*American Economic Journal: Applied Economics 2016, 8(4): 189–221*
*http://dx.doi.org/10.1257/app.20150373*

# Patents as a Spur to Subsequent Innovation? Evidence from Pharmaceuticals†

*By* Duncan S. Gilchrist*

*This paper examines how an incumbent's patent protection acts as an implicit subsidy toward non-infringing substitutes. I analyze whether classes of pharmaceuticals whose first entrant has a longer period of market exclusivity (time between approval and generic entry) see more subsequent entry. Instrumenting for exclusivity using plausibly exogenous delays in the development process, I find that a one-year increase in the first entrant's market exclusivity increases subsequent entry by 0.2 drugs. The effect is stronger for subsequent entrants that are lesser clinical advances, suggesting it is driven primarily by imitation.* (*JEL* K11, L65, O31, O34)

Research dating back at least as far as Nordhaus (1967) has analyzed the role of intellectual property protection in incentivizing innovation. The motivation for such protection tends to be the idea that without property rights, the free market might underinvest—relative to the socially optimal level—in the costly development of new technologies. The tradeoff is that bestowing a firm with market power leads to prices that are statically inefficient, driving a wedge between the allocation that is realized and that which is socially optimal.

*Wealthfront Inc., 900 Middlefield Road, Redwood City, CA 94063 (e-mail: duncan@wealthfront.com). I am especially grateful for the advice and encouragement of David Cutler, Michael Luca, and Andrei Shleifer. For helpful comments and suggestions, I additionally thank Philippe Aghion, Kate Baicker, Natalie Bau, Lanier Benkard, Ernie Berndt, Dan Carpenter, Michael Chernew, Thomas Covert, Liran Einav, Richard Freeman, Peter Ganong, Jerry Green, Shane Greenstein, Rob Huckman, Haiden Huskamp, Simon Jaeger, Larry Katz, Mark Lemley, Greg Lewis, Robin Lee, Josh Lerner, Jon Levin, Hong Luo, Eric Maskin, Ariel Pakes, Genevieve Pham-Kanter, Benjamin Roin, Martin Rotemberg, Al Roth, Bhaven Sampat, Emily Glassberg Sands, Vicki Sato, Ariel Dora Stern, Heidi Williams, Thomas Wollmann, and Ali Yurukoglu, as well as seminar participants at the Harvard IO lunch, the Harvard HCP Seminar, and the NBER Productivity Seminar. I thank David Friedman, Michael Lanthier, Kevin Sharer, Stuart Watt, and Jeff Way for sharing their insight into the FDA, the pharmaceutical industry, and patent law. I am thankful to Kevin Garewal for his assistance in obtaining the data for this project and I gratefully acknowledge financial support from the Harvard Business School Doctoral Program. Finally, I am grateful to three referees for helpful comments on an earlier draft. Data licensing disclaimers follow. The statements, findings, conclusions, views, and opinions contained and expressed herein are not necessarily those of IMS Health Incorporated or any of its affiliated or subsidiary entities. The statements, findings, conclusions, views, and opinions contained and expressed in this article are based in part on data obtained under license from the following IMS Health Incorporated or affiliate information service(s): IMS National Sales Perspectives, 1992–2013, IMS Health Incorporated. All Rights Reserved. AHFS® Pharmacologic/Therapeutic Classification used with permission. ©2014, the American Society of Health-System Pharmacists, Inc. (ASHP) The Data is a part of the AHFS Drug Information®; ASHP is not responsible for the accuracy of transpositions from the original context.

†Go to http://dx.doi.org/10.1257/app.20150373 to visit the article page for additional materials and author disclosure statement or to comment in the online discussion forum.

Traditional models think of an incumbent's intellectual property protection, such as a patent, as inhibiting entry in the same market by another firm.[1] Of course, at the same time, providing market power to an incumbent through a patent makes it valuable for rivals to invest themselves in finding ways to circumvent the incumbent's protection, thereby allowing them to imitate a product and share in the oligopoly rents. Existing economic models that consider this possibility, such as Gallini (1992), assume that such imitation is necessarily socially costly because it leads to excessive entry; they suggest that the optimal length of protection should be short so as not to encourage it.

However, a second school of thought argues that an incumbent's intellectual property protection actually spurs socially valuable innovation. For example, in the medical literature, Scannell et al. (2012) argue qualitatively that the loss of patent protection by blockbuster drugs is the foremost cause of the sustained decline in the productivity of pharmaceutical development: they use the music industry as an analogy, writing "[i]magine how hard it would be to achieve commercial success with new pop songs if any new song had to be better than the Beatles, [and] if the entire Beatles catalogue was available for free."[2] This reference to the music industry is actually no coincidence—in fact, a similar argument has been made in industries producing creative content. There, the specific idea is that copyright protection of existing content spurs the development of new, better content. This has long been a key argument made by proponents of copyright term extensions; it was used, perhaps most notably, by lobbyists asking US Congress to pass the Copyright Term Extension Act of 1998.[3]

The thrust of this argument is that if intellectual property protection is narrow, then although an incumbent's protection prevents competitors from selling an identical product, it does not prevent them from selling differentiated products. Yet, it may be the case that otherwise socially valuable differentiated products are not profitable to develop once the first product is available at competitive prices; e.g., a drug maker might not find it profitable to develop welfare-increasing drugs when an existing drug is already available as a generic. In the framing of Bulow, Geanakoplos, and Klemperer's (1985) prominent analysis of market power within oligopolies, the market power afforded to an incumbent by their patent protection might be an important strategic complement to an innovator's own market power.

I analyze this phenomenon in this paper. Specifically, I analyze both how one firm's patent protection spurs subsequent entry and also the composition of affected entry. My focus is within the pharmaceutical sector, the single largest in terms of

[1] In most of the theoretical literature, patents incentivize innovation specifically because they provide protection from imitation. See, e.g., Nordhaus (1967, 1969, 1972); Scherer (1972); Fudenberg et al. (1983); Klemperer (1990); and Gilbert and Shapiro (1990). Scotchmer (1991) provides a helpful review.

[2] The decline in productivity has been called a "crisis"—see Pammolli, Magazzini, and Riccaboni (2011) and the references cited therein. Cockburn (2007) provides an elegant analysis and discussion of some of the possible forces at work. The perplexing slowdown in productivity is sometimes called "Eroom's Law," contrasting with the well-known Moore's law in the semiconductor industry. Writers in the popular press have also identified existing generics as a prominent consideration for firms directing R&D investments toward new drug development: e.g., Parker (2013) suggests that in 2013 the impending availability of a generic version of Ambien (a blockbuster drug for insomnia) made it less appealing for competitors to develop new drugs targeting the same condition. A specific quotation is reproduced in online Appendix Section A.

[3] This act is sometimes called the Mickey Mouse Protection Act; see Martin (2002) for a discussion.

domestic R&D expenditure and a sector in which advances yield tremendous value to society.[4] Within this setting, the first question is the extent to which the generic version of an existing treatment impedes the development of a competing drug; the second question is the extent to which the affected drugs constitute socially wasteful imitation versus valuable innovation. Focusing within pharmacologic classes of closely related drugs, I first show that the length of time between the first in class drug's (FIC's) approval and its generic's entry has an economically and statistically significant causal impact on the number of new drugs that subsequently come to market in that class. That is, I show that the longer the period of time for which the FIC has market exclusivity versus its generic, the more subsequent entry in that class.[5] Second, I ask what kinds of drugs are most affected and find the effect is concentrated among drugs that are lesser clinical advances.

I begin by analyzing a simple theoretical model that motivates the setting and clarifies how an incumbent's patent impacts the investment decision of a subsequent entrant. As is the case in pharmaceuticals, in the model patents are narrow and do not inhibit the entry of differentiated substitutes. Instead, in the model a patent holder is a price-setter; after patents expire, generic entry is assumed to occur and prices fall to marginal cost. The model shows how the subsequent entrant's profitability, and thus initial investment decision, is affected by the incumbent's patent. When it comes to welfare, the model illustrates how entry is a double-edged sword: it helps consumers, who benefit from both the availability of a new drug as well as the price-competition its availability induces, but has an ambiguous effect on social surplus due to the potentially sizable fixed cost of entry. Finally, the model shows how the drugs that are most socially valuable—i.e., significant advances in quality—are those least likely to be deterred by the incumbent's loss of patent because these drugs are more likely to be able to stand alone. In other words, the model suggests the incumbent's loss of patent protection is most likely to deter "me-too" imitation than it is genuine innovation.

My empirical analysis, which makes up the remainder of the paper, analyzes the model's predictions. I focus on pharmacologic classes of new molecular entities (NMEs), groupings of pharmaceuticals that are closely related in chemical composition and in physiological effect but which are differentiated at the molecular level. This method of categorizing drugs provides me with a principled procedure with which to identify groups of drugs that are sufficiently similar to be substitutes in the eyes of prescribers and patients, yet are sufficiently differentiated so that each NME requires its own costly clinical trials to be marketed, and one NME's patent protection does not preclude the entry of the others.[6] I assemble a rich dataset that

[4] The pharmaceutical sector makes up nearly 20 percent of industrial R&D (National Science Board (NSB) 2012). Improvements to health can be immensely valuable. Murphy and Topel (2006) estimate that a 0.1 percent reduction in cancer mortality is worth $500 billion to current and future Americans.

[5] I refer to entry which is not first in class as "subsequent" throughout. Although the phrase "follow-on" is perhaps more grammatically convenient, it would be misleading because it is often used to describe innovation, which is cumulative, e.g., as in Sampat and Williams (2015). Similarly, innovation that is called "sequential" usually refers to innovation that builds on its predecessors, as in Bessen and Maskin (2009).

[6] Getting any NME to market is costly because the FDA requires proof of safety and efficacy; a commonly cited figure is that the total capitalized cost of a new drug is on the order of $1 billion (DiMasi, Hansen, and Grabowski 2003). This differs from the low cost of bringing a generic to market, which I return to below. The channel I describe

includes, for NMEs belonging to classes with first in class approval in the period 1987–2011, approval dates, annual sales, market sizes, dates of clinical development, dates of filing and expiration for key patents, and, importantly, whether a generic version has been approved and, if so, when.

The model requires that sales respond to competition within class (else the entry of one drug's generic would not erode the profits of another) and that entry is sequential (else drugs would enter the market simultaneously, leading patent lifetimes to be aligned), so I first look to the data for evidence of these characteristics in the market for pharmaceuticals. Though causal identification of the first factor is difficult, I provide anecdotal and descriptive evidence consistent with the hypothesis that the generic entry of one drug in a class reduces the sales of another. Next, I show that drug development is a highly staggered process: nearly 40 percent of approved drugs only begin the often decade-long process of clinical development *after* the first in class has been approved. Finally, I show that, consistent with the model's main prediction, very few new drugs come to market after the FIC has gone generic.

I turn next to my main empirical analysis, which examines how the length of FIC exclusivity impacts subsequent innovation in the same class. I start by discussing potential challenges to identification. While exclusivity typically ends after the expiry of a key patent, the specific timing of that patent's expiry relative to approval (i.e., exclusivity) is determined by many factors, some of which are idiosyncratic, while others are market-related. For example, as Budish, Roin, and Williams (2015) insightfully leverage to study the effect of commercialization lags on clinical trial investments, exclusivities tend to be eroded by the length of clinical trials. If the length of clinical trials is correlated among drugs in the same class, then exclusivities may be as well, and this would confound estimates generated by simply regressing entry on FIC exclusivity. My analysis exploits two strategies to handle the potential for endogeneity in market exclusivity.

First, I show that controlling directly for the two main ways in which endogeneity might enter—market size (and profitability) and the length of clinical trials —does little to change the estimates. This suggests the main effect is unlikely to be driven by factors such as drug profitability, which could lead to a spurious correlation between FIC exclusivity and entry since, in profitable markets, firms are incentivized to both enter as well as find ways to maintain longer exclusivity. Furthermore, if market exclusivities were endogenously related to class characteristics, then they should be correlated within class, and this is not born out in the data.

Second, I conduct an instrumental variables analysis that exploits uncertainty and disruptions in the development process as a source of exogenous variation to identify the effects of variation in the remaining length of patents on the number of molecules in the same drug class. Specifically, my IV analysis exploits the combined facts that (i) the path of drug development can be hard to predict, yet (ii)

could equally affect complements as well as substitutes, but I do not analyze complements because I am without a principled method for categorizing them.

patent filing is defensive.[7] The first means firms sometimes discover molecules before they know specifically what to do with them, and the second means that firms are incentivized to file patents early. The end result is that firms sometimes file patents on a molecule well in advance of its progression to clinical trials (contrasting with a strategy in which firms maximize market exclusivity by filing patents as late as possible).

Thus, I use delays between patent filing and the start of clinical review as an instrument for market exclusivity. I show empirically that these delays are not related to patent characteristics, such as claims and citations, nor are they related to market characteristics, such as ex post annual sales. They are, however, related to factors that are indicative of initial uncertainty and disruptions in the development process, such as the length of time between when the patent was filed and when it is first transferred to another entity (e.g., in a licensing agreement), as well as whether the patent was transferred to a new assignee in a merger transaction. Importantly, they are not correlated within pharmacologic classes. Since a story about the endogeneity of these delays requires that they are somehow related to a class-specific unobservable, that they are uncorrelated within class means a confounding factor would have to be at once idiosyncratic to FIC drugs, unrelated to subsequent entrants, yet also related to total entry. While I cannot rule out such a story, it is not an easy one to tell.

The results, across both IV and non-IV specifications with and without controls, consistently show that a one-year increase in first in class exclusivity yields a 25–30 percent increase in subsequent entry, or about 0.2 drugs. This is a large effect: it implies that a 1 standard deviation increase in first-in-class exclusivity *doubles* subsequent entry. The online Appendix presents a battery of robustness tests which show that the results are not particularly sensitive to sample restrictions, to specification choices, and that they survive a placebo test where I look at the effect of last in class exclusivity on prior entry.

Finally, I analyze how the effect of FIC exclusivity on subsequent entry is moderated by magnitude of quality improvements. I proxy for advances in quality by asking whether new drugs receive priority review status from the FDA, a status granted to molecules found to be exceptionally promising in clinical trials.[8] Consistent with the model's prediction, I find that FIC exclusivity has a significantly stronger impact on the number of subsequent entrants that do not receive priority review than on those that do (and I cannot distinguish the latter effect from zero). This suggests the effect I observe is more likely to be driven by imitation than it is socially valuable innovation.

Prior empirical work on patents has overlooked the effect I analyze. Indeed, the majority of empirical work on patents has focused on patenting as an outcome of

[7] A well-known example of the first phenomenon is the story of Viagra's discovery. The Viagra molecule was originally discovered as a treatment for angina, but in the course of clinical review (for angina) it was discovered that men taking the drug found themselves unexpectedly erect.

[8] A priority review designation accelerates approval decisions—the FDA's stated goal is to complete a priority review in six months. However, it does not affect the level of evidence as to a drug's safety and efficacy required for approval.

the innovative process.[9] Taking the reverse approach, I contribute to a sparse body of empirical research that considers and finds evidence for how the characteristics of patents manifest themselves in new innovations. Sakakibara and Branstetter (2001) and Lerner (2002) actually find patent laws have only limited impacts on innovation; with respect to research that finds evidence of the opposite, I am aware only of Moser (2005), who shows that nineteenth century patent laws influenced the geographic distribution of inventors.

Of course, the idea that competition in itself might affect new product introduction is not new, and other empirical work supports a link between competition and investment within the pharmaceutical context. Civan and Maloney (2009) show that that early stage investment is positively correlated with the price of approved drugs in the same therapeutic area, and concurrent work by Branstetter, Chatterjee, and Higgins (2014) shows generic entry in a therapeutic area is associated with declines in early stage investment. This research supports the thrust of my findings. However, to my knowledge, no existing research has articulated and analyzed the fundamental link between the characteristics of patent protection and these downstream effects.

The remainder of the paper proceeds as follows. I open in Section I with a model to formalize the effect that an incumbent's patent protection has on subsequent entry. Section II introduces the data, and Section III provides descriptive analysis and background on the pharmaceutical industry. In Section IV, I describe my empirical strategy before presenting estimates of the impact of an incumbent's patent protection on subsequent entry. Section V sheds light on the welfare implications of my findings and, finally, Section VI concludes.

## I. A Model

In this section, I develop a simple model to illustrate how the length of incumbent's patent protection affects subsequent entry and welfare. I first provide an overview and then add structure to the model before analyzing its implications.

### A. *Preliminaries*

There are two firms, an incumbent and a potential entrant, which have each discovered their own drugs. In the first period, the incumbent $I$ is already present in the market. The potential entrant $E$ has a candidate drug, but it must pay a fixed cost $F$, which is distributed exponentially with mean $\mu$, to enter.

If $E$ enters, it benefits from patent protection for $t_E$ subsequent periods, which means it is guaranteed to be the exclusive seller of its product during that time. After its patent protection expires, generic sellers flood the market, pushing its price down to its unit production costs, which are 0. Thus, in order for $E$ to enter, it must be able to make back its entry cost $F$ within the first $t_E$ periods.

However, $E$ must also compete with $I$, and $I$'s patent protection only remains in force for $t_I < t_E$ periods. As with $E$, once $I$'s protection expires, its drug becomes

[9] Griliches (1998) provides a review and a discussion of the evidence for and against this design.

available at its marginal cost, which is also 0. That means $E$'s time on patent can be thought of in two segments: during the first $t_I$ periods, it competes in a Bertrand-Nash duopoly with $I$, while during the last $t_E - t_I$ periods, $E$ sets prices knowing $I$ is available at marginal cost. I assume firms do not discount, so $E$'s entry decision is simple: it enters if its total profit less entry cost is at least 0, otherwise it abandons its drug.[10]

Before proceeding, I note that I am implicitly assuming that $I$'s patent protection is sufficiently narrow to not directly impact $E$'s ability to enter. This captures the reality in pharmaceutical development, where patents are narrow: pharmaceutical patents generally only provide protection against identical substitutes (e.g., the same molecule) and not differentiated substitutes (e.g., molecules that share features and treat the same condition, but are not identical). I do not model generic entry explicitly for the sake of simplicity, but my setup is motivated by the fact that once patents have expired, it is not costly for generic competitors to enter the market, and such entry tends to happen rapidly, reducing prices paid on the order of 90 percent in the year or two following patent expiry.[11] The channel through which patent protection impacts competition in my model—by not directly inhibiting the entry of a differentiated competitor, but rather by impacting that competitor's market power—means my model shares features with Gallini (1992), where rivals can pay a fee to imitate a patent-protected incumbent. However, Gallini (1992) does not consider the case where the rival's product is differentiated from that of the incumbent.

I model the demand side through a discrete choice framework, in which a unit mass of consumers may pick at most one drug to consume, and consumers differ according to their price sensitivities. Specifically, consumer $i$ receives utility from drug $j$ according to its quality level $\delta_j$ and price $p_j$:

$$u_{ij} = \delta_j - \frac{1}{\alpha_i} p_j. \tag{1}$$

For simplicity, I normalize the utility of the outside option $u_0$ to be the same across consumers and assume it is sufficiently low such that consumers always buy a drug when both $E$ and $I$ are available. Price sensitivities are parametrized by the $\alpha_i$, which are distributed exponentially in the population with mean $\lambda$; $\lambda$ captures the dispersion of price sensitivity in the population. If $\lambda$ is higher, then consumers are on average less price sensitive. Of course, in the United States, consumers are often covered by pharmaceutical insurance and so they do not bear the full cost of drugs directly out of pocket. Nevertheless, US individuals do bear part of the cost of pharmaceutical purchases, e.g., through tiered formularies, or even simply through a higher cost of pharmaceutical insurance; it is thus reasonable to assume that demand is downward sloping. Furthermore, this demand framework captures the idea that in the eyes of drug sellers, wealthier individuals are on the whole more price inelastic, which is likely to be true if they have higher willingness to pay for quality—and that

[10] Allowing for discounting complicates the algebra but does not change the comparative statics that follow.

[11] Reiffen and Ward (2005) estimate the fixed cost of generic entry is roughly $300,000. This is very small relative to the $1 billion cost of drug development cited earlier. Berndt (2002) provides a review of the evidence on generic entry.

willingness to pay shows up through some combination of more generous insurance policies and higher out of pocket payments.[12] Finally, I assume that $E$'s quality is exogenously higher than $I$'s, i.e., that $\delta_E > \delta_I$, and I write the difference in qualities $\Delta = \delta_E - \delta_I$. This demand framework captures the idea that drugs are differentiated and that consumers differ in how they value quality relative to price.

### B. *Implications*

I first solve for equilibrium profits before analyzing welfare and the determinants of both entry and welfare. Looking to demand, consumer $i$ prefers the entrant's drug if $\delta_E - \frac{1}{\alpha_i} p_E > \delta_I - \frac{1}{\alpha_i} p_I$, i.e., $\alpha_i > \frac{p_E - p_I}{\Delta}$. This means demand in any period for $E$'s drug is

$$D_E(p_E, p_I) = \int_{\frac{p_E - p_I}{\Delta}}^{\infty} \frac{1}{\lambda} \exp\left(-\frac{\alpha}{\lambda}\right) d\alpha = \exp\left(-\frac{p_E - p_I}{\lambda \Delta}\right), \tag{2}$$

and $D_I = 1 - D_E$. Taking first-order conditions, equilibrium prices during the period both drugs are on patent are $p_E^* = \lambda\Delta$ and $p_I^* = \lambda\Delta\left(W\left(e^2\right) - 1\right)$, where $W(\cdot)$ is the Lambert $W$ function, while they are simply $p_E^* = \lambda\Delta$ and $p_I^* = 0$ once $I$ is off patent. Altogether, this means $E$'s total profits conditional on entry are

$$\pi_E^* = t_I \lambda \Delta e^{W\left(e^2\right) - 1} + (t_E - t_I)\lambda\Delta e^{-1}, \tag{3}$$

and the ex ante probability that $E$ will enter is $\Pr(\pi_E^* \geq F) = 1 - \exp(-\mu\pi_E^*)$, which I denote by $\rho$. This expression yields the first model's first prediction.

**Prediction 1:** The probability that $E$ enters is increasing in the amount of time remaining on $I$'s patent:

$$\frac{\partial \rho}{\partial t_I} > 0.$$

This first comparative static is quite natural: $E$'s profits are higher while $I$ is on patent, so the likelihood that $E$ enters is higher if $I$ has more patent protection remaining. The majority of my empirical results focus precisely on analyzing how subsequent entry is affected by the length of the first entrant's remaining patent protection.

Of course, $E$'s entry need not be socially beneficial. From a policymaker's perspective, the question is not whether longer protection for the incumbent increases entry, but whether it is an effective lever for incentivizing valuable innovation and not wasteful imitation. I turn to this question next, and assume the social planner discounts the future at rate $\gamma < 1$. Consumer surplus in any period is simply the integral under the demand curve, so total consumer surplus in the state of the world

[12] The choice of exponential distribution is motivated by the skewness of real incomes. An alternative interpretation of $\alpha_i$ is that all consumers have the same price sensitivity, but that quality sensitivity varies.

in which there is only an incumbent $I$ and the state of the world in which $E$ enters is, respectively, given by the following expressions:[13]

$$CS_I = \frac{1-\gamma^{t_I}}{1-\gamma}\lambda(\delta_I - u_0)e^{-1} + \frac{\gamma^{t_I}}{1-\gamma}\lambda(\delta_I - u_0), \tag{4}$$

$$CS_{I,E} = \frac{1-\gamma^{t_I}}{1-\gamma}\lambda\Big(\Delta e^{W(e^2)-2} + (\delta_I - u_0)e^{-\frac{\Delta(W(e^2)-1)}{\delta_I - u_0}}\Big) + \frac{\gamma^{t_I}(1-\gamma^{t_E-t_I})}{1-\gamma}\lambda(e^{-1}\delta_E + (1-e^{-1})\delta_I - u_0) + \frac{\gamma^{t_E}}{1-\gamma}\lambda(\delta_E - u_0).$$

And social surplus is consumer surplus plus firm profits:

$$SS_I = CS_I + \frac{1-\gamma^{t_I}}{1-\gamma}\lambda(\delta_I - u_0)e^{-1}, \tag{5}$$

$$SS_{I,E} = CS_{I,E} - F + \frac{1-\gamma^{t_I}}{1-\gamma}\lambda\Delta\zeta + \frac{\gamma^{t_I}(1-\gamma^{t_E-t_I})}{1-\gamma}\lambda\Delta e^{-1},$$

where $\zeta$ is a constant. These expressions yield the following results.

**Prediction 2:** $E$'s entry improves consumer surplus but has an ambiguous effect on social surplus:

$$CS_{I,E} > CS_E,\ SS_{I,E} \overset{?}{\lessgtr} SS_E.$$

That $E$'s entry helps consumers is to be expected: entry means a new, higher quality drug is available, and that the price of the incumbent's drug is lower. But the overall effect of entry on social surplus is ambiguous. To see this, consider the fixed cost $F$: in the limit, if $F$ is zero then the entry of a higher quality product is necessarily socially beneficial, while if $F$ is infinite then all entry must be wasteful.

So far, the model has shown that the incumbent's patent will have some impact on subsequent entry, and that the impact of entry on social welfare is uncertain. But of course both of these effects depend on the magnitude of the quality increase (parametrized by $\Delta$). The third prediction analyzes the extent to which this parameter moderates the effect of $t_I$ on $E$'s entry.[14]

[13] See online Appendix Section B for the full derivation of these expressions.

[14] This result requires an assumption that the expected fixed cost $\mu$ is not too small; this ensures that equation (4) is sufficiently concave so that $E$'s profitability is not due entirely to the period during which both drugs have patent protection. A sufficient condition is that $\mu > \frac{3}{\lambda\Delta(3t_I + t_E)}$.

**Prediction 3:** The larger the increase in quality, $\Delta$, the more unlikely it is that $t_I$ influences $E$'s entry decision:

$$\frac{\partial^2 \rho}{\partial t_I \partial \Delta} < 0.$$

This comparative static shows that the influence of the length of $I$'s remaining patent protection on $E$'s likelihood of entry is moderated by quality differences between the two drugs. This implication is intuitive: drugs that are significant clinical advances should not be highly impacted by an incumbent's patent protection because the difference in quality boosts demand for the new drug. However, when it comes to social surplus we see the opposite effect.

**Prediction 4:** The larger the increase in quality, $\Delta$, the more $E$'s entry increases social surplus:

$$\frac{\partial (SS_{I,E} - SS_I)}{\partial \Delta} > 0.$$

The final comparative static says that, all else equal, social surplus increases more from entry when the new drug is of higher quality. This should not be surprising; higher quality drugs are more socially valuable.

We now arrive at the bottom line. Prediction 3 says the extent of the impact of $I$'s patent on $E$'s entry is decreasing in $\Delta$, while Prediction 4 says that the larger is $\Delta$, the more important the entrant is for social surplus. Putting these two predictions together, the situations in which the incumbent's patent is most important for entry are the situations in which entry is least likely to be socially valuable. I analyze the empirical implications of these last two predictions in the penultimate section of this paper.

Of course, the welfare implications of entry would be different under different assumptions. Cournot competition rather than Bertrand would provide more scope for business stealing and waste, while a Hotelling model with horizontal differentiation would increase the scope for consumer benefits relative to this vertical model. With this expository framework and its predictions in mind, I next discuss my data. My empirical analysis follows.

## II. Data

I analyze how the number of drugs approved in a pharmacologic class is influenced by the first in class drug's market exclusivity. In the following, I briefly explain each of the main data sources and its contribution before providing an overview of the sample's composition and some summary statistics. Online Appendix Section C provides an extended discussion of the data.

I begin with data from the FDA that classifies all NMEs approved in the United States between 1987 and 2011 into pharmacologic classes.[15] For all drugs, I then

[15] See Lanthier et al. (2013). I focus exclusively on non-biologic drugs approved in the United States, and in order to ensure data completeness, I exclude classes in which the FIC was approved prior to 1987. The United States

obtain the date of first approval and first generic entry (if generic entry has occurred) from the FDA's Drugs@FDA database and the FDA's Orange Book, respectively. Only 38 percent of drugs in the sample have gone generic as of the end of 2013, so in cases in which I do not have realized exclusivities, I supplement them with a measure of expected exclusivity.

My measure of expected exclusivity comes from identifying, for each drug, the single patent most likely to inhibit generic entry (typically pertaining to a drug's active ingredient). I then proxy for the date of expected generic entry using this patent's expiry date. Additional details are given in the online data Appendix, but my methodology for identifying this patent is inspired by Hemphill and Sampat (2012) and specifically exploits the fact that the Hatch-Waxman Act of 1984 allows firms to extend a single patent to make up for up to half the time spent in clinical trials. I use the expiry date of the extended patent. Indeed, Hemphill and Sampat (2012) codify patents for drugs that went generic in the 2000s and find that the extended patent represents the patent on the drug's active ingredient in 79 percent of drugs analyzed, and in my data, the expiration of extended patents is highly predictive of generic entry: for the 81 drugs in my sample that have gone generic and have a patent for which an application for extension was filed, the $R^2$ from a regression of realized exclusivity on exclusivity predicted by the extended patent is 0.95. I also note that in robustness tests described later and presented in the online Appendix, I show that my results are robust to restricting the sample to only those classes for which I have realized exclusivities. Altogether, of the 293 drugs in my sample that belong to 156 classes, I have a measure of exclusivity for 237, of which 127 are first in class.

Next, I collect data on the timing of clinical trials, market sizes, and sales.

*Timing of Clinical Trials*.—Details on when pharmaceuticals begin development is generally proprietary, but a prerequisite for a patent extension to make up time lost in clinical trials is that firms publicly certify the beginning and end of clinical development.[16] I collect these data from the Federal Register.

*Market Size*.—My main measure of market size is inspired by Acemoglu and Linn's (2004) analysis of the effect of market size on pharmaceutical development. In particular, using data from the Medical Expenditure Panel Survey, I identify the total annual US prevalence of each class's primary ICD-9 condition code. That is, I identify the number of US individuals who suffer from a specific condition, and call this prevalence market size.[17]

is not the only market for pharmaceuticals, but it is by far the largest, accounting for over 50 percent of global sales in 2000 (WHO 2004).

[16] The term *clinical* denotes development in humans, as opposed to laboratory or computational-based development. These data were first analyzed by Keyhani, Diener-West, and Powe (2006).

[17] As I note in the online Appendix, I have experimented with other measures of market size and find they have little influence on my results. Budish, Roin, and Williams (2015) similarly find their results are robust to conditioning on different measures of market size.

*Sales*.—To capture the price dimension of profitability, I use data from IMS Health's National Sales Perspective which detail total annual US sales for the 1,000 top selling drugs each year from 1992–2012.

Before turning to the summary statistics, I note that the categorization of "first in class" I use differs slightly from the FDA's formal categorization. This is motivated by the fact that, in some cases, multiple drugs are approved very early on and sometimes one of the early entrants is considerably more effective than the others. In these cases, the technical definition of first in class does not coincide with the market's focal early entrant. A poignant example is given by the class of atypical antipsychotics. The first entrant, Clozaril, was approved in 1989, but Clozaril's mass-market appeal was severely limited because it produces a potentially fatal side effect in 1–2 percent of patients (Alvir et al. 1993). The second atypical antipsychotic, Risperdal, had no such severe side effects and was granted priority review status; after its approval in 1993, Risperdal rapidly became the standard of care, with annual sales that were nearly 50 percent higher than Clozaril's just one year after approval (Finkel 2012). Though Clozaril was technically the FIC, it was largely irrelevant (as a competitor) for future entrants. Thus, in cases in which multiple drugs entered early on in short succession, my strategy is to focus on what I call the *effective* FIC: this is the authentic FIC unless the authentic FIC is immediately followed by a priority review drug, in which case the effective FIC is that priority review drug. The effective FIC differs from the authentic FIC in 32 cases and is approved an average of 3.0 years after the authentic FIC. I probe the sensitivity of my results to this definition in online Appendix F.A. It should come as no surprise that my results are strongest for classes in which the definitions of authentic and effective FIC coincide. For brevity, I refer to the effective FIC simply as the FIC.

To summarize, the sample consists of all non-biologic drugs that belong to classes with first approval in the period 1987–2011. Altogether, the sample consists of 156 classes representing 293 drugs. In practice, since my data on market exclusivities and the dates marking the beginning of clinical development (used to compute the instrument) are not complete, I restrict my analysis to the 111 classes for which I have that data, representing 252 drugs (although this too I probe in online Appendix F.A).

I present summary statistics on the sample in Table 1. The table consists of three panels which respectively present statistics by class (panel A), then by drug for FIC drugs only (panel B), and finally by drug, but for all drugs (panel C). Looking first to the by-class figures in panel A, the first row shows that class sizes are generally small but that they are skewed—the mean number of entrants subsequent to the FIC is 0.73 (and the median 0) while the maximum is 7. Mean clinical development times, in the second row, have a mean of 8.31 years and standard deviation of 3.81 years. Mirroring class size, market size, and maximum annual revenue are highly skewed: the mean market size is about 8 million, while the largest is 45.5 million.[18] The by-drug figures in panels B and C show similar patterns; mean exclusivities in both panels are just below 12 years and average times in clinical development

[18] There are fewer than 111 observations in the fifth row because my revenue data only contains revenues for one or more drugs in 88 of the 111 classes.

are about 8 years. I forgo discussion of the length of time between patent filing and the beginning of clinical development (my instrument for market exclusivity) until Section IVB.

## III. Facts about Competition and Entry

This section provides descriptive analysis of entry in the pharmaceutical industry. I first probe two of the model's main assumptions: that generics in class influence demand for other (differentiated) branded drugs, and that pharmaceutical development is sequential and not characterized by a single race to market. I then show that, consistent with the model's first prediction, subsequent entry is substantially less likely to occur after the FIC has gone generic.

### A. *Generic Competition in Class*

Although others have documented the fact that competition within a class of drugs—measured, e.g., by the number of distinct molecules—impacts sales (see, e.g., Lichtenberg and Philipson 2002), to my knowledge there is no evidence on the specific impact of generic competition in class on profitability. However, that the availability of one drug's generic negatively affects the sales of different drugs in the same class is crucial for my argument; if these drugs are not close substitutes, the mechanism exposited by the model could not be rationalized by the data.

Causally estimating the impact of one drug's generic on the sales of others is difficult. Although exploiting an event-study framework around the time of generic entry might seem appealing at first glance, it is not clear that such a framework is appropriate because the timing of generic entry is not random.[19] Thus, my strategy in this section is to provide descriptive and anecdotal evidence.

I begin with a case study. Lipitor was the fifth drug approved in the class of statin drugs (technically, HMG-CoA reductase inhibitors), which primarily treat conditions associated with cardiovascular disease. Lipitor came to market in 1996 and quickly achieved blockbuster status—by the time it went generic in November of 2011, Lipitor had sold more than any other medicine, with global sales of $125 billion over 14.5 years (Associated Press 2011). However, Lipitor's sales did not grow monotonically over the period of its branded life; instead, Lipitor sales peaked in 2006, when two of its closest competitors, the statins Zocor and Pravachol, went generic. The press expected this to be a turning point for Lipitor's sales (see online Appendix Section A.B), and indeed it was, as visualized in Figure 1, panel A.[20] Though it may not be causal, the relationship between Lipitor's sales and the generic entry of its competitors is striking, with sales falling 25 percent from peak from 2006–2010. Lipitor was the best-selling drug of all time; presumably if one drug

[19] Even on a relatively short timescale (e.g., one year in advance), generic entry may be anticipated in important ways (e.g., through changes in marketing strategy of the type analyzed by Ellison and Ellison (2011)).

[20] The source for this sales data is IMS Health's publicly available Top-Line Market Data, available at www.drugs.com.

Table 1—Summary Statistics

| | Mean | Median | SD | Min | Max | Observations |
|---|---|---|---|---|---|---|
| *Panel A. By class* | | | | | | |
| Number of subsequent entrants in class | 0.73 | 0 | 1.39 | 0 | 7 | 111 |
| Mean time in clinical development (yrs) | 8.31 | 7.53 | 3.81 | 0.78 | 26.97 | 111 |
| Market size (1,000,000s) | 7.88 | 3.22 | 10.51 | 0.08 | 45.55 | 111 |
| Maximum annual revenue ($1,000,000s) | 1,000.12 | 364.96 | 1,665.66 | 26.84 | 8,558.62 | 88 |
| *Panel B. By drug, first in class only* | | | | | | |
| Market exclusivity (yrs) | 11.81 | 12.65 | 3.16 | 5.00 | 18.32 | 111 |
| Time in clinical development (yrs) | 7.98 | 6.98 | 4.06 | 0.78 | 26.97 | 111 |
| Patent filing to clinical development (yrs) | 5.25 | 4.43 | 3.87 | 0.00 | 20.39 | 111 |
| Maximum annual revenue ($1,000,000s) | 708.79 | 245.35 | 1,227.84 | 24.51 | 7,064.03 | 85 |
| *Panel C. By drug, all* | | | | | | |
| Market exclusivity (yrs) | 11.61 | 12.66 | 3.25 | 5.00 | 18.32 | 196 |
| Time in clinical development (yrs) | 8.21 | 7.24 | 3.88 | 0.78 | 26.97 | 192 |
| Patent filing to clinical development (yrs) | 4.95 | 4.35 | 3.62 | 0.00 | 20.39 | 189 |
| Maximum annual revenue ($1,000,000s) | 841.37 | 287.82 | 1,418.35 | 13.80 | 8,558.62 | 166 |

*Notes:* This table presents summary statistics for the main sample of 111 drug classes as described in Section II. Panel A presents statistics by drug class, panel B by drug for FIC drugs only, and panel C presents statistics for all drugs. There are fewer than 111 observations in the rows describing maximum annual revenues because I only observe revenues for 88 classes and 85 drugs of the 111 classes analyzed. Panel C also faces sample restrictions: of the 252 drugs in the sample, I only observe market exclusivity for 196, the timing of clinical development for 192, the value of the instrument for 189 observations, and finally revenues for 166.

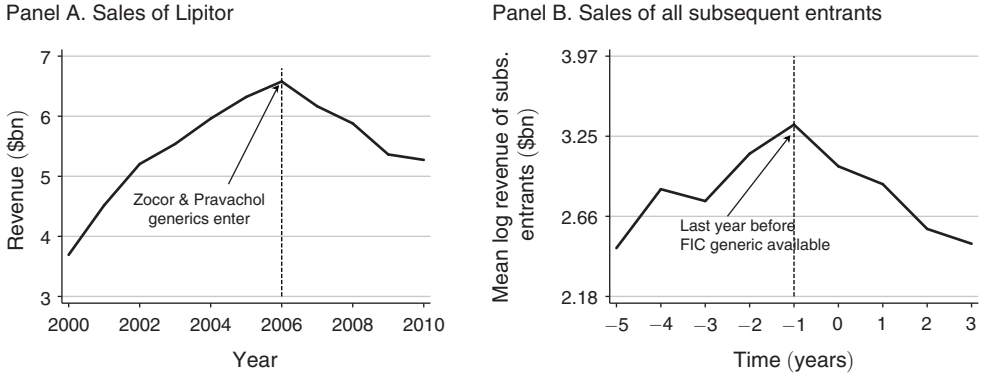


Figure 1. Generic Entry and Sales of Subsequent Entrants

*Notes:* These figures respectively show plots of annual sales over time. Panel A plots annual US sales for Lipitor, where the data come from IMS Health's publicly available Top-Line Market Data. Panel B shows the mean of log annual sales of subsequent entrants relative to the time of the FIC's generic entry. The sample is as described in Section II.

was to be able to defend itself against its competitors' generics, it would have been Lipitor.[21]

An analogous relationship is present across drug classes. In Figure 1, panel B, I plot the log of annual sales against the timing of first-in-class generic entry (so that year zero denotes the year of FIC generic entry), averaged across classes.[22]

[21] Aitken, Berndt, and Cutler (2009) provide an extended discussion of declines in spending on blockbuster drugs.

[22] I take logs before averaging because (as described in Section II) drug sales show substantial skew.

Here, we see a similar downward trend that starts when the FIC goes generic. While this is not a causal test, and other events affect the results (e.g., some subsequent entrants themselves go generic three years after the FIC), the sales data do suggest that generic competition in a class impacts the sales of other drugs.

### B. *The Timing of Development Decisions*

The model analyzed earlier assumed that the second firm's entry decision was made after the first firm had sunk its own fixed costs and entered. That development decisions are sometimes sequential (as opposed to simultaneous) is important for the theory advanced in this paper; otherwise, dates of generic entry would be aligned and it would be implausible that one drug's generic would deter the entry of another. My data, which detail the dates that drugs entered clinical trials for the first time, speaks to this point.

Figure 2 presents a histogram which examines the timing of the start of clinical development for subsequent entrants. The horizontal axis is normalized so that development start times are relative to the FIC's approval (at time $= -1$) and FIC generic entry (at time $= 0$).[23] The chart shows a large proportion, nearly 40 percent, of subsequent entrants do not begin clinical development until the FIC is approved. This is important because although the start date of clinical development is public after approval for approved drugs, the extent to which pre-approval development decisions are public information is difficult to quantify; in the period before FIC approval, market participants likely know something about their competitors actions but this is sensitive competitive intelligence, and so they are unlikely to know the whole story.[24] However, approval is highly publicized so firms starting trials then could not have mistakenly believed they could be first to market, and once a drug is approved it is likely that competitors are able to secure a list of its patents to forecast its exclusivity. It is additionally worth noting that only 11 percent of subsequent entrants come to market in the year after the FIC, i.e., even if 60 percent of subsequent entrants do start development before the FIC is approved, only a small fraction of subsequent entrants come to market at roughly the same time as the FIC.

Figure 2 also shows that no new drugs begin development after the FIC has gone generic (there is no mass to the right of time $= 0$). This is consistent with the model's prediction that the incumbent's generic reduces incentives for new firms to enter, as well as with what Kevin Sharer, the former CEO of Amgen, told me: "Once we enter generic land, there's no incentive [to develop new drugs]" (quoted with permission, April 15, 2014).

[23] This normalization is accomplished by dividing the length of time between the start of clinical development and the FIC's generic entry by the FIC's exclusivity.

[24] Development decisions are often described as competitive intelligence, although at some level, once a human subject pool has been recruited and used in testing, it is harder to keep secret, especially since investors may require notification. Databases like Adis R&D Insight, Pharmaprojects, and Thomson Reuters Cortellis purport to collect exactly this type of information, though the extent to which their data are collected contemporaneously versus after the fact is not clear, nor is the extent to which they are complete.

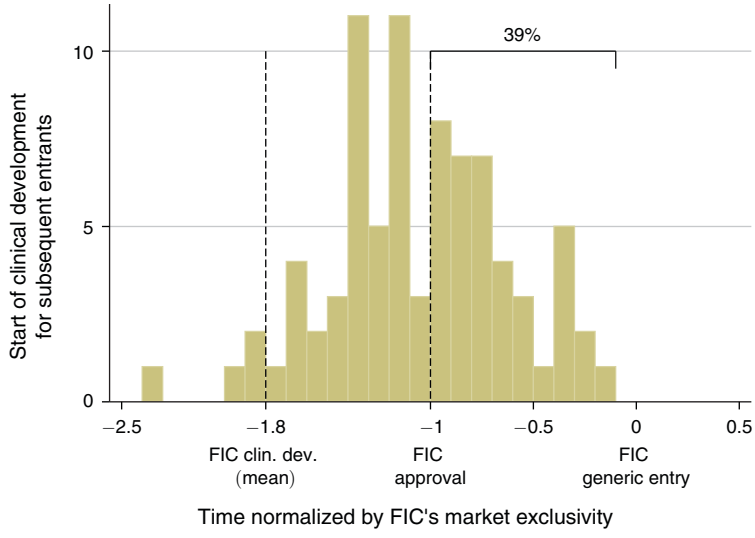


FIGURE 2. TIMING OF CLINICAL DEVELOPMENT FOR SUBSEQUENT ENTRANTS

*Notes:* This figure shows a histogram of the timing of the start of clinical development for subsequent entrants relative to FIC generic entry, where the $x$-axis is normalized by the FIC's market exclusivity so that $-1$ corresponds to FIC approval and 0 corresponds to FIC generic entry. This normalization is accomplished by dividing the length of time between the start of clinical development and the FIC's generic entry by the FIC's exclusivity. The start of clinical development is defined as the date on which an Investigational New Drug Application, a required precursor to the start of human clinical trials, is approved. The sample is as described in Section II.

### C. *The Timing of Subsequent Entry*

If new drugs tend to be less profitable once the FIC has gone generic, then the long-run profitability of any given drug should be lower the later it enters (as the clock ticks on the FIC's exclusivity). All else equal, this should create a downward gradient in new drug approvals over time, with few approvals after the FIC has lost exclusivity. I analyze this prediction in Figure 3, which shows a histogram of the timing of non-FIC approvals relative to the FIC's generic entry, where as before the horizontal axis is normalized within classes so that time $= -1$ corresponds to when the FIC enters and time $= 0$ corresponds to when the FIC goes generic. The plot shows that entry timing is negatively correlated with FIC exclusivity, with nearly two thirds of non-FIC drug approvals occurring before just half of the FIC's branded lifetime has expired. The right half of the histogram is almost empty; very few new drugs come to market in classes in which the FIC has gone generic. In online Appendix Section E, I analyze this relationship in regression form and find that the correlation between the number of new entrants and remaining FIC exclusivity is statistically significant.[25]

[25] I do not have an instrument for the time remaining on the FIC's patent, so I do not interpret this analysis causally. The ideal instrument for this context would be something that unpredictably and randomly invalidates a branded incumbent's patents. Such a shock would enable an event-study analysis, where the strategy would be to ask if entry is reduced after an incumbent's patents are randomly invalidated. One possibility for an instrument, which has been exploited in prior research analyzing the financial impact of generic entry, see Panattoni (2011)

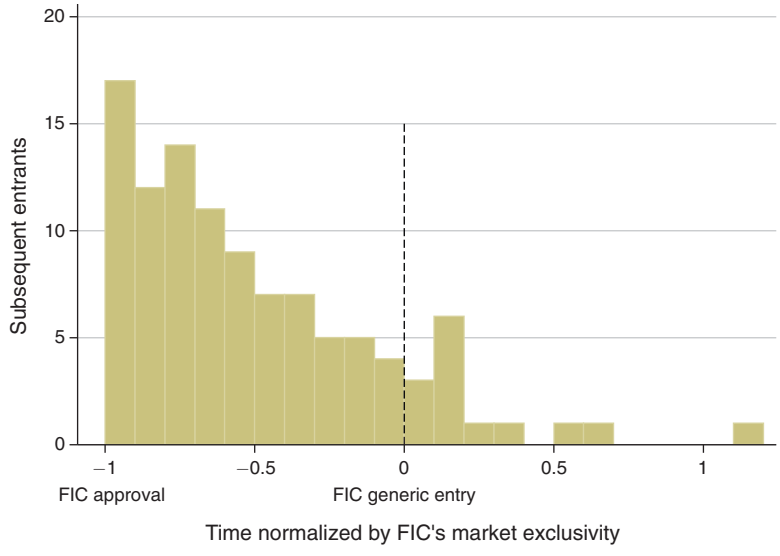


FIGURE 3. TIMING OF SUBSEQUENT ENTRY

*Notes:* This figure shows a histogram of the timing of subsequent entry relative to FIC generic entry, where the $x$-axis is normalized by the FIC's market exclusivity so that $-1$ corresponds to FIC approval and 0 corresponds to FIC generic entry. This normalization is accomplished by dividing the length of time between the entry of the subsequent entrants and the FIC's generic entry by the FIC's exclusivity. The sample is as described in Section II. The relationship visualized here is analyzed empirically in online Appendix Section E.

## IV. Main Results

I have presented evidence that drugs in the same class appear to be substitutes, that entry decisions are made sequentially, and that subsequent entry rarely occurs once the FIC has gone generic. These findings lay the foundation for this paper's thesis, which is that longer FIC exclusivity increases subsequent entry. This section presents my main results. I first describe my empirical framework before discussing potential confounders, my instrument, and the results.

### A. *FIC Exclusivity and Subsequent Entry*

My empirical strategy examines how the exclusivity of FIC drugs impacts subsequent innovation. Formally, with $j$ indexing drug classes, I estimate regressions of the form

$$SubsEntrants_j = \alpha + \beta FICExcl_j + \gamma' X_j + \varepsilon_j, \tag{6}$$

and Jacobo-Rubio, Turner, and Williams (2014), is the outcomes of Paragraph IV lawsuits, which are sometimes precursors to generic entry. However, the sample of such lawsuits is small and neither the timing nor the outcomes of Paragraph IV lawsuits appear to be random—they tend to occur shortly after a drug's active ingredient patent has expired and often result in generic entry.

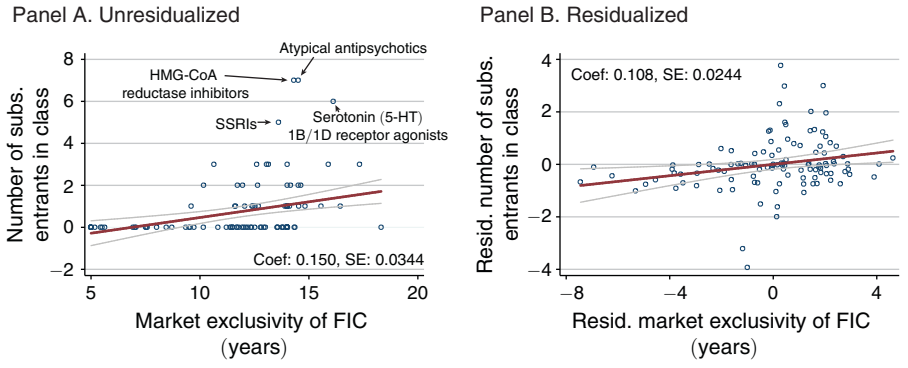


FIGURE 4. SUBSEQUENT ENTRY AND FIRST IN CLASS EXCLUSIVITY

*Notes:* Panel A shows a plot of the number of entrants subsequent to the FIC drug against the FIC drug's market exclusivity. Classes with at least 5 subsequent entrants are labeled. Panel B shows a plot of the residualized number of entrants subsequent to the FIC against the FIC's residualized market exclusivity. Residuals are conditional on year of FIC approval fixed effects, mean development time, and market size, and the specification of the controls and sample are as described in Section II. The slope of the line of best-fit and its associated robust standard error are presented in the corners of the figures.

where $SubsEntrants_j$ is the number of entrants subsequent to the FIC, $FICExcl_j$ is the FIC's market exclusivity, and $X_j$ is a vector of controls. Given the count nature of drug launches, I estimate quasi-maximum likelihood (QML) Poisson models, and since my data are as of 2011 and some classes started earlier than others, I include FIC approval year fixed effects throughout.[26]

Figure 4, panel A illustrates the raw relationship between subsequent entry and FIC exclusivity without controls, and shows that there is indeed a strong positive correlation. However, the slope of that graph should not necessarily be interpreted as causal; understanding specifically what determines a drug's market exclusivity is helpful for understanding where biases could enter. Figure 5 visualizes the typical sequence of events in drug development for a drug that progresses from discovery all the way to patent expiry. A prospective drug starts in preclinical development and testing, which usually takes place in the lab. After an Investigational New Drug (IND) Application is approved by the FDA, the drug progresses into the period of regulatory review characterized by human clinical trials. If trials are viewed as successful in proving safety and efficacy, a New Drug Application (NDA) is submitted to the FDA. In the case of approval, the drug is marketed exclusively by its developer until it "goes generic" when the first generic enters the market.[27] Federal law guarantees that NMEs are protected from generic entry for a minimum of five years after a drug is first approved, but, in general, entry is restricted for longer because of patent

[26] Some classes have had very few years to allow for subsequent entry, so in online Appendix F.3 I show that restricting the sample to only those classes in which the data include at least ten years after FIC entry does not affect the results. I also note here that including fixed effects in a nonlinear model typically leads to an incidental parameters problem. In Poisson models, this can be corrected using Hausman, Hall, and Griliches' (1984) transformation, which I apply throughout.

[27] Though I focus only on approved drugs, the likelihood that any individual drug candidate progresses all the way from the start of clinical review to approval is not high, on average. Hay et al. (2011) estimate it is less than 10 percent.

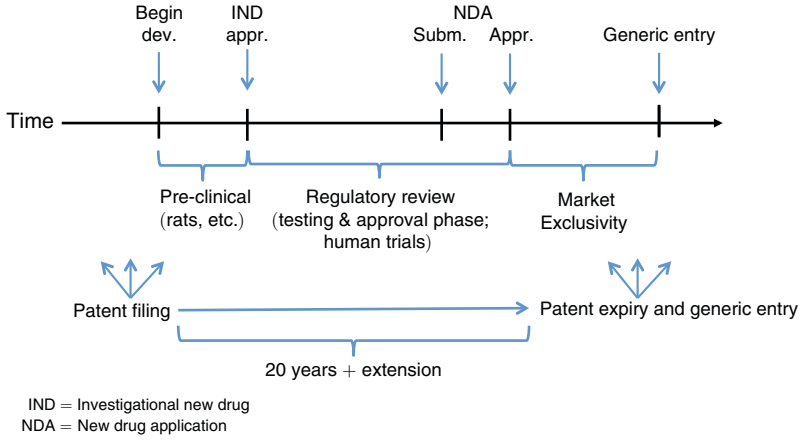


FIGURE 5. TIMELINE OF DRUG DEVELOPMENT

*Notes:* This figure shows the typical timeline of drug development. In particular, it depicts how market exclusivity, which is defined as the time between NDA approval and generic entry, is impacted by the timing of patent filing.

protection. As a legal matter, patents pertaining to active ingredients generally need to be filed before the beginning of clinical trials. Once the drug has been administered to humans, its novelty (required for enforceable patenting) can be questioned.

Altogether, this means there are three main factors that influence the length of exclusivity. The first factor pertains to the timing of approval, in the middle of Figure 5. As Budish, Roin, and Williams (2015) point out, since patent lengths are fixed at 20 years (plus applicable extensions, which tend to be minor), an important component of market exclusivity is simply the time a drug spends in trials ("commercialization lag," in their words). From the start of clinical trials, the clock is ticking toward patent expiry, so long trials tend to reduce exclusivity. Thus, to the extent that trial lengths are correlated within classes, then they could bias $\beta$ in the estimation of equation (6).[28] This means it is important that I control for the length of clinical review.

The second factor influencing exclusivity is the timing of generic entry. This factor pertains to the end point on the right-hand side of Figure 5. For example, the innovator may be able to file new auxiliary (yet enforceable) patents, e.g., on the molecule's salt forms or the manufacturing process, which make it difficult for generics to enter even after the AI patent has expired. In some cases, the firm selling a branded drug might even contract with generic entrants to stay out of the market in what is called a "reverse payment settlement."[29] Since the extent of any such manipulation is presumably driven by profitability, I control for market sizes and additionally present an analysis which controls for sales.

[28] For example, if trials in class $j$ are short, then drugs in class $j$ all have long exclusivities, so that $FICExcl_j$ is larger. Keeping with the example, long exclusivities in class $j$ presumably mean it is more profitable, so if fixed costs are similar across classes, then total entry in $j$ is higher, meaning $SubsEntrants_j$ is larger. Then the positive relationship I observe between $FICExcl_j$ and $SubsEntrants_j$ could be due to the fact that classes with shorter trials are more attractive.

[29] Hemphill (2007, 2006) provide extensive description and analysis of reverse payment settlements.

Third and finally, the timing of the AI patent's filing relative to the start of clinical development is not always the same. Variation in the filing date affects the left-hand side of Figure 5. As I explain below, scientific uncertainty leads the gap between patent filing and the start of trials to exhibit substantial variation, which is orthogonal to other class characteristics. This provides the basis for my instrument.

However, before proceeding to the IV analysis, I present results which control for the first two factors: time in trials and market size. Figure 4, panel B shows that conditioning on these controls, as well as fixed effects for the year of FIC approval (since some classes are older than others), does not remove the positive relationship between subsequent entry and first in class exclusivity.[30]

As an additional test for endogeneity in market exclusivity, I ask whether market exclusivities are positively related within classes. The motivation underlying this test is that if some factor confounds the estimated relationship between $FICExcl_j$ and $SubsEntrants_j$, then presumably that factor would also influence the exclusivity of other drugs in class $j$. This test can only be operationalized for classes in which there is at least one subsequent entrant, but it still provides insight into how much exclusivity is determined by factors that are idiosyncratic to a drug versus common to a class. I implement this test by estimating an OLS model where the outcome is the exclusivity of drug $k$ and the independent variables are the exclusivity of drug $l$, as well as my baseline controls for market size, time in trials, and year of FIC approval fixed effects, for all pairwise combinations of drugs $k$ and $l$ belonging to the same class. The coefficient estimate on the exclusivity of drug $l$ is $-0.0089$ with standard error 0.0366, where the standard error is robust and clustered by class.[31] Overall, this suggests that after conditioning on time in development and market size, the potential for endogeneity in naive estimation of equation (6) is likely to be limited. Nevertheless, I pursue clean identification in the IV analysis below.

### B. *IV Strategy*

I have shown that there is a strong relationship between FIC exclusivity and subsequent entry, and that this relationship is not diminished by conditioning on time in trials and market size. I have also shown that exclusivity is not significantly correlated among drugs within the same class; this means an endogeneity story would require some external factor which impacts both subsequent entry and FIC exclusivity but not the exclusivity of other drugs. However, clean identification in this setting is possible through use of an instrument that exploits the third principal factor determining market exclusivity—the timing of patent filing. This section explains the logic underlying this instrument before presenting an empirical analysis of the instrument itself.

[30] The relationship is slightly weakened in the linear model presented in the figure, but it is actually strengthened in the corresponding Poisson estimates, which are presented, with controls added sequentially, in the first three columns of Table 3. However, I forgo an extended discussion of these estimates until I take up all of my main results in Section IVC.

[31] Online Appendix Figure D.1 visualizes a similar relationship by plotting the exclusivities of subsequent entrants against the exclusivities of FIC drugs (not all pairwise exclusivities). Panel A shows that without controls the relationship is weak and statistically insignificant. Conditioning on mean development time in class, market size, and year of FIC approval fixed effects leads to a weak negative association, as visualized in panel B.

*The Timing of Patent Filing.*—Patent lifetimes are fixed, so the date of a drug's AI patent filing has a strong influence on the drug's exclusivity. So what determines the timing of patent filing?

Firms are incentivized to file patents as late as possible (which, to ensure novelty, is the day before the beginning of clinical trials) because doing so maximizes future exclusivity. However, the reality is that patents are sometimes filed substantially in advance of the start of clinical trials. The summary statistics presented in Section II show that, for FIC drugs, patents are filed at median 4.43 years before clinical trials begin.[32] Interviews with patents experts and R&D directors suggest that this is no surprise. They point me toward three primary factors that contribute to it.

First, there can be agency problems within firms. In particular, the career motives of scientists and the motives of the larger organization are not necessarily aligned. Scientists receive acclaim through research presented in papers and in conferences. Firms support these endeavors, but they place a high value on intellectual property and tend to require that patents be filed before new results are made public. Together, this means patents may be filed at a time that suits the scientist—before the firm is prepared to undertake clinical review.

Second, there can be larger organizational frictions. The pharmaceutical industry has been characterized by a tremendous amount of merger and acquisition activity in recent decades (see, e.g., Deloitte 2009). A major organizational shock like an acquisition presumably complicates strategic decision-making and might slow the progression from the lab to clinical review. Below, I show empirically that mergers are associated with longer delays from patent filing to the start of trials.

Third, and perhaps most importantly, the science of drug development is highly uncertain while the patent system forces firms to act defensively. That is, the trajectory of a drug's development is not necessarily linear—a molecule is sometimes initially thought to treat one condition when it later turns out to be better at treating another.[33] Yet the owner of a patent is whichever firm filed it first, and the discounted gain of waiting an extra year to file, i.e., an extra year of exclusivity 20 years in the future, may be low relative to the loss incurred if another firm files the patent first. This means firms file patents as soon as it is known that a molecule *may* be of commercial interest, but before it is necessarily clear what that commercial interest will be.

Consider, e.g., the path to approval of Arava, a drug presently marketed for rheumatoid arthritis. The active ingredient in Arava was first discovered and patented in the late 1970s by researchers searching for new agricultural pesticides (Oh and O'Connor 2013). In 1985, researchers experimenting with rats believed Arava might have potential for use in transplantation (Hoi and Littlejohn 2005). However, it was not until 1993 that the drug first entered clinical development—for arthritis. By the time Arava was approved in 1998, its patent was already about to expire. The FDA's exclusivity minimums kept generics out for almost 7 years, about half of the

[32] The date on which intellectual property protection is first sought is actually called a patent's priority date. Priority dates are sometimes called "effective filing dates" and determine expiration dates. More detail on the differences between priority and filing dates are found in online Appendix Section C.D. For simplicity, I refer to the priority date throughout as the filing date.

[33] Recall the case of Viagra in footnote 7.

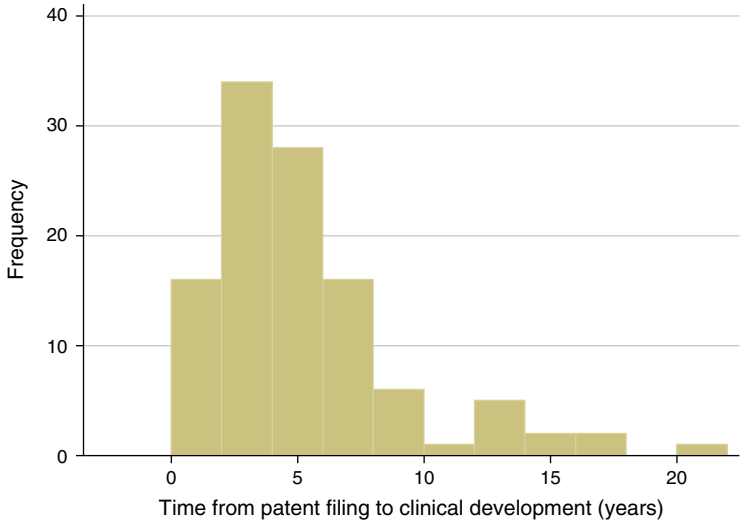


FIGURE 6. HISTOGRAM OF THE TIME FROM PATENT FILING TO CLINICAL DEVELOPMENT

*Notes:* This figure shows a histogram of the time from patent filing to the start of clinical development for FIC drugs. The start of clinical development is defined as the date on which an Investigational New Drug Application, a required precursor to the start of human clinical trials, is approved. The sample is as described in Section II.

mean exclusivity of almost 12 years in my sample. However, that Arava's development was delayed well past its initial discovery does not appear to be the result of low sales expectations. At the time of approval, financial analysts estimated that Arava's annual sales could exceed $800 million by 2003 (Wall Street Journal 1998). Consistent with the thesis of this paper, Arava has now been off patent for almost a decade and there has been no entry aside from Arava in its class.

*Identification Strategy.*—My IV strategy exploits gaps between when patents are first established on a molecule and the beginning of clinical trials. Specifically, my instrument for market exclusivity is the time between the date on which intellectual property protection was first sought and the date marking the beginning of clinical development. I denote the instrument for drug $i$ by $z_i = t_{DevStart,i} - t_{PatentFiling,i}$. Large, positive values of $z_i$ reflect molecules that are discovered substantially before they are commercialized. For example, the gap between Arava's patent filing date and the start of clinical development was 15 years (1978 through 1993), and its market exclusivity was subsequently just 7 years. Figure 6 shows a histogram of the instrument for FIC drugs. Much of the mass is just above zero, reflecting patent filing shortly before the beginning of trials.[34] However, there is also a right tail in which drugs like Arava are found.

[34] In fewer than 4 percent of cases, calculation of the instrument yields a negative value, reflecting a filing date that occurs after the beginning of clinical trials. Since this might reflect that the patent identified by my methodology is not an active-ingredient patent, I am hesitant to use these values, and set the value of the instrument in these cases to zero. However, leaving the instrument at its original values or dropping these observations altogether does

Formally, identification requires that $Corr(z_j, \varepsilon_j) = 0$. That is, identification requires that delays between patent filing and the start of clinical development for FIC drugs are unrelated to other factors that affect entry in the class. Although this cannot be tested directly, in the following two sections I provide two pieces of evidence to its effect. First, I ask descriptively how the instrument relates to patent characteristics. I find no evidence that delays between patent filing and the start of development signal anything about patent characteristics like claims or citations, but they are related to factors like merger activity. Second, to dispel concerns that the instrument might be picking something up about unobserved profitability, I ask whether the instrument is related to sales in class, and, in a test like that used in my analysis of market exclusivities, I ask also whether the instrument is significantly correlated within class. I find no evidence for either.

*The Instrument and Patent Characteristics.*—I analyze the relationship between the instrument and patent characteristics in Table 2. Columns 1–3, respectively, ask whether the instrument is correlated with the patent's number of claims, the number of citations the patent itself lists, and the number of other patents that cite it. The sample is the full sample of 189 drugs for which I observe the date on which clinical development begins (and thus am able to compute a value for the instrument). Reflecting the count nature of these outcome variables, I estimate QML Poisson models. The point estimates are neither economically nor statistically significant in all cases.[35] This suggests these patents were not in some sense less "important" ex ante, as judged by claims and citations, nor ex post, as judged by the number of other patents referencing the patent in question.[36]

Next, in columns 4–6, I ask whether the instrument is related to reassignment of patent ownership, which is tracked by Thomson Innovation.[37] Column 4 estimates an OLS model of a dummy for reassignment on the instrument and shows that the instrument is not statistically nor economically related to reassignment. In column 5, I ask whether the instrument is related specifically to reassignment involving a merger, and this time the estimate is positive and statistically significant at the 10 percent level. Finally, in column 6, I analyze how the instrument relates to the length of the time between filing and first reassignment, a measure which I view as a proxy for the degree of initial uncertainty. The estimated coefficient is strong and positive.[38] Altogether, Table 2 shows that drugs with large $z_i$'s appear to have patents that are similar to those of other drugs, except that they are more likely to have been involved in a merger and that they took longer to make it from the filer's hands into those new of a owner. These results suggest large $z_i$ are likely to originate from organizational changes and scientific uncertainty, rather than from particularities of the patent itself.

not affect the main results. I note also that the mass in the histogram would be located closer to zero if I focused on true filing dates, as these are always weakly greater than priority dates.

[35] The results are similar if I focus on FIC drugs only.

[36] Patent importance is frequently measured by patent citations, e.g., as in Hall, Jaffe, and Trajtenberg (2005).

[37] Technically, I examine changes in patent assignee. Such changes can represent licensing agreements in addition to changes in ownership.

[38] In columns 5 and 6, the sample is conditional on any reassignment, so the number of observations drops from 189 to 165.

Table 2—Patent Characteristics and the Time between Patent Filing and Clinical Development

| Dependent variable is | Number of claims (1) | Number of citations (2) | Number of patents citing (3) | (1/0): Any patent reassign. (4) | (1/0): Reassigned in merger (5) | Time from filing to first reass. (yrs) (6) |
|---|---|---|---|---|---|---|
| Patent filing to clinical dev. (yrs) | 0.00476 | −0.0348 | −0.0155 | −0.00279 | 0.0169 | 0.338 |
| | (0.0188) | (0.0277) | (0.0187) | (0.00703) | (0.00955) | (0.143) |
| Estimation | Poisson | Poisson | Poisson | OLS | OLS | OLS |
| Mean of dependent variable | 16.19 | 18.93 | 79.70 | 0.885 | 0.182 | 5.358 |
| Observations | 189 | 189 | 189 | 189 | 165 | 165 |

*Notes:* This table examines the relationship between patent characteristics and the instrument, which is the time between patent filing and clinical development. The start of clinical development is defined as the date on which an Investigational New Drug Application, a required precursor to the start of human clinical trials, is approved. Columns 1–3, respectively, present Poisson models of the relationship between the instrument and the patent's number of claims, the number of citations the patent itself lists, and number of other patents that cite it. Columns 4–6 examine whether the instrument is related to reassignment by estimating OLS models. Column 4 regresses a dummy for any reassignment on the instrument, column 5 regresses a dummy for whether that patent was ever reassigned in a merger transaction on the instrument, and column 6 regresses the time from filing to first reassignment on the instrument. The sample is the full of drugs for which I observe the start of clinical development. Specifications (5) and (6) are conditional on any reassignment, so the number of observations drops from 189 to 165. Poisson models are estimated by quasi-maximum likelihood and robust standard errors are reported in parentheses.

*The Instrument and Class Characteristics.*—The evidence analyzed above provides context for the instrument, but it does not speak directly to the question of exogeneity: the instrument might still reflect something about a class that influences both FIC exclusivity and total entry. For example, capacity-constrained firms might put lower priority on less profitable markets, leading delays to be negatively correlated with profitability. Or, even if the instrument is randomly assigned, then a longer delay should require a higher per year expected profitability in order for the firm to break even; this type of selection effect would lead delays to be positively correlated with profitability. Finally, it may simply be more difficult to develop new drugs in some classes than others.

I examine the first two possibilities in Figure 7, panel A, where I proxy for profitability using the log of the maximum annual revenue earned by any drug in a class, and ask how that relates to the instrument for FIC drugs. (I condition on my baseline controls: FIC approval year fixed effects, market size, and mean development time in class.) The figure shows there is no significant relationship between between class sales and the time between patent filing and the start of development for the FIC.[39]

Though the instrument is not related to class sales does not mean it is not related to other class-specific unobservables, and this possibility is difficult to rule out completely. To address it, I implement a test similar to that presented earlier: I ask whether the instrument is correlated within classes. In particular, I estimate an OLS

[39] I choose the maximum annual revenue earned by any drug in the class as a proxy for profitability because this figure reflects whether a given class could possibly be a top seller. However, the results are similar if I instead proxy for profitability with the maximum of the FIC drug's annual sales. I additionally show in online Appendix Section F.E that controlling directly for sales in my main regressions does not affect the results.

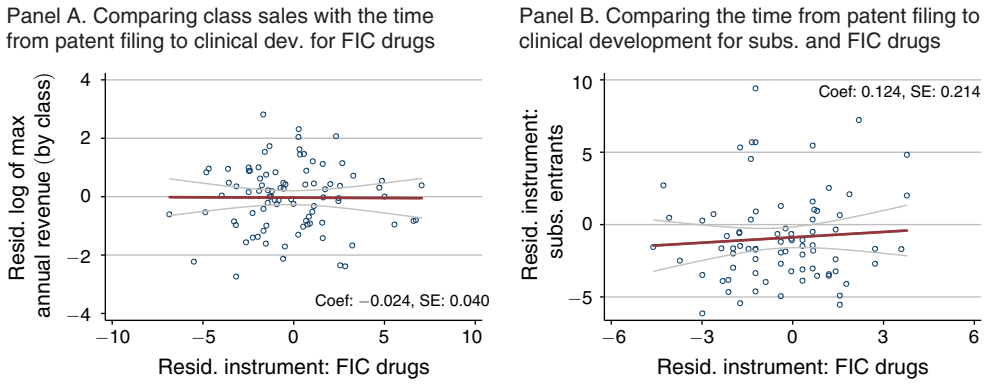


FIGURE 7. EVIDENCE ON THE VALIDITY OF THE INSTRUMENT

*Notes:* These figures respectively plot (i) class sales against the instrument for FIC drugs and (ii) the instrument for drugs subsequent to the FIC against the instrument for FIC drugs. The instrument is the time between patent filing and clinical development. Residuals are conditional on year of FIC approval fixed effects, mean development time, and market size. The measure of class sales is the log of the maximum of annual sales for all drugs in the class. The sample is as described in Section II, and the slope of the line of best-fit and its associated robust standard error clustered by drug class are presented in the corner of each figure.

model where the outcome is $z_k$, the independent variables are $z_l$ and my baseline controls, and there is an observation for all pairwise combinations of drugs $k$ and $l$ belonging to the same class. The estimated coefficient on the market exclusivity of drug $l$ is $-0.00375$ with standard error 0.0471, where the standard error is robust and clustered by class. In Figure 7, panel B, I visualize a similar analysis, but this time restrict the sample to ask only how the instrument for subsequent entrants compares with the instrument for FIC drugs (again, conditional on my baseline controls). The figure shows the estimated relationship is economically and statistically insignificant. I conclude that the instrument is unlikely to be picking up class-specific unobservables. I next present the first stage and then turn to the results.

*The First Stage.*—Figure 8 presents the first stage, where both axes are conditional on my baseline controls. The relationship between the instrument and FIC exclusivity is strong—a one year change in the value of the instrument is associated with 0.45 fewer years of exclusivity—and the plot shows that the relationship remains strong across the support of the instrument. The associated $F$-statistic is 43.52.

## C. *Results*

In the previous subsections, I explained the sources of variation underlying a drug's market exclusivity and I presented evidence analyzing the potential for endogeneity in exclusivities. I then proceeded to explain the logic behind my instrumental variable and presented evidence analyzing its relationship with patent and market characteristics, before finally presenting the first stage. This section presents my main results on how FIC exclusivity affects subsequent entry.

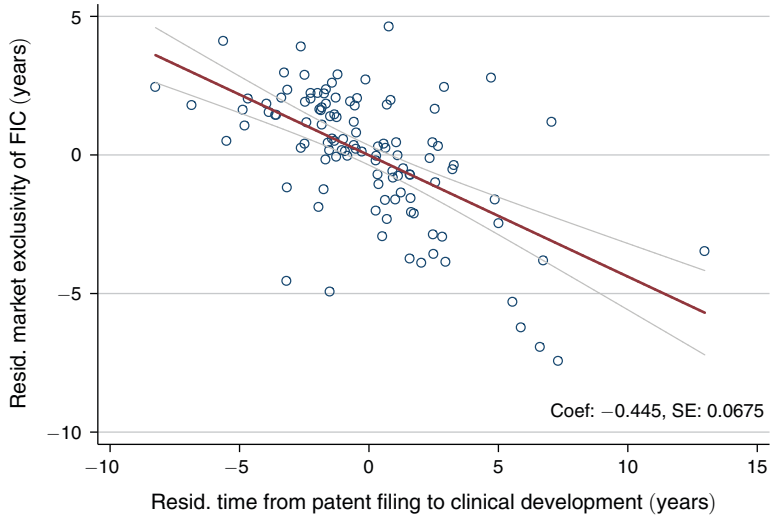


FIGURE 8. RELATIONSHIP BETWEEN MARKET EXCLUSIVITY AND TIME FROM PATENT FILING TO CLINICAL DEVELOPMENT

*Notes:* This figure shows a plot of the residualized first-stage relationship between a FIC drug's market exclusivity and the time between its patent filing and start of clinical development. Residuals are conditional on year of FIC approval fixed effects, mean development time, and market size, and the specification of the controls and sample are as described in Section II. The slope of the line of best-fit and its associated robust standard error are presented in the bottom right of the figure.

Table 3 presents the estimates, with results from Poisson models in the first set of three columns and from Poisson-IV models in the second set of three columns. Because of the Poisson's nonlinearity, I implement the IV using two-stage residual inclusion.[40] All standard errors are robust and the second-stage IV standard errors are corrected for the two-stage design using the methodology of Murphy and Topel (1985) and Hardin (2002). All specifications include fixed effects for the year of FIC approval. The first specification in each set of columns includes only the year of FIC approval fixed effects, and then in subsequent columns I incrementally add my baseline controls. Corresponding first stage estimates are reported below the IV results.

In all cases, I estimate a strong, positive, and statistically significant effect of the FIC's exclusivity on subsequent entry. The point estimates remain relatively consistent as controls are added. The Poisson-IV results are slightly larger in magnitude than the Poisson results, but they are also less precise, and the differences between the coefficients are not statistically significant. Overall, the estimates show a 1-year increase in FIC exclusivity yields a 25–30 percent increase in subsequent entry in class. Poisson models estimate proportional effects, so to ease interpretation of the results, I present in the fifth row of the table the main exclusivity coefficient multiplied by the mean of the dependent variable. These estimates, which I refer to as

[40] The Poisson's nonlinearity means that simply plugging predicted values from the first stage into the second stage yields inconsistent estimates. I implement a control function methodology as detailed in Wooldridge (2002) and Imbens and Wooldridge (2007).

TABLE 3—ESTIMATES OF THE EFFECT OF FIRST IN CLASS EXCLUSIVITY ON SUBSEQUENT ENTRY

| | Dependent variable is number of subsequent entrants in class (mean = 0.73) | | | | | |
|---|---|---|---|---|---|---|
| | QML Poisson models | | | QML Poisson-IV models | | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Market exclusivity of FIC (years) | 0.229 | 0.243 | 0.251 | 0.411 | 0.335 | 0.337 |
| | (0.0604) | (0.0631) | (0.0613) | (0.181) | (0.128) | (0.128) |
| Mean development time | No | Yes | Yes | No | Yes | Yes |
| Market size | No | No | Yes | No | No | Yes |
| Year of FIC approval fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Exclusivity coef × mean of dep variable | 0.167 | 0.177 | 0.183 | 0.300 | 0.245 | 0.246 |
| Estimates from the first stage: | | | | | | |
| Patent filing to clinical dev. (years) | – | – | – | −0.284 | −0.442 | −0.445 |
| | | | | (0.0653) | (0.0669) | (0.0675) |
| *F*-Statistic | – | – | – | 18.84 | 43.55 | 43.52 |
| Observations | 111 | 111 | 111 | 111 | 111 | 111 |

*Notes:* This table presents estimates of the effect of first in class exclusivity on subsequent entry. Columns 1–3 present results from quasi-maximum likelihood Poisson models which incrementally add controls for mean development time in class and market size. Columns 4–6 replicate columns 1–3 but instrument for FIC market exclusivity using the time between patent filing and the start of clinical development, where the start of clinical development is defined as the date on which an Investigational New Drug Application, a required precursor to the start of human clinical trials, is approved. All models are conditional on year of FIC approval fixed effects. The IV models are implemented using a control function and first-stage estimates are presented in the final rows of the table. The sample and controls are specified as described in Section II. Robust standard errors are reported in parentheses and standard errors for the IV models are corrected for the two-stage design as described in the text.

"scaled," are in units of subsequent drugs per year of first-in-class exclusivity. The estimates presented in columns 3 and 6 show one additional year of first in class exclusivity yields approximately 0.18 to 0.25 more subsequent drugs. This is a relatively large effect. For example, it implies that if FIC exclusivities were just 1 standard deviation shorter (about three years), the number of subsequent entrants in the average class would be reduced to 0. Or, in the other direction, the estimates imply that increasing FIC exclusivities by 1 standard deviation would double the number of subsequent entrants.

### D. *Robustness Tests*

Online Appendix Section F presents five sets of robustness checks that verify the robustness of my main results; these are omitted here for brevity. First, I show that the main Poisson results remain similar when the sample restrictions described in Section II are relaxed or modified. Second, I show that the main Poisson analysis survives placebo tests that analyze the role of the exclusivities of non-FIC drugs on entry. Third, I investigate the robustness of the main IV estimates. In particular, I present the reduced form of the second stage and show the IV estimates are robust to a linear specification. Fourth, since some classes in my data have had only a few years to accumulate new entrants, I show that the Poisson and Poisson-IV results are unchanged when the sample is restricted to only those classes that have had at least ten years for subsequent to entry to occur. Fifth, to ensure my results are not

confounded by class profitability, I show that the Poisson and Poisson-IV results are unchanged when I control directly for sales.

## V. Imitation or Innovation?

I have shown the exclusivity of the FIC has an important effect on subsequent entry. In this section, I analyze whether the main effect I observe is more likely to be driven by imitation or innovation. Recall that the third prediction of the model was that the FIC's patent should have a greater impact on subsequent entry when the entrant is a less significant clinical advance, while the fourth prediction was that these are specifically the circumstances under which subsequent entry is less valuable for society. In other words, the cases in which the patent matters are likely to be less socially beneficial; they are more likely to be imitation than innovation. I ask now whether whether the third prediction carries weight in the data.

Systematically quantifying differences in quality between drugs tends to be difficult; as Kesselheim, Wang, and Avorn (2013) write, prior work has looked at a variety of measures including the number of citations a drug receives in the medical literature, the number of citations its main patent receives, or simply its sales. My approach is to follow Lanthier et al. (2013) and Dranove, Garthwaite, and Hermosilla (2014) in asking whether a drug received a priority review designation from the FDA.[41]

I perform this analysis by estimating equation (6) once more, but this time I separate subsequent entrants according to whether they received priority review status. The results are reported in Table 4. The first two columns, respectively, report Poisson and Poisson-IV estimates analyzing the effect of FIC market exclusivity on the number of subsequent entrants which are priority review, while the third and fourth columns do the same for non-priority review drugs. The results in the fourth to last row, which reports the exclusivity coefficients scaled by the mean of the dependent variable, show that the estimated effect of FIC exclusivity on subsequent priority review entrants is roughly one-twentieth of that on subsequent non-priority review entrants. The differences between these effects are significant at the 1 percent level. In other words, the FIC's exclusivity only impacts subsequent entry for non-priority review drugs.[42]

Altogether, this evidence suggests the bulk of the effect I observe is driven by drugs with below average social returns; i.e., that another year of FIC's exclusivity does more to foster imitation than it does innovation. However, there are at least three reasons not to overreach with these results.

[41] A different approach would be to estimate changes in consumer surplus using a full structural model, e.g., with a framework as in Petrin (2002), who evaluates the welfare impact of the introduction of the minivan. Others have tried to estimate the welfare impact of drug introductions (e.g., Arcidiacono et al. 2013), but it is difficult to capture horizontal differences across drugs, and the value of such differences to consumers, with the data that are available.

[42] Although I focus in this section on how the impact of the FIC's exclusivity is moderated by quality differences, it can be shown that the model's third and fourth predictions are equally true for average price sensitivities, $\lambda$. In online Appendix Section G, I analyze how the main effect interacts with the price sensitivity of demand and find that, consistent with the model, the main effect is stronger when demand is more price-inelastic.

Table 4—Priority Review Designations

| | Dependent variable is number of subs. priority entrants in class | | Dependent variable is number of subs. non-priority entrants in class | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Market exclusivity of FIC (years) | 0.101 | 0.168 | 0.298 | 0.425 |
| | (0.0761) | (0.256) | −0.0779 | (0.157) |
| Mean development time | Yes | Yes | Yes | Yes |
| Market size | Yes | Yes | Yes | Yes |
| Year of FIC approval fixed effects | Yes | Yes | Yes | Yes |
| Mean of dependent variable | 0.09 | 0.09 | 0.64 | 0.64 |
| Exclusivity coef × mean of dep variable | 0.009 | 0.015 | 0.191 | 0.272 |
| Estimation | Poisson | Poisson-IV | Poisson | Poisson-IV |
| *F*-Statistic from the first stage | – | 43.52 | – | 43.52 |
| Observations | 111 | 111 | 111 | 111 |

*Notes:* This table replicates the base analysis in Table 3 but splitting new drugs by Priority-Review status, an expedited review status granted by the FDA for drugs that are expected to present significant clinical advances. Results are from Poisson and Poisson-IV models estimated by quasi-maximum likelihood. The instrument, which is the difference in time between the patent's filing date and the start of clinical development, is implemented as a control function. The sample and controls are specified as described in Section II. Robust standard errors are reported in parentheses and standard errors for the IV models are corrected for the two-stage design as described in the text.

First, priority review is a measure of what is known about drug quality prior to the drug reaching the market.[43] Drug quality is uncertain, and there certainly are examples of drugs that did not receive priority review yet have turned out to be immensely popular. For example, Singulair and Celexa were, respectively, fourth and fifth to market in their classes, and did not receive priority review. Yet both drugs' annual sales surpassed $4 billion, putting them into an elite group of top sellers.[44] By focusing on priority review as a measure of quality, my analysis does not capture any difference between expected and realized drug qualities.[45]

Second, there may be important consumer benefits to having a range of treatment options for a particular condition, even when a single drug does not stand above the rest. For example, medical research suggests that having a menu of options is particularly important in the treatment of depression and anxiety disorders because different patients respond differently to different drugs (Huskamp 2003). Insurer formularies also present evidence that variety matters, for example, when it comes to dipeptidyl peptidase-4 inhibitors, a class of drugs for diabetes in which no generics are available, Aetna's 2015 individual formulary covers both Januvia and Onglyza in its second of three tiers.[46] Neither of these drugs received priority review, yet clearly Aetna places them on the same level, and no substitutes are available in the first year. Of course, that drugs within the same class are not necessarily perfect substitutes is clearly demonstrated by Figure 1, panel B, which shows how revenues

[43] In the FDA's own words, priority review drugs are drugs which, "if approved, would provide a significant improvement in safety or effectiveness." See the FDA's guidance document, HHS (2014) for more details.

[44] This figure comes from publicly available IMS Health data available at www.drugs.com.

[45] To the extent that there are important differences between expected and realized drug qualities, it may be socially valuable to incentivize firms to pursue drugs which do not look to be blockbusters during clinical trials.

[46] See Aetna (2015).

of subsequent entrants change after the FIC goes generic. As described earlier, revenues of subsequent entrants fall on the FIC generic entry, but the figure shows they do not fall to zero. In fact, prices of subsequent entrants increase shortly before the FIC goes generic, suggesting that the subsequent entrants are pricing to the marginal consumer who benefits from differentiation (online Appendix Figure D.2 visualizes price dynamics). The analysis presented in this section does not fully capture the benefits of horizontal differentiation, and these benefits may be substantial.

Third, my very unit of analysis—pharmacologic classes—may limit the extent to which I can measure the effect of the FIC's generic on genuine subsequent innovation. The FDA purposefully defines classes narrowly, based on a combination of mechanism of action, physiologic effect, and chemical structure. That classes are narrow ensures that drugs within the same class are substitutes. But of course, in limiting my analysis to sets of relatively close substitutes, I am unable to measure the possibility that generic entry in one class deters real innovation in another.

## VI. Conclusion

This paper shows that patents have important effects on the direction of pharmaceutical innovation. Specifically, I show that the length of an incumbent's patent protection has a positive impact on subsequent entry: an extra year of FIC exclusivity leads to an estimated 25–30 percent increase in subsequent drug approvals in the same class, which equates to about 0.2 drugs.

This fact has implications for how researchers and policymakers weigh the benefits and disadvantages of patents. A simple framework that determines the optimal level of patent protection by comparing a single would-be innovator's incentives with the (subsequent) social cost imposed by that innovator's patent protection fails to account for the dynamic incentives that the initial innovator's patent protection provides to subsequent entrants. This paper shows that properly weighing the pros and cons of patent protection means taking into account not only how a single firm's patents impact its own incentives but also how those incentives are shaped by other firms' patents.

Furthermore, I analyze how an incumbent's patent protection influences the composition of innovation. Consistent with theory, I find the effect of the incumbent's patent protection is concentrated among drugs that the FDA classified at the time of approval as lesser clinical advances, suggesting that these drugs may have lower than average returns to society. However, there is also significant uncertainty within the process of pharmaceutical development and the benefits of more entry to consumers are frequently substantial, so it is unlikely that the affected R&D spending is purely wasteful. How large the benefits are, and how they compare to the costs, remain topics for future research.

More generally, this paper has implications for how intellectual property is characterized. Traditional research identifies two independent levers in patent design: length—how long a patent remains in force, and breadth—the extent to which a patent "covers" a field. This paper shows that this characterization is too simple because it fails to account for the fact that length and breadth are linked through elasticities of substitution. For example, if increasing the length of an incumbent's patent

protection increases the value of entry for differentiated substitutes, then shorter patents are, effectively, more broad.[47] Deeper understanding of this competition-induced connection between length and breadth may yield new insights into how intellectual property protection should be designed by the "economist-as-engineer" (Roth 2002) to shape innovation.

Finally, the focus in this paper has been on small molecule drugs and the paper's logic hinges on the fact that generics are chemically equivalent to approved drugs, yet they are cheap to introduce once patents have expired. However, the logic of the paper may apply to other fields, from biologic drugs (which are more considerably more complex at the molecular level, and do not yet have a notion of "generic" established) to creative media (as described briefly in the introduction). My work suggests there is considerable nuance to fully understanding the impact of intellectual property protection on present and future innovation.

## REFERENCES


**Acemoglu, Daron, and Joshua Linn.** 2004. "Market Size in Innovation: Theory and Evidence from the Pharmaceutical Industry." *Quarterly Journal of Economics* 119 (3): 1049–90.

**Ackerberg, Daniel, C. Lanier Benkard, Steven Berry, and Ariel Pakes.** 2007. "Econometric Tools for Analyzing Market Outcomes." In *Handbook of Econometrics*, Vol. 6A, edited by James J. Heckman and Edward E. Leamer, 4171–4276. Amsterdam: North-Holland.

**Aetna.** 2015. "2015 Aetna Pharmacy Plan Drug List: Three Tier Open Individual Plan." https://fm.formularynavigator.com/MemberPages/pdf/2015AetnaThreeTierIndividualOpenNJFormulary_4932_Full_757.pdf.

**Aitken, Murray, Ernst R. Berndt, and David M. Cutler.** 2009. "Prescription Drug Spending Trends In The United States: Looking Beyond The Turning Point." *Health Affairs* 28 (1): 151–60.

**Alvir, Jose Ma J., Jeffrey A. Lieberman, Allan Z. Safferman, Jeffrey L. Schwimmer, and John A. Schaaf.** 1993. "Clozapine-Induced Agranulocytosis." *New England Journal of Medicine* 329 (3): 162–67.

**Arcidiacono, Peter, Paul B. Ellickson, Peter Landry, and David B. Ridley.** 2013. "Pharmaceutical followers." *International Journal of Industrial Organization* 31 (5): 538–53.

**Associated Press.** 2011. "Lipitor becomes world's top-selling drug." *Crain's New York Business*, December 28. http://www.crainsnewyork.com/article/20111228/HEALTH_CARE/111229902/lipitor-becomes-worlds-top-selling-drug.

**Berndt, Ernst R.** 2002. "Pharmaceuticals in U.S. Health Care: Determinants of Quantity and Price." *Journal of Economic Perspectives* 16 (4): 45–66.

**Bessen, James, and Eric Maskin.** 2009. "Sequential innovation, patents, and imitation." *RAND Journal of Economics* 40 (4): 611–35.

**Branstetter, Lee, Chirantan Chatterjee, and Matthew J. Higgins.** 2014. "Killing the Golden Goose or Just Chasing it Around the Farmyard? Generic Entry and the Incentives for Early-Stage Pharmaceutical Innovation." http://funginstitute.berkeley.edu/wp-content/uploads/2013/12/Golden_Goose.pdf.

**Budish, Eric, Benjamin N. Roin, and Heidi Williams.** 2015. "Do Firms Underinvest in Long-Term Research? Evidence from Cancer Clinical Trials." *American Economic Review* 105 (7): 2044–85.

**Bulow, Jeremy I., John D. Geanakoplos, and Paul D. Klemperer.** 1985. "Multimarket Oligopoly: Strategic Substitutes and Complements." *Journal of Political Economy* 93 (3): 488–511.

**Civan, Abdulkadir, and Michael T. Maloney.** 2009. "The Effect of Price on Pharmaceutical R&D." *B.E. Journal of Economic Analysis and Policy* 9 (1).

**Cockburn, Ian M.** 2007. "Is the Pharmaceutical Industry in a Productivity Crisis?" In *Innovation Policy and the Economy*, Vol. 7, edited by Adam B. Jaffe, Josh Lerner, and Scott Stern, 1–32. Cambridge: MIT Press.


[47] This line of reasoning bears some resemblance to that articulated by O'Donoghue, Scotchmer, and Thisse (1998), who argue that in certain cases, what matters is not true patent life but rather the length of time a product has on patent until a new product is developed which completely displaces the old product. They call this length of time "effective" patent life. However, they do not consider how the development of the new product might actually be deterred by the expiration of the old product's patent, i.e., the subject of this paper.

**Deloitte.** 2009. "Acquisitions versus Product Development: An Emerging Trend in Life Sciences." (accessed January 12, 2014) www.deloitte.com/us/acquisitionsvsproductdevelopment.

**DiMasi, Joseph A., Ronald W. Hansen, and Henry G. Grabowski.** 2003. "The price of innovation: New estimates of drug development costs." *Journal of Health Economics* 22 (2): 151–85.

**Dranove, David, Craig Garthwaite, and Manuel Hermosilla.** 2014. "Breakthrough or Me-Too? The Impact of Medicare Part D on Biotech Innovation." Unpublished.

**Ellison, Glenn, and Sara Fisher Ellison.** 2011. "Strategic Entry Deterrence and the Behavior of Pharmaceutical Incumbents Prior to Patent Expiration." *American Economic Journal: Microeconomics* 3 (1): 1–36.

**Finkel, Rhona.** 2012. "The Atypical History of Atypical Antipsychotics." *Drug Information and Side Effects Database*, September 9. http://www.drugsdb.com/blog/history-of-atypical-antipsychotics.html.

**Fudenberg, Drew, Richard Gilbert, Joseph Stiglitz, and Jean Tirole.** 1983. "Preemption, leapfrogging and competition in patent races." *European Economic Review* 22 (1): 3–31.

**Gallini, Nancy T.** 1992. "Patent Policy and Costly Imitation." *RAND Journal of Economics* 23 (1): 52–63.

**Gilbert, Richard, and Carl Shapiro.** 1990. "Optimal Patent Length and Breadth." *RAND Journal of Economics* 21 (1): 106–12.

**Gilchrist, Duncan S.** 2016. "Patents as a Spur to Subsequent Innovation? Evidence from Pharmaceuticals: Dataset." *American Economic Journal: Applied Economics*. http://dx.doi.org/10.1257/app.20150373.

**Griliches, Zvi.** 1998. "Patent Statistics as Economic Indicators: A Survey." In *R&D and Productivity: The Econometric Evidence*, edited by Zvi Griliches, 287–343. Chicago: University of Chicago Press.

**Hall, Bronwyn H., Adam Jaffe, and Manuel Trajtenberg.** 2005. "Market Value and Patent Citations." *RAND Journal of Economics* 36 (1): 16–38.

**Hardin, James W.** 2002. "The robust variance estimator for two-stage models." *Stata Journal* 2 (3): 253–66.

**Hausman, Jerry A., Bronwyn H. Hall, and Zvi Griliches.** 1984. "Econometric Models for Count Data with an Application to the Patents-R&D Relationship." *Econometrica* 52 (4): 909–38.

**Hay, Michael, Jesse Rosenthal, David Thomas, and John Craighead.** 2011. "BIO/BioMedTracker: Clinical Trial Success Rates Study." Paper presented at the BIO CEO and Investor Conference, The Waldorf Astoria, New York, NY, February 15.

**Hemphill, C. Scott.** 2006. "Paying for Delay: Pharmaceutical Settlement as a Regulatory Design Problem." *New York University Law Review* 81: 1553–1623.

**Hemphill, C. Scott.** 2007. "Drug Patent Settlements Between Rivals: A Survey." Unpublished.

**Hemphill, C. Scott, and Bhaven N. Sampat.** 2012. "Evergreening, Patent Challenges, and Effective Market Life in Pharmaceuticals." *Journal of Health Economics* 31 (2): 327–39.

**Hoi, Albert A., and Geoffrey Littlejohn.** 2005. "Leflunomide." In *Antirheumatic Therapy: Actions and Outcomes Progress in Inflammation Research*, edited by Richard O. Day, Daniel E. Fürst, Piet L. C. M. van Riel, and Barry Bresnihan, 199–219. Basel, Switzerland: Birkhäuser Verlag.

**Huskamp, Haiden A.** 2003. "Managing Psychotropic Drug Costs: Will Formularies Work?" *Health Affairs* 22 (5): 84–96.

**Imbens, Guido W., and Jeffrey M. Wooldridge.** 2007. "Control Function and Related Methods." Presented at "What's New in Econometrics?" NBER Summer Institute Econometric Lectures, Harvard University, July 31.

**Jacobo-Rubio, Ruben, John L. Turner, and Jonathan Williams.** 2014. "The Private Value of Entry and Deterrence in the US Pharmaceutical Industry." https://media.terry.uga.edu/socrates/contact/documents/2014/01/24/value-entry-deterrence_Jan-11-2014.pdf.

**Kesselheim, A. S., B. Wang, and J. Avorn.** 2013. "Defining 'innovativeness' in drug development: A systematic review." *Clinical Pharmacology and Therapeutics* 94 (3): 336–48.

**Keyhani, Salomeh, Marie Diener-West, and Neil Powe.** 2006. "Are Development Times for Pharmaceuticals Increasing or Decreasing?" *Health Affairs* 25 (2): 461–68.

**Klemperer, Paul.** 1990. "How Broad Should the Scope of Patent Protection Be?" *RAND Journal of Economics* 21 (1): 113–30.

**Lanthier, Michael, Kathleen L. Miller, Clark Nardinelli, and Janet Woodcock.** 2013. "An Improved Approach to Measuring Drug Innovation Finds Steady Rates of First-in-Class Pharmaceuticals, 1987–2011." *Health Affairs* 32 (8): 1433–39.

**Lerner, Josh.** 2002. "Patent Protection and Innovation over 150 Years." National Bureau of Economic Research (NBER) Working Paper 8977.

**Lichtenberg, Frank R., and Tomas J. Philipson.** 2002. "The Dual Effects of Intellectual Property Regulations: Within- and Between-Patent Competition in the U.S. Pharmaceuticals Industry." *Journal of Law and Economics* 45 (S2): 643–72.

**Martin, Scott M.** 2002. "The Mythology of the Public Domain: Exploring the Myths Behind Attacks on the Duration of Copyright Protection." *Loyola of Los Angeles Law Review* 36 (1): 253–322.

**Moser, Petra.** 2005. "How Do Patent Laws Influence Innovation? Evidence from Nineteenth-Century World's Fairs." *American Economic Review* 95 (4): 1214–36.

**Murphy, Kevin M., and Robert H. Topel.** 1985. "Estimation and Inference in Two-Step Econometric Models." *Journal of Business and Economic Statistics* 3 (4): 370–79.

**Murphy, Kevin M., and Robert H. Topel.** 2006. "The Value of Health and Longevity." *Journal of Political Economy* 114 (5): 871–904.

**Newhouse, Joseph P., and the Insurance Experiment Group.** 1993. *Free for all? Lessons from the RAND Health Insurance Experiment.* Cambridge: Harvard University Press.

**Nordhaus, William D.** 1967. "The Optimal Life of a Patent." Cowles Foundation for Research in Economics Discussion Papers 241.

**Nordhaus, William D.** 1969. "An Economic Theory of Technical Change." *American Economic Review* 59 (2): 18–28.

**Nordhaus, William D.** 1972. "The Optimal Life of a Patent: Reply." *American Economic Review* 62 (3): 428–31.

**National Science Board** (**NSB**). 2012. *Science and Engineering Indicators 2012.* Arlington: National Science Foundation.

**O'Donoghue, Ted, Suzanne Scotchmer, and Jacques-François Thisse.** 1998. "Patent Breadth, Patent Life, and the Pace of Technological Progress." *Journal of Economics and Management Strategy* 7 (1): 1–32.

**Oh, Jiwon, and Paul W. O'Connor.** 2013. "An Update of Teriflunomide for Treatment of Multiple Sclerosis." *Therapeutics and Clinical Risk Management* 9: 177–90.

**Pammolli, Fabio, Laura Magazzini, and Massimo Riccaboni.** 2011. "The productivity crisis in pharmaceutical R&D." *Nature Reviews Drug Discovery* 10 (6): 428–38.

**Panattoni, Laura E.** 2011. "The effect of Paragraph IV decisions and generic entry before patent expiration on brand pharmaceutical firms." *Journal of Health Economics* 30 (1): 126–45.

**Parker, Ian.** 2013. "The Big Sleep." *New Yorker*, December 9. http://www.newyorker.com/magazine/2013/12/09/the-big-sleep-2.

**Petrin, Amil.** 2002. "Quantifying the Benefits of New Products: The Case of the Minivan." *Journal of Political Economy* 110 (4): 705–29.

**Reiffen, David, and Michael R. Ward.** 2005. "Generic Drug Industry Dynamics." *Review of Economics and Statistics* 87 (1): 37–49.

**Roth, Alvin E.** 2002. "The Economist as Engineer: Game Theory, Experimentation, and Computation as Tools for Design Economics." *Econometrica* 70 (4): 1341–78.

**Sakakibara, Mariko, and Lee Branstetter.** 2001. "Do Stronger Patents Induce More Innovation? Evidence from the 1998 Japanese Patent Law Reforms." *RAND Journal of Economics* 32 (1): 77–100.

**Sampat, Bhaven, and Heidi L. Williams.** 2015. "How Do Patents Affect Follow-On Innovation: Evidence From the Human Genome." National Bureau of Economic Research (NBER) Working Paper 21666.

**Scannell, Jack W., Alex Blankley, Helen Boldon, and Brian Warrington.** 2012. "Diagnosing the decline in pharmaceutical R&D efficiency." *Nature Reviews Drug Discovery* 11 (March): 191–200.

**Scherer, F. M.** 1972. "Nordhaus' Theory of Optimal Patent Life: A Geometric Reinterpretation." *American Economic Review* 62 (3): 422–27.

**Scotchmer, Suzanne.** 1991. "Standing on the Shoulders of Giants: Cumulative Research and the Patent Law." *Journal of Economic Perspectives* 5 (1): 29–41.

**U.S. Department of Health and Human Services (HHS)**. 2014. "Guidance for Industry: Expedited Programs for Serious Conditions—Drugs and Biologics." U.S. Department of Health and Human Services Food and Drug Administration. Silver Spring, MD, May.

**World Health Organization (WHO)**. 2004. "The World Medicines Situation." http://apps.who.int/medicinedocs/pdf/s6160e/s6160e.pdf.

**Wooldridge, Jeffrey M.** 2002. *Econometric Analysis of Cross Section and Panel Data.* Cambridge: MIT Press.

***Wall Street Journal*** **(WSJ)**. 1998. "FDA Advisory Panel Recommends Approving Hoechst Arthritis Drug." *Wall Street Journal*, August 7. http://www.wsj.com/articles/SB902522891901774000.