One product, many patents: Imperfect intellectual property rights in the pharmaceutical industry

Charu N. Gupta*
UCLA Anderson School of Management

February 28, 2022

Latest version here
Please do not circulate or cite without permission

Abstract

Economists' standard notion of intellectual property rights considers a single patent per product, with a clearly defined scope, certain enforcement, and a fixed term of monopoly protection. Yet common across industries are "imperfect" intellectual property rights: More than one patent may cover a single product, with the scope and enforcement of each uncertain, contributing to an indeterminate period of monopoly protection. Using data on the pharmaceutical industry, I systematically document the presence of imperfect intellectual property rights at the product level and provide the first evidence on the extent to which they impact competition. In a sample of novel drugs, I show that roughly 70 percent of drugs are covered by multiple intellectual property rights. I offer evidence on two mechanisms by which the accumulation of such rights for a single drug may delay generic entry: by introducing a binding later patent expiration and by increasing uncertainty in the scope and enforceability of remaining patents. In an instrumental variables analysis, I determine that the accumulation of patents for a single drug product delays generic entry by over 4 years per drug (amounting to more than 30 percent of mean monopoly life), well beyond the expiration of the drug's initial molecule patent. This research suggests large consequences for consumer welfare in terms of drug pricing and new molecule development and offers an important nuance for future work on optimal patent policy and innovation—that intellectual property rights are less rigid than we typically assume.

*charu.gupta@anderson.ucla.edu. I am grateful to my advisors Matthew Grennan, Daniel Levinthal, and Noam Yuchtman for their continual guidance and support. I thank Francesco Agostinelli, Juan Pablo Atal, Aymeric Bellon, Laura Boudreau, Stuart Craig, Andre Victor Doherty Luduvice, Hanming Fang, Mallick Hossain, Kathleen Hui, Hyo Kang, Margaret Kyle, Darius Lakdawalla, Aviv Nevo, Alex Olssen, Oren Reshef, Lindsey Patterson, Stefano Pietrosanti, Bhaven Sampat, Alejandro Sanchez Becerra, Carlos Segura-Rodriguez, Amanda Starc, Scott Stern, Filipe Stona, and conference and seminar participants at the American Law and Economics Association (ALEA) Annual Meeting, CCC Doctoral Conference, International Industrial Organization Conference, Laboratory of Innovation Science at Harvard Research Seminar, LMU Munich-Max Planck Institute for Innovation and Competition joint seminar, Mack Innovation Doctoral Association Seminar, NBER Aging and Health Trainee Seminar, NBER Productivity Seminar, NBER Summer Institute (Innovation), Northwestern Kellogg-USPTO Annual Conference on Innovation Economics, Penn Empirical Micro Lunch, Penn Health Economics Workshop, Program on Regulation, Therapeutics, and Law (PORTAL) at Harvard Medical School, Purdue Krannert, Technology & Policy Research Initiative IP Days, UCLA Anderson, Wharton Technology and Innovation Conference, and Wharton Health Economics Seminar for helpful comments and discussions. Financial support for this research was provided by the Mack Institute for Innovation Management. Any errors are my own.

1 Introduction

Standard economic theory views intellectual property (IP) rights as "well-defined" and "enforceable" (Coase, 1960; Mas-Colell et al., 1995), meaning a single patent covers a product, that patent has a certain scope, and there is certain enforcement of monopoly rights while the patent is in effect. Under these assumptions, an innovator enjoys a fixed period of monopoly protection meant to recoup its costs of research and development. Once the fixed term expires, competitors may enter the market with an imitation product. Existing theoretical literature on patent races and optimal IP policy relies on this "one-patent-one-product" model (e.g., Denicolo, 1996; Gilbert and Shapiro, 1990; Klemperer, 1990; Nordhaus, 1969, 1972; Scherer, 1972), and recent empirical evidence has highlighted the disincentives the fixed-term system creates for products requiring longer times to commercialization (Budish et al., 2015).

In reality, the assumptions of the standard model rarely hold (Bessen, 2009; Gans et al., 2008; Lemley and Shapiro, 2005). I define "imperfect" intellectual property rights as encompassing three possible characteristics: (1) Multiple patents may cover a single product and the (2) scope and (3) enforcement of each patent may be uncertain. These imperfections contribute to an indeterminate period of monopoly protection. For example, innovators in high-technology industries—such as software, semiconductors, information technology, and pharmaceuticals—often take out numerous patents on different aspects of the same product (Cohen et al., 2000; Noel and Schankerman, 2013). This practice can feasibly extend a product's actual monopoly term well beyond its initial patent life. While oft-discussed among industry leaders, policy makers, and legal scholars, the prevalence of such imperfect intellectual property rights at the product level and their impact on competition and innovation has yet to be quantified.

This paper provides the first systematic evidence on the existence and importance of imperfect intellectual property rights using data from the pharmaceutical industry. In the first part of the paper, I document the presence of imperfect IP and of monopoly power beyond the initial patent term. In a sample of novel drugs, I find that 70 percent of drugs are covered by multiple intellectual property rights and that mean actual monopoly term (13.3 years) is more than 2 years longer than mean initial protection term (11.1 years); the final part of the paper determines the extent to which such extensions in monopoly life are attributable to the accumulation of intellectual property protection. In the second part, I discuss the mechanisms by which IP accumulation for a single drug may influence the timing of generic entry: by introducing a binding later IP expiration for generic entry and by increasing uncertainty in the scope and enforceability of remaining IP rights. Consistent with these mechanisms, I document non-monotonic generic entry patterns relative to IP expirations. In the third part, I use data on patent applications and an instrumental variables analysis to evaluate the causal impact such IP accumulation has on a drug's monopoly life. This analysis corrects for endogeneity concerns stemming from branded firms' patenting decisions and generic firms' decisions on which markets to enter: Branded firms seek to obtain additional patents

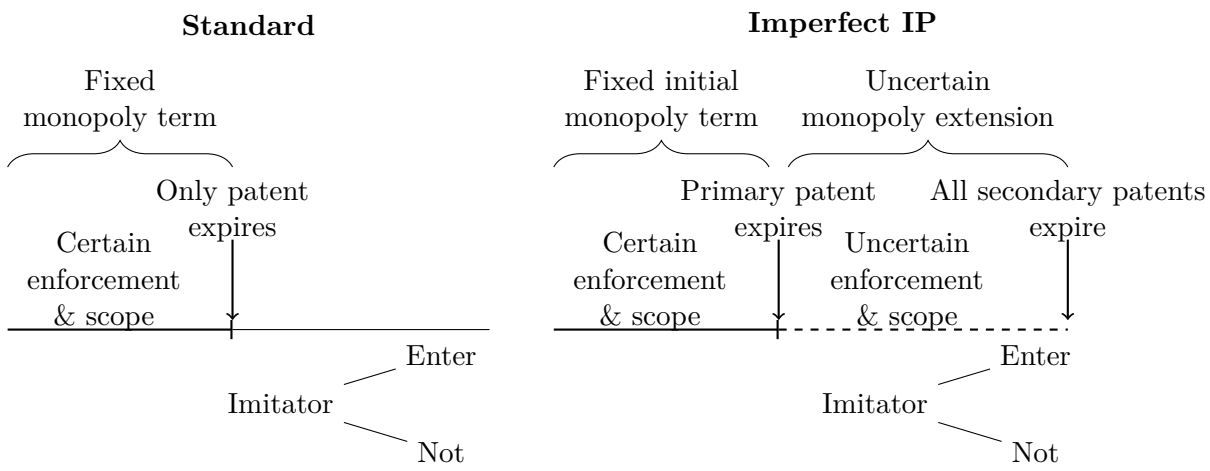
on their most valuable drugs; these are the same drugs that may experience shorter monopoly terms due to their attractiveness to generic entrants. Note that by accumulation I am referring to the combination of the three imperfect IP characteristics (many patents per product and uncertain scope and enforcement of each). My results indicate that the accumulation of patents extends the average drug's monopoly life by over 4 years (equal to more than 30 percent of mean monopoly term), suggesting large potential consequences for consumer welfare in terms of drug pricing.

The pharmaceutical industry offers a noteworthy setting for the study of imperfect intellectual property rights. First, the size of the industry alone—$485 billion in the United States in 2018—makes it an important market for analysis (IQVIA, 2019). Second, debate is ongoing on the strategic use of patenting for branded drug products—referred to as "life-cycle management" within the industry and as "evergreening" by critics. Branded manufacturers often obtain additional patents beyond a drug's initial patent for its active ingredient (the primary patent). These additional patents, called secondary patents, cover auxiliary features such as a drug's formulation (i.e., physical form and route of administration), manufacturing process, or use in treating a particular condition. One view holds that, given the high costs of pharmaceutical research and development (R&D), these additional patents are necessary to incentivize innovation (Gaudry, 2011; Voet, 2020); critics, however, argue that this practice serves only to extend a drug's monopoly life and maintain high drug prices (Angell, 2005; Tribble, 2019). Empirical evidence has shown that length of monopoly protection is an important determinant of innovation in the pharmaceutical sector (Budish et al., 2015), and thus, estimating the causal impact of an additional patent on monopoly life becomes a salient exercise.

Empirically, the pharmaceutical industry also offers advantages that simplify a study of imperfect intellectual property rights. First, there is a clear distinction between the primary and secondary patents, and the primary patent generally has certain enforcement. As such, the setting allows measurement of any extension in monopoly term beyond the initial (or primary) patent term. This may not be the case in other industries where a single product may not have an explicit primary patent. Further, the notion of a product (a given drug) is simpler than in other industries and there are limits to the possible scope of patents that can influence a drug's monopoly term. The Food and Drug Administration (FDA) requires relevant patents on a drug's substance, formulation, and approved uses to be listed in its Orange Book; these are the only patents that may extend a drug's period of monopoly protection. With these simplifications, my estimates of imperfect intellectual property and its impact on competition are likely conservative; the imperfections may be even larger in industries where the definition of a product is more complex and involves patents on numerous component products with nebulous scope (consider, e.g., a smartphone).

Figure 1 offers a visual depiction of the differences between the standard and imperfect models of intellectual property rights under the simplifications of the pharmaceutical sector. Under the standard model, there is a fixed period of protection while the only patent is in effect. After

Figure 1: Standard vs. imperfect IP models



its expiration, competitors may choose to enter or not with an imitation product. Under the imperfect model, there is a fixed initial period of protection for the primary patent that is fully enforced and an uncertain monopoly extension afterwards, due to uncertain scope and enforcement of the secondary patents. This extension term becomes the empirical outcome of interest in the instrumental variables analysis.

As a motivating example, consider the antiretroviral drug Kaletra—used in the treatment of HIV and recently studied as a possible therapy for COVID-19 (Cao et al., 2020). Kaletra, a combination of the active ingredients lopinavir and ritonavir, is protected by 28 separate Orange Book patents and 5 regulatory exclusivities—another form of IP protection granted to pharmaceuticals by the FDA, described in more detail in Section 2. The drug was launched in the U.S. in September 2002, and its primary patent expired in May 2014. Generic entry, however, did not occur until December 2016, after the expiration of several secondary patents and more than 2.5 years after the primary patent expiration. This paper aims to quantify the degree to which such extensions in a drug's monopoly life are attributable to secondary patent accumulation.

My empirical analysis combines several sources of data for the pharmaceutical industry, including information on each drug patent and exclusivity from the FDA, approval dates for branded drugs, and corresponding launch dates for each drug's first generic entrant. I focus on new chemical entities (NCEs) approved between 1985 and 2010. New chemical entities are those drugs with an entirely new active ingredient never before approved by the FDA (Thomas, 2017). For example, Prozac, launched in 1987 for the treatment of depression, is the first drug approved with active ingredient fluoxetine and an NCE; its product extensions with the same active ingredient—Prozac Weekly and Sarafem—are not. Focusing on NCEs means that my estimates will offer a lower bound on the extent to which secondary patenting lengthens monopoly life in that they do not include the effects of product extensions under new brand names.[1]

[1]That is, a branded manufacturer may take out additional patents on its planned product extensions and launch

From these sources, I construct three novel analytic datasets. The first is a drug-level dataset of 370 NCEs, used in descriptive analyses. The second is a drug-month-level dataset following each NCE from branded launch through generic entry. Drugs are observed for 159 months on average (from branded launch through generic entry), yielding a total of 58,898 drug-month observations. I use this dataset to explore how secondary IP rights influence first generic entry timing. The third dataset is a drug-application-level dataset merging each NCE with its secondary patent applications (both granted and rejected) from the United States Patent and Trade Office (USPTO).[2] The matched dataset, restricted to drugs that obtain secondary patent applications and have non-missing values of control variables, covers 248 NCEs and 3,723 drug-application observations and is used in the instrumental variables analysis.[3]

In the first part of the paper, I provide descriptive evidence on the presence of imperfect intellectual property rights and the duration of monopoly protection. First, I show that multiple IP rights cover a single product. In particular, the average drug is covered by 4 Orange Book patents and 3.4 regulatory exclusivities.[4] Second, I show that actual monopoly terms last beyond initial protection terms. On average, NCEs have an initial monopoly life—meaning time from branded drug launch to primary IP expiration—of 11.1 years. Average effective monopoly life—meaning time from branded drug launch to first generic entry—is 13.3 years. Of course, this 2.2-year extension in monopoly life cannot be considered a causal result of secondary patents, due to systematic differences in drugs that have more patents relative to those that do not and the endogenous nature of generic firms' decisions on which markets to enter. I determine a causal estimate in the final part of the paper. Third, I show that not all IP are enforced. The mean potential monopoly life—time from branded launch to the final secondary IP expiration—is 17.9 years. This is the monopoly term the average NCE would receive if all its secondary IP were fully enforced.

In the second part of the paper, I discuss the mechanisms through which secondary IP influences first generic entry timing. I present a simple conceptual framework suggesting that secondary IP delays generic entry in two ways: by introducing a binding later IP expiration and by increasing uncertainty in the scope or enforceability of remaining IP rights. I then provide a series of descriptive evidence consistent with these mechanisms. Visual bunching evidence confirms that IP extensions are binding in their delay of generic entry: While 20 percent of branded drugs see their first generic

these as the NCE's monopoly protection is set to expire. In recent work, Fowler (2019) confirms such launch timing of product extensions.

[2]Technically, patent applications are never rejected by the USPTO but rather abandoned by the applicant. I use the term "rejected" throughout the paper to refer to these applications.

[3]Specifically, I include as controls a drug's market size and therapeutic area, discussed more fully in Section 6. I determine drug market size using the Medical Expenditure Panel Survey's (MEPS) Prescribed Medicines Files and therapeutic area using the World Health Organization's Anatomical Therapeutic Chemical (ATC) classification system. The restriction to NCEs with non-missing market size means the exclusion of (1) drugs with low market sizes and hence not captured by survey data and (2) drugs that are received in hospital settings, e.g., chemotherapy drugs.

[4]I focus on patents listed in the FDA's Orange Book as these are the patents which generic firms must account for when they seek to enter a market.

entrant at primary IP expiry, another 10 percent see generic entry at exactly the time of a later secondary IP expiration. For those drugs with non-binding IP expirations—whose first generic enters either after all IP have expired or between IP expirations—I exploit variation in the number of secondary IP on a drug to disentangle the sources of delayed entry. I separate uncertainty in market profitability from uncertainty in IP scope/enforceability and find that the latter is an important factor.

In the final part of the paper, I estimate the causal impact of an additional patent on a drug's monopoly term. Underlying endogeneity concerns stem from the interrelatedness of branded manufacturers' patenting decisions and generic manufacturers' entry decisions: Branded firms want to protect their most valuable drugs with additional patents; these are the same markets generic firms want to enter earlier, putting the drugs at risk of shorter monopoly terms. I employ an instrumental variables analysis using secondary patent applications for a drug; I evaluate the impact of whether a secondary application is granted on a drug's *extra* monopoly life (meaning time from primary IP expiry to generic entry). Looking at patent applications accounts for endogenous selection by branded firms into *applying* for a patent. Assuming that conditional on being included in an application, whether a drug is granted a secondary patent is not as good as random, it is necessary to find an instrument for whether the patent application is *granted*. For this purpose, I construct a measure of examiner leniency; such examiner designs have become increasingly popular in the economics literature across a variety of settings (e.g., Arnold et al., 2018; Farre-Mensa et al., 2020; Kling, 2006; Sampat and Williams, 2019). The relevance condition for a valid instrument is readily satisfied: A 10-percentage-point increase in an examiner's average grant rate is associated with an 6.7-percentage-point increase in the likelihood a secondary drug patent application is granted. The exclusion restriction requires that examiner leniency only affects extra monopoly life through the likelihood a patent application is granted, which is a reasonable assumption. The IV analysis indicates that a randomly granted secondary patent extends monopoly life by 18.3 months. Given that the average drug has 2.7 secondary Orange Book patents obtained prior to generic entry, this would suggest 49 months, or more than 4 years, of extra monopoly life per drug attributable to patent accumulation. Given a mean actual monopoly life of 13.3 years, this suggests secondary IP rights account for over 30 percent of a drug's monopoly term.

This paper contributes most directly to the literature on intellectual property protection and incentives for pharmaceutical innovation (Branstetter et al., 2016; Budish et al., 2015; Gaessler and Wagner, 2019; Gilchrist, 2016; Higgins et al., 2020; Mohapatra and Zhang, 2020). These papers evaluate how intellectual property duration affects outcomes such as new product introductions, R&D investment, or project continuation decisions but do not explore how firms can influence that IP duration. In particular, Budish et al. (2015) suggest that the fixed-term patent system reduces incentives for longer-term research due to the lag between discovery (when a drug is first patented) and commercialization. I add to this research by showing that an innovating firm may extend its product's monopoly term beyond the initial fixed patent term by accumulating patents on the

product. If firms with products requiring longer development times offset reductions in the initial term by acquiring more patents, then the distortions of the fixed-term system may be less than previously assumed. Conversely, if the products acquiring monopoly extensions are those that do not require greater development costs or time, then patent accumulation on a single product may exacerbate the distortions of intellectual property design. My work thus offers an important nuance that intellectual property rights are less rigid than we typically assume and that these imperfections can alter our discussions on optimal patent policy and innovation.

The extent to which IP accumulation extends monopoly life is also important due to the trade-off inherent in intellectual property protection: A longer patent life encourages firms' ex-ante incentives to innovate but also increases the deadweight loss resulting from monopoly pricing (Nordhaus, 1969). Quantifying the causal impact of an additional patent on monopoly extension is thus a requisite first step in any evaluation of intellectual property policy. Within the pharmaceutical industry, related papers have explored the patenting propensity of branded pharmaceutical firms and corresponding patent challenges by generics (Grabowski et al., 2017; Grabowski and Kyle, 2007; Hemphill and Sampat, 2011, 2012). I build on this literature by estimating the causal impact of secondary patenting on monopoly life. To my knowledge, I provide the first such causal estimate of the extent to which an incremental patent extends innovators' monopoly protection in any industry.

My research on imperfect intellectual property rights relates to other work on non-standard intellectual property. Lemley and Shapiro (2005) provide an overview on the economics of "probabilistic patents"—including uncertainty in the commercial significance of a granted patent or in the scope and enforcement of the patent—and Bessen (2009) offers a theoretical treatment of contracting in the presence of imperfect IP rights. In their study of patent grant delays and cooperative licensing, Gans et al. (2008) detail the various types of uncertainty that exist in intellectual property rights. I offer a contribution by empirically documenting the extent and consequences of imperfect intellectual property rights at the product level in a key sector of the U.S. economy.[5]

The remainder of this paper proceeds as follows: Section 2 provides institutional details related to the pharmaceutical industry. Section 3 describes the data. Section 4 documents the existence of imperfect IP and of monopoly power beyond the initial patent term. Section 5 explores the mechanisms through which secondary IP impacts generic entry timing via a simple conceptual framework and a series of empirical evidence. Section 6 estimates a causal impact of secondary IP on monopoly term extension. Section 7 concludes.

[5]My work relates to empirical studies in other sectors on cross-firm patent thickets (Galasso and Schankerman, 2010; Hall and Ziedonis, 2001; Noel and Schankerman, 2013; Ziedonis, 2004) and within-firm strategic patenting (Abrams et al., 2018; Cockburn and MacGarvie, 2011; Hegde et al., 2009; Kurakina, 2020; Noel and Schankerman, 2013; Righi and Simcoe, 2020). In the pharmaceutical sector, Frakes and Wasserman (2020) study how examiner time allocations relate to the granting of patents of weaker validity. The analyses of these papers occur at either the industry, firm, or patent level. By explicitly attaching patents to products, my paper considers imperfections at a different unit of analysis (the product level) and corresponding impacts to monopoly protection.

2 Background: The pharmaceutical industry

The Drug Price Competition and Patent Term Restoration Act of 1984, more commonly known as the Hatch-Waxman Act, established today's system of both drug regulation and intellectual property protection (Eisenberg, 2001). Meant to balance the competing interests of generic and big pharma manufacturers, the act was a "complex legislative compromise" that had dual goals of simultaneously fostering the generic drug industry while providing additional protections to branded drugs (Angell, 2005; Eisenberg, 2001, p. 121). This section briefly describes branded drug development and approval, the intellectual property protection available for small-molecule drugs, the generic entry process, and proposed reforms to pharmaceutical IP protection. For an overview of the pharmaceutical industry more generally, see Lakdawalla (2018).

2.1 Branded drug development and approval

Drug development begins with discovery of a new molecule. Once a molecule is designed, the firm will patent it (see more on drug patenting below), and the molecule begins preclinical, nonhuman testing. Successful drug candidates will then move through successive phases of clinical trials in human volunteers to establish safety and efficacy. Preclinical and clinical trials for a new molecule can take 10 years on average, upon the conclusion of which a manufacturer will submit a New Drug Application (NDA) to the FDA for approval, including the results of all phases of testing (PhRMA, 2015). Only after FDA approval can a manufacturer market a drug to consumers. While estimates vary, development of a brand new molecule is costly, ranging from $500 million to $2 billion per drug, depending on the therapy and manufacturer (Adams and Brantner, 2006).[6] As such, intellectual property protection is considered necessary to incentivize firms to undertake these high costs of development.

2.2 Intellectual property protection

The pharmaceutical industry is unique in that it receives intellectual property protection not only via patents granted by the USPTO but also via regulatory exclusivities by the FDA (Eisenberg, 2001). I describe here the two types of IP protection provided for pharmaceutical manufacturers.

[6] Development costs for new drugs that make use of an already approved molecule are substantially reduced (according to Chong and Sullivan (2007), by about 40 percent). As existing molecules have already demonstrated safety in clinical trials, producers can typically obtain approvals for new uses of those drugs with fewer resources, i.e. time, money, and trial subjects (Hernandez et al., 2017).

2.2.1 Drug patents

Drug patents extend 20 years from the date of USPTO application and cover one of four characteristics: (1) the drug substance, or active ingredient; (2) its use for treating a certain indication; (3) its formulation, including both the physical form (e.g. liquid or capsule) and administration route (e.g. by mouth or injection); and (4) the manufacturing process to make the drug (Angell, 2005). Once a molecule is developed, the branded firm will patent its active ingredient; this patent offers the strongest form of protection and is considered the primary patent. We think of primary patents as having certain enforcement. Secondary patents, covering one of the other three characteristics, are filed over the course of a drug's development and even after its launch. Multiple patents covering the same product can extend effective patent life, and Angell (2005) notes that even the pill colors and coatings can be patented. Hatch-Waxman also provides for a single USPTO patent term extension per new active ingredient of up to 5 years to account for time in which a drug is in development or under FDA review; branded manufacturers typically choose to apply this extension to the primary, or strongest, patent. Note that while patents last from 20 years after their initial filing, the *effective* patent term is typically smaller, due to the lag between when a patent is obtained (during the drug discovery/development process) and when a product is launched.

Hatch-Waxman requires branded manufacturers to list relevant patents (covering either the drug substance, formulation, or approved uses) in the FDA's Orange Book (Angell, 2005).[7] When a generic firm wants to enter a market, it must account for those patents listed in the Orange Book. As a drug substance patent is the first filed, it often expires before the secondary ones. The scope of secondary patents are described in the Orange Book under use codes or descriptions provided by manufacturers; the FDA relies on these descriptions rather than an actual reading of the patent claims when making generic approval determinations (Rai, 2012). As such, branded firms have previously employed a tactic of listing overly broad use codes in the Orange Book that exceed the actual scope of the patent (Rai, 2012). Under Hatch-Waxman, generic manufacturers, meanwhile, can file a "Paragraph IV certification" or challenge to a branded drug's remaining patents (described in more detail below).

2.2.2 FDA regulatory exclusivities

The Hatch-Waxman Act established a second pathway for IP protection for branded drugs—regulatory exclusivities from the FDA. They are granted once a drug is approved for marketing, and during periods of exclusivity, generic manufacturers cannot enter the market.[8] While patents

[7] Unapproved uses and manufacturing processes should not be listed in the Orange Book.

[8] These exclusivities actually take two different forms: (1) data exclusivity and (2) marketing exclusivity. Data exclusivity covers a firm's clinical trial data on safety and efficacy. During periods of data exclusivity, generic manufacturers may not rely on the original firm's data to submit an abbreviated new drug application (ANDA, described in Section 2.3; however, generic firms are free to develop and submit their own full NDA. Marketing

are submitted to the USPTO for approval, exclusivities are submitted to the FDA. In applying for exclusivities, applicants may submit studies that were conducted by another entity—meaning manufacturers can rely on (1) clinical data or literature from other companies and/or (2) the FDA's prior safety and efficacy findings from an approved drug—thus lowering their burden of proof. Patent terms and exclusivity periods may or may not run concurrently, depending on the type of exclusivity (FDA, 2018).

Numerous different exclusivities exist to protect branded drugs. New chemical entity (NCE) exclusivity is a 5-year protection starting from drug launch for a new active ingredient never before approved by the FDA (Thomas, 2017). Exclusivities may also be given to existing drug products. Changes to approved drugs result in a three-year clinical investigation exclusivity (CIE), including changes in dosage form or dosing regimen, administration route, strength, disease indication, and change from a prescription (Rx) to an over-the-counter (OTC) indication, among others (FDA, 1999; Thomas, 2017). Other exclusivities include those for orphan drugs, for approved medicines that undergo subsequent pediatric clinical trials, and for infectious disease products. Appendix Table 6 provides more detail on these key exclusivities; see Thomas (2017) for a description of all 16 FDA exclusivities that currently exist.

2.3 Generic entry

The Hatch-Waxman Act also provided a pathway for generic entry—the Abbreviated New Drug Application (ANDA)—which lowered the burden of evidence required for generic drug approval. Prior to the act, generic manufacturers were required to submit full NDAs proving safety and efficacy of their drugs. An ANDA allows the generic manufacturer to rely on clinical trial data of the original manufacturer and instead show only bioequivalence to the original branded drug. Note that generic drugs contain the same active ingredient, in the same dosage form, same strength, and route of administration as the original drug. The time and costs of generic development are nontrivial. Copying a branded drug and conducting bioequivalence studies can take a few years, and the ANDA approval process requires both factory inspections by the FDA and laboratory tests of the generic product. Thus, although substantially lower than the costs required for branded drug development, there are still sizeable sunk costs of entry for a generic manufacturer, estimated to be between $250,000 and $20 million per ANDA (Scott Morton, 1999).

A generic firm may submit an ANDA to the FDA four years after NCE launch, although actual entry cannot occur until at least one year later, after the NCE exclusivity has expired. At the time of application, a generic firm must account for a drug's patents listed in the Orange Book and all remaining exclusivities. That is, the generic entrant must specify its intended market entry date and show that, by that date, any unexpired IP on a drug will either be invalid or that the generic

exclusivity prevents generic firms from submitting either application to the FDA irrespective of where the clinical trial data come from (Thomas, 2017). Effectively, exclusivities—whether data or market—forestall generic entry.

can enter without infringing upon those remaining IP. The FDA may then tentatively approve a generic application, with official approval coming only after the scope/enforceability of remaining IP has been fully determined.

Hatch-Waxman created a specific process by which such determinations occur. Specifically, generic firms may submit a Paragraph-IV challenge of a branded drug's remaining patents on the grounds of either non-infringement or invalidity.[9] When a generic manufacturer files an ANDA with a Paragraph-IV certification, the branded firm can either (1) do nothing or (2) sue (within 45 days) the generic manufacturer for patent infringement. If the branded firm does nothing, then the FDA may approve the generic drug.

If the branded manufacturer sues, then there is an automatic stay on FDA approval to allow for ensuing litigation or a settlement. If the courts decide in favor of the branded firm, then the patent challenge fails and the FDA will not approve generic entry. For the generic firm that chooses to submit an ANDA while patents remain on a drug, they thus risk potentially large litigation expenses (and the previously incurred development expenses) without guarantee of approval. Whenever a first court ruling decides in favor of the generic (that either all or some remaining secondary patents are unenforceable), the stay ends and the FDA will officially approve generic entry. Generic entry can occur immediately if all remaining patents are deemed unenforceable. If there is partial enforcement, i.e. at least one remaining secondary patent is fully enforced while others are deemed unenforceable, then generic entry can occur at the point at which those fully enforced patents expire. The first generic firm to win a patent challenge receives 180 days of generic exclusivity, during which time the branded firm and generic challenger constitute a duopoly. After the 180 days, other generic manufacturers may submit ANDAs and enter the market. (Branstetter et al., 2016). Thus, a trade-off exists for the first generic entrant for a drug: Earlier entry gives it the opportunity to earn duopoly profits but comes with greater uncertainty about IP enforceability, raising potential litigation costs.

Scholars disagree on the impact of additional IP to generic entry. One view holds that secondary patents and exclusivities would have little effect in forestalling generic entry since generic manufacturers theoretically can enter after the primary IP expiration (i.e., the latter expiration of either the active ingredient patent or NCE exclusivity). For instance, if a brand new molecule is initially approved by the FDA for the treatment of depression and then subsequently obtains secondary IP for the treatment of bipolar disorder, a generic manufacturer theoretically should be able to enter the market for depression at the primary IP expiration. Yet other scholars note that strategic patenting practices by branded manufacturers still create a considerable barrier to generic manufacturers for entry for any use, as they must navigate an "infringement minefield" of all potential infringement cases that could result (Furrow, 2008; Tribble, 2019). That is, there is considerable

[9] Under non-infringement, generic firms seek approval for the generic with a narrower labeling that excludes on-patent uses.

uncertainty in the scope and enforceability of secondary IP that must be resolved. I settle this debate by providing the first systematic evidence on when generic entry occurs relative to a drug's primary and secondary IP expirations.

2.4 Proposed reforms to pharmaceutical IP

Various legislative proposals have suggested reforms to IP policy in the pharmaceutical industry to limit the extent to which IP accumulation extends a drug's monopoly life. These include reducing exclusivity periods and restricting CIE to drugs showing significant clinical benefit relative to existing therapies (Thomas, 2017). Other proposals similarly suggest raising the threshold for patentability, as is the case in other countries. India, for example, requires higher standards for nonobviousness, and any patent applications that claim improvements over existing drugs must show evidence of clinical benefit; Brazil and Argentina have similar requirements for secondary patents (Amin and Kesselheim, 2012; Sampat and Shadlen, 2017).[10] A first step to evaluating any such change to IP policy is to quantify the impact of secondary intellectual property protections for pharmaceuticals on resulting monopoly terms and generic entry.

3 Data

This project draws from several data sources. The main source of information on drug approvals, patents, and exclusivities is the FDA. Historical FDA Orange Books list all relevant patents and exclusivities and their expiration dates for each small molecule drug in the US. I supplement these data with information from the Drugs@FDA database, including approval dates for branded and generic drugs. Due to inconsistencies in the FDA data in generic approval dates, I manually review news articles and press releases from generic manufacturers to confirm precise launch dates for all drugs in my sample. More details on each granted patent as well as all patent applications (both granted and rejected) come from research datasets at the USPTO.[11] I determine drug market size using the Medical Expenditure Panel Survey's (MEPS) Prescribed Medicines Files and therapeutic area using the World Health Organization's Anatomical Therapeutic Chemical (ATC) classification system.

From these sources, I create three novel analytic datasets. The first is a drug-level dataset of 370 new chemical entities (NCEs) approved between 1985 and 2010.[12] Focusing on NCEs means that my

[10] Sampat and Shadlen (2017) find little effect of India's and Brazil's policies to limit secondary grant rates and greater impact of Argentina's policies. They focus on the granting of these patents; I estimate the impact of granted secondary patents on monopoly terms in the US.

[11] Specifically, I use the following USPTO datasets: Patent Examination Research Dataset, Patent Assignment Dataset, Patent Claims Research Dataset, and Historical Patent Data Files.

[12] These 370 NCEs are those that have a generic entrant as of July 2020. Empirically, the restriction to NCEs with a generic entrant is necessary so that I have a measure of each drug's actual monopoly life. NCEs which are

estimates will offer a lower bound on the extent to which secondary patenting lengthens monopoly life (by excluding the effects of product extensions under new brand names). Key variables include branded launch date, generic entry date, and the start and expiry of each intellectual property right. One challenge is determining which of a drug's IP is primary and which are secondary. I consider a drug's molecule patent to be the one receiving a USPTO patent term extension and in the case of no such extension, the earliest patent filed. I take the primary IP expiry to then be the latter of either the molecule patent or NCE exclusivity expiry. All subsequent patents and exclusivities are secondary.

The second is a drug-month-level dataset of these NCEs from branded launch through the month of first generic entry. Drugs are observed for 159 months on average, yielding a total of 58,898 drug-month observations. This dataset tracks the periods during which each patent or exclusivity is in effect, and I use it to explore how secondary IP influence first generic entry timing.

The third dataset is a drug-application-level dataset merging each NCE with its secondary patent applications, both granted and rejected. Attaching patent applications to products is not a straightforward task and existing work on patent applications generally employs industry-, firm-, or patent-level analyses. I thus offer a methodological contribution that could potentially be applied to future work in other sectors. Construction of this dataset requires restricting the universe of all USPTO patent applications to the set of drug applications and determining which of these applications correspond to each NCE. I attach a patent application to a drug if its claim text includes its active ingredient and if its assignee is one of the assignees for the drug's granted patents listed in the Orange Book; I also include parent and child applications of Orange Book patents. I restrict to drugs that apply for secondary patents and have non-missing values of control variables. The matched dataset includes 248 NCEs and 3,723 drug-application observations.[13] I use this final dataset in the OLS and IV analyses of the impact of secondary patents on extra monopoly life. More details on the construction of this dataset are available in Appendix A.2.

3.1 Definitions

I list below definitions for key terms used throughout the paper.

Primary versus secondary IP:

- *Primary IP*: Either the drug's molecule/active ingredient patent or NCE exclusivity (whichever has a later expiration date). Typically the strongest IP on a drug, with certain enforcement

not included in the sample are those approved by 2010 that have not yet experienced generic competition and those brought to market but then discontinued; this restriction has minimal impact on the sample.

[13]The reduction in sample size from 370 to 248 NCEs is primarily due to the restrictions to drugs with secondary patent applications and to drugs with non-missing market size. Due to the survey nature of the MEPS data, drugs with low market sizes or received in hospital settings will have missing values for market size.

- *Secondary IP*: All of a drug's other patents/exclusivities obtained after the primary IP, covering auxiliary features and uses

Measures of monopoly duration:

- *Initial monopoly term*: Time from branded launch to primary IP expiry
- *Potential monopoly term*: Time from branded launch to final IP expiry
- *Actual/effective monopoly term*: Time from branded launch to generic entry
- *Extra monopoly life/monopoly extension*: Time from primary IP expiry to generic entry (dependent variable in Section 6 analyses)

4 Existence of imperfect IP

I present in this section a series of descriptive statistics invalidating the standard "one-patent-one-product" model. I show that pharmaceutical firms obtain multiple IP rights per drug product, that not all IP rights are enforced, and that monopoly terms across drugs vary, lasting beyond the initial term indicated by the primary IP.

Table 1: Summary statistics

Variable	Mean	Std. Dev.	Min	Max
Number of Orange Book patents	4	3.8	0	28
Number of FDA exclusivities	3.4	2.7	1	17
Initial monopoly term (years)	11.1	3.8	1.1	19.4
Potential monopoly term (years)	17.9	6.7	3.3	44.9
Actual monopoly term (years)	13.3	4.3	3.4	32.6

Total number of drugs is 370. All monopoly terms are relative to branded launch.

Table 1 provides summary statistics on IP accumulation and monopoly life. The first key takeaway from this table is that multiple IP may cover a single drug. The average drug is protected by 4 Orange Book patents and 3.4 FDA exclusivities, with substantial cross-drug heterogeneity in that IP acquisition.[14] For example, HIV treatment Kaletra, discussed previously, is covered by a maximum 28 Orange Book patents, and Gleevec (imatinib), an oral chemotherapy medication used for leukemia, is covered by a maximum 17 FDA exclusivities.

The bottom three rows of Table 1 look at different measures of IP duration. If drugs were restricted to solely their primary (or initial) IP, they would experience an average monopoly life of 11.1 years on the market.[15] (Note that this measure restricts to time on the market post-launch—from

[14] The patent numbers focus on patents specifically listed in the Orange Book; these patents cover drug substance, formulation, or approved uses and are the ones generic firms must account for when they seek to enter a market. A branded firm may also take out other patents that are not listed in the Orange Book, such as patents on manufacturing process.

[15] All drugs in the sample are new chemical entities (NCEs) and thus receive a 5-year NCE exclusivity. The

branded launch to primary IP expiry—and does not include time when the drug may have been protected by a primary patent but still in development). The fourth row gives statistics on potential monopoly term—time from branded launch to final IP expiry. This represents the monopoly life a branded firm would experience if all secondary IP were fully enforced, 17.9 years on average. Comparing this row to the final row indicates that for at least some drugs, not all IP are enforced. That is, when generic entry occurs, additional patents or exclusivities may still be in place but are deemed unenforceable in terms of preventing generic entry. We infer this because mean actual effective monopoly life—time from branded launch to generic entry—is 13.3 years, more than 4 years shorter than the mean potential term. Note that this actual monopoly term, or time to generic entry, is more than two years longer than the initial monopoly term. Subsequent analyses will evaluate the extent to which such extensions in monopoly life represent a causal impact of secondary IP rights.

Figure 2: Distributions of initial and actual monopoly terms

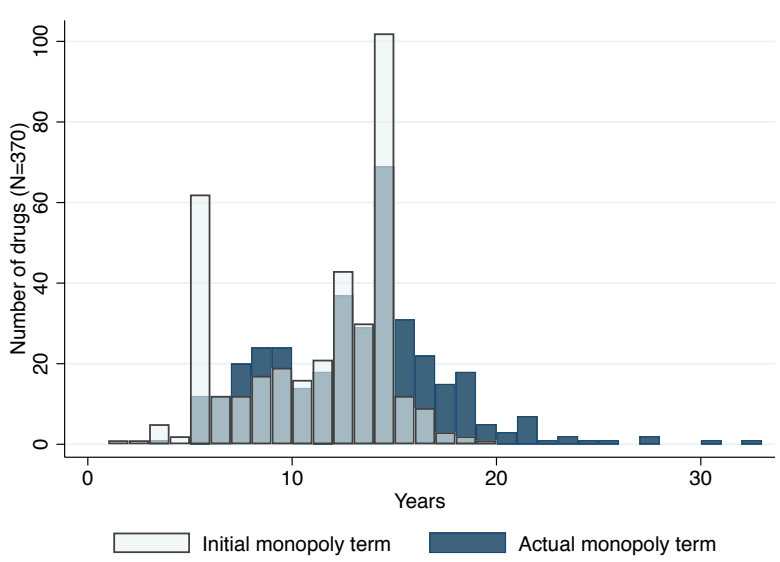


Figure 2 compares the distributions of initial and actual monopoly terms.[16] Two noteworthy facts emerge from this figure. First, there is substantial heterogeneity across drugs in both time to primary IP expiry and in time to generic entry. Second, the distribution of monopoly life shifts right once secondary IP is included, i.e. the distribution of time to generic entry is to the right of that of time to primary IP expiry. This figure suggests that considering only the fixed initial term

minimum initial monopoly term in Table 1 (1.1 years) is below 5 years because in a handful of cases, the branded manufacturer discontinued its original product and transferred the remaining exclusivity period to another product with the same active ingredient. In other cases, a branded manufacturer licensed rights to its drug to different manufacturer who then received the remaining exclusivity.

[16]Two spikes are present in the distribution of initial monopoly terms, one at 5 years and one at 14 years. The first spike at 5 years represents those drugs for which the 5-year NCE exclusivity is the primary IP. The second spike at 14 years results from a limit on the USPTO patent term extension to 14 years past the drug's launch date.

of the primary IP may result in an incomplete measure, or underestimate, of a product's monopoly term. Future research on optimal IP design and innovation should incorporate this potential for products to experience monopoly extensions.

The section confirms that intellectual property rights in the pharmaceutical industry are imperfect: multiple IP cover a single drug, not all of those IP are enforced, and drugs experience varying terms of monopoly protection that can last beyond the initial term. The next section explores the channels through which secondary IP rights influence generic entry timing, and Section 6 estimates the causal effect of those rights on monopoly extension.

5 Mechanisms: How secondary IP impacts generic entry timing

This section discusses the mechanisms through which the accumulation of IP may delay generic entry. I present a simple conceptual framework that produces testable implications I can take to my data. In particular, I posit that secondary IP rights work in two ways: by introducing a binding later expiration for some drugs and by increasing uncertainty in overall IP scope and enforceability for others. I then present a series of empirical evidence that is consistent with the conceptual framework.

5.1 Conceptual framework

I consider a generic firm's decision to enter a pharmaceutical market (where each market corresponds to a branded drug), taking the level of intellectual property protection on the drug as given. The timing of first generic entry in a market exhibits a trade-off between incentives to enter early and preempt competitors on the one hand, and uncertainty in market profitability and/or IP enforceability on the other hand. These two types of uncertainty have different implications. Uncertainty in market conditions leads the generic to actively delay entry until some signal is received that reduces or resolves that uncertainty (as in real options theory). Uncertainty in IP enforceability delays entry (even if the generic manufacturer wants to enter) until that uncertainty is resolved via established legal institutions. I discuss this distinction in the conceptual framework that follows. In particular, the timing of generic entry and number of secondary IP on a drug will help disentangle these types of uncertainty. I consider two separate scenarios: when a branded drug is protected only by primary IP (as in the standard model) and when it is protected by both primary and secondary IP (in the imperfect model).

Suppose first that the branded drug is covered only by primary IP. We generally think of the primary IP as a strong type of protection covering the drug's molecule with certain enforcement. For a generic firm deciding when to enter, attempting entry prior to that expiration will likely lead

to huge costs in litigation that outweigh the benefits. Thus, the generic firm will delay its entry until that primary IP expires. Indeed, while generic firms may prepare and submit applications to the FDA in advance of the actual primary IP expiration, they will specify their anticipated entry date as when the primary IP expires. The FDA may then tentatively approve the generic with an official approval date occurring upon primary IP expiration. We would thus expect the primary IP to be binding for many generic entrants, particularly in those markets with large potential revenues, with first generic entry occurring as soon as that primary IP expires.

Implication 1 (standard model): There will exist a spike in entry at the time of primary IP expiration, due to binding primary IP.

Note that not all generic firms will enter immediately upon primary IP expiration. The firm may wait to learn more information that resolves some uncertainty in market profitability. For example, the generic firm may delay entry in markets offering less revenue or where there may be concern about new treatments in development that would make the drug in question obsolete. In such cases, the generic firm may delay undertaking the costs of bioequivalence trials and its application to the FDA until receiving some positive signal on market profitability.

Implication 2 (standard model): There will exist gradual entry delayed past primary IP expiration, due to learning about market conditions.

Now let us consider the imperfect model where a branded drug has both primary and secondary IP protection. For such drugs, binding primary IP and uncertainty in market profitability remain factors that the generic entrant must consider. However, the accumulation of IP on a single drug can also lengthen the branded manufacturer's protection term in two other ways—by introducing a binding secondary IP expiration and by increasing uncertainty in overall scope/enforceability of remaining IP rights. Due to such uncertainty, generic entry for many drugs, as described in Section 2.3, occurs following patent challenges in which a generic entrant seeks to enter the market prior to the expiration of all secondary patents. The generic firm will submit an application to the FDA specifying that remaining patents are either invalid or that the generic can enter without infringing upon those patents. The uncertainty is then resolved via litigation in the courts, with final FDA approval for the generic occurring only following a court decision in its favor or a settlement agreement between the branded and generic manufacturers.

Suppose the court rules one or more remaining patents are fully enforceable. In this case, the FDA will approve generic entry can occur at the expiration of the final enforceable patent, and we would expect a binding secondary IP for these entrants. As in the case of binding primary IP, we would anticipate such a scenario in more attractive markets, e.g. that treat chronic conditions or generate higher revenues, and where the generic manufacturer is willing to incur litigation expenses for earlier entry. Thus, if one or more secondary IP rights are perfectly enforced, then IP accumulation can extend a branded drug's monopoly term by changing the binding expiration for generic entry—from

the primary IP to a secondary one.

Implication 3 (imperfect model): There will exist a spike in entry delayed to the time of a secondary IP expiration, due to binding secondary IP.

We may also see intermediate entry occurring between IP expirations, i.e. after the expiration of at least one (primary or secondary) IP but before the expiration of all IP. This would occur if the courts rule that all remaining patents are unenforceable. Note that in this case, the delayed or non-binding entry arises due to the need to wait for legal institutions to resolve questions on patent enforceability rather than from the generic firm's active decision to wait until some additional learning occurs (as suggested by Implication 2). When entry does occur, we would expect it to be faster (with respect to the last enforced IP expiration) than in the case of delayed entry under market uncertainty. Here, as in the case under binding expirations, the generic manufacturers expect the market to be profitable. They have completed their bioequivalence studies and submitted ANDAs to the FDA but must now wait for a court decision. As soon as a favorable decision to the generic is made not to enforce remaining patents, they enter the market.

Implication 4 (imperfect model): There will exist intermediate entry occurring between two IP expirations, due to nonenforcement of remaining secondary IP. All else equal, this delayed entry will occur at a faster rate than the delayed entry that occurs with learning about market conditions.

In summary, this conceptual framework suggests that under the standard model, we would expect to see either immediate entry at expiration of the primary patent (Implication 1) or delayed entry under learning about market conditions (Implication 2). Under the imperfect model, two additional sources of delay are possible. We will see entry at the expiration of a binding later secondary patent (Implication 3) or intermediate entry occurring between expirations but at a faster rate than occurs under market learning (Implication 4).

5.2 Empirical evidence

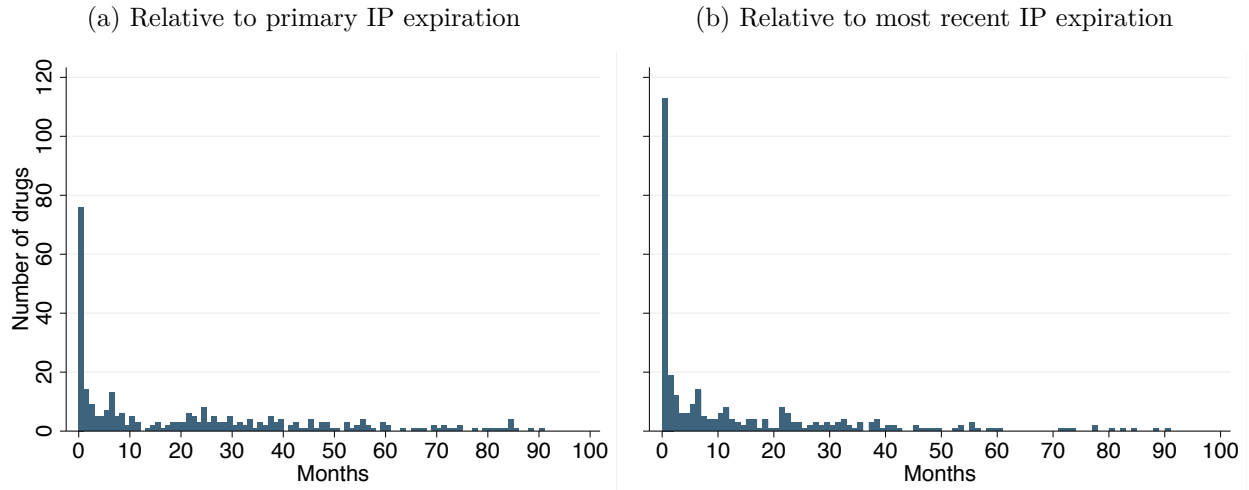
I present a series of empirical evidence consistent with the above conceptual framework. First, I test for Implications 1 and 3. I explore the extent to which IP rights—both primary and secondary—are binding in blocking generic entry. Figure 3 displays the timing of generic entry in months for each drug relative to its IP expirations.[17] Panel (a) gives generic entry relative to the primary IP expiration, i.e. a drug's *extra* monopoly life relative to the standard model. The spike at 0 months confirms Implication 1: A drug's primary IP is binding for 74 out of 370 drugs, or 20 percent of

[17] For visual purposes, I limit the display of this and subsequent figures to those drugs with first generic entry timing within 100 months following the relevant IP expiration. Appendix A.3 gives the same figures without limiting the display, i.e. including the few drugs with generic entry occurring after the 100-month windows. Any sample sizes and proportions given include drugs with entry outside these windows.

the sample.[18] But looking at a drug's primary IP does not give a complete picture. Panel (b) thus considers generic entry relative to the branded drug's most recent (primary or secondary) IP expiration prior to generic entry. This panel confirms Implication 3: With secondary IP included, the spike at month 0 rises dramatically, with another 37 drugs, or 10 percent of the sample, experiencing generic entry at exactly when a secondary IP expires. Thus, binding secondary IP rights are important, accounting for half as much binding entry as primary IP.

Two questions emerge from this figure: First, which secondary IP matters? Is it the first secondary IP that is binding, or do later secondary IP similarly block entry? Second, we see that for many drugs, IP expirations are non-binding—for these drugs, the first generic enters either after the drug's final IP expiration or between expirations, as indicated by Implications 2 and 4. How do we separate uncertainty about market conditions versus uncertainty about IP enforceability as the sources of non-binding entry? I address each of these questions in turn.

Figure 3: Timing of generic entry (in months)



Restricted to NCEs with generic entry at or after primary IP expiration, i.e. 329 drugs out of total sample of 370. Panel (a) gives timing of generic entry relative to primary expiration while panel (b) gives timing relative to most recent (primary or secondary) IP expiration prior to generic entry.

To answer the first question as to which secondary IP is binding, I estimate a linear probability model of first generic entry and include as explanatory variables indicators for the month of primary and secondary IP expirations and the periods between those expirations. I use the drug-month-level dataset tracking drugs from branded launch through initial generic entry. There are 370 drugs observed for 159 months on average, for 58,898 drug-month observations total. I estimate the

[18] Note that a small portion of drugs (11 percent) see generic entry prior to primary IP expiration. A manual review of these drugs suggests these entrants are authorized generics, marketed by either a separate company that has received permission by the branded firm to enter early in exchange for a portion of its profits or by the branded manufacturer itself. (The branded manufacturer may choose to launch its own generic to preempt future outside generic competition.) I exclude these drugs from Figure 3.

following regression:

$$\mathbb{1}\{Entry\}_{it} = X_{it}\beta + \delta_i + \varepsilon_{it} \tag{1}$$

The outcome of interest $\mathbb{1}\{Entry\}_{it}$ is a binary variable indicating whether or not first generic entry for drug i occurs in month t following branded launch. X_{it} is a vector of indicators for month of primary IP expiration, time between primary and first secondary IP expirations, month of first secondary IP expiration, time between first and second secondary IP expirations, and so forth. For fourth and subsequent secondary IP, I collapse the periods during which they are in effect into a single indicator and their expirations into another. A final indicator specifies when all IP has expired. Note that as multiple patents or exclusivities may expire in the same month (or even on the same day), I form these indicators based not on individual patents and exclusivities but rather on periods of IP protection. That is, multiple secondary patents may expire in the "month of first secondary IP expiration." With the inclusion of these indicators, the excluded reference period is the time after branded launch and before the primary IP expires. I include drug fixed effects and robust standard errors clustered at the drug level.

Figure 4: Linear probability model of first generic entry

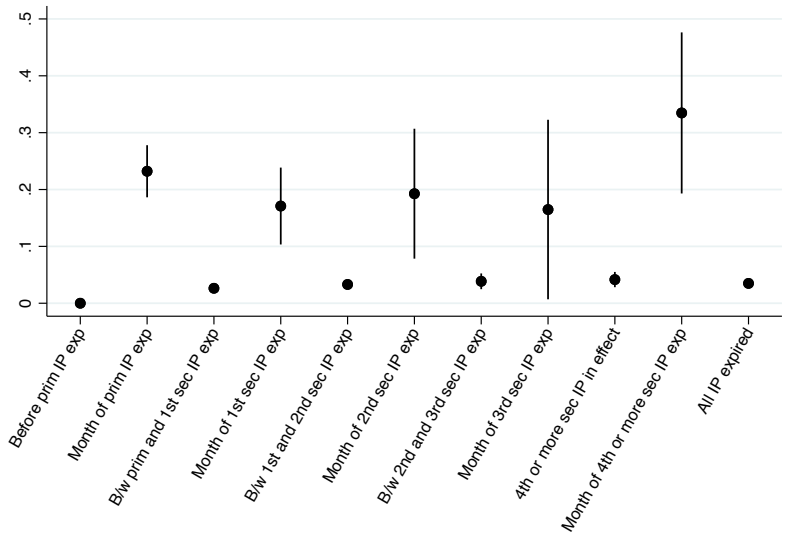


Figure plots β coefficients from linear probability model; all are significant at the 1-percent level except coefficient on indicator for "month of third secondary IP expiration," which is significant at the 5-percent level. Reference period is time before the primary IP expires.

Figure 4 plots the coefficients of the linear probability model. Generic entry is 23.2 percentage points higher in the month of a drug's primary IP expiration than in the months prior. After primary IP expiry, the relative probability of generic entry then drops dramatically (to merely 2.6 percentage points higher) until the first secondary IP expiry, at which point it rises again (to 17.1 percentage points higher). This non-monotonic pattern continues with subsequent secondary IP, suggesting the binding nature of these IP in delaying generic entry.

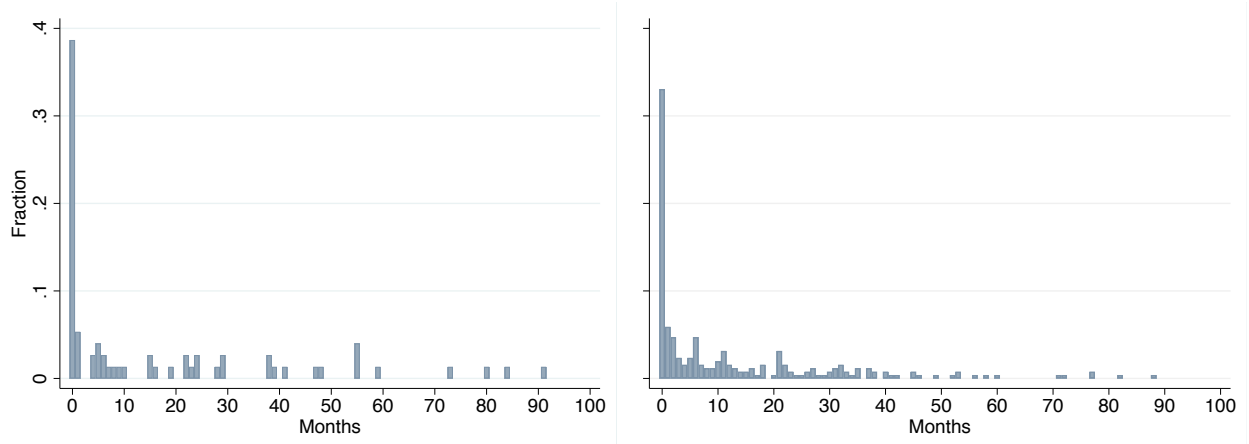
A key takeaway from this figure is that not only does the first secondary IP matter but later secondary IP also introduce binding entry points. It is also important to note that this figure plots relative entry in a *given* month. As the periods between two IP expirations for a drug may last for many months, when we consider total entry, there is more non-binding entry (occurring between or after all expirations) than binding entry (occurring exactly when an IP expires). Appendix Figure 10 shows the raw proportions of cumulative entry occurring in each period of IP protection. I next turn to answering the second question of disentangling the sources of non-binding entry.

To separately test for Implications 2 and 4, I exploit variation across drugs in their number of secondary IP. Figure 5 repeats the analyses of Figure 3 on timing of generic entry but separated by whether or not a drug has any secondary IP. For drugs without secondary IP, entry timing is determined by either binding primary IP (Implication 1) or uncertainty in market profitability (Implication 2). Panel (a) of Figure 5 shows the generic entry timing for these drugs relative to primary IP expiration. We see, as would be expected, that primary IP is binding for a much larger portion of these drugs (39 percent) relative to the entire sample (20 percent in Figure 3). Nonetheless, many drugs in this sample still have delayed entry (despite no secondary IP to forestall it), meaning that for these drugs it is the uncertainty in market profitability that matters, as suggested by Implication 2.

Figure 5: Timing of generic entry (in months)

(a) Drugs without secondary IP; relative to primary IP expiration

(b) Drugs with secondary IP; relative to most recent IP expiration



Restricted to NCEs with generic entry at or after primary IP expiration. Panel (a) gives timing of generic entry relative to primary expiration for drugs without secondary IP; number of drugs is 75. Panel (b) gives timing relative to most recent (primary or secondary) IP expiration prior to generic entry for drugs with secondary IP; number of drugs is 254.

Panel (b) of Figure 5 restricts to drugs with secondary IP and depicts generic entry timing relative to the branded drug's most recent (primary or secondary) IP expiration prior to entry. For this set of drugs, intellectual property protection is binding in blocking generic entry for 33.1 percent (or

84 out of 254 drugs). Separating this binding IP expiry by type, for 18.5 percent the primary IP is binding and for the other 14.6 percent, secondary IP is binding. For the remaining drugs with non-binding or delayed entry, two sources of uncertainty are possible, in either market profitability or IP scope/enforceability.

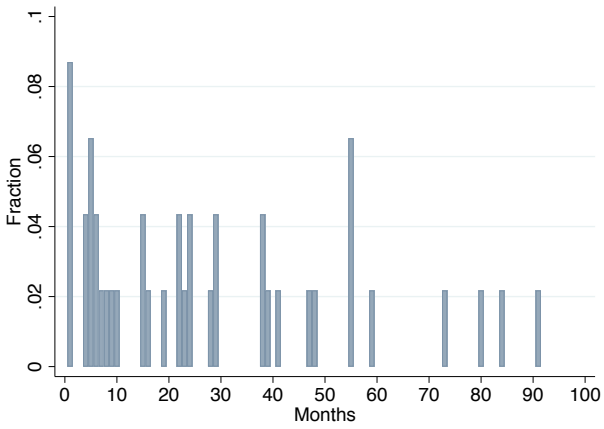
To disentangle these sources of uncertainty, I narrow in on exactly the set of drugs with non-binding generic entry; Figure 6 presents their generic entry timing. Panel (a) depicts drugs without secondary IP. We know that delay for these drugs must reflect learning in market conditions (Implication 2). For drugs with secondary IP, I further divide the sample into those where the generic enters after all secondary IP have expired and those with intermediate entry between IP expirations. Panel (b) looks at this first group. For these drugs, IP enforceability is no longer an issue and market profitability is the only source of uncertainty. We would expect these drugs to have similar entry patterns as those in panel (a). Indeed, the entry patterns look visually similar. Out of the drugs in panel (a) without secondary IP and delayed entry, 30 percent enter by 10 months after their primary IP expiration, 52 percent by 25 months, and 72 percent by 50 months. For the drugs in panel (b) with secondary IP and delayed entry after all secondary IP have expired, the comparable statistics are similar: 29 percent by 10 months post-final secondary IP expiration, 46 percent by 25 months, and 68 percent by 50 months. To further assess their similarity, I apply the non-parametric Kolmogorov-Smirnov test of the equality of two distributions; with a p-value of 0.897, the test fails to reject the null hypothesis that these two entry distributions are the same.

Panel (c) looks at entry timing for generics with secondary IP that enter while at least one secondary IP still remains on a drug; entry is graphed relative to the most recent (primary or secondary) IP expiration. By Implication 4, uncertainty in IP scope/enforceability is an important contributor, with intermediate entry occurring once the nonenforcement of remaining IP is determined. Consider the counterargument: There is no uncertainty about which secondary patents are enforced (i.e., there is certain unenforcement of these patents) and the only source of uncertainty is in market profitability. If this were the case, we would expect the entry timing distribution for this group to match those in panels (a) and (b). But this is not what we see—the entry distribution looks remarkably different. As suggested by Implication 4, entry is faster: 40 percent enter by 10 months post-expiration; 70 percent by 25 months; and 92 percent by 50 months. This makes sense if we consider the different types of drugs in these samples. Drugs that have delayed entry after all IP expired are likely to be drugs less attractive to generic entry, e.g., that may be in lower revenue markets or at risk of becoming obsolete pending results of clinical trials on other drugs in development. Rather than a situation where the generic firm is waiting to enter as soon as the binding IP expires, entry occurs only once a positive signal has been received on market profitability—e.g., a medical study that suggests a new use for an older off-patent drug.

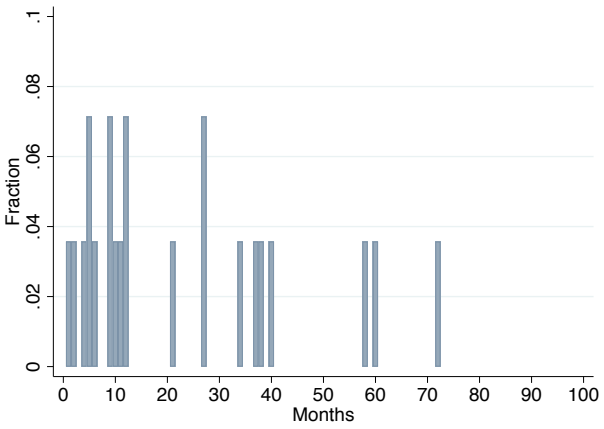
In contrast, a generic with intermediate entry between IP expirations is likely a more attractive drug, e.g., in a higher revenue market, and the generic must wait for a determination on the

Figure 6: Timing of *delayed* generic entry (in months)

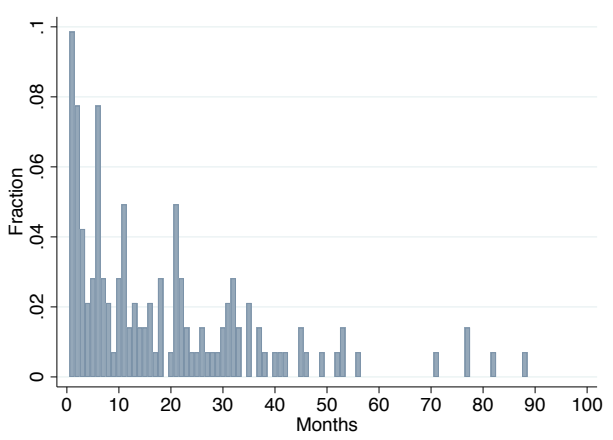
(a) Drugs with no secondary IP; relative to primary IP expiration



(b) Drugs with secondary IP & all have expired; relative to final IP expiration



(c) Drugs with secondary IP & at least one still unexpired; relative to most recent IP expiration



Restricted to NCEs with non-binding IP expirations, i.e., delayed generic entry. Panel (a) gives timing of delayed generic entry relative to primary expiration for drugs without secondary IP; number of drugs is 46. Panel (b) gives delayed timing relative to final IP expiration for drugs with secondary IP and all secondary IP expired; number of drugs is 28. Panel (c) gives delayed timing relative to most recent (primary or secondary) IP expiration prior to generic entry for drugs with at least one secondary IP still unexpired; number of drugs is 142.

remaining secondary IP scope and enforceability. Under such a situation, we would expect faster entry for these drugs, occurring as soon as nonenforcement of remaining IP is established. To confirm, I again apply the Kolmogorov-Smirnov test to assess the equality of the distribution in panel (c) with those in panels (a) and (b). These tests reject the hypotheses that the generic entry distribution for drugs with secondary IP still remaining is the same as that of drugs with no secondary IP (with a p-value of 0.001) or that of drugs with all secondary IP expired (p-value of 0.009). I also conduct a Cox regression of entry timing in which I compare the hazard or probability of generic entry for drugs with secondary IP still unexpired to those with all IP expired.[19] The regression indicates that the probability a drug sees its first generic entrant in any given month after its final *enforced* IP expiration (conditional on not having had a generic entrant prior) is twice as likely for drugs with at least one IP still unexpired than for those with all IP expired.[20]

Taken together, this evidence can be summarized as follows: Secondary IP are important for influencing first generic entry timing and they work in two ways. First, secondary IP rights introduce later binding expirations—secondary IP account for half as many binding expirations (10 percent) as primary IP (20 percent). Second, they also influence generic entry by increasing uncertainty in the scope or enforceability of the remaining IP on a drug. Nearly 40 percent of drugs in the sample see their first generic enter between two IP expirations. Comparing the distribution of generic entry timing for these drugs relative to those whose first generic enters after all IP have expired suggests that market profitability alone is not driving delayed entry for these drugs. Rather, uncertainty in IP scope/enforceability must also be a factor.[21]

6 Causal impacts to monopoly term

6.1 Naive analysis: Comparing drugs by number of secondary patents

Estimating a causal impact of secondary IP on monopoly life is difficult due to endogeneity concerns stemming from both the branded firm's decision to patent and the generic firm's decision to enter a market. The naive analysis would simply compare drugs with secondary IP to those without

[19] Specifically, I run the following regression: $h_{it} = h_0(t)\exp\big(\beta \mathbb{1}\{IP\ unexpired\}_{it}\big)$, where the hazard rate h_{it} is the conditional probability a drug sees its first generic entrant in month $t+\delta$ post-final enforced IP expiration given that it has not had a generic entrant through month t; $h_0(t)$ gives the baseline hazard, i.e. the probability of generic entry for any drug t months post-final enforced IP expiration; and $\mathbb{1}\{IP\ unexpired\}_{it}$ is an indicator for having at least one secondary IP still unexpired. I include robust standard errors clustered at the drug level. The sample includes 170 drugs and 4,480 drug-month observations.

[20] The unexponentiated coefficient on $\mathbb{1}\{IP\ unexpired\}_{it}$ is 0.686, significant at the 1-percent level. The exponentiated coefficient, or hazard ratio, is 1.985–indicating a hazard (probability of generic entry) twice as large for drugs with secondary IP still unexpired as those with all IP expired.

[21] Considering the results of the Cox regression—which shows a hazard twice as large for this group as for those with all IP expired—if we suppose that certain unenforcement is the case for one-third of this group (and hence uncertainty in market profitability the contributing factor behind their delayed entry), that would suggest for $\frac{2}{3} \times 40 = 27$ percent of the sample, uncertainty in IP scope/enforceability is the limiting factor.

and observe resulting monopoly terms. If intellectual property rights were as good as randomly assigned, this would be enough to estimate a causal effect of secondary IP. Two main endogeneity concerns make such an analysis problematic. First, there is the potential for omitted variable bias: Characteristics of a drug that are unobservable to the researcher may influence both the branded firm's decision to patent and the generic firm's decision to enter (and hence, the monopoly life of a drug). Stated differently, there is selection into patenting in that drugs with secondary IP are systematically different from those without. Second, there is a reverse causality concern: The branded firm may anticipate the end of its monopoly life and as a result, attempt to obtain more intellectual property to forestall entry.

Intuitively, branded firms want to protect their most valuable or promising drugs with additional IP, and these are exactly the drugs that generics want to copy (which would shorten a drug's effective monopoly life). By not accounting for this selection into patenting or unobserved attractiveness to generic entry, then the estimated coefficient on a simple regression of monopoly life on number of secondary IP would be too small (a negative bias). That is, such a regression would underestimate the causal effect of secondary IP. I present the naive analysis in Table 2. For comparison with subsequent analyses, I focus specifically on secondary patents.[22]

I regress a drug's *extra* monopoly life in months (meaning time between primary IP expiry and generic entry) on its number of secondary Orange Book patents obtained prior to generic entry; I restrict to Orange Book patents as these are the ones which generic firms must account for when they seek to enter a market. Column (1) shows this basic regression; the coefficient of interest is insignificant. A first solution to improve the naive analysis would be to add controls to proxy for a drug's unobserved characteristics. Column (2) adds fixed effects for branded drug approval year, Column (3) adds fixed effects for therapeutic area, and Column (4) adds a control for market size. Scott Morton (1999) shows that therapeutic area and market size are important for generic entry decisions; their inclusion helps control for attractiveness to a generic firm.[23] Moving left to right across the table we see that the addition of these controls does help explain the total variation in extra monopoly life and that their inclusion further raises the coefficient on the number of secondary patents. Finally, as expected, drugs with larger markets have shorter extra monopoly terms and controlling for this raises the estimated coefficient on number of secondary patents. Looking at the specification in Column (4), with both sets of fixed effects and the control for market size, suggests that each secondary Orange Book patent is associated with an additional 1.8 months of monopoly life.

[22] Appendix Table 7 gives the naive analysis using number of all secondary IP (both Orange Book patents and FDA exclusivities) as the main explanatory variable.

[23] Requiring drugs with non-missing market size reduces the sample size because the MEPS survey data do not capture (1) drugs with low market sizes and (2) drugs that are received in hospital settings, e.g., chemotherapy drugs.

Table 2: Secondary patents and extra monopoly life: Drug-level analysis

	(1)	(2)	(3)	(4)
# secondary patents	0.327	0.890	1.328	1.823**
	(0.857)	(0.894)	(0.871)	(0.874)
ln(market size)				-3.813**
				(1.568)
Observations	288	288	288	288
Adj. R^2	-0.003	0.127	0.176	0.202
Approval year FE		yes	yes	yes
Therapeutic area FE			yes	yes

Dependent variable in all regressions is extra monopoly life in months (i.e., from a drug's primary IP expiry to generic entry). Analysis at the drug level. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, and * p<0.1.

6.2 Comparing granted versus rejected secondary patent applications

Even with the inclusion of these controls, it is plausible that observed differences in monopoly life across drugs may still reflect selection of which drugs are patented rather than a causal effect of secondary patents on delaying generic entry. Thus, a next step would be to compare monopoly terms for drugs in accepted versus rejected secondary patent applications. If conditional on being included in an application, whether a patent is granted is as good as random, then such an analysis would imply a causal effect. I match drugs to their patent applications and carry out a drug-application-level analysis.[24] The dependent variable is again a drug's extra monopoly life and the main explanatory variable is an indicator for whether or not each secondary patent application for that drug is granted. Because the treatment (whether a patent is granted or not) is at the application level, I follow Abadie et al. (2017) and use robust standard errors. Appendix A.6 repeats the analyses that follow with standard errors clustered at the drug level; the coefficients of interest remain significant.

Table 3 presents the OLS results of the application-level analysis, with Columns (1) through (4) adding the same controls as before. The increase in the coefficient on $\mathbb{1}_{Granted}$ relative to the estimates of the naive analysis confirms a selection into patent applications. Moving across specifications from left to right, the increase in the adjusted R-squared indicates that these controls soak up variation in the model, but the stability of the coefficient on $\mathbb{1}_{Granted}$ suggests that they not correlated with whether or not a given patent application is granted, conditional on an application being submitted to the USPTO. The results of Table 3 suggest that each secondary Orange Book patent application granted is associated with 5.5 months of additional monopoly life.

[24] Appendix A.2 describes this matching process in detail.

Table 3: Secondary patents and extra monopoly life for drugs in granted versus rejected applications: Application-level analysis

	(1)	(2)	(3)	(4)
$\mathbb{1}_{Granted}$	4.080**	4.843***	5.617***	5.525***
	(1.731)	(1.571)	(1.455)	(1.450)
ln(market size)				-1.107***
				(0.342)
Observations	3,723	3,723	3,723	3,723
Adj. R^2	0.050	0.246	0.388	0.390
Approval year FE		yes	yes	yes
Therapeutic area FE			yes	yes

Dependent variable in all regressions is extra monopoly life in months (i.e., from a drug's primary IP expiry to generic entry). Analysis at the drug-application level; number of drugs is 248. All regressions include patent application/examiner art unit fixed effects. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, and * p<0.1.

6.3 Exploiting variation in examiner leniency

Comparing granted and rejected patent applications above controlled for selection into obtaining a patent *application*. It is plausible that there still remains a selection issue into whether or not a patent is *granted*. This may be the case, e.g., if branded firms devote more resources to revising to a secondary patent application for certain drugs that are more valuable or more at risk of earlier generic entry. Assuming that conditional on being included in an application, whether a drug is granted a secondary patent is *not* as good as random, the natural solution is to find an instrument for which patent applications are granted patents. I follow previous research and use the "leniency" of the assigned patent examiner as an instrument (Farre-Mensa et al., 2020; Sampat and Williams, 2019). Prior work confirms that patent examiners have considerable discretion in whether or not to grant a patent (Cockburn et al., 2003; Lemley and Sampat, 2012; Sampat and Lemley, 2010) and that the assignment of patent applications is random conditional on application year and technology area (Sampat and Williams, 2019).[25] This type of analysis thus exploits both the variation in cross-examiner leniency and the quasi-random assignment of patent applications to examiners.

I construct a measure of examiner leniency—the share of applications granted by an examiner—using a separate first-stage sample of applications within the NBER drugs and organic compounds subcategories that are *not* attached to the NCEs in my analytic sample.[26] Three criteria must be satisfied for a valid instrument: the relevance condition, exclusion restriction, and monotonicity

[25] All regressions control for technology area with the inclusion of patent application/examiner art unit fixed effects.

[26] Appendix A.7 repeats the subsequent analyses using a leave-out examiner leniency as the instrument instead. Magnitudes on the coefficients of interest remain similar.

assumption. The relevance condition requires that this measure of examiner leniency is relevant to explaining whether a patent application is granted. Table 4 gives first-stage results regressing whether a patent application is granted on the leniency measure; the table demonstrates the relevance condition is satisfied. Across specifications, the 0.67 point estimate suggests that a 10-percentage-point increase in an examiner's average grant rate is associated with a 6.7-percentage-point increase in the likelihood a secondary patent application is granted. The F-statistic of these regressions is roughly 560, substantially above the threshold for weak instruments. The exclusion restriction requires that examiner leniency only affects extra monopoly life through the likelihood a drug's secondary patent application is granted, which is a reasonable assumption. Finally, the monotonicity assumption requires that patents granted by a stricter examiner would also be granted by a more lenient examiner, and vice versa.

Table 4: First-stage results

	(1)	(2)	(3)	(4)
leniency	0.673***	0.672***	0.667***	0.668***
	(0.0289)	(0.0285)	(0.0283)	(0.0283)
ln(market size)				-0.00530
				(0.00380)
Observations	3,370	3,370	3,370	3,370
Approval year FE		yes	yes	yes
Therapeutic area FE			yes	yes
F-statistic	564.1	569.6	558.1	558.2

Dependent variable in all regressions is an indicator for whether a patent application is granted. Analysis at the drug-application level; number of drugs is 244. All regressions include patent application/examiner art unit fixed effects. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, and * p<0.1.

Table 5 presents the IV results. The estimates in all four specifications are substantially larger than the OLS estimates, suggesting that, as expected, selection into whether a patent is granted indeed biases the OLS estimates downward. The preferred specification in Column (4) indicates that a randomly granted secondary patent extends monopoly life by 18.3 months. Given that the average drug has 2.7 secondary patents obtained prior to generic entry, this would suggest over 49 months, or more than 4 years, of extra monopoly life attributable to secondary patent acquisition. It is important to note that the IV estimate is not giving the average treatment effect but rather the local average treatment effect (LATE)—the effect of an additional patent granted on extra monopoly life for those drugs where changing examiner leniency would change whether the patent application was granted. It does not apply to those applications that would always be granted irrespective of examiner leniency or never granted. That is, we can consider the 18.3 months as an average treatment effect specifically for those patents that are granted by a more lenient examiner but not by a stricter one. If we think that such patents are less likely to be enforced later by the courts (and hence less likely to contribute to delays in generic entry), then the LATE will

underestimate the average effect of a secondary patent and this per-drug extension in monopoly life of 49 months will be a conservative estimate of the extent to which secondary patents delay generic entry.

Table 5: Secondary patents and extra monopoly life using variation in examiner leniency: IV analysis

	(1)	(2)	(3)	(4)
$\mathbb{1}_{Granted}$	14.155***	13.817***	18.266***	18.273***
	(4.889)	(4.282)	(3.935)	(3.937)
ln(market size)				-1.191***
				(0.341)
Observations	3,370	3,370	3,370	3,370
Approval year FE		yes	yes	yes
Therapeutic area FE			yes	yes

Dependent variable in all regressions is extra monopoly life in months (i.e., from a drug's primary IP expiry to generic entry). Examiner leniency instruments for whether a secondary patent application is granted or not. Analysis at the drug-application level; number of drugs is 244. All regressions include patent application/examiner art unit fixed effects. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, and * p<0.1.

7 Conclusion

Despite the standard "one-patent-one-product" model in intellectual property rights discourse, a common practice by innovators is to acquire numerous patents on different features of the same product to extend its monopoly life. In this paper, I systematically document the presence of imperfect intellectual property rights in the pharmaceutical industry. Multiple patents may cover a single product, with the scope and/or enforceability of each patent potentially uncertain, contributing to indeterminate periods of monopoly protection. In a sample of novel drugs, I find that 70 percent of drugs are covered by multiple intellectual property rights and that actual monopoly terms often last well beyond the fixed term of a drug's initial molecule patent. I present a simple conceptual framework to explain the mechanisms by which the accumulation of intellectual property rights on a single drug may delay generic entry: by introducing a binding later IP expiration and by increasing uncertainty in the scope and enforceability of remaining IP rights. Consistent with this framework, I provide empirical evidence of non-monotonic generic entry patterns relative to IP expirations. I also provide, to my knowledge, the first evidence in any industry on the extent to which an incremental patent extends monopoly life. By exploiting the quasi-random assignment of patent applications to examiners and variation in cross-examiner leniency, I determine that the average drug receives more than 4 years of additional monopoly life due to the accumulation of secondary Orange Book patents, equivalent to over 30 percent of mean monopoly term.

The paper's analysis demonstrates that the strategic use of patents by innovating firms can substantially increase entry barriers into a market and lengthen monopoly durations.[27] These results have important implications for consumer welfare, innovation incentives, and the design of intellectual property rights both broadly across sectors and within the pharmaceutical industry in particular. A lot of attention has been paid recently to rising US prescription drug prices (Hernandez et al., 2019; Talmadge, 2020). In a back-of-the-envelope calculation outlined in Appendix A.8.1, I determine that the 49-month delay in generic entry leads to an average consumer welfare loss of $275 million per drug due to monopoly pricing.[28] Of course, the consumer welfare loss estimated here is a transfer from consumers to producers under monopoly pricing, and the potential for monopoly extension influences firms' ex-ante incentives to innovate. Thus, I cannot make a statement on the overall dynamic welfare impacts to consumers.

While I leave a complete welfare analysis to future research, existing evidence indicates that the patent-term extension I observe here should have had a sizable impact on pharmaceutical innovation. In a second back-of-the-envelope calculation detailed in Appendix A.8.2, I determine that the 49-month extension in monopoly life should have increased pharmaceutical R&D investment (as measured by clinical trials) from 28 to 94 percent relative to a counterfactual where innovating firms are restricted to solely the fixed term of their initial molecule patent. This increase in R&D investment translates to an increase in the number of approved NCEs of between 2 and 7 percent (i.e., between 7 and 24 new drugs) during my sample period. Although the counterfactual is unobservable, the imperfections in our patent system mean that firms may be able to counteract distortions of a fixed-term system by accumulating patents on a single product and this can have large implications for innovation incentives. As such, my work offers an important nuance that intellectual property rights are less rigid than we typically assume and that these imperfections matter for discussions on optimal patent policy and innovation.

In particular, some observers argue that the ability of branded firms to make minor tweaks to existing drugs in dosage or formulation disproportionately extends their effective patent terms relative to the R&D costs of these improvements.[29] Others point out that, given the high costs of drug development, these extensions are necessary to incentivize a drug's initial development. One question that naturally arises from this paper is whether the imperfections in our intellectual property system lead to resulting monopoly terms that more closely match those of a first-best system, where terms are differentiated according to innovation size or measures of quality. This question becomes more salient if we consider all drugs approved by the FDA beyond the truly new molecules studied in this paper. I leave a full exploration of this as an avenue for future research.

[27] In recent work, Kyle et al. (2021) note that primary drug patents are the key patents impacting generic competition in India. My results are complementary to theirs. I show that in addition to primary patents, secondary Orange Book patents play a significant role in delaying competition in the US drug market.

[28] Note that the term "consumer" refers to all "downstream" entities, i.e., patients, drug retailers, and insurance companies.

[29] This relates to the antitrust issue of "product hopping" whereby a manufacturer will cease production of prior dosages or formulations and switch consumers to the newer one at higher monopoly prices.

References

Abadie, A., S. Athey, G. Imbens, and J. Wooldridge (2017). When should you adjust standard errors for clustering? NBER Working Paper No. 24003.

Abrams, D. S., U. Akcigit, and J. Grennan (2018). Patent value and citations: Creative destruction or strategic disruption? NBER Working Paper No. 19647.

Adams, C. P. and V. V. Brantner (2006). Estimating the cost of new drug development: Is it really $802 million? *Health Affairs* 25(2)**:** 420–428.

Amin, T. and A. S. Kesselheim (2012). Secondary patenting of branded pharmaceuticals: A case study of how patents on two HIV drugs could be extended for decades. *Health Affairs* 31(10)**:** 2286–2294.

Angell, M. (2005). *The Truth About the Drug Companies: How They Deceive Us and What to Do About It*. New York: Random House.

Arnold, D., W. Dobbie, and C. S. Yang (2018). Racial bias in bail decisions. *Quarterly Journal of Economics* 133(4)**:** 1885–1932.

Berndt, E. R. and M. L. Aitken (2011). Brand loyalty, generic entry and price competition in pharmaceuticals in the quarter century after the 1984 Waxman-Hatch legislation. *International Journal of the Economics of Business* 18(2)**:** 177–201.

Berndt, E. R., R. M. Conti, and S. J. Murphy (2017). The landscape of US generic prescription drug markets, 2004-2016. NBER Working Paper No. 23640.

Bessen, J. E. (2009). Imperfect property rights. Boston University School of Law Working Paper No. 09-46.

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

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

Cao, B., Y. Wang, D. Wen, W. Liu, J. Wang, G. Fan, L. Ruan, B. Song, Y. Cai, M. Wei, X. Li, J. Xia, N. Chen, J. Xiang, T. Yu, T. Bai, X. Xie, L. Zhang, C. Li, Y. Yuan, H. Chen, H. Li, H. Huang, S. Tu, F. Gong, Y. Liu, Y. Wei, C. Dong, F. Zhou, X. Gu, J. Xu, Z. Liu, Y. Zhang, H. Li, L. Shang, K. Wang, K. Li, X. Zhou, X. Dong, Z. Qu, S. Lu, X. Hu, S. Ruan, S. Luo, J. Wu, L. Peng, F. Cheng, L. Pan, J. Zou, C. Jia, J. Wang, X. Liu, S. Wang, X. Wu, Q. Ge, J. He, H. Zhan, F. Qiu, L. Guo, C. Huang, T. Jaki, F. G. Hayden, P. W. Horby, D. Zhang, and C. Wang (2020). A trial of lopinavir-ritonavir in adults hospitalized with severe Covid-19. *New England Journal of Medicine* 382(19)**:** 1787–1799. PMID: 32187464.

Chong, C. R. and D. J. Sullivan (2007). New uses for old drugs. *Nature* 448(7154)**:** 645–646.

Coase, R. (1960). The problem of social cost. *Journal of Law and Economics* 3**:** 1–44.

Cockburn, I. M., S. Kortum, and S. Stern (2003). Are all patent examiners equal? Examiners, patent characteristics, and litigation outcomes. In W. M. Cohen and S. A. Merrill (Eds.), *Patents in the Knowledge-Based Economy*, pp. 19–53. National Academies Press.

Cockburn, I. M. and M. J. MacGarvie (2011). Entry and patenting in the software industry. *Management Science* 57(5)**:** 915–933.

Cohen, W. M., R. R. Nelson, and J. P. Walsh (2000). Protecting their intellectual assets: Appropriability conditions and why U.S. manufacturing firms patent (or not). NBER Working Paper No. 7552.

Denicolo, V. (1996). Patent races and optimal patent breadth and length. *Journal of Industrial Economics***:** 249–265.

Eisenberg, R. S. (2001). The shifting functional balance of patents and drug regulation. *Health Affairs* 20(5)**:** 119–135.

Farre-Mensa, J., D. Hegde, and A. Ljungqvist (2020). What is a patent worth? Evidence from the U.S. patent "lottery". *Journal of Finance* 75(2)**:** 639–682.

FDA (1999). Guidance for industry: Applications covered by Section 505(b)(2). Center for Drug Evaluation Research, U.S. Department of Health and Human Services. `https://www.fda.gov/downloads/Drugs/Guidances/ucm079345.pdf`. Accessed Jan. 19, 2019.

FDA (2018). Frequently asked questions on patents and exclusivity. https://www.fda.gov/drugs/developmentapprovalprocess/ucm079031.htm. Accessed Mar. 17, 2019.

Fowler, A. C. (2019). Hurry up or wait?: Strategic delay in the introduction of pharmaceutical line extensions. Working paper.

Frakes, M. D. and M. F. Wasserman (2020). Investing in ex ante regulation: Evidence from pharmaceutical patent examination. NBER Working Paper No. 27579.

Furrow, M. E. (2008). Pharmaceutical patent life-cycle management after KSR v. Teleflex. *Food & Drug Law Journal* 63**:** 275.

Gaessler, F. and S. Wagner (2019). Patents, data exclusivity, and the development of new drugs. CRC TRR 190 Discussion Paper No. 176.

Galasso, A. and M. Schankerman (2010). Patent thickets, courts, and the market for innovation. *RAND Journal of Economics* 41(3)**:** 472–503.

Gans, J. S., D. H. Hsu, and S. Stern (2008). The impact of uncertain intellectual property rights on the market for ideas: Evidence from patent grant delays. *Management Science* 54(5)**:** 982–997.

Gaudry, K. S. (2011). Evergreening: A common practice to protect new drugs. *Nature Biotechnology* 29(10)**:** 876–878.

Gilbert, R. and C. Shapiro (1990). Optimal patent length and breadth. *RAND Journal of Economics***:** 106–112.

Gilchrist, D. S. (2016). Patents as a spur to subsequent innovation? Evidence from pharmaceuticals. *American Economic Journal: Applied Economics* 8(4)**:** 189–221.

Grabowski, H., C. Brain, A. Taub, and R. Guha (2017). Pharmaceutical patent challenges: Company strategies and litigation outcomes. *American Journal of Health Economics* 3(1)**:** 33–59.

Grabowski, H. G. and M. Kyle (2007). Generic competition and market exclusivity periods in pharmaceuticals. *Managerial and Decision Economics* 28(4-5)**:** 491–502.

Hall, B. H. and R. H. Ziedonis (2001). The patent paradox revisited: An empirical study of patenting in the U.S. semiconductor industry, 1979-1995. *RAND Journal of Economics***:** 101–128.

Hegde, D., D. C. Mowery, and S. J. Graham (2009). Pioneering inventors or thicket builders: Which U.S. firms use continuations in patenting? *Management Science* 55(7)**:** 1214–1226.

Hemphill, C. S. and B. N. Sampat (2011). When do generics challenge drug patents? *Journal of Empirical Legal Studies* 8(4)**:** 613–649.

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

Hernandez, I., C. B. Good, D. M. Cutler, W. F. Gellad, N. Parekh, and W. H. Shrank (2019). The contribution of new product entry versus existing product inflation in the rising costs of drugs. *Health Affairs* 38(1)**:** 76–83.

Hernandez, J. J., M. Pryszlak, L. Smith, C. Yanchus, N. Kurji, V. M. Shahani, and S. V. Molinski (2017). Giving drugs a second chance: Overcoming regulatory and financial hurdles in repurposing approved drugs as cancer therapeutics. *Frontiers in Oncology* 7**:** 273.

Higgins, M. J., M. J. Kronlund, J. Park, and J. Pollet (2020). The role of assets in place: Loss of market exclusivity and investment. NBER Working Paper No. 27588.

IQVIA (2019). The global use of medicine in 2019 and outlook to 2023: Forecasts and areas to watch.

Klemperer, P. (1990). How broad should the scope of patent protection be? *RAND Journal of Economics***:** 113–130.

Kling, J. R. (2006). Incarceration length, employment, and earnings. *American Economic Review* 96(3)**:** 863–876.

Kurakina, M. (2020). The dark side of patents: Effects of strategic patenting on firms and their peers. Working paper.

Kyle, M., B. Sampat, and K. Shadlen (2021). TRIPS, patents, and drug prices in India. Working paper.

Lakdawalla, D. N. (2018). Economics of the pharmaceutical industry. *Journal of Economic Literature* 56(2)**:** 397–449.

Lemley, M. A. and B. Sampat (2012). Examiner characteristics and patent office outcomes. *Review of Economics and Statistics* 94(3): 817–827.

Lemley, M. A. and C. Shapiro (2005). Probabilistic patents. *Journal of Economic Perspectives* 19(2): 75–98.

Mas-Colell, A., M. D. Whinston, and J. R. Green (1995). *Microeconomic Theory*. New York.

Mohapatra, D. P. and Y. Zhang (2020). Does market exclusivity improve access to drugs? The case of US anti-ulcer drug market. Working paper.

Morton, F. S. and M. Kyle (2011). Markets for pharmaceutical products. In M. V. Pauly, T. G. McGuire, and P. P. Barros (Eds.), *Handbook of Health Economics*, Volume 2, pp. 763–823. Elsevier.

Noel, M. and M. Schankerman (2013). Strategic patenting and software innovation. *Journal of Industrial Economics* 61(3): 481–520.

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

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

PhRMA (2015). Biopharmaceutical research & development: The process behind new medicines.

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

Righi, C. and T. Simcoe (2020). Patenting inventions or inventing patents? Strategic use of continuations at the USPTO. NBER Working Paper No. 27686.

Sampat, B. and H. L. Williams (2019). How do patents affect follow-on innovation? Evidence from the human genome. *American Economic Review* 109(1): 203–36.

Sampat, B. N. and M. A. Lemley (2010). Examining patent examination. *Stanford Technology Law Review* 2010: 2.

Sampat, B. N. and K. C. Shadlen (2017). Secondary pharmaceutical patenting: A global perspective. *Research Policy* 46(3): 693–707.

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

Scott Morton, F. M. (1999). Entry decisions in the generic pharmaceutical industry. *RAND Journal of Economics* 30(3): 421–440.

Talmadge, S. (2020). Sticker shock in the pharmacy. *New York Times*. `https://www.nytimes.com/2020/09/23/well/live/sticker-shock-in-the-pharmacy.html`. Accessed Oct. 6, 2020.

Thomas, J. R. (2017). Regulatory exclusivity reform in the 115th Congress. Congressional Research Service Report No. R44951.

Tribble, S. J. (2019). Drugmakers play the patent game to ward off competitors. `https://www.nbcnews.com/health/health-news/drugmakers-play-patent-game-ward-competitors-n915911`. Accessed May 25, 2019.

Voet, M. A. (2020). *The Generic Challenge: Understanding Patents, FDA and Pharmaceutical Life-Cycle Management*. BrownWalker Press.

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

Ziedonis, R. H. (2004). Don't fence me in: Fragmented markets for technology and the patent acquisition strategies of firms. *Management Science* 50(6): 804–820.

A Appendix

A.1 FDA regulatory exclusivities

Table 6 provides term lengths for several key exclusivities; see Thomas (2017) for a description of all 16 FDA exclusivities that currently exist.

Table 6: Key FDA exclusivities by type and term length

Exclusivity type	Term length
New chemical entity (NCE)	5 years
Clinical investigation exclusivity (CIE)	3 years
–New dosage form or dosing regimen	
–New administration	
–New strength	
–New indication	
–Switch from Rx to OTC	
Orphan drug exclusivity (ODE)	7 years
Pediatric (PED)	6 months
Qualified infectious disease products (QIDP)	5 years

A.2 Data construction

I describe here the construction of the drug-application-level dataset. Granted drug patents are listed in historical FDA Orange Books. I merge these granted patents to their application, examiner, and assignee (i.e., patent owner/drug manufacturer) information using the USPTO Patent Assignment and Examination Research Datasets. Note that drugs may have additional granted patents beyond what is listed in the Orange Book, however generic manufacturers must only account for Orange Book-listed patents when they attempt entry. Hence, this set of patents is the relevant subset of granted patents to be used in estimating the effect of an incremental patent on monopoly life.

Attaching rejected patent applications to drugs is a more complex process. I begin with the universe of patent applications from the USPTO Historical Patent Data Files and restrict to those applications belonging to either of two NBER technology subcategories: drugs or organic compounds. I merge these applications with their full claims text available from the USPTO Patent Claims Research Dataset. Using the complete list of all active ingredients for the 370 NCEs in my analytic sample, I attach applications to individual drugs by first restricting to those applications with a claim including at least one of the drug's active ingredients. Because an active ingredient may be mentioned by another company who does not manufacturer the drug in question, I obtain the list of all assignees for each NCE's granted patents in the Orange Book and then further restrict appli-

cations for each NCE to the subset with at least one of the same assignees. Because of instances in the data where a patent application for a given drug does not include the active ingredient name, I also include parent and child applications of Orange Book patents.

One limitation of the USPTO data is that rejected patent applications are only available if filed on or after November 29, 2000. Granted patent applications are available before and after this date. One approach to addressing this limitation would be to restrict to patent applications filed on or after November 29, 2000. However, such an approach would produce an overestimate of the effect of an incremental patent on monopoly duration (by excluding from the analysis some granted patents). Thus, for each drug, I include all granted secondary patents in the Orange Book regardless of filing date and rejected secondary patent applications on or after November 29, 2000, provided they were filed in advance of generic entry. I rely on an assumption that the unobserved rejected patent applications are "missing at random" and have characteristics similar to the observed rejected applications. Restricting to those drugs with non-missing market size and at least one patent application, the resulting drug-application-level dataset has 248 drugs and 3,723 applications.

To instrument for whether a given patent application is granted or not, I construct a time-invariant measure of examiner leniency using a separate first-stage sample. The first-stage sample includes all applications from the NBER subcategories for drugs or organic compounds that are NOT attached to any NCE in my sample and for which examiner information is available. The measure of examiner leniency used in the main text is the share of applications granted over time by an examiner in this first-stage sample. In the analytic sample, I then exclude applications where the examiner is missing (leading to 247 drugs/3,631 applications) and following Sampat and Williams (2019), applications where the examiner reviews less than ten applications in the first-stage sample (leading to 246 drugs/3,564 applications). I also exclude applications where the examiner is only present in the USPTO data in years up to 2000, as these examiners will be attached solely to granted patent applications and will artificially have leniency measures of 1 (leading to 244 drugs/3,372 applications). Excluding two applications that are singleton observations within a given examiner art unit (as these are dropped with the inclusion of art unit fixed effects) results in a final sample for the IV analysis of 244 drugs and 3,370 applications.

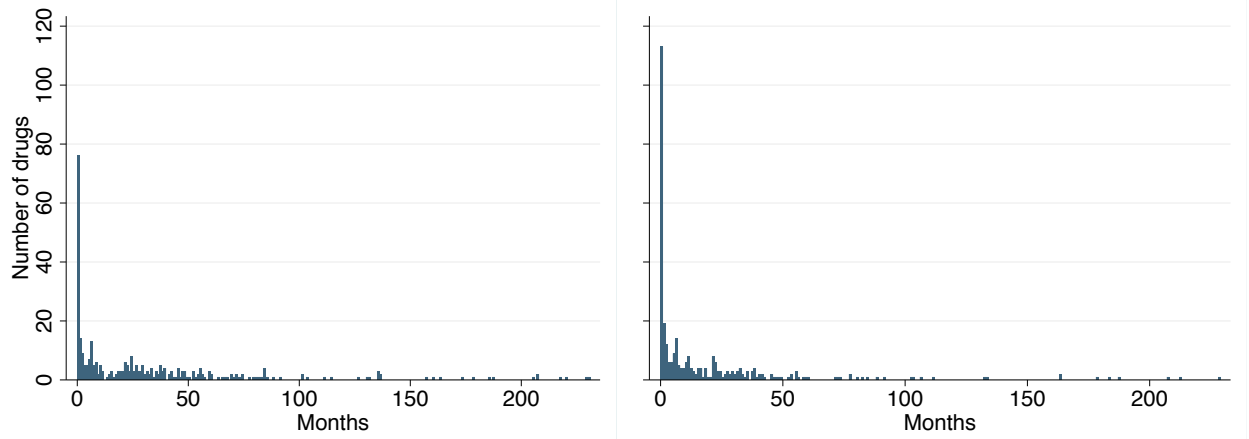
A.3 Complete distributions of generic entry timing

Figures 7 – 9 display the full windows of generic entry timing for Figures 3 – 6, i.e., including drugs with generic entry occurring after 100 months since the relevant IP expiration.

Figure 7: Timing of generic entry (in months)

(a) Relative to primary IP expiration

(b) Relative to most recent IP expiration

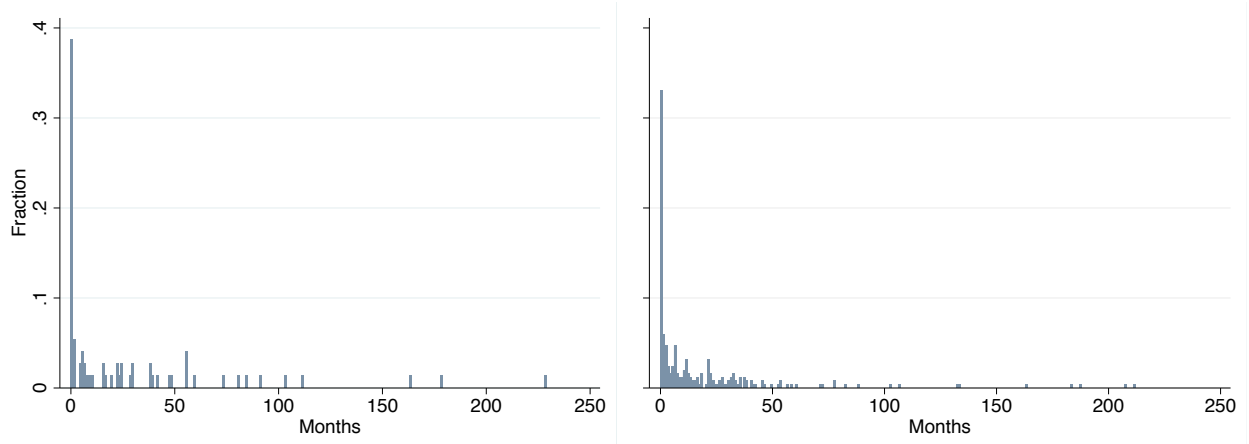


Restricted to NCEs with generic entry at or after primary IP expiration, i.e. 329 drugs out of total sample of 370. Panel (a) gives timing of generic entry relative to primary expiration while panel (b) gives timing relative to most recent (primary or secondary) IP expiration prior to generic entry.

Figure 8: Timing of generic entry (in months)

(a) Drugs without secondary IP; relative to primary IP expiration

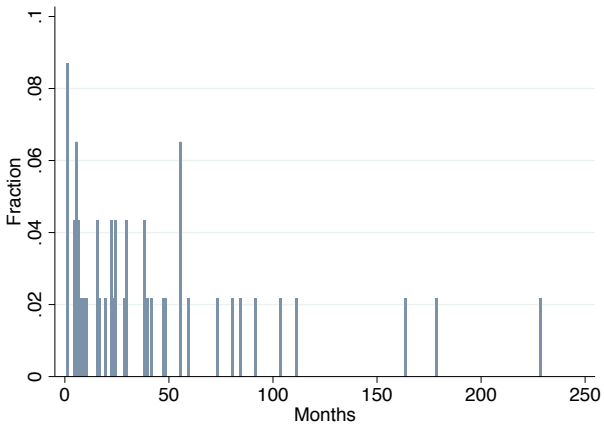
(b) Drugs with secondary IP; relative to most recent IP expiration



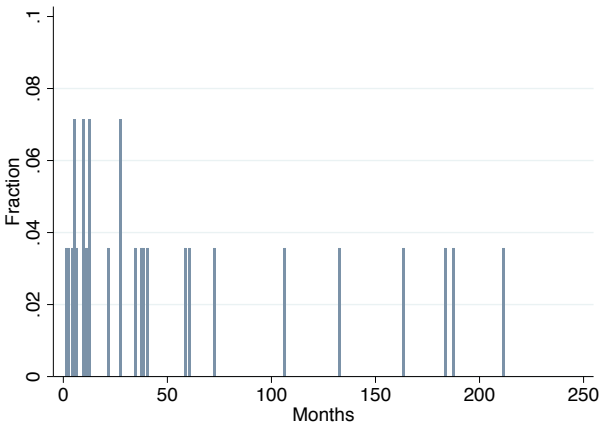
Restricted to NCEs with generic entry at or after primary IP expiration. Panel (a) gives timing of generic entry relative to primary expiration for drugs without secondary IP; number of drugs is 75. Panel (b) gives timing relative to most recent (primary or secondary) IP expiration prior to generic entry for drugs with secondary IP; number of drugs is 254.

Figure 9: Timing of *delayed* generic entry (in months)

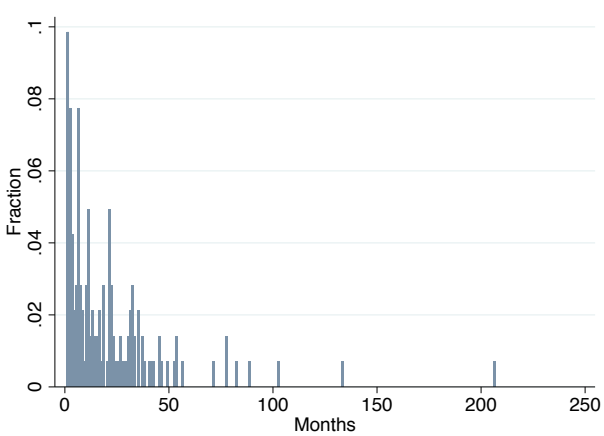
(a) Drugs without secondary IP; relative to primary IP expiration



(b) Drugs with secondary IP & all have expired; relative to final IP expiration



(c) Drugs with secondary IP & at least one still unexpired; relative to most recent IP expiration



Restricted to NCEs with non-binding IP expirations, i.e., delayed generic entry. Panel (a) gives timing of delayed generic entry relative to primary expiration for drugs without secondary IP; number of drugs is 46. Panel (b) gives delayed timing relative to final IP expiration for drugs with secondary IP and all secondary IP expired; number of drugs is 28. Panel (c) gives delayed timing relative to most recent (primary or secondary) IP expiration prior to generic entry for drugs with at least one secondary IP still unexpired; number of drugs is 142.

A.4 Cumulative entry relative to IP expirations

Figure 10 shows the raw proportions of first generic entry occurring in each period of IP protection. The figure confirms binding IP expirations for some drugs and uncertainty in market profitability and/or IP enforceability delaying entry for others.

The leftmost bar shows that a small portion of drugs (11 percent) see generic entry prior to primary IP expiration. A manual review of these drugs suggests these entrants are authorized generics, marketed by either a separate company that has received permission by the branded firm to enter early in exchange for a portion of its profits or by the branded manufacturer itself. (The branded manufacturer may choose to launch its own generic to preempt future outside generic competition.)

The dark blue bars confirm that 20 percent of the sample has a binding primary IP and another 10 percent total has a binding secondary IP. The rightmost bar shows that 19 percent of drugs see their first generic entrant after all IP have expired, meaning that for these drugs, uncertainty in market conditions is the only limiting factor. The remaining drugs see intermediate entry occurring between two IP expirations; for these drugs, either market conditions or IP enforceability may be the source of delay. In sum, total non-binding entry (occurring between or after all expirations) is larger than binding entry.

Figure 10: Cumulative entry relative to IP expirations

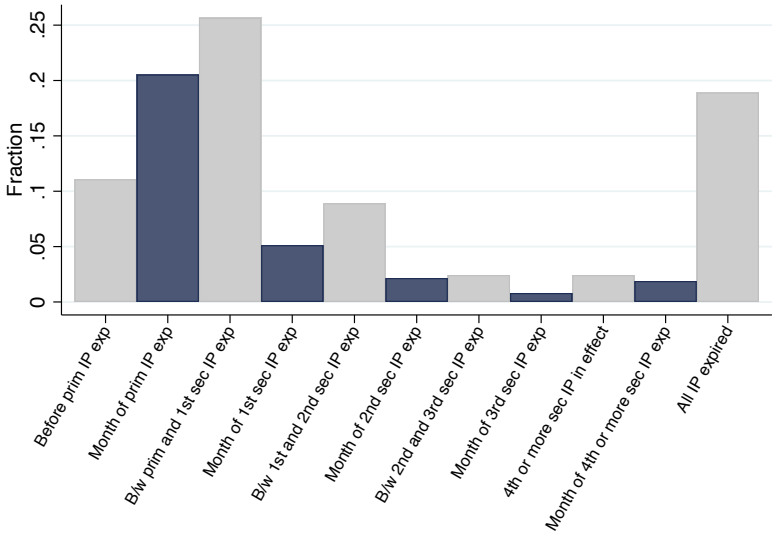


Figure gives proportion of sample experiencing generic entry in each period of IP protection. Number of drugs is 370.

A.5 Naive analysis: Comparing drugs by number of secondary IP

Table 7 repeats the analysis of Table 2 using the number of all secondary IP as the main explanatory variable. On average, each NCE has 2.7 secondary Orange Book patents and 5.9 secondary exclusivities obtained prior to generic entry. The coefficient in the preferred specification here (0.83) is lower relative to that on *# secondary patents* (1.82) in Table 2 and insignificant, suggesting that manufacturers may obtain more exclusivities for patents at risk of shorter monopoly terms, but on average, they are not associated with monopoly extension.

Table 7: Secondary IP and extra monopoly life: Drug-level analysis

	(1)	(2)	(3)	(4)
# secondary IP	-0.0433	0.209	0.456	0.826
	(0.519)	(0.551)	(0.551)	(0.554)
ln(market size)				-3.656**
				(1.557)
Observations	288	288	288	288
Adj. R^2	-0.003	0.123	0.168	0.191
Approval year FE		yes	yes	yes
Therapeutic area FE			yes	yes

Dependent variable in all regressions is extra monopoly life in months (i.e., from a drug's iprimary IP expiry to generic entry). Analysis at the drug level. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, and * p<0.1.

A.6 Application-level analyses with clustered standard errors

Tables 8 – 10 repeat the analyses of Tables 3 – 5 using robust standard errors clustered at the drug level. The coefficient on $\mathbb{1}_{Granted}$ remains significant in the preferred specification for both the OLS and IV analyses.

Table 8: Secondary patents and extra monopoly life for drugs in granted versus rejected applications: Application-level analysis

	(1)	(2)	(3)	(4)
$\mathbb{1}_{Granted}$	4.080	4.843	5.617*	5.525*
	(4.024)	(3.684)	(3.053)	(3.033)
ln(market size)				-1.107
				(1.214)
Observations	3,723	3,723	3,723	3,723
Adj. R^2	0.050	0.246	0.388	0.390
Num. of (drug) clusters	248	248	248	248
Approval year FE		yes	yes	yes
Therapeutic area FE			yes	yes

Dependent variable in all regressions is extra monopoly life in months (i.e., from a drug's primary IP expiry to generic entry). Analysis at the drug-application level. All regressions include patent application/examiner art unit fixed effects. Robust standard errors clustered at the drug level in parentheses. *** p<0.01, ** p<0.05, and * p<0.1.

Table 9: First-stage results

	(1)	(2)	(3)	(4)
leniency	0.673***	0.672***	0.667***	0.668***
	(0.0372)	(0.0367)	(0.0366)	(0.0367)
ln(market size)				-0.00530
				(0.00538)
Observations	3,370	3,370	3,370	3,370
Num. of (drug) clusters	244	244	244	244
Approval year FE		yes	yes	yes
Therapeutic area FE			yes	yes
F-statistic	564.1	569.6	558.1	558.2

Dependent variable in all regressions is an indicator for whether a patent application is granted. Analysis at the drug-application level. All regressions include patent application/examiner art unit fixed effects. Robust standard errors clustered at the drug level in parentheses. *** p<0.01, ** p<0.05, and * p<0.1.

Table 10: Secondary patents and extra monopoly life using variation in examiner leniency: IV analysis

	(1)	(2)	(3)	(4)
$\mathbb{1}_{Granted}$	14.155	13.817	18.266**	18.273**
	(11.292)	(8.659)	(8.012)	(8.060)
ln(market size)				-1.191
				(1.174)
Observations	3,370	3,370	3,370	3,370
Num. of (drug) clusters	244	244	244	244
Approval year FE		yes	yes	yes
Therapeutic area FE			yes	yes

Dependent variable in all regressions is extra monopoly life in months (i.e., from a drug's primary IP expiry to generic entry). Examiner leniency instruments for whether a secondary patent application is granted or not. Analysis at the drug-application level. All regressions include patent application/examiner art unit fixed effects. Robust standard errors clustered at the drug level in parentheses. *** p<0.01, ** p<0.05, and * p<0.1.

A.7 Leave-out examiner leniency as instrument

Tables 11 – 12 repeat the analyses of Tables 4 – 5 using a leave-out measure of examiner leniency as the instrument. Rather than determining examiner leniency from the separate first-stage sample, I calculate it using the analytic sample as the share of all other applications granted by the examiner excluding the application in question.

Table 11: First-stage results (leave-out examiner leniency)

	(1)	(2)	(3)	(4)
leave-out leniency	0.451***	0.443***	0.432***	0.432***
	(0.0316)	(0.0312)	(0.0312)	(0.0312)
ln(market size)				-0.00468
				(0.00397)
Observations	3,310	3,310	3,310	3,310
Approval year FE		yes	yes	yes
Therapeutic area FE			yes	yes
F-statistic	202.2	199.2	188.2	187.6

Dependent variable in all regressions is an indicator for whether a patent application is granted. Analysis at the drug-application level; number of drugs is 239. All regressions include patent application/examiner art unit fixed effects. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, and * p<0.1.

Table 12: Secondary patents and extra monopoly life using variation in leave-out examiner leniency: IV analysis

	(1)	(2)	(3)	(4)
$\mathbb{1}_{Granted}$	11.744*	18.216***	21.818***	21.490***
	(7.073)	(6.591)	(6.258)	(6.230)
ln(market size)				-1.176***
				(0.339)
Observations	3,310	3,310	3,310	3,310
Approval year FE		yes	yes	yes
Therapeutic area FE			yes	yes

Dependent variable in all regressions is extra monopoly life in months (i.e., from a drug's primary IP expiry to generic entry). Leave-out examiner leniency instruments for whether a secondary patent application is granted or not. Analysis at the drug-application level; number of drugs is 239. All regressions include patent application/examiner art unit fixed effects. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, and * p<0.1.

A.8 Back-of-the-envelope calculations

A.8.1 Consumer welfare loss from monopoly pricing

There are two ways by which the introduction of a generic may expand consumer surplus: first, by lowering costs for those who were already purchasing the branded drug and switch to the generic, and second, by expanding the market for a drug. I consider solely the first channel as I calculate the consumer welfare loss from the 49-month per-drug delay in generic entry attributable to secondary patenting. I determine the difference in total dollars spent by consumers who were taking the average branded drug and would have switched to a generic version if it were available earlier. Key inputs into this calculation are thus mean per-drug revenues for branded versus generic versions of a molecule and the percent of consumers who switch from branded to generic. As noted by Branstetter et al. (2016), the term "consumer" in the pharmaceutical setting captures not just patients, but all "downstream" entities, including drug retailers and insurance companies.

To determine per-drug revenues under monopoly pricing, I rely on Table 1 of Berndt et al. (2017), which gives total pharmaceutical revenues in 2016 across all molecules (branded and generic) as \$446,491 million in 2016Q1 dollars. Converted to 2020Q1 dollars, this equals \$482,880 million.[30] This total covers 2,158 molecules, amounting to a per-drug revenue of $\$482,880 \text{ million} \div 2,158 =$ \$224 million. In 2016, 74 percent of these revenues accrued to branded manufacturers, suggesting a mean branded per-drug revenue under monopoly pricing of $\$224 \text{ million} \times 0.74 = \$165.6$ million. This is likely an underestimate of monopoly revenues for two reasons: First, Berndt et al.'s definition of a molecule aggregates across dosage strengths but not dosage forms. For example, the tablet form of a molecule is considered a separate drug from the capsule form. As such, a single branded drug may be counted as multiple molecules, reducing the mean per-drug revenue. Second, by averaging across both branded and generic versions, the per-drug revenue obtained here will be smaller than if only branded versions were included. As such, my calculations as to the welfare loss to consumers from delayed generic entry will be conservative, as actual monopoly pricing is likely higher.

To determine per-drug revenues for generic molecules, I use estimates by Berndt and Aitken (2011), who find that generic prices fall to 78 percent of their initial value by six months post-generic entry, to 50 percent by one year post-generic entry, and to 23 percent by two years. I make the simplifying assumptions that the first generic entrant launches its product at the same price as the branded version and that these price declines occur at exactly six months, one year, and two years post-generic entry. Again, these assumptions will contribute to an underestimate of the welfare loss to consumers, as generic price at launch may be lower than the branded price and gradual price declines will have likely occurred in advance of these specific endpoints.

[30] I convert 2016Q1 dollars to 2020Q1 dollars using the US Bureau of Economic Analysis's gross domestic product implicit price deflator, retrieved on December 5, 2020 from the FRED data service of the Federal Reserve Bank of St. Louis at `https://fred.stlousfed.org/series/GDPDEF`. 2016Q1 has an index value of 104.865 and 2020Q1 has an index value of 113.415, leading to a conversion ratio of 1.0815.

If all consumers taking the branded version switched to the generic once it were available, they would spend \$165.6 million $\times \frac{6}{12} =$ \$82.8 million in the first six months post-generic entry, \$165.6 million $\times \frac{6}{12} \times 0.78 =$ \$64.6 million in the next six months, \$165.6 million $\times \frac{12}{12} \times 0.5 =$ \$82.8 million from year one to year two post-generic entry, and \$165.6 million $\times \frac{25}{12} \times 0.23 =$ \$79.4 million in months 25 through 49. This equals a total of \$309.6 million spent per drug in the 49 months after generic entry, if all consumers used the generic version. In contrast, if all consumers used the branded version, they would spend \$165.6 $\times \frac{49}{12} =$ \$676.2 million. This amounts to a difference of \$676.2 million − \$309.3 million = \$366.6 million per drug attributable to monopoly pricing over 49 months.

The final input to determine is the degree of consumer switching from branded to generic versions of a drug upon generic availability. Berndt and Aitken (2011) find a generic efficiency rate, meaning share of generic prescriptions for molecules where a generic version is available, of 92 percent and an overall generic share (including molecules without an available generic) of 75 percent as of 2009. Morton and Kyle (2011) note that aggressive formulary management by pharmacy benefit managers and mandatory substitution laws result in a branded product often losing 75 percent or more of its market share very quickly upon generic entry. Using this figure, I arrive at a final welfare loss to consumers of \$366.6 million $\times$ 0.75 = \$275 million per drug resulting from the 49-month extension in monopoly pricing. This calculation represents a lower bound on the welfare loss consumers experience from delayed generic entry not only due to the choice of assumptions above but also in that it ignores any market expansion that may occur with generic entry. Note that this consumer welfare loss does not contribute to any changes in total social welfare but instead represents a transfer from consumers to producers under monopoly pricing.

A.8.2 Impacts to innovation

I carry out a second back-of-the-envelope calculation to determine how the 49-month patent extension may have influenced new drug development during my sample period. I begin with Budish et al.'s (2015) elasticity of R&D investment with respect to an additional year of patent life. They find a semi-elasticity ranging between 7 and 23 percent, where R&D investment refers to clinical trials. Applying their estimate to my results implies that the average 49-month extension in monopoly term should have increased number of clinical trials by between 28 to 94 percent relative to a counterfactual where innovating firms are restricted to solely the fixed term of their initial molecule patent.

Of course, not all clinical trials result in new molecules actually approved by the FDA. To translate this elasticity into one of how approved NCEs vary with the observed patent extension, I require estimates of the proportion of clinical trials that are successful (i.e., resulting in an approved drug) and the proportion of total approved drugs that are NCEs. My elasticity of interest $\frac{\partial \text{ (NCEs)}}{\partial \text{ (patent term)}}$ can thus be calculated as follows:

$$\frac{\partial \text{ (NCEs)}}{\partial \text{ (patent term)}} = \frac{\partial \text{ (clinical trials)}}{\partial \text{ (patent term)}} \times \frac{\partial \text{ (total approved drugs)}}{\partial \text{ (clinical trials)}} \times \frac{\partial \text{ (NCEs)}}{\partial \text{ (total approved drugs)}} \quad (2)$$

The first term on the right-hand side of Equation 2 is the Budish et al. elasticity. For the second term, I rely on a comprehensive study by Wong et al. (2019) of drug development success rates across therapeutic indications. They find that 13.8 percent of all Phase 1 clinical trials lead to an approved drug. Finally, my data shows that roughly 50 percent of all drugs approved by the FDA between 1985 and 2010 were NCEs, providing an estimate for the third term.

As a lower bound, I estimate a $28 \times 0.138 \times 0.5 = 1.9\%$ increase in NCEs during my sample period attributable to the average 49-month patent extension, equivalent to $370 \times 0.019 \approx 7$ NCEs out of the total sample of 370. As an upper bound, I estimate a $94 \times 0.138 \times 0.5 = 6.5\%$ increase, equal to $370 \times 0.065 \approx 24$ NCEs. Given new molecules often represent significant therapeutic advances relative to existing treatments, this increase in NCEs may correspond to a sizeable expansion in social welfare relative to the counterfactual where drugs are limited to their molecule patent. I leave a complete welfare analysis to future research.