Economics and Policy*, 2010, *13* (1), 4.

Tse, Tony, Kevin M Fain, and Deborah A Zarin, "How to avoid common problems when using ClinicalTrials. gov in research: 10 issues to consider," *Bmj*, 2018, *361*.

U.S. Food and Drug Administration, "Approved Drug Products with Therapeutic Equivalence Evaluations (Orange Book)," Database 1984-2016. Accessed February 2026.

_ , "Drugs@FDA: FDA-Approved Drugs," Database 1985-2014. Accessed February 2026.

_ , "FDALabel: Full-Text Drug Product Labeling," Database 2005-2016. Accessed February 2026.

US Institute of Medicine, *Improving the quality of cancer clinical trials: Workshop summary*, The National Academies Press, 2008.

Vokinger, Kerstin N, Camille EG Glaus, Aaron S Kesselheim, Miquel Serra-Burriel, Joseph S Ross, and Thomas J Hwang, "Therapeutic value of first versus supplemental indications of drugs in US and Europe (2011-20): retrospective cohort study," *bmj*, 2023, *382*.

Ward, Erin H, "Defining Active Ingredient: The US Food and Drug Administration's Legal Interpretation of Regulatory Exclusivities.," *Congressional Research Service (CRS) Reports and Issue Briefs*, 2019.

Wermuth, Camille G., "Selective Optimization of Side Activities: the SOSA Approach," *Drug Discovery Today*, 2006, *11*, 160–161.

Williams, Heidi L, "How do patents affect research investments?," *Annual Review of Economics*, 2017, *9*, 441–469.

Wood, R. A. et al., "Omalizumab for the Treatment of Multiple Food Allergies," *New England Journal of Medicine*, 2024. Phase III OUtMATCH trial showing superiority over placebo in reducing allergic reactions.

World Health Organization, "International Classification of Diseases, Tenth Revision (ICD-10)," Classification system 2019. Accessed February 2026.

Yin, Wesley, John R Penrod, Ross Maclean, Darius N Lakdawalla, and Tomas Philipson, "Value of survival gains in chronic myeloid leukemia," *Am J Manag Care*, 2012, *18* (Suppl 11), S257–S264.

Yurukoglu, Ali, Eli Liebman, and David B Ridley, "The role of government reimbursement in drug shortages," *American Economic Journal: Economic Policy*, 2017, *9* (2), 348–382.

Zarin, Deborah A, Tony Tse, Rebecca J Williams, and Thiyagu Rajakannan, "Update on trial registration 11 years after the ICMJE policy was established," *New England Journal of Medicine*, 2017, *376* (4), 383–391.

A Data Construction

This Appendix provides additional detail on the construction of the data used in the paper. Section A.1 describes the FDA approval records used to define innovator drugs, generic competitors, and approvals for new uses. Section A.2 describes the PubMed data used to measure scientific publications and new disease associations. Section A.3 describes the four clinical trial datasets. Section A.4 describes the drug–indication crosswalk. Section A.5 presents the distributions of the two exclusivity measures used in our empirical analysis.

A.1 Drug Approvals

Our primary source of data on drug approvals and generic entry is **Drugs@FDA**, the U.S. Food and Drug Administration's administrative database of approved drug products. With limited exceptions, Drugs@FDA contains products approved for sale in the United States through a New Drug Application, Abbreviated New Drug Application, or Biologic License Application. It does not include, among other categories, drugs that have not been approved by the FDA, products still under review, dietary supplements, most over-the-counter products approved outside the NDA process, animal drugs, or biologic products regulated outside the Center for Drug Evaluation and Research.

We use information from three Drugs@FDA files:

1. *Application.txt*, containing application-level information;
2. *Product.txt*, containing product-level information; and
3. *RegActionDate.txt*, containing initial and supplemental approval actions.

The structure of the publicly available Drugs@FDA files changed on November 6, 2016. After that date, the FDA stopped publishing several fields used to identify the types of supplemental approval actions associated with each application. As a result, we use the latest available version of the earlier data structure, dated October 25, 2016 and obtained through the Internet Archive. For the construction of our active moiety crosswalk described in Section A.1.2, we instead use the January 2023 version of Drugs@FDA.

A.1.1 Background on the FDA approval process and data files

With few exceptions, drugs that are prescribed in the United States must first be approved by the FDA. FDA approval indicates that, for its intended population, a drug's benefits outweigh both known and potential risks.[66] A brief primer on the vocabulary of drug applications and key features of the approval process provides helpful context for the data in this paper.

Application numbers: The FDA assigns six-digit application numbers to three types of applications: new drug applications (NDAs); abbreviated new drug applications (ANDAs), and biologic

[66] See https://www.fda.gov/drugs/development-approval-process-drugs for details.

license applications (BLAs). Each application may contain multiple drug products that differ in dosage form or strength.

New drug applications (NDAs): An NDA is submitted when there exists sufficient evidence on the safety and effectiveness of a drug to meet the FDA's requirements for new drug marketing. NDAs are typically submitted for all novel, brand-name drugs.

Abbreviated new drug applications (ANDAs): An ANDA is submitted for the approval of a generic drug that is the "therapeutic equivalent" of any previously approved drug. Drug products are considered to be therapeutic equivalents only if they are pharmaceutical equivalents and if they can be expected to have the same clinical effect and safety profile when administered to patients under the conditions specified in the labeling.

Biologic license applications (BLAs): A BLA is an application submitted for the approval of a biological product.

Application-level data The Application data file contains information on 20,560 drug approvals; 5,082 are NDAs, 15,341 are ANDAs, and 137 are BLAs.

Product-level data The Product data file contains information for each application that includes—among other variables—a record of drug ingredients that are associated with each application. Specifically, the file includes the following variables:

Active ingredients: An active ingredient in a drug is defined by the FDA as any component that provides pharmacological activity or another direct effect in the diagnosis, cure, mitigation, treatment, or prevention of disease. Several NDAs and ANDAs may have the same active ingredients.

Dosage form: A dosage form is the physical form in which a drug is both produced and dispensed. These include tablets, capsules, and injectable forms.

Strength: The strength of a particular drug product captures how much of the active ingredient is present in each dosage form.

Supplemental approval data The RegActionDate data file contains records of all approvals associated with each NDA and ANDA. This file contains 128,906 approvals, which correspond to 17,299 unique application numbers. In addition to initial marketing approvals, this file contains records of "supplementary new drug applications" (sNDAs).

An sNDA is an application to the FDA requesting a change to the agency's approval for an existing drug. Changes may include new or modified indications, revisions to prescribing information, new routes of administration, or new patient populations.

There are more than 50 categories of approval actions that can be granted using an sNDA. These approval types include standard approvals (which appear in our data as "N" type approvals), presubmissions ("M"), and manufacturing changes ("SCM"). For each approval, the FDA includes its action date, the date on which the application was approved.

A.1.2 Identifying innovator drugs

We use data from Drugs@FDA to define our base sample of new medicines. We call a drug an **innovator** if it meets the following criteria:[67]

1. It was approved under a New Drug Application (NDA) to the FDA.

2. It is the first approval of a unique combination of components.

Determining whether a drug meets the first criterion is straightforward. Records in the Drugs@FDA database (applications file) are assigned one of three application types: "A" (abbreviated new drug application), "B" (biologics license application), or "N" (new drug application). We restrict consideration to applications of type "N."

Determining whether a drug meets the second criterion is more challenging—and foreshadows the medication name matching procedures discussed throughout this appendix. The core challenge is determining what constitutes a *unique component* in a drug. Many drugs are part of "substance sets" that include closely related products, which are pharmacologically and physiologically equivalent. Consider a drug molecule that contains a hydrogen atom. Replacing this hydrogen with "a metal or its equivalent" will form a salt, while replacing it with "an organic radical" will form an ester (Ward (2019), citing Amarin Pharm. Ireland Ltd. v. FDA, 106 F. Supp. 3d 196, 199 n.1 (D.D.C. 2015)). In many cases, these slight modifications have no clinically significant effects. However, these modified forms often appear as distinct active ingredients in the Drugs@FDA database. For example, our data include both *acetazolamide* and its salt form *acetazolamide sodium* as active ingredients. If we treated unique active ingredient strings as unique drug components for the purpose of constructing this sample, we would overcount the number of innovator products by incorrectly categorizing salts, esters, and other small chemical modifications to drugs as entirely new inventions.

Instead, we follow the FDA's own practice when evaluating the novelty of products and define innovator drugs in terms of unique *active moieties*: the part of the larger molecule that is responsible for the physiological or pharmacological action of the drug substance.[68] For the example above, *acetazolamide* is the active moiety in both *acetazolamide* and *acetazolamide sodium.* Thus, we construct our sample of innovator drugs by identifying the initial approvals of unique combinations of *active moieties* by the FDA.

[67] This is conceptually similar to, but not the same as, the FDA classification "new molecular entity" (NME). NMEs are defined as new drugs whose active ingredients are chemical substances that are marketed for the first time in the United States. Put differently, NMEs are drugs for which the FDA has not previously approved some component. NMEs are a strict subset of our sample definition. Our modification is to include first approvals of unique combinations of ingredients, even if every component has been previously approved by the FDA.

[68] Debates about whether the FDA should refer to "active ingredients" or "active moieties" when, for example, making determinations about various types of exclusivity have been ongoing since the 1980s. See Ward (2019) for a more detailed discussion.

Active moiety crosswalk To the best of our knowledge, there is no standard crosswalk from active ingredient strings, as appear in Drugs@FDA, to active moieties.[69] We construct such a crosswalk, using the Drugs@FDA Product file as our input data.[70]

We begin by standardizing strings of active ingredients recorded in the database. We ensure text is lower case, components (for drugs that are combinations of multiple ingredients) are alphabetized and separated by semicolons, and extraneous punctuation is removed. Next, we separate each active ingredient string into its component parts. For example, *abacavir; lamivudine; zidovudine* is split into its three pieces. For each component in a drug, we manually review PubChem, chemistry textbooks, and online resources to identify the active moiety in the drug. We reconstruct strings for combination drugs, again ensuring that the component active moieties are alphabetized and formatted consistently.

Of the 25,976 unique active ingredient strings in our data, we determine that the active moiety and active ingredient are identical in 12,510 cases.

Defining our sample We drop six application numbers that are present in the Application file that lack corresponding data in the Product file. All application numbers in the Product file have matching data in the Application file.

We begin by constructing a sample of initial drug approvals. We focus on records with an "N" (standard approval) doctype designation in the Supplement data file. There are 140 unique applications with multiple associated "N" approvals. For these 140 cases, we select the earliest approval recorded as the initial approval date for that application. We also drop 1,126 applications for which we never observe an "N" approval in our data. Thus, we construct a preliminary dataset that includes 16,171 initial approvals.

We merge these records onto the Application and Product files to yield a dataset with information on initial drug approval and product characteristics for 15,361 applications. Excluded from this sample are 836 applications without records in the Product and Application files, and 4,895 applications that have records in the Product and Application files but no corresponding approval histories in the Supplement file.

Finally, we impose the two restrictions described above to construct a sample of innovator drugs, which are unique at the active moiety level. We begin by dropping 122 drugs approved under

[69] The International Nonproprietary Names (INN) initiative maintained by the World Health Organization attempts to deal with exactly this problem. The description of the INN program reads: "When the INN Programme was initiated, it was decided that in such situations, in order to limit the number of published INNs, an INN should be selected for one member of such a group only. This approach, which concerns especially substance sets formed by salts or esters of the same active moiety was validated in the 20th report of the WHO Expert Committee on Nonproprietary Names for Pharmaceutical Substances (Technical Report Series No. 581)." See https://www.who.int/publications/i/item/inn-05-167-3 for details. INNs, however, may not always map onto drug names as used in the United States. For example, "paracetamol" (commonly used in countries where British English is used) is the INN for "acetaminophen," the drug often marketed under the brand-name Tylenol.

[70] In the interest of developing resources that can be re-used in research, we use the most current version of the Drugs@FDA database at the time this work was completed, dated 05 January 2023, for this particular exercise. All other analyses use an older version of Drugs@FDA (Drugs@FDA 1.0) that contains information on product re-approvals central to our analysis.

BLAs.[71] We drop 77 instances in which active moiety groups have no products that were approved under NDAs.[72] We drop 737 drugs based on their dates of approval: 250 instances in which drugs were initially approved before 1962, when the Kefauver-Harris Drug Amendment was signed into law; and 487 drugs approved between 1962 and 1985, when the regulatory structure introduced by the 1984 Hatch-Waxman Act was implemented.[73]

We identify **1,102** unique combinations of active moieties that meet our criteria for "innovator" drugs. Our analyses restrict consideration to the **990** drugs with at least one patent or regulatory exclusivity.

A.1.3 Linking "innovator" drugs to generic competitors

We designate any drug that is approved under an ANDA a "generic."[74] There are 8,322 application numbers in our dataset approved under ANDAs. To determine the first generic entrant for an innovator drug, we consider the set of all ANDAs with identical sets of active moieties and select the drug product with the earliest approval date. If there are multiple generic drugs within an active ingredient group being approved on the earliest date—which occurs for 186 unique products—we designate the drug with the lowest application number as the earliest generic entrant. This selection procedure has no impact on our analysis, which does not depend on the identity of the first generic entrant.

A.1.4 Counting approvals and re-approvals

We construct three FDA approval outcomes.

Initial approvals For each innovator drug, the initial approval is the earliest standard "N" action associated with the innovator NDA.

New use approvals An "SE1" action is an FDA approval for a new or modified indication. We refer to these actions as new use approvals. Our primary outcome assigns a drug-year an indicator equal to one if the innovator NDA receives at least one "SE1" action in that year.

[71] Most BLAs are approved through the FDA's Center for Biologics Evaluation and Research (CBER). The sample of BLAs approved through CDER and indexed in Drugs@FDA is not representative of biologic drugs as a class.

[72] Examination of these cases indicates that these are instances where products were either (i) never approved as FDA brand-name drugs (e.g., application number 077394 was processed as an abbreviated new drug application in 2005, corresponding to "sodium bicarbonate" - baking soda); or (ii) correspond to innovator drugs that were approved before our dataset begins in 1962 (e.g. various applications related to penicillin, which was widely used beginning in the 1940s).

[73] The Kefauver-Harris Drug Amendment, often referred to as the "Drug Efficacy Amendment," introduced the requirement that drug manufacturers demonstrate the effectiveness and safety of drugs before approval. Manufacturers were also required, after this point, to disclose accurate information about side effects. Cheap generic drugs could no longer be re-branded as brand-name (expensive) drugs and re-marketed as "breakthroughs."

[74] This designation is standard and consistent with FDA language. In FDA documents, ANDA and 'generic' are often used interchangeably. See https://www.fda.gov/Drugs/DevelopmentApprovalProcess/HowDrugsareDevelopedandApproved/ApprovalApplications/AbbreviatedNewDrugApplicationANDAGenerics/default.htm for Drugs@FDA descriptions of ANDAs for generic products

Our primary analyses record re-approvals associated with an innovator drug (i.e., regulatory actions under the same new drug application associated with the innovator approval). In robustness checks, we instead collect re-approvals associated with all NDAs that include the same active moiety. Concretely, consider the case of ranitidine, most commonly marketed under the brand-name Zantac. Ranitidine was first approved by the FDA in 1983 as a treatment for peptic ulcers.[75] It was subsequently re-approved under 11 additional NDAs, with distinct brand-names. Our primary analysis focuses on re-approvals for new indications associated with only the first NDA. In robustness analyses, we consider re-approvals associated with any ranitidine NDA. Note that ranitidine is an outlier in terms of total number of associated NDAs: the median innovator drug in our data is approved under just one NDA and only 43 drugs are approved under more than five NDAs.

All supplemental approvals We separately identify all non-initial supplemental actions associated with an innovator NDA. This broader category includes approvals for new uses, strengths, dosage forms, routes, formulations, patient populations, and labeling changes.

A.2 Scientific publications

We collect biomedical publication records from PubMed, maintained by the National Library of Medicine. PubMed indexes biomedical publications and assigns standardized Medical Subject Headings (MeSH) to the chemicals, diseases, and other concepts discussed in each publication. We use these data to construct three measures:

1. the number of publications associated with each drug;
2. whether a publication acknowledges public research support; and
3. the first appearance of new drug-disease associations in the indexed literature.

A.2.1 Linking active moieties to MeSH identifiers

As of 2022, there were more than 30,000 entries in MeSH. Each MeSH term has an associated unique ID (UI). Although each term corresponds to at most one UI, one UI can be associated with multiple synonymous terms. If chemical names are tagged with UIs, then, the numerical identifiers can allow us to identify and properly group synonyms.

For each active moiety in our innovator drug sample, we identified the corresponding MeSH term and UI. We used a crosswalk between MeSH terms and UIs drawn from the PubMed database.[76] When active moiety strings matched the text description of MeSH terms precisely, we assigned both

[75] Note that its approval date, 1983, means that Zantac is not technically part of our analytical sample. We include it here for the sake of illustration.

[76] Specifically, we used the "chemicals" table from the bulk version of PubMed. We constructed a list of all unique MeSH terms reported in the "substname" field and their associated UIs reported in the "substUI" field. As these fields are added by indexers at the National Library of Medicine (and not constructed by authors themselves), we did not verify the accuracy of these substname-substUI links. However, we did note that, as expected, each substname is associated with one substUI and that the same substUI is associated with multiple substnames.

the MeSH term and UI to the active moiety. When there was no precise match, we searched the web-based MeSH browser, to determine if there may be an alternative formatting, spelling, etc. that resulted in a match. In our sample of 1,102 innovator products, we successfully identified a MeSH UI for 810 (74 percent).

Comparatively few combination drugs have associated MeSH UIs. When MeSH terms exist for combination drugs, formatting is standardized (e.g. "hydrocortisone, urea drug combination" (MeSH UI: C010017)). However, of the 351 combination drugs in our sample of innovator products, only 78 have a MeSH UI.

For single-component drugs without a matched MeSH chemical identifier, we assign publication counts of zero. For combination drugs, which are less frequently associated with unique MeSH identifiers, we instead collect publications that reference every component of the combination. These procedures are used to construct our measures of publication activity. (e.g., for "hydrocortisone, urea drug combination," we collect all publications that reference both "hydrocortisone" and "urea"). Our analyses of new drug–disease associations, described in Section A.2.3, are restricted to single-component drugs with matched MeSH chemical identifiers.

A.2.2 Publication counts and recorded public funding

We collect flags from PubMed that indicate whether a publication received any funding from a public agency. We draw on information from the "grants" table of PubMed and designate a publication as having any public funding if there is any entry associated with one of three variables in PubMed: Grant ID, Acronym (Grant Agency), Grant Agency.

A.2.3 New disease associations

We use the disease descriptors assigned by PubMed to measure the arrival of newly documented links between drugs and diseases. For each single-component innovator drug with a matched MeSH chemical identifier, we identify all associated scientific publications and collect all indexed disease descriptors in the MeSH disease tree. To avoid classifying a link between a drug and a disease as "new" each time the database's indexing terminology changes, we exclude a drug–disease pair if its first observed co-occurrence is in the same year that the disease descriptor was introduced into the MeSH vocabulary. For each unique drug–disease pair, we identify the first publication year in which the disease descriptor appears alongside the drug. We call the first observed co-occurrence of a drug and a disease descriptor a *new disease association*. We count the number of new disease associations in each relative drug-year. We interpret this measure as a coarse proxy for the arrival of potential scientific leads. Of course, the co-occurrence of a disease and a drug in a scientific publication is a noisy measure.

A.3 Clinical trials

We combine four sources of clinical trial and drug development records that differ in temporal coverage, unit of observation, and reporting conventions. Given these differences, our primary analyses using clinical trial records use a binary variable that indicates whether *any* clinical trial involving a drug was active in the year of interest.

A.3.1 Datasets

Table A1 summarizes the coverage of our four clinical trial datasets.

Table A1: Clinical Trials Datasets

Data source	Version	Coverage	Principal Unit	Proprietary
NDA Pipeline	Received April 2007	1982–2001	Annual update	No
Pharmaprojects	Current through May 2011	1995–2010	Annual update	Yes
Cortellis	Received 2016	1995–2015	Trial registration	Yes
ClinicalTrials.gov	AACT download dated July 28, 2022	2010–2016	Trial registration	No

NDA Pipeline

We collect records from the NDA Pipeline dataset, maintained by F-D-C Reports. The NDA Pipeline is a reference document covering drug developments during the previous calendar year. Each observation includes a text description of the stage of development for a given drug compound. We use data from 1982 to 2001.

Pharmaprojects

We collect additional information on drug development between 1995 and 2010 from Pharmaprojects, a proprietary dataset that records events associated with drugs at various stages of the development pipeline.[77] Pharmaprojects is compiled and sold by Citeline. Employees gather information from company websites, reports, and press releases. Every company in the database verifies information related to their listed drugs. Pharmaprojects reports trials that are newly initiated and ongoing in each annual report.

Citeline's documentation associated with the dataset notes: "There is continual two-way communication between Pharmaprojects staff and their contacts in the pharmaceutical and biotechnology industries; both to gather new data and importantly, to verify information obtained from other sources."

[77] Our version of Pharmaprojects is current through May 2011.

Cortellis

We also collect records on drug development between 1995 and 2015 from Cortellis, a proprietary dataset that, like Pharmaprojects, records details on drug development activity.[78]

ClinicalTrials.gov

We supplement these records with data drawn from ClinicalTrials.gov, the largest repository of clinical trial records in the world, for clinical trials initiated between 2010 and 2016. Since 2007, sponsors of trials regulated by the U.S. Food and Drug Administration are required, by law, to register clinical trials with ClinicalTrials.gov. Since 2005, registration in a global registry has been required as a condition of publication in any journal associated with the International Committee of Medical Journal Editors.

A.3.2 Data construction

NDA Pipeline

NDA Pipeline records include the following information: document publication year, firm name, trade name of drug, generic name of drug, abbreviation of drug name, description of study, status, and miscellaneous notes. We identify developments about clinical trials by searching for keywords within the "status" variable. Specifically, we keep records where the status variable includes either "phase" (as in, "Phase I trial") or "clinical." We drop records where the description includes "preclinical."

As these records are intended to provide annual updates on drug development pipelines, clinical trial records are best interpreted as reports of trials that were active in the year prior to publication. That is, they reflect firm decisions about project *continuation*, rather than project *initiation*. Since these documents were published one year after the relevant clinical developments, we subtract one from the year of publication to generate a variable capturing the year in which a trial was conducted.

We link records to our baseline sample of FDA-approved active moieties using the generic name of the drug. We standardize the formatting of the string (remove punctuation, alphabetize components, remove extraneous characters, etc.) and attempt two merges. First, we attempt to merge the generic name of the drug in this dataset to the active ingredients associated with our baseline sample. For observations that do not successfully merge, we next attempt to link generic names to the active moieties associated with our baseline sample. We impute annual trial counts of zero for all active moieties approved during this time period.

Our final dataset includes records associated with 543 FDA-approved active moieties (or combinations) tested between 1981 and 2000. The most studied drugs in our sample are paclitaxel and thalidomide.[79]

[78] Details on the contents of Cortellis are available here: https://cortellislabs.com/api/clinical/. There are a small number of records for trials initiated before 1995, which appear to be incomplete (e.g., there is one trial in the Cortellis data that is active in the year 1981). We do not include these records in our sample.

[79] Paclitaxel is a type of chemotherapy that can treat various forms of cancer, originally approved in 1992. See https:

Pharmaprojects

We use data from Pharmaprojects on the status of drug development projects. We collect information on clinical trial developments by, similarly, searching within a variable about project status ("originator status"). As with NDA Pipeline data, we keep records where the status variable includes either "phase" (as in, "Phase I trial") or "clinical." We drop records where the description includes "preclinical." Pharmaprojects records provide annual updates on drug development projects and, thus, also capture project *continuation*, rather than project *initiation*.

Our linking procedure with Pharmaprojects data is identical to the linking project for NDA Pipeline. We link records to our baseline sample of FDA-approved active moieties using the generic name of the drug. We standardize the formatting of the string (remove punctuation, alphabetize components, remove extraneous characters, etc.) and attempt two merges. First, we attempt to merge the generic name of the drug in this dataset to the active ingredients associated with our baseline sample. For observations that do not successfully merge, we next attempt to link generic names to the active moieties associated with our baseline sample. We impute annual trial counts of zero for all active moieties approved during this time period.

Our final dataset includes records associated with 613 FDA-approved active moieties (or combinations) tested between 1995 and 2011. The most studied drug in our sample is paclitaxel.

Cortellis

We also collect data from Cortellis that—like the Pharmaprojects data—collects data on drug development typically sold for the purposes of "competitive intelligence." Cortellis includes one record per clinical trial and includes variables that capture start year and end year. We transform these data into records that indicate whether a given trial on a drug was active in each year.

Our linking procedure with Cortellis data is identical to that used for NDA Pipeline and Pharmaprojects. We, again, impute annual trial counts of zero for all active moieties approved during this time period.

Our final dataset includes records associated with 990 FDA-approved active moieties (or combinations) tested between 1995 and 2016. The most studied drug in our sample is docetaxel.

ClinicalTrials.gov

We use data from ClinicalTrials.gov to supplement records from NDA Pipeline, Pharmaprojects, and Cortellis. Although ClinicalTrials.gov was made available to the public in 2000, we collect records beginning in 2010, by which time trial registration had become more common.[80] We collect

//www.nytimes.com/2005/01/08/business/new-form-of-breast-cancer-drug-approved.html. Thalidomide was approved by the FDA in 1998 to treat leprosy and has since been reapproved to treat various forms of cancer. Given its (notorious) history of causing birth defects, its distribution is tightly controlled. See https://www.nytimes.com/1998/07/17/us/thalidomide-approved-to-treat-leprosy-with-other-uses-seen.html.

[80] In 2005, the International Committee of Medical Journal Editors (ICMJE) began requiring trial registration as a condition of publication. In 2007 the requirements for submission to ClinicalTrials.gov were expanded after Congress passed the Food and Drug Administration Amendments Act of 2007 (FDAAA) (PDF). Section 801 of FDAAA (FDAAA

records of all trials, regardless of trial phase, that were active between 2010 and 2016. ClinicalTrials.gov allows us to construct two measures of trial activity: trial initiation decisions (using trial start years) and trial continuation decisions (using trial active years). To facilitate integration with NDA Pipeline and Pharmaprojects data, our main analyses use trial continuation decisions measured using trial active years. There are a number of concerns about the completeness and accuracy of ClinicalTrials.gov records (see, for example, Zarin et al. (2017); Tse et al. (2018)). Given these limitations, we report analyses that consider the likelihood of any clinical trial associated with a given drug in each year, rather than considering the total number of trials.

As ClinicalTrials.gov is administered by the National Library of Medicine, we follow the same procedure to link records used for PubMed data. Trial sponsors report the interventions (therapies, behavioral treatments, drugs, etc.) associated with each study using standardized MeSH terms, in the file *browse_interventions*. We link MeSH terms associated with clinical trials to the MeSH terms associated with each drug in our baseline sample. We impute annual trial counts of zero for all active moieties approved during this time period with no matching trial records.

Our final dataset includes records associated with 1,013 FDA-approved active moieties (or combinations) tested between 2010 and 2016. The most studied drug in our sample is docetaxel.

A.3.3 Clinical trials sample

We construct a sample of clinical trials by appending records drawn from NDA Pipeline, Pharmaprojects, Cortellis, and ClinicalTrials.gov. Our primary measure of trial activity is an indicator variable that takes a value of one if there was any active clinical trial associated with an innovator drug in a given year. Although our data enable us to consider the total volume of clinical trials, we prefer to focus on a binary measure to avoid introducing concerns about differences in measurement and coverage across the four datasets.

A.4 Drug-indication crosswalk

We construct data on the specific therapeutic uses (indications) associated with each drug approval and re-approval in our sample. For each approval/re-approval event, we collect text descriptions of the associated indications from a variety of sources. Where possible, we use FDA application number and date of (re-)approval to retrieve the relevant drug label from the FDA's full-text drug product labeling service, FDALabel.[81] We supplement data on labels—which are primarily available from 2005 forward—with data from drug approval packets[82], dated company press releases, scientific articles, and other internet searches.

For each text description, we identify a set of associated alphanumeric codes from the International Classification of Diseases, Tenth Revision (ICD-10), a classification system of diagnosis

801) required more types of trials to be registered and additional trial registration information. For more details, see https://clinicaltrials.gov/ct2/about-site/history.

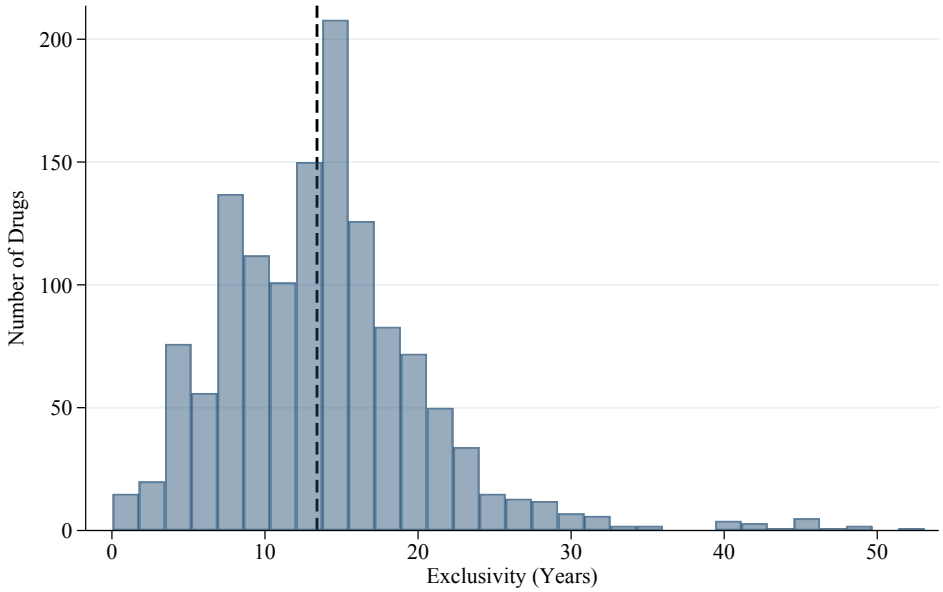
[81] https://www.fda.gov/science-research/bioinformatics-tools/fdalabel-full-text-search-drug-product-labeling

[82] Drug approval records are available for many drugs beginning in 1995 through the Drugs@FDA interactive search tool: https://www.accessdata.fda.gov/scripts/cder/daf/index.cfm

codes representing conditions and diseases, related health problems, abnormal findings, signs and symptoms, injuries, and external causes of injuries and diseases. We use online searches to identify an initial set of "likely" ICD-10 codes. We verify the accuracy of these codes using online code look-up tools, intended for hospital staff who are completing billing tasks.[83] A second researcher verified the accuracy of a random 20 percent sample of codes. We collected the most granular ICD-10 code available, but use truncated versions in some analyses. Concretely, instead of using **G43.1** (Migraine with Aura), we often use **G** (Diseases of the Nervous System) and **G43** (Migraine).

A.5 Distribution of exclusivity measures

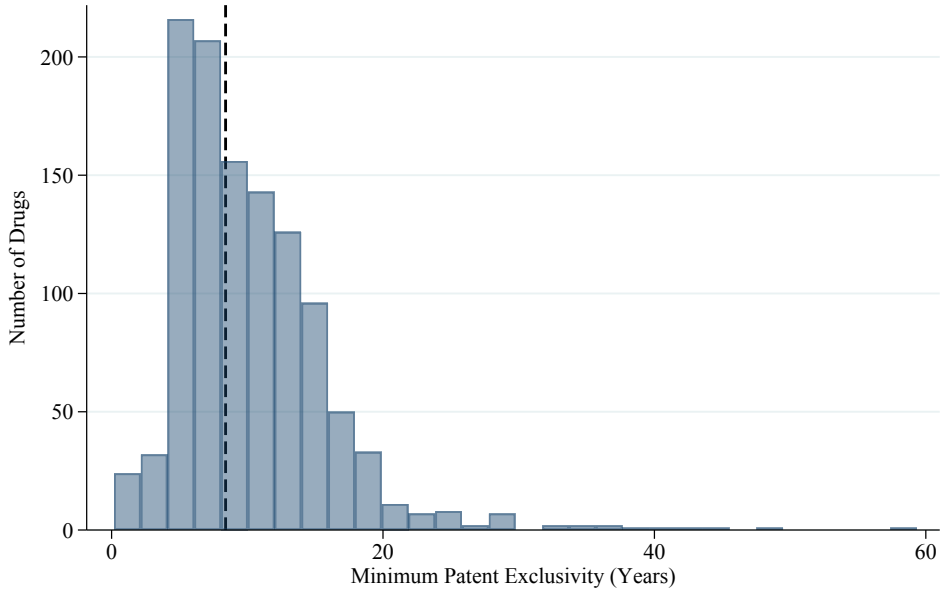
Figure A1: Distribution of exclusivity



Notes: Figure is a histogram that plots the distribution of the exclusivity measure defined in the text as the number of years between an innovator drug's initial approval and the earlier of the expiration of all patents and regulatory exclusivities and generic entry. The distribution of this measure is plotted for the sample of innovator drugs defined in Section 4.5.

[83] For example, https://icdcodelookup.com/icd-10/codes

Figure A2: Distribution of minimum patent exclusivity



Notes: Figure is a histogram that plots the distribution of the minimum patent exclusivity measure defined in the text as the number of years between an innovator drug's initial approval and the expiration of all regulatory exclusivities and the first-filed patent. The distribution of this measure is plotted for the sample of innovator drugs defined in Section 4.5.

B Additional Results

B.1 Lifecycle Patterns

Figures 3 and 4 plot trends in the probability that an innovator drug is re-approved for a new use around exclusivity expiration using two different samples. In Figure 3, we include each of the 990 innovator drugs in our primary sample for every year in which data are available. This means that the sample of drugs shifts across points in Figure 3. This changing sample composition generates the non-monotonic trend between $t = -20$ and $t = -15$. Consider, for instance, a drug approved in the year 2000 that receives ten years of exclusivity, and recall that our data include observations through 2014. In Figure 3, this drug appears in the sample plotted at each point between $t = -10$ and $t = 4$. This same drug will not appear in Figure 4, as the balanced panel specifications require, respectively, five and ten years of data both before and after exclusivity expiration.

This Section presents additional cuts of trends across each drug's lifecycle, follows approvals across all NDAs containing the same active moiety, and reviews the small number of approvals appearing in calendar years after the expiration year. These exercises confirm that the three empirical signatures introduced in Section 3.3 are robust to alternative cuts of our data.

B.1.1 Fractional exclusivity periods

Drugs in our sample receive different durations of exclusivity. We therefore construct a normalized event-time measure that maps each drug's period from initial approval to exclusivity expiration onto the interval $[-1,0)$.

For eachf drug, we divide this interval into five equal segments:

$$[-1.0,-0.8),\ [-0.8,-0.6),\ [-0.6,-0.4),\ [-0.4,-0.2),\ [-0.2,0).$$

Under this scaling, the date of initial approval is assigned $t=-1$ and the date of exclusivity expiration occurs at time $t=0$.[84] We add a sixth post-expiration interval, $[0,0.2)$, with the same duration as each of the five pre-expiration segments for that drug. For a drug with twenty years of exclusivity, each segment contains four years. For a drug with five years of exclusivity, each contains one year.

Figure A3 plots the probability that a drug is re-approved for a new use in each of these normalized periods. The probability that a drug is re-approved for a new use is highest in the second period (the first period that does not include initial approval), at which time nearly seven percent of drugs are re-approved at least once. The probability of re-approval declines monotonically in each subsequent period and falls below one percent in the first period after exclusivity expiration. These trends are consistent with the three empirical signatures of a missing market for new uses introduced in Section 3.3.

B.1.2 Cohorts with equal exclusivity duration

Next, we inspect trends in the probability of re-approval for a new use within cohorts of drugs that receive the same number of years of exclusivity.

For each integer exclusivity duration x, we follow drugs from initial approval through five years after expiration. We report the four cohorts nearest the center of the exclusivity distribution:

$$x \in \{12,13,14,15\}.$$

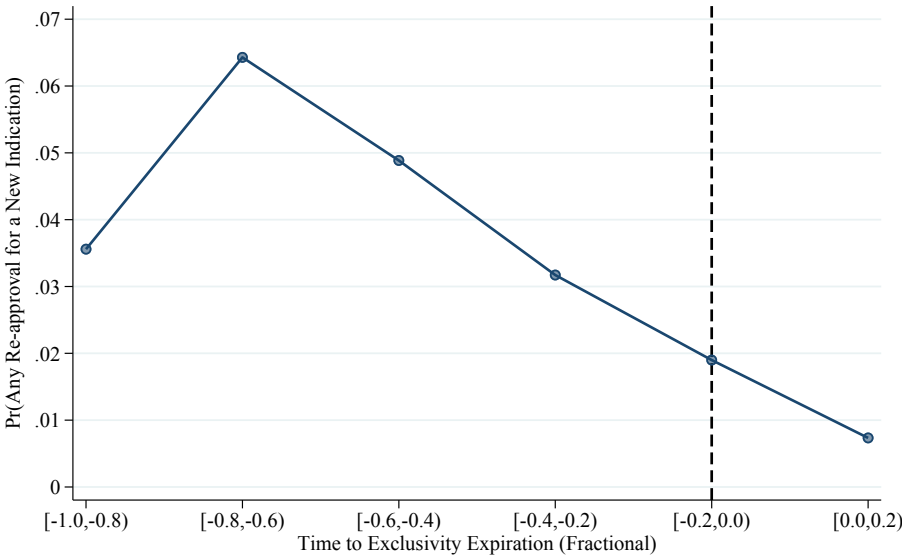
The samples include 59 drugs with twelve years of exclusivity, 70 with thirteen years, 120 with fourteen years, and 85 with fifteen years.

Restricting attention to a common exclusivity duration eliminates changes in composition generated by differences in lifecycle length, but substantially reduces sample size. The resulting series are consequently noisy.

Nonetheless, the trends in Figure A4 for drugs with exactly $x=\{12,13,14,15\}$ years of exclusivity remain consistent with the three empirical signatures of a missing market. For each cohort, the probability that a drug is re-approved for a new use increases sharply in the years following initial approval (marked with a dotted grey line), peaks soon thereafter, then declines. In each case, the probability of re-approval for a new use falls to zero within two years of exclusivity expiration.

[84] Note that our data record the precise dates of initial approval, exclusivity expiration, and re-approval, so we are able to construct these periods precisely.

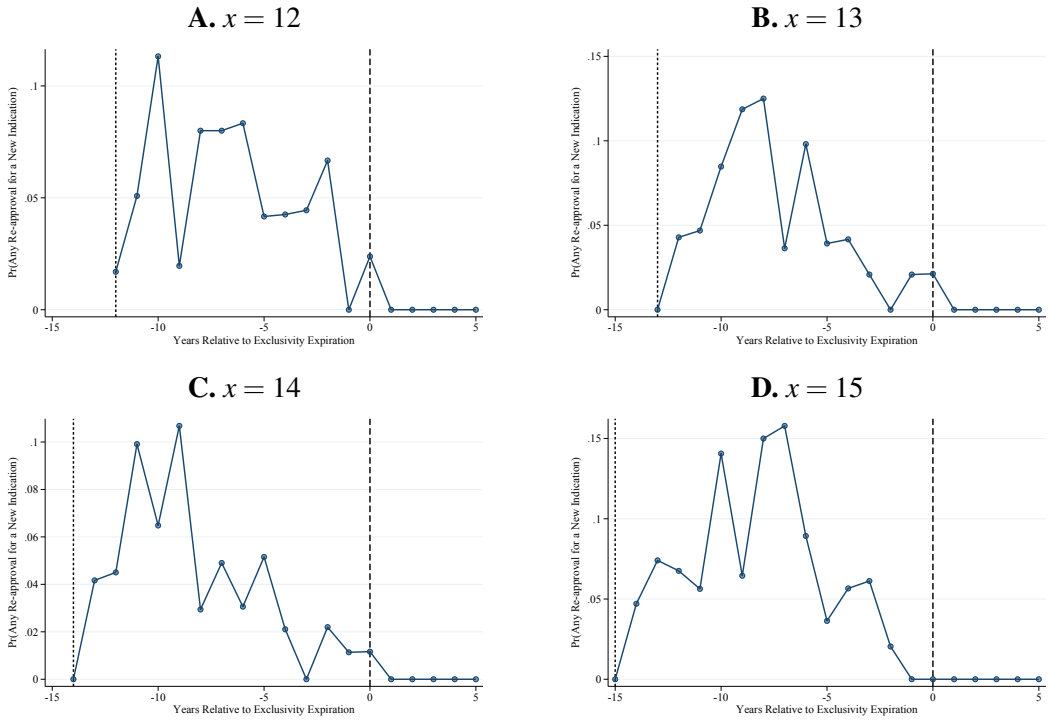
Figure A3: Fractional Exclusivity Periods



Notes: Figure plots the probability of re-approval for a new use for the sample of innovator drugs defined in Section 4.5 in each of six periods relative to exclusivity expiration ($N = 990$). Exclusivity expiration is defined as the earlier of two events: expiration of all patents and regulatory exclusivities linked to the innovator drug in the FDA's Orange Book, or FDA approval of a generic drug with an identical set of active moieties. We split each drug's period of exclusivity into five periods of equal length and plot the probability of re-approval in each of these periods; we consider trends in a sixth period following exclusivity expiration, set to the same length.

Unsurprisingly, given the comparatively small samples, these trends are noisier than those in Figures 2, 3, and 4, but each is consistent with the idea that incentives for investment are high but declining prior to exclusivity expiration and zero afterward.

Figure A4: Exclusivity Cohorts

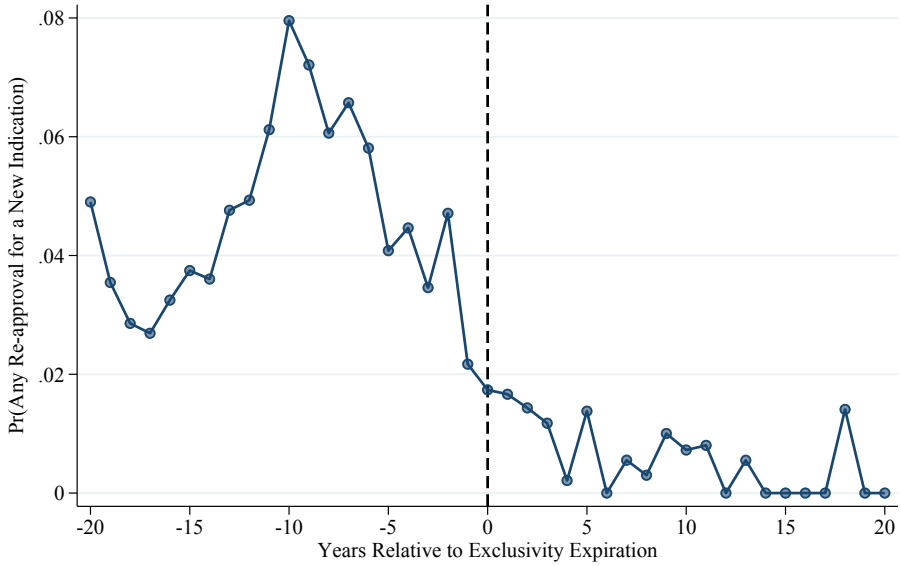


Notes: Panels plot the probability of re-approval for a new use for samples of innovator drugs with exactly x years of exclusivity, where $x = \{12, 13, 14, 15\}$. Trends are relative to exclusivity expiration, defined as the earlier of two events: expiration of all patents and regulatory exclusivities linked to the innovator drug in the FDA's Orange Book, or FDA approval of a generic drug with an identical set of active moieties. We select $x = 12, 13, 14, 15$, which capture the period around median exclusivity (13.3 years).

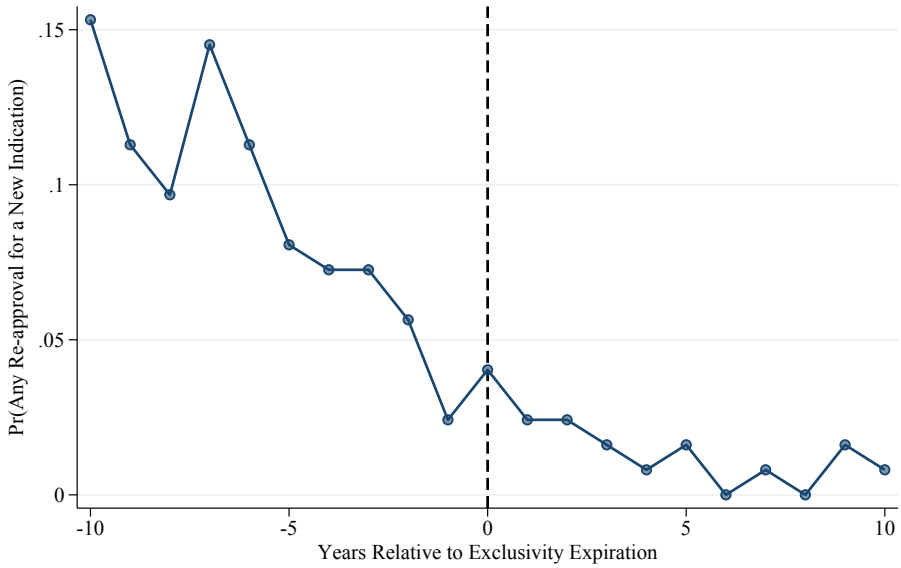
B.1.3 New use approvals associated with any NDA

Figure A5: Probability of Re-Approval (All NDAs)

A. Pooled Sample



B. Balanced Panel (10 years)



Notes: Panel A plots the probability of re-approval for a new use for all active moieties associated with the sample of innovator drugs defined in Section 4.5 ($N = 990$). Panel B is a ten-year balanced panel ($N = 124$). Trends are relative to exclusivity expiration for the innovator NDA, defined as the earlier of two events: expiration of all patents and regulatory exclusivities linked to the innovator drug in the FDA's Orange Book, or FDA approval of a generic drug with an identical set of active moieties.

B.1.4 Hand review of new use approvals after exclusivity expiration

Of the 511 new use approvals in our data, only 17 occur in $t \geq 1$, the calendar years following the year of exclusivity expiration. We review each of these 17 approvals by hand. Details are listed below.

- Levofloxacin (S035). The FDA approved a new use as a treatment for anthrax 40 days after exclusivity expiration, in response to a sudden change in public health needs.
- Ticarcillin/clavulanate (S010). The FDA approved an additional indication associated with serious bacterial infections 308 days after exclusivity expiration. Exclusivity expiration was triggered, however, by approval of an ANDA held by GlaxoSmithKline itself rather than entry by a competing generic manufacturer.
- Ticarcillin/clavulanate (S016). The FDA approved a new use for lower respiratory tract infections, including pneumonia, 5 years after measured exclusivity expiration. Again, exclusivity expiration was triggered by approval of an ANDA held by GlaxoSmithKline itself rather than entry by a competing generic manufacturer.
- Guanfacine (S010). The FDA approved a new use for ADHD 1.5 years after exclusivity expiration. This re-approval was part of a post-marketing requirement imposed by the agency.
- Ursodiol (S016). The FDA approved a new use for gallstone dissolution 2.2 years after exclusivity expiration. No generic version of ursodiol was yet available when this new use was approved.
- Oxiconazole (S003). The FDA approved a new use for tinea versicolor 234 days after exclusivity expiration. No generic version was yet available when this new use was approved, and regulatory documents suggest that re-approval was delayed.
- Propafenone (S002). The FDA approved a new use for maintaining normal heart rhythm in patients with symptomatic atrial fibrillation or flutter 3.1 years after exclusivity expiration. No generic version was yet available when this new use was approved.
- Ticlopidine (S018). The FDA approved a new use to reduce clotting after implantation of a coronary-stent 1.8 years after exclusivity expiration. FDA records suggest that the re-approval was a response to public health concerns about potentially dangerous off-label use.
- Clozapine (S047). The FDA approved a new use for reducing recurrent suicidal behavior in patients with schizophrenia or schizoaffective disorder 8.2 years after exclusivity expiration. Clozapine's re-approval was unusually late and reflects a distinctive public-health context: treatment options in this high-risk patient population are quite limited and the manufacturer had agreed to specific post-marketing monitoring and research obligations as a condition of initial approval.

- Loratadine/pseudoephedrine (S019). The FDA approved a new use for temporary relief of nasal congestion due to the common cold 1.5 years after exclusivity expiration. Regulatory documents suggest that re-approval was delayed.

- Topiramate (S022). The FDA approved a new use for migraine prevention 238 days after exclusivity expiration. Given the short lag, development and FDA review were almost certainly underway before exclusivity expiration.

- Topiramate (S018). The FDA approved the drug as a monotherapy for seizure 1.5 years after exclusivity expiration. Regulatory documents indicate, however, that the innovator firm had submitted materials for re-approval several years prior to exclusivity expiration, and that the agency had requested additional information several times before finally approving the new use.

- Dalteparin (S035). The FDA approved extended treatment and prevention of recurrent blood clots in cancer patients 2.3 years after exclusivity expiration. No generic version had yet been approved.

- Iopromide (S033). The FDA approved use for contrast-enhanced CT imaging 4.8 years after exclusivity expiration. No generic version had been approved by the action date, and the approval concerned a diagnostic procedure rather than a new therapeutic use.

- Dutasteride (S023). The FDA approved a new use 170 days after exclusivity expiration that, on inspection, narrowed an earlier approval given new evidence from longer-term clinical trials.

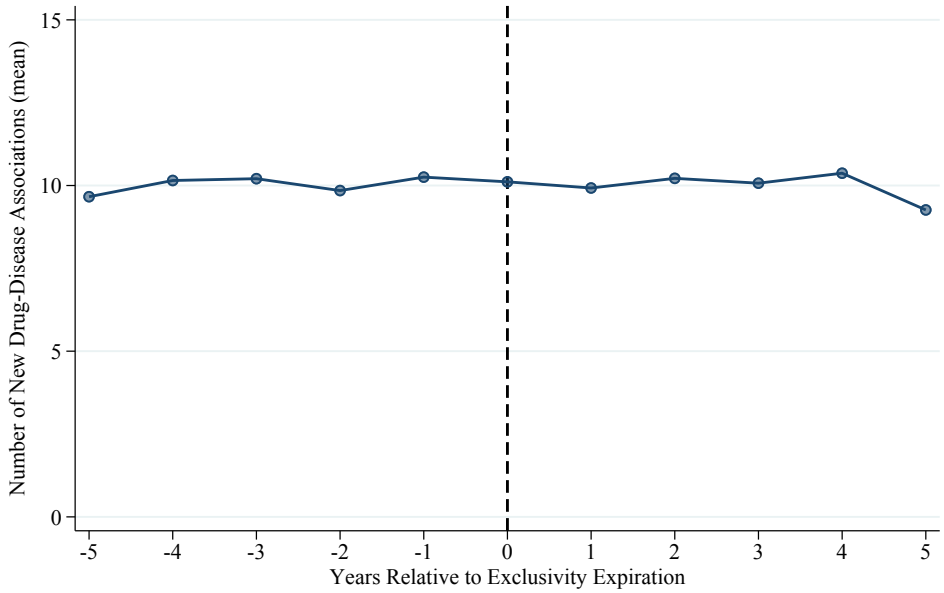
- Ciprofloxacin (S083). The FDA approved ciprofloxacin for complicated urinary tract infections and pyelonephritis 10.7 years after exclusivity expiration. The studies underlying the approval were conducted in response to an FDA Written Request for additional research.

- Erlotinib (S025). The FDA approved new use for different lung cancer indication 2.4 years after exclusivity expiration. On inspection, the re-approval narrowed a previous approval for non-small cell lung cancer in light of new evidence.

B.2 Scientific Information and Clinical Development

B.2.1 Balanced panel estimates

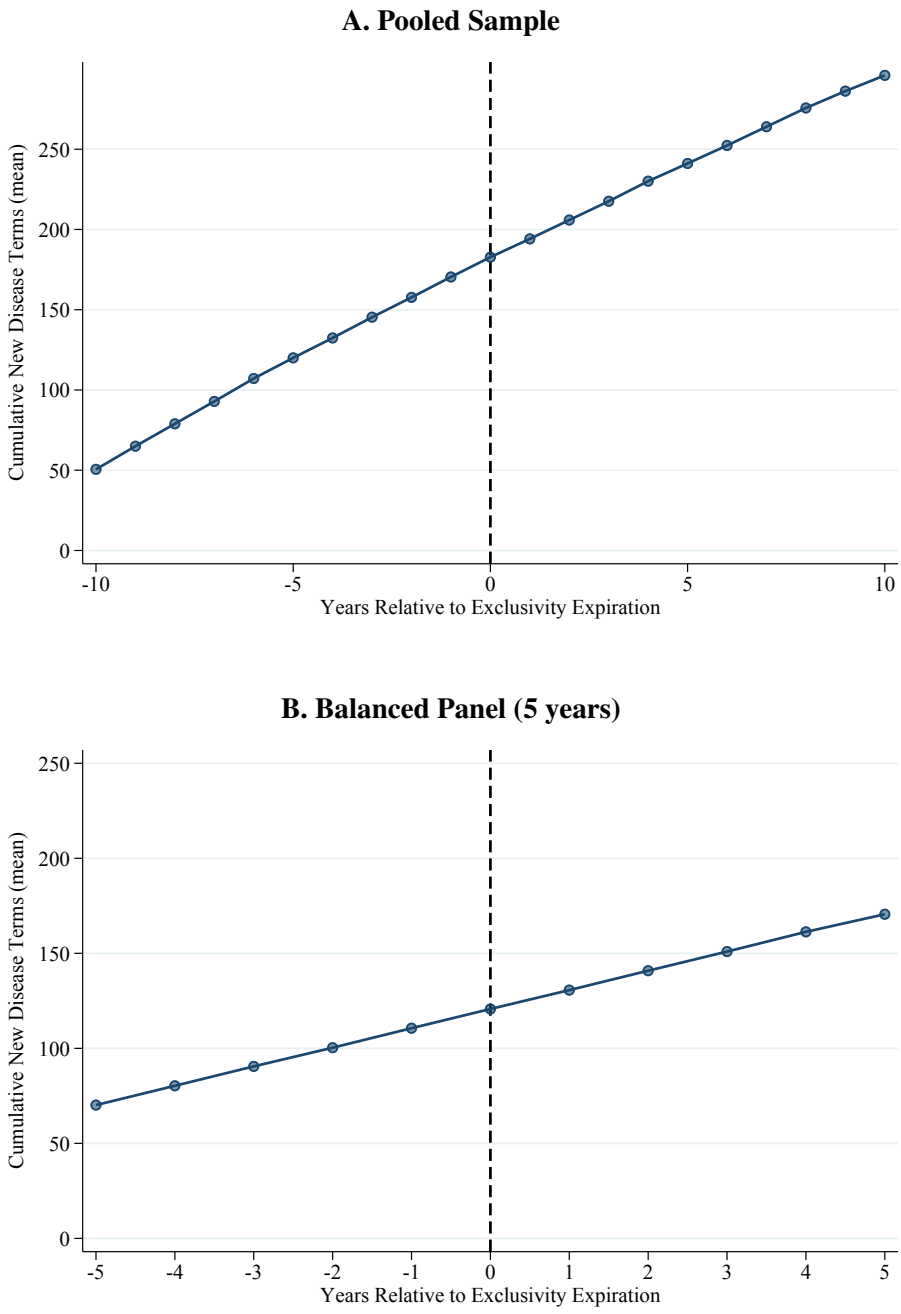
Figure A6: New Disease Associations, Balanced Panel



Notes: Figure plots the arrival of new disease associations in the scientific literature in years relative to exclusivity expiration, for single-ingredient innovator drugs using a five-year balanced panel. A new disease association is the first year in which a MeSH disease descriptor co-occurs with the drug.

B.2.2 Cumulative new disease associations

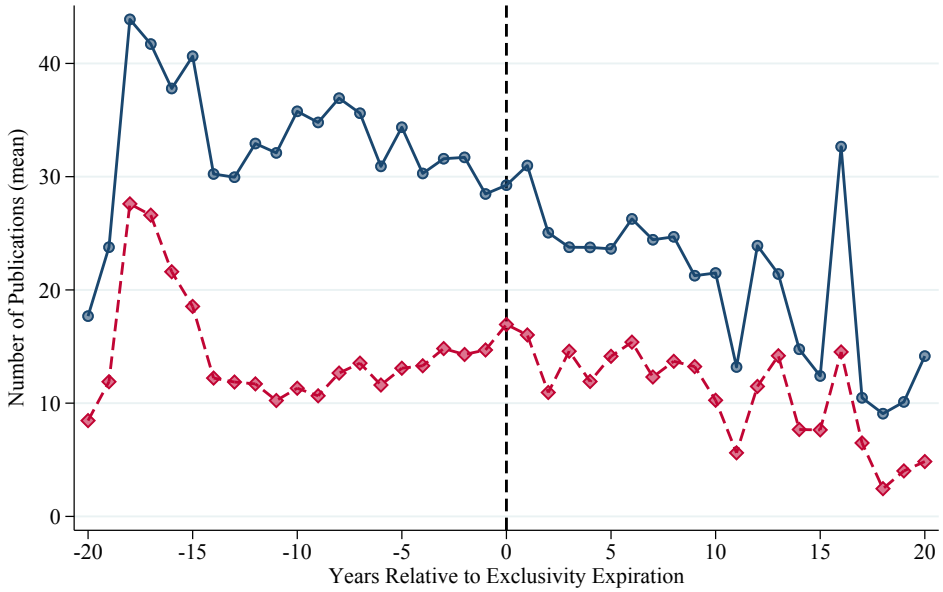
Figure A7: Cumulative New Disease Associations



Notes: Figure plots the arrival of new disease associations in the scientific literature in years relative to exclusivity expiration, for single-ingredient innovator drugs. Panel A uses all available observations; Panel B uses a five-year balanced panel. A new disease association is the first year in which a MeSH disease descriptor co-occurs with the drug. Event year zero is shown but is omitted from the pre/post estimates.

B.2.3 Scientific publications relative to exclusivity expiration

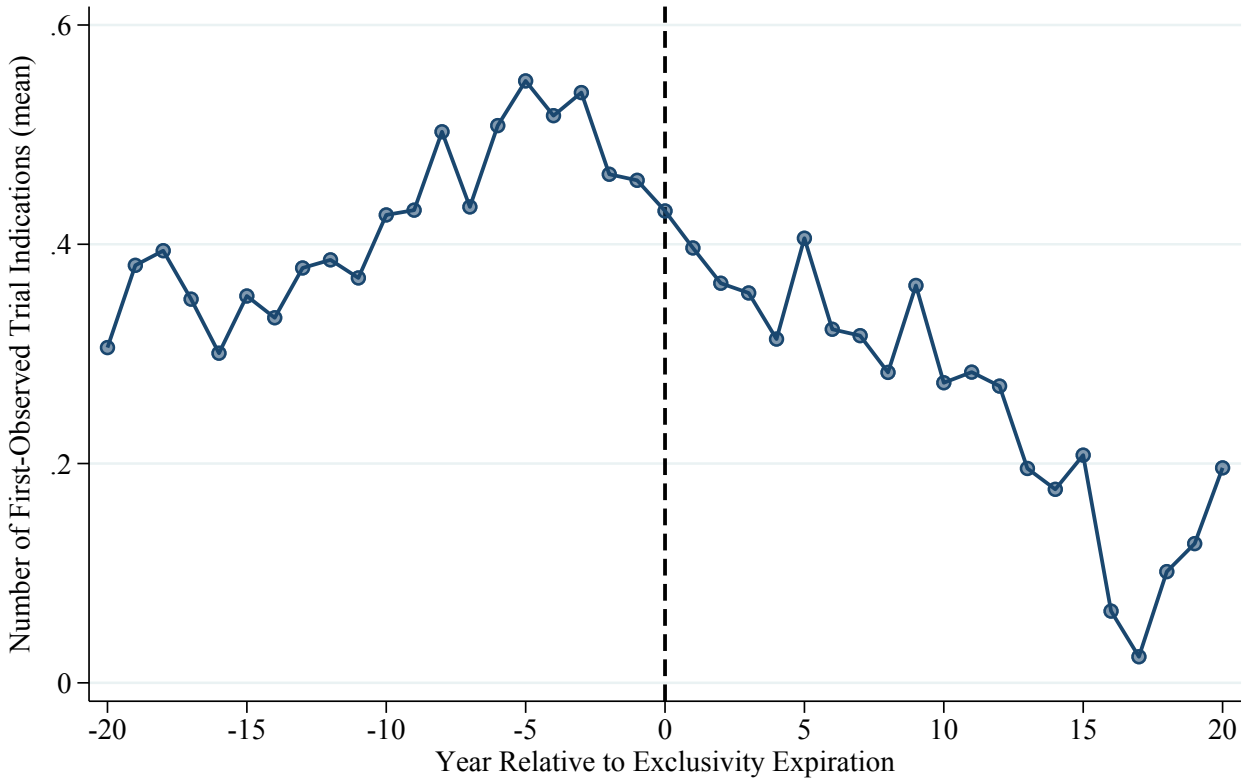
Figure A8: Scientific Publications Relative to Exclusivity Expiration



Notes: Figure plots the average number of scientific publications published in each relative year for the sample of innovator drugs defined in Section 4.5 ($N = 990$). Publication counts are plotted separately for publications that acknowledge receiving any public research support (blue solid line) and publications that do not (red dashed line). Exclusivity expiration dates are defined as the earlier of two events: expiration of all patents and regulatory exclusivities linked to the innovator drug in the FDA's Orange Book, or FDA approval of a generic drug with an identical set of active moieties.

B.2.4 Clinical development

Figure A9: New Clinical Trial Indications Relative to Exclusivity Expiration



Notes: Figure plots the mean number of indications first observed in clinical trial data for each innovator drug-year, in years relative to exclusivity expiration. For each drug–indication pair, we define the first-observed year as the earliest year in which the normalized indication appears in our clinical trial data. We exclude unusable or vague indication descriptions. Drug-years with no first-observed indication are assigned a value of zero. The vertical line marks exclusivity expiration; $t = 0$ is the calendar year in which exclusivity expires.

B.3 Expiration Margin Comparisons

Section 5.2 documents that drugs with longer periods of enforceable protection are ultimately approved for more new uses. Because these estimates capture averages over drugs' full periods of enforceable intellectual property protection, they may obscure discrete changes in commercialization at the time of exclusivity expiration. We therefore present a complementary test that focuses on this margin.

To fix ideas, consider two drugs that are ten years out from initial approval $t = 10$. Suppose that one drug receives ten years of exclusivity $T = 10$, while the other receives eleven $T = 11$. If the duration of enforceable intellectual property protection affects the level of commercialization, we

Table A2: Expiration-Margin Comparisons

	Minimum Patent Exclusivity		Exclusivity	
	Exact	Still Protected	Exact	Still Protected
Still protected in $T+1$	0.0044	0.0185	0.0053	0.0193
	(0.0096)	(0.0060)	(0.0065)	(0.0049)
Control Mean	—	0.015	—	0.008
Approval-year Fixed Effects	Yes	Yes	Yes	Yes
Stack Fixed Effects	Yes	Yes	Yes	Yes

Notes: The outcome is an indicator for at least one new use approval in year $T+1$ after initial approval. *Exact comparison* compares drugs with exactly T years of protection to drugs with exactly $T+1$ years of protection. *Still-protected comparison* compares drugs with T years of protection to all drugs that remain protected in year $T+1$. Standard errors, clustered at the drug-application level, are reported in parentheses.

should expect to see a difference in the probability of a new use approval in lifecycle year $t = T+1$.

More generally, for each drug i, let D_i denote its exclusivity duration in years. For each value of exclusivity duration T, we compare drugs with $D_i = T$ to drugs with longer periods of measured exclusivity and ask whether they differ in the probability that they are reapproved for a new use in year $T+1$. We stack these comparisons across values of T and estimate

$$y_{iT} = \beta \cdot \text{StillProtected}iT + \alpha_T + \gamma a(i) + \varepsilon_{iT}$$

where y_{iT} is an indicator for at least one new use reapproval for drug i in year $T+1$, StillProtectediT is an indicator for remaining protected in year $T+1$, α_T denotes stack fixed effects, and $\gamma_{a(i)}$ denotes fixed effects for the year of initial approval. Standard errors are clustered at the drug-application level.

Table A2 reports two versions of this exercise. The first compares drugs with $D_i = T$ to drugs that receive exactly $T+1$ years of measured exclusivity. The estimated difference is positive but, given that approvals for new uses are rare and the sample sizes in these cohort-level comparisons are small, imprecisely estimated. To increase the power of this test, we broaden the comparison group and compare drugs with $D_i = T$ to all drugs with $D_i \geq T+1$. In this broader comparison, drugs with longer measured exclusivity are approximately 1.9 percentage points more likely to receive a new use approval in lifecycle year $T+1$, using either measure of exclusivity. Using minimum patent exclusivity, the estimate is 1.85 percentage points, with a standard error of 0.60 percentage points. Using our broader exclusivity measure, the corresponding estimate is 1.93 percentage points, with a standard error of 0.49 percentage points.

B.4 Scientific Opportunities

Interpreting the estimates in Section 5.2 as the effect of longer enforceable protection requires assuming that variation in a drug's effective exclusivity at approval is not systematically related to its scientific potential for new uses. This appendix examines that assumption using two proxies for scientific potential derived from the biological characteristics of drugs in our sample.

The first proxy is the number of human proteins that each drug is known to act on in producing its therapeutic effects. The second is the number of diseases linked, through human-genetic evidence, to at least one of those proteins. Intuitively, drugs that act on more proteins, or on proteins implicated in more diseases, likely have greater potential for additional therapeutic uses.

These measures are useful for our purposes because they do not mechanically assign greater scientific potential to drugs that have already attracted more research or commercial investment. Of course, neither measure perfectly captures the latent set of scientifically feasible new uses, which is inherently unobservable. But both are useful in shifting the source of variation away from observed drug–disease development histories and toward features of the drug's biology.

It is worth emphasizing a set of limitations inherent to any exercise of this kind. The latent set of scientifically feasible new uses for a given drug is unobservable; indeed, our analysis turns in part on the fact that new scientific opportunities arise in ways that are difficult to predict given the existing state of scientific knowledge. Although there exist many databases that map drugs to potential diseases, including more recent efforts that use artificial intelligence to identify opportunities for repurposing, estimates of underlying scientific potential constructed using these sources are likely to be endogenous to prior research and commercialization decisions.[85]

B.4.1 Measuring scientific opportunity

We construct our first measure using DrugBank, which catalogs drugs alongside information about the molecular targets through which they act. DrugBank distinguishes pharmacologically active targets—proteins implicated in the drug's therapeutic mechanism—from a broader set of targets. We count the number of distinct pharmacologically active human protein targets associated with each drug.

[85] Given this concern, we ruled out a number of resources that are often used to identify opportunities for repurposing. The non-profit Every Cure, for instance, publishes machine-learning derived "MATRIX" scores, which predict diseases for which existing drugs are likely to be effective. Early iterations of these scores relied on so-called "drug-target-disease" (DTD) models that mechanistically evaluate how drugs interact with biological targets and, in turn, how these targets affect the presentation and evolution of disease; although the linkages predicted by DTD models necessarily reflect the state of scientific understanding and, in turn, the stock of existing research investment, they are anchored in mechanistic reasoning about drug–target and target–disease relationships rather than in records of which drug–disease pairs have already been studied together. The current iteration of EveryCure MATRIX scores instead integrates evidence from a broad set of sources—drug–target databases as well as scientific publications, clinical trials, and records of reapproval—so drugs that have already attracted more commercial attention will receive higher scores by construction. Other candidate databases, including SwissProt/UniProtKB protein annotations and the Therapeutic Target Database (TTD), suffer from the same problem—that they are, essentially, circular measures of opportunity. SwissProt/UniProtKB annotations are manually curated based on links in existing literature; annotations are more detailed for proteins that have attracted more research attention. TTD similarly catalogs targets and indications drawn from the literature.

We construct our second measure using data from the Open Targets Platform, which collects evidence on associations between protein targets and human diseases. For each drug, we identify the diseases linked to any of its pharmacologically active DrugBank targets and count the number of distinct diseases for which at least one target has a human-genetic association score of at least 0.3, Open Targets' default cutoff for "moderate" evidence. This evidence is drawn primarily from human-genetic studies, including genome-wide association studies and rare-variant analyses, which ask whether naturally occurring genetic variation is associated with disease risk or presentation.

We link DrugBank entries to active moieties in our sample using a string-matching procedure that proceeds in four steps. We first standardize drug names in both datasets and match identical names. For drugs that remain unmatched, we next allow one standardized name to begin with the other, which captures many differences arising from salts and prodrug forms. We then match names that contain the same set of words in a different order. Finally, we convert salts and related forms to their parent-acid names and attempt a final match. These procedures allow us, for example, to match azilsartan to azilsartan medoxomil, calcitonin salmon to salmon calcitonin, and alendronate to alendronic acid.

Of the 2,037 single-component active moieties in our baseline approval data, we successfully match 1,835 (90 percent) to a DrugBank entry. We restrict the analyses below to single-component innovator drugs that satisfy the restrictions in the main analysis, matched to DrugBank, and with non-missing measures of both minimum patent exclusivity and exclusivity. This yields a sample of 673 drugs.

B.4.2 Empirical Exercises

We first ask whether drugs that score higher on these measures are, in fact, approved for more new uses. Panel A of Table A3 regresses the total number of reapprovals for new uses on each proxy, with fixed effects for initial year of approval. The relationship is positive for both measures, although it is estimated more precisely for the human-genetic measure.

We next ask whether drugs with longer periods of effective exclusivity also score higher on either measure of scientific potential. For each proxy, we estimate

$$y_i = \beta \cdot \mathsf{exclusivity}_i + \mu_{a(i)} + \varepsilon_i$$

where the dependent variable y_i is one of our proxies for measured scientific opportunity, $\mathsf{exclusivity}_i$ is one of our two measures of enforceable intellectual property protection, and $\mu_{a(i)}$ is a fixed effect for drug i's year of initial approval. Panel B of Table A3 reports these estimates.

Across both proxies and both measures of exclusivity, the estimated relationships are small and statistically indistinguishable from zero. That is, we find no evidence that drugs with longer periods of effective protection have systematically greater measured scientific potential for additional uses.

Figure A10 presents the corresponding relationships graphically for minimum patent exclusivity. For both measures, the relationship between measured scientific potential and effective exclu-

sivity duration is essentially flat, consistent with the estimates in Table A3.

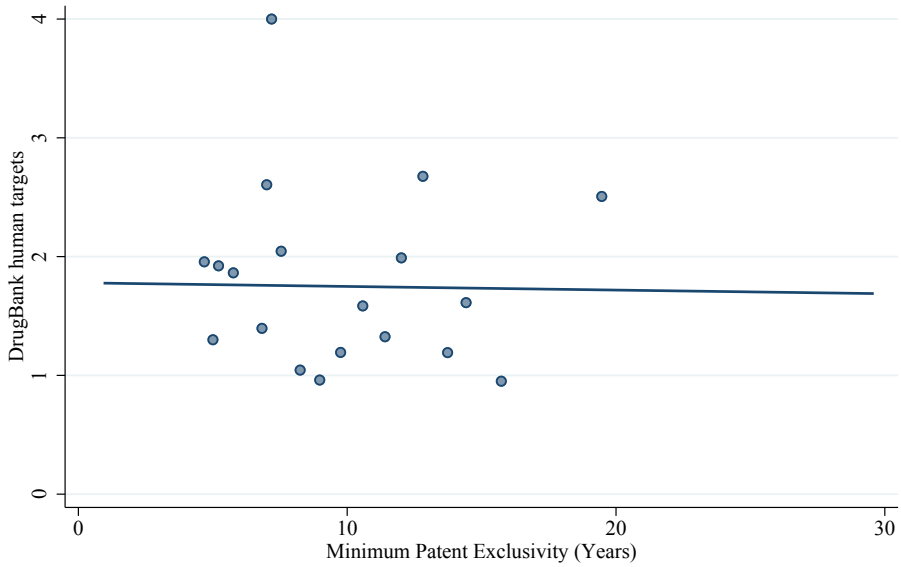
Table A3: Scientific Opportunity and Exclusivity Duration

	DrugBank Human Targets:	Open Targets: Diseases (human-genetic)
Panel A. *Relationship with new use approvals*		
New/modified indication approvals	0.0096	0.0030
(quantity)	(0.0178)	(0.0017)
Panel B. *Relationship with exclusivity duration*		
Minimum Patent Exclusivity	0.0056	0.0015
(years)	(0.0463)	(0.0035)
Exclusivity	0.0656	0.0065
(years)	(0.1007)	(0.0056)
Approval-year Fixed Effects	Yes	Yes
Observations	673	673

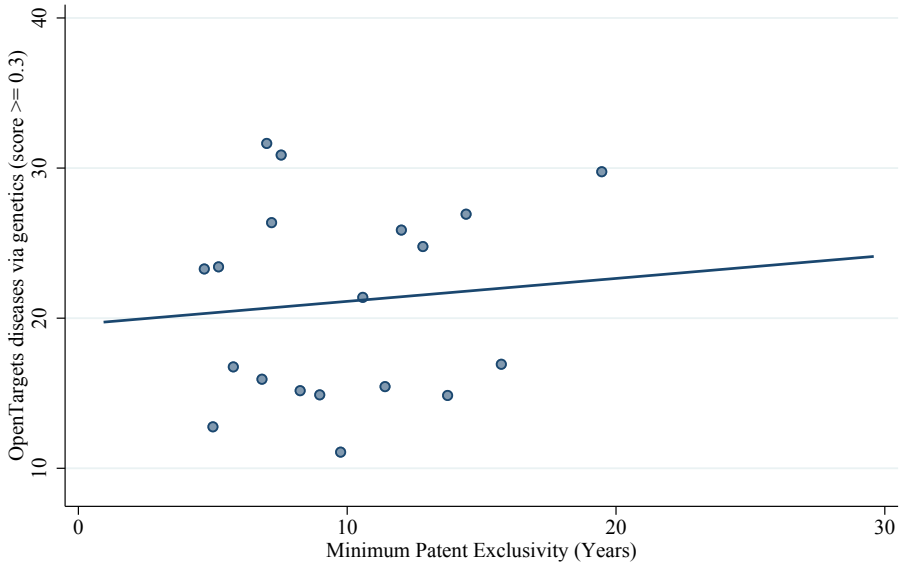
Notes: Each cell reports the coefficient from a separate OLS regression. Panel A regresses the total number of new use approvals on each proxy for scientific opportunity. Panel B regresses each proxy on the indicated measure of exclusivity duration. All specifications include fixed effects for the year of initial FDA approval. HC1 robust standard errors are reported in parentheses. Our sample includes all single-component innovator drugs with a match in DrugBank and for which both exclusivity measures are defined ($N = 673$).

Figure A10: Scientific Opportunity and Exclusivity Duration

A. Pharmacologically Active Human Targets (DrugBank)



B. Diseases Linked Through Human-Genetic Evidence (Open Targets)



Notes: Figure plots the relationship between two proxies for scientific opportunity and minimum patent exclusivity for the sample of single-component innovator drugs that match to DrugBank and have both exclusivity measures defined ($N = 673$). Panel A uses the number of pharmacologically active human protein targets associated with each drug. Panel B uses the number of distinct diseases linked through human-genetic evidence to at least one of those targets. Both axes are residualized with respect to initial-approval-year fixed effects.

C Stylized Analysis of Pull Incentives

When changes to p_{eligible}, p_{enforce}, and p_{secret} are infeasible or unattractive, "pull mechanisms"—which reward successful outcomes ex post—can substitute for enforceable intellectual property rights (see, for example, Barder et al., 2005 and Snyder et al., 2020). Any pull mechanism must replicate the functions of the parameters p_{eligible}, p_{enforce}, and π in our framework. A funder must determine the conditions under which an inventor is eligible for, say, a prize or advanced market commitment payout; develop a credible mechanism that guarantees compensation to the first inventor who satisfies the requisite conditions (thereby enforcing against copycats); and offer a payment that substitutes for the quasi-rents promised by intellectual property rights.[86]

Setting the reward size is the core design challenge. If the reward is too small, few (if any) innovators will participate; if it is too large, the funder overpays. Optimal design therefore involves balancing the innovation gains from higher rewards against the fiscal cost of overpayment. We formalize this tradeoff with two stylized models below.

C.1 Full Information Case

Let V denote the expected social value of successful investment, i.e., the prize to society if the pull mechanism indeed successfully elicits the hoped-for invention. Let there be a single representative risk-neutral firm considering whether to proceed with research and development. As in the main text, let c denote the expected cost of investment, and let p_{comm} denote the probability of success. We assume throughout that $V > c$, i.e., the social value of successful investment exceeds its cost; else, the social planner clearly should not bother.

Consider first the case where society has full information about the probability of success p_{comm} and the cost of investment c. Society wants the representative firm to invest if and only if

$$p_{\text{comm}} V - c \geq 0. \tag{6}$$

Let Π denote the prize society pays for a successful invention. The representative firm will invest if and only if

$$p_{\text{comm}} \Pi - c \geq 0. \tag{7}$$

Under the maintained assumption of full information about c and p_{comm}, the optimal pull mechanism takes a very simple form: if condition (6) obtains and society wants to elicit the investment, then society sets a prize such that condition (7) obtains with equality. The optimal prize is

$$\Pi^* = \frac{c}{p_{\text{comm}}}. \tag{8}$$

Intuitively, society elicits the invention by offering a prize equal to the firm's risk-adjusted cost. Next, we consider a stylized version of incomplete information.

[86] For details on how these three elements were operationalized to incentivize investment in the pneumococcal vaccine, see Kremer et al. (2020).

C.2 Incomplete Information Case

Now assume that society knows V and c, as above, but not the probability of success p_{comm}. This may capture cases in which the social planner can estimate research and development costs ex ante or audit costs ex post, but nonetheless intrinsically knows far less about the probability of successful investment than the firm itself. We can solve for the choice of Π that minimizes the worst-case loss relative to the first best under full information, considered over all possible realizations of p_{comm}.

If the planner sets Π too low, it loses the social value of the invention, $p_{\mathsf{comm}}V - c$. In the worst case, its choice is too low by just an epsilon (i.e., $p_{\mathsf{comm}}\Pi - c = -\varepsilon < 0$), and its loss is therefore $p_{\mathsf{comm}}V - c = c(\frac{V}{\Pi} - 1)$ as we take ε to zero.

If the planner sets Π too high, it overpays relative to what was needed. That is, it pays Π but could have paid $\frac{c}{p_{\mathsf{comm}}}$. In the worst case of overpayment, $p_{\mathsf{comm}} = 1$ and society has overpaid by $\Pi - c$. Society elicits the invention for a social gain of $V - \Pi$ but could have gotten a larger social gain of $V - c$.

The worst-case optimal choice of Π takes the following simple form:

Proposition 3. *The prize that minimizes the worst-case loss relative to first best is given by*

$$\Pi^* = \sqrt{cV}. \tag{9}$$

Proof. The worst case loss from setting Π too low is $p_{\mathsf{comm}}V - c = \frac{c}{\Pi}V - c$, while the worst case loss from setting Π too high is $(V - c) - (V - \Pi) = \Pi - c$. Hence we seek to choose Π to solve

$$\min_{\Pi} \max \left[\frac{c}{\Pi}V - c, \Pi - c \right] \tag{10}$$

Observe that the worst case loss from underpaying is strictly decreasing in Π (and is infinite at $\Pi = 0$) while the worst case loss from overpaying is strictly increasing in Π (and is negative at $\Pi = 0$). Additionally, both worst cases are continuous in Π. Therefore, the solution to (10) equates the two worst case losses:

$$\frac{c}{\Pi^*}V - c = \Pi^* - c.$$

Algebra then yields

$$\Pi^* = \sqrt{cV},$$

as claimed. □

Intuitively, the optimal prize is larger than the firm's costs c to reflect uncertainty about eventual success; the amount of this "overpayment" is increasing in the ratio of social value V to firm cost c. Precisely, society overpays relative to costs by a factor of $\sqrt{\frac{V}{c}}$ to balance the harm from potentially overpaying against the harm from underinvestment.

For example, if the social value of the realized invention is 100 times its cost (i.e., $V = 100c$), then the worst-case optimal prize is ten times cost (i.e., $\Pi = 10c$). This prize overpays by $\Pi - c = 9c$

if successful commercialization was guaranteed ($p_{\mathsf{comm}} = 1$), and it fails to elicit the invention if $p_{\mathsf{comm}} < 0.10$, for a social loss as large as $p_{\mathsf{comm}} V - c = 9c$.

C.3 Discussion

The simple analysis above considers the case of a single representative firm. One additional consideration in the case where multiple firms could potentially invest is that a too-large prize Π might induce a socially wasteful race. This risk is especially salient in the case where V is much higher than c and where firms understand that p_{comm} is high.

An auction or "batching" mechanism might help mitigate overpayment in this instance by inducing potential innovators to compete on price (i.e., a lower prize Π) rather than speed (see Budish et al. (2015b) for a discussion). The design of optimal pull mechanisms with multiple potential innovator firms, and a richer informational and strategic environment than is considered here, is an excellent topic for future research.