Innovation, Litigation, and New Drugs

Ramsi A. Woodcock*

August 30, 2016

Abstract

A generic drug maker that has been sued for infringing a patent on a branded drug will sometimes promise, as part of an agreement settling the litigation, to delay selling its drug until as late as the expiration of the patent term. I adapt the standard optimal patent term model to determine whether the delay in competition caused by such agreements raises consumer welfare by increasing rewards for innovation. Calibrating the model with U.S. drug market data, I find that, for all but the patents with the greatest probability of surviving litigation, settlements that delay entry by more than fifteen months harm consumers. I find that generic drug makers are in some cases willing to agree to enough delay to harm consumers, even in the absence of a reverse payment, and that harm can reach $1.3 billion for the average drug. JEL Codes: K41; L41; L65; O34.

*Assistant Professor, Department of Risk Management and Insurance, Robinson College of Business, Secondary Appointment, College of Law, Georgia State University. Email: rwoodcock@gsu.edu. I am grateful to Kangoh Lee, F.M. Scherer, Glenn W. Harrison, Alan G. Isaac, Walter G. Park, Maria Arbatskaya, and participants at the 2015 Roundtable for Engineering Entrepreneurship Research (REER) at Georgia Tech for comments on this work.

1 Introduction

U.S. law permits a generic drug maker that believes the patents protecting a branded drug are invalid to bring a generic version of the drug to market before expiration of the official lives of the patents. The branded drug maker that holds the patents may sue to block the generic's entry. Branded makers sometimes choose to settle this litigation by signing settlement agreements with generic makers that obligate the generic makers to stay out of the market for a period of time, sometimes until patent expiry. Hemphill (2009) lists settlements of this kind for a number of blockbuster drugs, perhaps the most famous of which is the heart drug Lipitor.

There can be no objection to these settlements if the branded maker's ("Brand's") patents are in fact valid. But if they are not, then the agreements amount to collusion between drug makers to deprive consumers of competing products and lower prices, in violation of U.S. antitrust law. In some of these "entry settlements," Brand makes a payment, sometimes in the hundreds of millions of dollars, in exchange for Generic's agreement to defer entry. These payments are particularly suggestive of collusion because, if Brand's patents were valid, one would expect any payment to flow in the opposite direction, from Generic to Brand, as compensation for infringement. Federal Trade Commission (2010) therefore labels these payments "pay for delay" and even their defenders, such as Harris et al. (2014), call them "reverse payments."

Lemley (2001) observes that the U.S. Patent and Trademark Office puts limited effort into reviewing patent applications before approving them. As a result, many patents underlying entry settlements may be invalid; indeed, Greene and Steadman (2010) find a 52% litigation success rate for patentholders in the drug market. It is for this reason that the U.S. Federal Trade Commission ("FTC") has sought to use antitrust law to block entry settlements, at least when they involve a reverse payment. After a string of lower court rulings to the contrary, U.S. Supreme Court (2013) accepted the FTC's argument that reverse payment settlements can violate antitrust law; however, the Court refused to ban them outright,

preferring to subject them instead to case-by-case review for harm to consumers.

Detractors of reverse payment settlements, such as Edlin et al. (2015), argue that they harm consumers by extending Brand's market exclusivity beyond the point at which Brand would be forced to compete with a generic entrant ("Generic") if Brand were to lose its patent suit against Generic. On this account, longer exclusivity leads to higher prices, which reduce consumer welfare. Using this approach, Federal Trade Commission (2010) estimates that reverse payment settlements cost consumers $35 billion over ten years. Supporters of the settlements, such as Kobayashi et al. (2015) and Langenfeld and Li (2003), argue that they make consumers better off by increasing the reward to Brand for creating new drugs. On this account, such "dynamic" consumer welfare benefits arise because firms that can expect to enjoy exclusivity and higher prices for a longer period of time have more to gain from making better drugs, and will therefore invest in doing so. This in turn drives up the value of new drugs for consumers, making them better off, even after adjusting for the losses they suffer from higher prices.

The dynamic argument amounts to the claim that the effective period of exclusivity enjoyed by Brand in the absence of an entry settlement is too short. But no study has yet actually built a dynamic model and attempted to measure the effect of gains from innovation on consumer welfare in these settlements. This article fills that gap. I adapt the standard optimal patent term ("OPT") model, as presented by Denicolò (2007), to patent entry settlements, and calibrate it with drug market data. I find that even after taking the innovation incentive effects of higher prices into account, settlements that delay entry by little more than a year harm consumers for all patents except those that are most likely to be upheld anyway in court. I consider both the case in which only a small number of generic makers are able to enter the market, leading to duopoly-like conditions until patent expiry (the period of "capped duopoly" case), and the case in which entry leads to a short fixed period of duopoly, followed by full entry and competition (the period of "fixed duopoly" case). I find that in the capped duopoly case consumer harm from these settlements is

relatively low, on the order of millions of dollars. I find that in the fixed case, consistent with the prediction of Hemphill (2006), it is much larger, on the order of billions of dollars. Under fixed duopoly, delay in entry deprives consumers of a competitive, rather than just a duopoly, market.

I follow Shapiro (2003) in referring to the probability that a patent will prevent entry of competitors until the expiration of its term as the patent's "strength." An invalid patent has a low strength because it will not be effective at barring entry by competitors before the expiration of its term unless a judge mistakenly upholds it. Woodcock (2016) argues that policymakers should pretend that patent law chooses patent strength to maximize consumer welfare. I show that settlements that delay entry by more than 15 months, relative to the date of entry implied by this welfare-maximizing strength, harm consumers. This result is important because it shows that, even after gains from innovation are taken into account, entry settlements can still harm consumers. I also show that in some cases Generic is willing to agree to enough delay to harm consumers, even in the absence of a reverse payment. This result suggests that the focus of enforcers and courts only on entry settlements that involve a reverse payment is misplaced.

Sensitivity analysis in this model yields few insights. One area of clarity, however, is the role of the portion of the consumer welfare created by a drug that consumers enjoy regardless of the amount of delay in entry, which I call "monopoly consumer welfare." I show that when it is large, delay is unlikely to harm consumers. The stronger a patent, the less likely it is that a settlement that delays entry will harm consumers, because the patent would be upheld anyway in court. Monopoly consumer welfare is large when a drug has high value to consumers, even when little has been invested in its improvement. For such drugs, small improvements create large welfare gains, justify greater patent strength, and therefore imply a lower likelihood of harm through delay. My results are therefore sensitive to my choice of elasticity of demand for drugs, which I use to estimate monopoly consumer welfare. In my calibration, I use an elasticity of -0.35 to estimate $1.7 billion in monopoly consumer welfare

for the average drug.

The mere fact of settlement of patent disputes itself has been celebrated as offering benefits to consumers, quite independently of any benefits arising from extension of the period of exclusivity. Shapiro (2003) argues that by allowing Brand and Generic to achieve compromises between the extremes of patent validity and invalidity attainable under litigation, settlement is always capable of making consumers better off. For example, by agreeing to a licensing fee, the parties might indirectly set a market price for a drug for the entire patent term that is intermediate between the extremes of monopoly and competitive pricing. This compromise price, unobtainable under a litigated outcome, could raise consumer welfare. I show in the Appendix that such gains do not exist when the variable fixed through settlement is the date of entry, as opposed to price. The only drivers of consumer benefits from entry settlements in my model are therefore gains from extension of exclusivity and gains from the avoidance of litigation cost.

My results suggest that entry settlements can harm consumers, even after returns to innovation are taken into account. This has policy significance, as Woodcock (2016) argues that in the patent context the mere possibility of harm implies that patent entry settlements should be banned under antitrust law. However, I do not wish to suggest that my results here settle the question whether entry settlements harm consumers. Instead, I mean to show that it cannot be taken for granted that reverse payment settlements benefit consumers if one accounts for innovation. Further research both to improve estimates of model parameters and better to adapt the model to the peculiarities of the drug market is required before research in this area can move from sowing doubt to reaping certainty.

I provide some background on the drug regulatory process, describe my model, adapt it to the U.S. drug market, and then discuss my results. Although I focus on settlements that involve no reverse payment, I also use my model to estimate the magnitude of such payments and the additional harm they may cause.

2 Background

Brand and Generic strike patent entry settlements against the background of a highly-regulated drug development process. The U.S. Food and Drug Administration ("FDA") must determine, based on clinical trials, that a new drug is safe and effective before approving it for sale.[1] Typically, a drug company does basic research to identify a promising compound, then patents it and commences clinical trials. If the compound fails the trials, the process repeats until the FDA grants approval.

If the drug is based on a new compound (i.e., it meets the definition under the law of a "new chemical entity"), then the Hatch-Waxman Act guarantees unchallengeable exclusivity in marketing to the drug maker for five years. Thereafter, if there is life remaining on the patent, generic drug companies may challenge the patent in court. Such challenges are sometimes referred to as "Paragraph IV challenges" after the portion of the Hatch-Waxman Act that regulates them. The Hatch-Waxman Act provides that the patentholder can obtain a stay of 30 months on entry by the challenger during patent litigation arising from a Paragraph IV challenge. Thus the exclusivity permitted by the Hatch-Waxman Act may extend to 7.5 years if litigation is not resolved before the end of the stay. Thereafter, the challenger may enter if it wins its case. I assume away the existence of these exclusivity periods in the initial exposition of my model, but add them in later.

The Hatch-Waxman Act permits the first Paragraph IV challenger 180 days of exclusivity with respect to other generic makers that might want to enter the market. It is possible for several generic makers to challenge a patent at the same time, but usually no more than three do so. Thus in the period after challenge-based entry there are usually no more than three generics in the market, as compared to many more upon patent expiry and, possibly, upon expiration of the 180-day first-filer exclusivity period.[2] The capped duopoly form of

[1] The discussion in this section is based on Kelly (2011), Grabowski and Kyle (2007), and DiMasi et al. (1991).

[2] Exclusivity rules are different for biologics. As discussed by Grabowski et al. (2011), it is not clear whether in practice the different exclusivity rules translate into different exclusivity durations.

my model captures the former case and the fixed duopoly form captures the latter.

3 The Model

3.1 Overview

I start with the OPT model of Denicolò (2007), which was pioneered by Nordhaus (1969) and refined by Gilbert and Shapiro (1990), among many others. In it, innovators invest in research and development ("R&D") effort Y in order to increase the potential value of a product. Although Denicolò (2007) and Nordhaus (1969) focus on improvements that increase value by reducing cost, I consider the more common case of improvements that increase demand, which better fits my new drug development context. I follow Denicolò (2007) in assuming that Y has constant marginal cost of α and gives the average probability that a new drug with value V will be created in a given unit of time. As Y increases, the probability that the new drug will be created sooner rather than later increases, driving up the value of the drug in present value terms. This Poisson innovation production function obeys the general OPT model requirement that there be diminishing returns in Y.

The incentive to invest in Y as well as the benefits of the drug to consumers are determined by the way in which the value created by the drug is divided between firms and consumers. In the OPT model, it is standard to assume that the successful innovator obtains monopoly profits on its creation for the duration of the patent term, T, after which competition drives price and profit to zero. Meanwhile, consumers enjoy limited value during the patent term and the full value of the new product only after patent expiry. The longer the patent term, the greater Brand's reward for innovation and therefore the more Brand may invest in its creation. But the longer the term, the fewer the fruits enjoyed by consumers. A balance must be struck. The problem in the OPT model is to find the duration of the patent term that maximizes consumer welfare.[3]

[3]Often the focus is on total, rather than consumer, welfare. I share with Salop (2009) the view that the

I am interested in consumer welfare resulting from entry dates set through litigation and settlement, instead of in the optimal duration of the formal patent term. I call the duration of exclusivity ordered by a court after litigation T_E^{lit}. It can be either zero, if Brand loses, or T, if Brand wins. Every patent has a probability p of being upheld in litigation, so the expected patent term is $E[T_E^{lit}] = pT$, which is less than or equal to the full patent term. I call pT the "litigation entry date." If entry is followed by duopoly until patent expiry, then, ex ante, Brand enjoys monopoly profit for period pT and duopoly profit for period $(1-p)T$, after which profit falls to zero.[4] When Brand and Generic settle, they choose an exclusivity duration T_E^{set}, which I loosely refer to as the "settlement entry date," after which Generic enters the market. As in the case of litigation, entry is followed by duopoly until patent expiry, thus Brand enjoys monopoly profits for period T_E^{set} followed by duopoly profits for period $T - T_E^{set}$, and nothing thereafter. I use the term "duopoly" loosely; sometimes two or three generic makers enter at once after a challenge or settlement. The important thing is that not so many enter as to drive price immediately to competitive levels. By "Generic" I mean one, or a small number of, generic drug makers.

Consumer and firm welfare are determined ex ante in my model at the same instant at which Brand decides on and executes a one-off lump sum investment in Y. If settlement is not possible and Brand must litigate, Brand decides how much Y to purchase by determining the present value of its monopoly and duopoly profits given a litigation entry date pT. If Brand knows that it will settle for entry on a particular date, Brand uses that date instead in deciding how much Y to purchase. A period of R&D ensues until there is success and the drug is created. If the patent is litigated, then Brand incurs a lump sum litigation cost L instantaneously at the drug creation date, and monopoly and duopoly profits immediately start to flow in. If the litigation is settled, then it settles instantaneously at the drug creation date at zero cost (i.e., there is no litigation cost), and profits immediately start to flow in.

current goal of antitrust is the maximization of consumer welfare. Accordingly, I consider only consumer welfare here.

[4] I reserve discussion of the fixed duopoly case for Section 3.5.

The problem in which I am interested is determining the threshold T_E^{set} beyond which consumer welfare in settlement is less than consumer welfare in litigation at pT. From this, I can determine how much delay in entry a settlement may bring about without harming consumers. To solve the problem I must determine p. Shapiro (2003) suggests that p is determined in part by patent law through its choice of rules. If it wants to increase p, for example, patent law might weaken the requirement that an invention be nonobvious in order to be patentable. Woodcock (2016) and Elhauge and Krueger (2012) argue that policymakers must assume that patent law chooses p to maximize consumer welfare. I therefore find the p that maximizes consumer welfare, call it p^{max}, and use it to find the threshold for T_E^{set}. I call the associated delay in entry in settlement relative to the litigation entry date, $T_E^{set} - p^{max}T$, "permissible delay." Delay beyond this level harms consumers.

I consider both the case in which Brand has the power to choose Y and the case, pioneered by Scherer (1967) and outlined in Denicolò (2007), of a patent race. In this latter scenario, firm i in the patent race contributes y_i in innovation effort and Y is the aggregate effort of all firms in the race. Each firm has probability $\frac{y_i}{Y}$ that it will be the first to succeed at creating and patenting the new drug. Thus individual firms expect average returns from investing in innovation even though the patent racers as a group experience diminishing returns, as required by the Poisson innovation production function. This drives Y above the level that a monopolist would choose and expected profits to zero. I find that my drug market estimates do not support the existence of a monopoly in research and so I focus on this patent race case.

3.2 Structure

As shown in Figure 1, firms spend αY on R&D at time $\tau = 0$; a new drug is created at some $\tau = \mathcal{T}$; at $\mathcal{T}$, a single firm receives a patent on the new drug that allows it to exclude all competitors from the market for patent term T. Brand is that single firm. If it litigates, Brand spends L in litigation costs at that same moment. If it settles, there is no cost.

Starting immediately, value flows in to Brand, Generic, and consumers over time $t \in [0, \infty)$, for which I take $t = 0$ to occur at $\tau = \mathcal{T}$.

Let T_E be the duration of monopoly pricing, measured from $t = 0$, the date of innovation, until Generic enters. In terms of τ, entry occurs at $\mathcal{T} + T_E$. When it represents the entry date in litigation, T_E is the random variable T_E^{lit} that can either take on the value zero or the value T, because courts rule patents either invalid or valid. They do not fix particular intermediate dates of entry. When T_E represents the entry date in settlement, it is the deterministic variable T_E^{set} and can take on any value between zero and T, inclusive, because the parties can settle for any entry date up to patent expiry. Settling for a date after T is an illegal private extension of the patent term. T_E is deterministic in settlement because I assume that policymakers may fix it and that firms know the T_E that has been fixed.

Let V be the total welfare created by a drug, as determined by a partial equilibrium analysis that ignores innovation. V_p is the portion of that value distributed to producers and V_c is what remains. To capture the traditional antitrust concern that monopoly harms consumers and reduces total welfare, an increase in T_E must drive V_c down by more than it increases V_p, reducing V. Because V_c ignores innovation, I call it "static welfare." My goal is to find expected consumer welfare,

$$\mathcal{V}_c \equiv E[f(w)|Y]E[V_c(T_E)], \tag{1}$$

in litigation and in settlement. w is a random innovation state variable and f is a nonnegative function that scales static welfare based on the innovation state. Y determines the probability that a given w will appear. To capture the positive, diminishing-returns, effect of R&D on welfare, $E[f(w)|Y]$ must increase in Y at a decreasing rate. I assume that T_E^{lit} is independent of the innovation state, w, which explains the separability of the expectations in (1).[5]

Firms in the patent race choose Y. In a competitive race, aggregate profits are zero. This condition determines firms' choice of Y. I therefore solve the zero-profit condition for

[5]I also assume risk neutrality throughout.

aggregate ex ante profits for firms in the patent race,

$$E[f(w)|Y]\left(E[v^p]-L\right)-\alpha Y=0, \tag{2}$$

for $E[f(w)|Y]$. $v^p < V_p$ is the static private value that the firm that wins the patent race enjoys; like V_p, it is a function of T_E. I use it in (2) instead of V_p because Generic takes part of V_p after it enters and generic drug makers do not typically compete in the patent race, so firms do not expect to enjoy Generic's profits. In the case of litigation, the static value enjoyed by the successful patent racer must be reduced by litigation cost L. L is zero in the case of settlement. Because firms choose Y, it is their expectations regarding w and T_E that determine (2). I assume that firms all have the same expectations and that I know what they are.

In the case of a monopoly in research, the monopolist chooses Y to maximize profit, and I solve the following first order condition, instead of (2), for $E[f(w)|Y]$:

$$\frac{dE[f(w)|Y]}{dY}\left(E[v^p]-L\right)=\alpha. \tag{3}$$

I call the $E[f(w)|Y]$ determined by either (2) or (3) x. In the Appendix, I show that x is increasing in T_E, v^p, and $E[v^p]$. These results capture the incentive effect of market exclusivity. As the duration of exclusivity increases, the expectation of quasi-profit resulting therefrom increases, causing firms to increase their R&D effort, and thereby to increase x. I sometimes refer to x as "the dynamic factor." Substituting x into (1), I expect welfare

$$\mathcal{V}_c^* \equiv xE[V_c]. \tag{4}$$

Settlements are free, so they permit Brand to save on litigation costs. I follow Willig and Bigelow (2004) and Edlin et al. (2015) in assuming that in monopoly and duopoly

product markets litigation costs are not passed on directly to consumers.[6] Litigation costs nevertheless reduce consumer welfare indirectly by reducing innovation rewards for firms. I define $\mathcal{V}_c^{*L}$ to be $\mathcal{V}_c^*$ when $T_E = T_E^{lit}$ and L is nonnegative and $\mathcal{V}_c^{*S}$ when $T_E = T_E^{set}$ and L is zero. Brand pays its litigation costs, L, out of v^p. I show in the Appendix that x is falling in L. Because litigation costs are not paid directly by consumers, V_c is unchanging in L. It follows from (4) that $\mathcal{V}_c^{*L}$ is falling in L.

As Denicolò (2007) describes, the OPT model usually defines the innovation state as time $\tau \in [0, \infty)$ (i.e., $w = \tau$) and $f(w)$ as $e^{-r\tau}$, where r is a discount rate that is shared by consumers and firms. It assumes that the probability of invention success follows a Poisson distribution and takes Y to be the invention success rate per unit time. I therefore have

$$E[f(w)|Y] = \frac{Y}{Y + r} \tag{5}$$

for the dynamic factor.[7]

Plugging (5) into (2) and (3) and solving for $\frac{Y}{Y+r} = x$, I obtain

$$x = 1 - \left[\frac{\alpha r}{E[v^p] - L}\right]^n. \tag{6}$$

$n = 1$ in the competitive patent race case and $n = \frac{1}{2}$ in the research monopoly case.

I now give more structure to v^p and V_c. I follow the approach of Denicolò (2007) in defining static value. The total potential value to consumers per unit time created by the invention is divided into consumer welfare per unit time under monopoly pricing in the drug market, $CS_M > 0$, monopoly profit per unit time, $\pi > 0$, and monopoly deadweight loss per unit time, $D > 0$. Under the monopoly that prevails for period T_E, consumers enjoy CS_M

[6] Indeed, because the additional generic makers that enter after patent expiry do not pay litigation costs, the parties to a patent challenge will not be able to pass their costs on to consumers even in the competitive market that prevails after patent expiry.

[7] Specifically, I have $E[f(w)|Y] = \int_0^\infty e^{-r\tau} e^{-Y\tau} Y d\tau = \frac{Y}{Y+r}$, in which $e^{-Y\tau}$ gives the chance of failure until time τ, $Y d\tau$ gives the chance of success in any given moment $d\tau$, such as that at time τ, $e^{-r\tau}$ gives the discounted value of success at time τ, and success is possible at any moment between $\tau = 0$ and the end of time.

and Brand π. Under the duopoly that prevails for $T - T_E$, the aggregate profit of all firms in the market (both Generic and Brand) is reduced to $\beta\pi$, where $\beta \in [0, 1]$, and Brand's profit is reduced to $s\beta\pi$, where $s \in (0, 1)$ is the share of aggregate duopoly profit enjoyed by Brand. Deadweight loss falls to $D(\beta) \in [0, D]$ under duopoly, and consumers enjoy the rest, which is the CS_M that they started with plus $(1 - \beta)\pi$ and $D - D(\beta) \equiv d(\beta)$. Under the competition that prevails for $\infty - T$, profit and deadweight loss are zero and consumers take all: $CS_M + \pi + D$. Figure 2 shows consumer and firm value over these periods.

The present value of consumers at the date of innovation is $V_c =$

$$\underbrace{\int_0^{T_E} CS_M e^{-rt} dt}_{\text{innovation to entry}} + \underbrace{\int_{T_E}^{T} \left[CS_M + (1-\beta)\pi + d(\beta)\right] e^{-rt} dt}_{\text{entry to patent expiry}} + \underbrace{\int_{T}^{\infty} \left[CS_M + \pi + D\right] e^{-rt} dt}_{\text{after patent expiry}}, \quad (7)$$

which becomes $\frac{1}{r}\left(CS_M + (z - z_E)\left[(1-\beta)\pi + d(\beta)\right] + (1 - z)\left[\pi + D\right]\right)$ after integrating and making the substitutions $z \equiv 1 - e^{-Tr} \in [0, 1)$ and $z_E \equiv 1 - e^{-T_E r} \in [0, z]$. For later use, I define z_E^{lit} and z_E^{set} to be z_E when it is a function of T_E^{lit} and T_E^{set} respectively. Further substituting the labels $u \equiv CS_M + (1 - z)\left[\pi + D\right]$ and $k \equiv (1 - \beta)\pi + d(\beta)$, I obtain

$$V_c = \frac{1}{r}\left(u + (z - z_E)k\right). \quad (8)$$

I obtain the following expression for Brand's value at the moment of innovation:

$$v^p = \int_0^{T_E} \pi e^{-rt} dt + \int_{T_E}^{T} s\beta\pi e^{-rt} dt = \frac{1}{r}\left[z_E \pi(1 - s\beta) + zs\beta\pi\right] = \frac{1}{r}\left[z_E m + n\right], \quad (9)$$

in which I employ the labels $m \equiv \pi(1 - s\beta)$ and $n \equiv zs\beta\pi$.

In (8) and (9), T_E appears only in z_E and z_E enters linearly into v^p and V_c. $E[v^p]$ and $E[V_c]$ therefore both depend on the expected value of a single variable, z_E. Substituting (9)

into (6), and (6) and (8) into (4), I therefore have

$$\mathcal{V}_c^* = \frac{1}{r}\left[1 - \left(\frac{\alpha r}{\frac{1}{r}\left[E[z_E]m + n\right] - L}\right)^n\right][u + (z - E[z_E])k] \equiv {}^a\mathcal{V}_c^*. \tag{10}$$

(10) reflects the tradeoff between innovation and static value that arises because T_E decreases V_c but increases x. As $E[T_E]$ increases, so too must the expectation of T_E's monotonic transformation z_E, driving up the left bracketed term but driving down the right bracketed term. As shown in Figure 3, in the competitive case (i.e., $n = 1$), this may result in a hump shape for ${}^a\mathcal{V}_c^*$ in $E[z_E]$, and indeed in $E[T_E]$ as well (not shown). Let ${}^a\mathcal{V}_c^{*L}$ be (10) when T_E is the entry date under litigation and $L > 0$ and let ${}^a\mathcal{V}_c^{*S}$ be (10) when T_E is the entry date under settlement and L is zero. The negative relation between $\mathcal{V}_c^{*L}$ and L causes ${}^a\mathcal{V}_c^{*L}$ to lie below ${}^a\mathcal{V}_c^{*S}$ in Figure 3. I note that in settlement $E[z_E] = E[z_E^{set}] = z_E^{set}$ because T_E^{set} is deterministic. In litigation, $E[z_E] = E[z_E^{lit}] = pz$. Going forward, I default to $n = 1$.

To determine permissible delay, I want to find the z_E^{set} for which settlement value just equals maximum litigation value, which latter I call ${}^a\mathcal{V}_c^{*Lmax}$. If z_E^{set*} is this z_E^{set}, then the "delay" in settlement in $E[z_E]$-space is $\Delta \equiv z_E^{set*} - p^{max}z$, where p^{max} is the patent strength that yields ${}^a\mathcal{V}_c^{*Lmax}$. Permissible delay is $\Delta_T \equiv -\frac{1}{r}ln(1 - z_E^{set*}) - p^{max}T$. Because of the hump shape of the curves, there may be two z_E^{set*}s. I am interested in the larger of the two because I want the z_E^{set} above which settlement value falls below litigation value. I solve

$${}^a\mathcal{V}_c^{*S}(z_E^{set*}) = \mathcal{V}_c^{*Lmax} \tag{11}$$

to obtain z_E^{set*}. p^{max}, ${}^a\mathcal{V}_c^{*Lmax}$ and z_E^{set*} involve messy solutions to quadratics; I outline them, in the case of a competitive patent race, in the Appendix.

All else equal, z_E^{set*} is increasing in the magnitude of the welfare loss caused by litigation. The greater the gain from settlement, the more may be squandered on delay without harming consumers. Let x^S be the dynamic factor under settlement (i.e., the first bracketed term in (10)), when $L = 0$, and let it be x^L in litigation. The difference in magnitude between

settlement and litigation welfare is therefore $\frac{1}{r}(x^S - x^L)(u + (z - E[z_E])k)$. If $x^S - x^L$ is fixed, an increase in $u + (z - E[z_E])k$, which may result from an increase in CS_M, drives up the gap between settlement and litigation welfare, and increases z_E^{set*}.

This effect may be counteracted, however, if increasing the gains from settlement also makes delay more costly. In that case, delay squanders the surplus from settlement at a greater rate, and more delay may not still leave consumers better off. This happens if driving up $u + (z - E[z_E])k$ also increases the rate at which settlement welfare falls in z_E^{set*}. Let $a \equiv \frac{1}{r}m$, $b = \frac{1}{r}n$, $e = u + zk$, and $d = \alpha r$; then the first derivative of (10) is

$$\frac{d^a \mathcal{V}_c^*}{dx} = \frac{1}{r}\left[\frac{d(bk + ae)}{(ax + b - L)^2} - k\right]. \tag{12}$$

All components of $u + (z - E[z_E])k$ appear in (12). This indicates that the counteracting effect is possible for these components and their overall effect on z_E^{set*} ambiguous at this level of generality. When static welfare is large, innovation is more fruitful, but there is also potentially more to lose through excessive delay.

z_E^{set*} is not the only factor that determines whether delay is harmful. For strong patents, the delay required to harm consumers, even when very small, may exceed the time between the expected entry date under litigation and patent expiry (i.e., between $p^{max}T$ and T) and consumers cannot therefore be harmed by delay. Variables that tend to drive up the patent strength that maximizes litigation welfare, which determines patent strength, therefore also determine whether delay is harmful.

CS_M is such a variable. Increasing CS_M unambiguously drives up (12), which when set equal to zero is the first order condition for maximizing litigation welfare. $E[z_E]$, and therefore p^{max}, must increase in response to satisfy the condition. (12) shows that delay increases consumer welfare so long as the static welfare loss to consumers associated with delay, k, is less than the welfare gain to consumers, which is given by $\frac{d(bk+ae)}{(ax+b-L)^2}$. CS_M is not a component of the loss because CS_M is the welfare that consumers enjoy even under

monopoly. But it is a component of gain because innovation magnifies its value. The gain is appropriately decreasing in the entry date, which is proxied by $E[z_E]$, because delay in entry permits Brand to arrogate more of the gain to itself. As CS_M increases in size, driving the gain up, but not the loss, Brand may enjoy more of the gain. Thus the equilibrating $E[z_E]$, and therefore patent strength, increase.

3.3 The Relevance Strength

It may be useful to depart from the assumption that the litigation entry date is p^{max}. Permissible delay is the threshold level of delay beyond which delay harms consumers, measured at p^{max}. I call the threshold level of delay beyond which delay harms consumers, measured at any other p, the "delay threshold" for that p. I consider the case in which patent law errs on the side of excessive patent strength and therefore $p \in [p^{max}, 1]$. I show in the Appendix that, in the capped case, the delay threshold is falling for all $p \in [p^{max}, 1]$ over my estimates for the parameters of the model. It would therefore seem that delay in excess of permissible delay harms consumers for all patent strengths in excess of p^{max}.

This claim must be qualified, however, because, as I have already observed, for strong patents the delay required to harm consumers exceeds the time until patent expiry. I call the strength beyond which this is true the "relevance strength" because permissible delay represents an upper bound on the delay threshold only for strengths that do not exceed this strength. If p^R is the relevance strength, then for $p \in [p^{max}, p^R)$ delay that exceeds permissible delay must harm consumers.[8]

The foregoing holds in the capped duopoly case. I show in the Appendix that in the fixed duopoly case with low litigation costs expected consumer welfare under litigation must exceed that under settlement for a range of patent strengths immediately below the relevance strength. This is pictured in Figure 4. I also find that p^{max} is zero in the fixed duopoly case for my parameter estimates. This means that the delay threshold must eventually fall to

[8]Figure 3 shows $z_E^R = p^R z$, the counterpart of $p^R T$ in $E[z_E]$-space.

zero and indeed become negative as p increases from zero to 1. When the delay threshold is negative, no delay may be tolerated, unless the relevance strength has been reached.

The "relevance strength" may be determined by finding the pz for which consumer value under litigation equals consumer value under settlement at patent expiry. Thus I solve for p in

$$ {}^{a}\mathcal{V}_c^{*S}(z) = {}^{a}\mathcal{V}_c^{*L}(pz). \tag{13}$$

Relevance strength has another useful interpretation in both the fixed and capped cases; it gives the greatest $p \in [p^{max}, 1]$ for which delay until patent expiry must harm consumers. My relevance strength results assume no reverse payment. Because, as I discuss more below, a reverse payment reduces consumer welfare in settlement, the relevance strengths that I report are lower than those for reverse payment settlements.

3.4 Uncertainty Regarding Challenge and Settlement Occurrence

I treat litigation and settlement as sole, mutually exclusive, alternatives. I treat them as sole alternatives because I am interested in the set of drugs for which patents are challenged. Entry dates for these drugs are determined either through settlement or litigation. As described in more detail in the Appendix, the estimates upon which I rely in calibrating the model relate to new drugs generally, and not the subset that are challenged. My results are therefore accurate only if challenged drugs are a representative subset of new drugs. This assumption is reasonable given that Grabowski et al. (2014) find that 81% of drugs with initial generic entry in September 2012 faced a legal challenge, with the number trending upward.

Treating the alternatives of settlement and litigation as mutually exclusive is somewhat more problematic because there is no reason to suppose that whenever settlements are possible the parties will in fact settle. Indeed, Greene and Steadman (2010) find that only 47%

of filed drug exclusivity challenges settle. There are two solutions to this problem. The first is to accept that my results only apply when policymakers force the parties to settle. The second is to insist that my results apply when settlement is optional for the parties, but only if the following condition holds. Firms always know ex ante whether or not they will settle, even if consumers do not. This condition is required to ensure that uncertainty about settlement does not cause firms to alter their decisions about how much Y to buy. I show in the Appendix that subject to this condition my results hold when settlement is optional.

3.5 Extension to Fixed Duopoly

Edlin et al. (2015) and Kobayashi et al. (2015) consider an extension of the standard model in which the period of duopoly is limited to a short fixed period. They believe a fixed period better captures the reality of the drug market.

I extend my model to duopoly of fixed duration H. When $T_E \in [T-H, T]$, this adjustment has no effect on value because duopoly cannot extend beyond patent expiry under any circumstance. Duopoly remains capped at T and the capped model continues to hold. However, for shorter T_E, the duopoly period no longer extends from T_E to T, but instead from T_E to $T_E + H$, and is followed by a competitive period from $T_E + H$ to ∞. This is a departure from the capped model.

New expressions for static Brand, Generic, and consumer welfare, as well as expected consumer welfare, are given in Section 8.11 of the Appendix. Static consumer welfare now falls faster before $z_E = \frac{z-z_H}{1-z_H}$ (which corresponds to entry dates before $T - H$ and I call the "early" settlement region) and Brand's value increases faster over the same region. Consumers have more to lose from delay because delay now eliminates a period of competition; Brand has more to gain for the same reason.

Static litigation value now falls below static settlement value for Brand, even in the absence of litigation costs, and static litigation value exceeds static settlement value for consumers. When Brand wins in litigation there is no fixed period of duopoly for it to enjoy

(this period cannot exceed patent expiry). Litigation victory therefore lacks some of the benefit associated with delay in settlement.

One would expect consumers to be better off in the fixed case than in the capped case because in the capped case they may have access to competitive pricing before T. I show in Section 8.11 in the Appendix that if $s < \frac{\pi}{\pi+D}$, then in settlement expected consumer welfare indeed exceeds expected consumer welfare under capped duopoly for at least some $E[z_E] < \frac{z-z_H}{1-z_H}$ (for $E[z_E] > \frac{z-z_H}{1-z_H}$, static values are the same in both cases and so expected consumer welfare is also identical). I also show that, in the fixed case, for a given patent strength within a large interior range of patent strengths, and low litigation costs, consumers may actually be better off under litigation, as shown in Figure 4. In Section 8.11 in the Appendix, I identify the range of strengths for which litigation is preferred by consumers. I also determine permissible delay in this case.

3.6 The Standard Entry Model as a Special Case

Edlin et al. (2015), Elhauge and Krueger (2012), and Willig and Bigelow (2004) model entry settlements without accounting for gains from innovation. This "standard entry model," appears in my model when value is measured at the moment of innovation rather than at the earlier date of the R&D investment decision (i.e., at $\tau = \mathcal{T}$ or $t = 0$, rather than at $\tau = 0$). At the moment of innovation the patent race is over, so $f(w)$ is a constant and αY is sunk. Measuring value from the date of innovation and normalizing $f(w)$ to one for convenience, $\mathcal{V}_c = V_c$ and Brand's expected revenue, $E[f(w)|Y]v^p$, becomes v^p. The expressions for V_c in (8) and v^p in (9) are consumer welfare and Brand's value, respectively, in the standard entry model.

In this scenario, Generic enjoys value

$$\frac{1}{r}\left[(z - E[z_E])(1-s)\beta\pi\right] - L_G, \tag{14}$$

which is also Generic's value in the standard model. L_G is Generic's litigation cost. During the period after entry and before patent expiry, which is represented in (14) by the term $z - E[z_E]$, Generic obtains the share of duopoly profit $1 - s$ that remains after Brand takes share s. I have ignored Generic's value so far because I assume that Generic does not take part in the patent race and does not therefore invest in innovation. I use it now to consider Generic's willingness to agree to delay.

3.7 Generic's Willingness to Delay

So far I have not sought to model bargaining between Brand and Generic, treating T_E^{set} as a policy variable. I now consider briefly the maximum delay to which Generic may agree. I follow the standard entry model in assuming that bargaining takes place after the R&D investment decision. I place bargaining at the moment of innovation (i.e., $\tau = \mathcal{T}$ and $t = 0$). The standard entry model identifies the maximum delay to which Generic may agree as that for which Generic's static settlement value equals its static litigation value. Using the expression for Generic's static value in (14), the discount factor associated with Generic's maximum settlement delay, Δ_G, is determined by

$$\underbrace{\frac{1}{r}(z - (z_E + \Delta_G))(1 - s)\beta\pi + \mathscr{T}}_{\text{settlement}} = \underbrace{\frac{1}{r}(z - p^{max}z)(1 - s)\beta\pi - L_G}_{\text{litigation}}, \tag{15}$$

where $\mathscr{T}$ is any transfer of value from Brand to Generic in exchange for settlement. In the case of fixed duopoly, expressions for Generic's static value for $z_E \in [z_{min}, \frac{z - z_H}{1 - z_H}]$ and $z_E \in (\frac{z - z_H}{1 - z_H}, z]$ in Section 8.11 of the Appendix may be used instead on the left-hand side of (15) (with a similar adjustment for $\mathscr{T}$) and the expression for Generic's static value under litigation in the Appendix used instead on the right-hand side to obtain the condition in that case.

To determine willingness to delay in the absence of a reverse payment, I set $\mathscr{T} = 0$. I solve (15) for Δ_G, and convert it to its equivalent in terms of years of delay, Δ_G^T. In calibrating

my model, I assume that Generic's litigation costs equal Brand's. I also determine the maximum extent to which Generic will permit delay to exceed permissible delay without a reverse payment, expressed as a share of the total extent to which delay may exceed permissible delay, $\frac{\Delta_G^T - \Delta_T}{T-(T_E+\Delta_T)}$. This provides a sense of the extent to which excessive delay can take place in the absence of any payment from Brand to Generic.

I note that the limits on delay in the absence of a reverse payment defined by (15) with $\mathscr{T} = 0$ should not be treated as the last word on delay. It follows from the work of Willig and Bigelow (2004) that if the parties do not have the same beliefs regarding p, any settlement entry date is possible even in the absence of a reverse payment.

3.8 Reverse Payments and Their Cost

Hemphill (2009) and Edlin et al. (2015) suggest that absent regulation Brand and Generic will tend to use a reverse payment to negotiate delay until patent expiry. I determine the size of the reverse payment, if any, required at p^{max} to make Generic willing to delay until patent expiry. I do this by substituting z for $z_E + \Delta_G$ on the left-hand side of (15). This reduces the first term on that side to zero; I then determine the value on the right-hand side to arrive at the necessary payment, $\mathscr{T}$.

A reverse payment reduces the incentive of Brand to innovate, relative to a settlement without a reverse payment, because the reverse payment reduces Brand's quasi-profit. This harm to consumers comes on top of the harm associated with delay. A reverse payment drives down expected consumer welfare under settlement and therefore reduces permissible delay. Kobayashi et al. (2015) and Langenfeld and Li (2003) fail to recognize this in defending reverse payment settlements on innovation grounds.

I model this effect as follows. Instead of assuming that L is zero in ${}^a\mathcal{V}^{*S}$, I redefine L as $\mathscr{T} > 0$, the amount of the reverse payment. I identify the "dynamic" harm to consumers of a reverse payment by comparing consumer harm from delay until patent expiry in the absence of a reverse payment with consumer harm when Brand makes the smallest reverse

payment necessary to make the same amount of delay worthwhile for Generic to accept.

4 Application to Drug Settlements

I now adapt my model to the U.S. drug development process and identify a number of assumptions required for it to describe that process.

4.1 Model Adjustments

I make three principal changes to my model in order better to reflect the drug market. First, to capture the effect of the Hatch-Waxman exclusivities described in Section 2, I allow Brand to enjoy an initial period of unchallengeable exclusivity. Second, I take account of non-R&D costs, which are significant in the drug industry. I assume that such costs are in fixed proportion to revenues. Finally, I find that estimates of R&D cost from the literature are too high for monopoly in innovation to work in the model. I therefore apply only the patent race version of the model to drug settlements.

4.1.1 A Prechallenge Period

The existence of a prechallenge period means that the earliest possible entry date, T_{min}, is no longer zero. This requires (1) a change in the way in which expectations are calculated, (2) a change in the minimum value that $E[z_E]$ can take in ${}^a\mathcal{V}_c^*$, and (3) that I discount litigation cost, L, over the prechallenge period.

With respect to (2), $E[z_E]$ must not fall below $z_{min} = 1 - e^{-rT_{min}}$, the discount factor corresponding to T_{min}. With respect to (3), I assume that litigation is resolved and L spent at T_{min}, and therefore Brand's litigation cost must be discounted by the factor $Z_{min} \equiv e^{-rT_{min}} = 1 - z_{min}$. Thus I must substitute $Z_{min}L$ for L. $\mathscr{T}$ and L_G in (15) each must also be multiplied by Z_{min}. With respect to (1), the calculation of litigation expectations must now treat T_{min} as the earliest possible entry date. $E[z_E^{lit}]$ now equals $p(z - z_{min}) + z_{min}$ and

$E[T_E^{lit}]$ equals $p(T - T_{min}) + T_{min}$.

4.1.2 Other costs

In the pharmaceutical industry, marketing costs and non-R&D capital expenditures can be substantial. In order to include these in the model, I treat π as revenues, rather than profits. I assume that non-R&D costs are always a fixed proportion of π because doing so approximates the variation in marketing costs based on market size described by Grabowski et al. (2002).

Let c be this fixed proportion of non-R&D cost to revenues. Because profit and not revenue triggers R&D effort, I must discount π by the factor $1 - c$ whenever π is enjoyed by the winner of the patent race. Thus (2), which gives Brand's expected profit, becomes

$$E[f(w)|Y]\left(E[v^p(1-c)] - Z_{min}L\right) - \alpha Y. \tag{16}$$

No such use of the factor $1 - c$ is required for static consumer welfare because consumers benefit from a reduction in π regardless what proportion of π is ultimately spent by firms on costs. (8) therefore remains unchanged, as does static consumer welfare in the case of fixed duopoly.

4.1.3 Monopoly and α

I infer α from an estimate of αY for firms that could expect not to face a Paragraph IV challenge and always to enjoy entry at T. DiMasi et al. (2003) and Grabowski et al. (2002), from whom I derive my estimates for αY and π, base their estimates on data from the 1990s. Grabowski et al. (2014) suggest that during that time fewer than 40% of drugs experienced Paragraph IV challenges. I therefore take my estimate for αY to be based on data from firms that expected to enjoy exclusivity until patent expiry without paying litigation costs.

For these firms, I have $z_E = z$ and, from (9),

$$E[v^p] = E[\frac{1}{r} z\pi(1-c)] = \frac{1}{r} z\pi(1-c). \tag{17}$$

I include the factor $1-c$ for the reason discussed in Section 4.1.2.

In the case of a patent race, I substitute (17) and (5) into (2) and, remarking that $Y = \frac{\alpha Y}{\alpha}$, I obtain

$$\alpha = \frac{\frac{z\pi(1-c)}{r} - \alpha Y}{r}. \tag{18}$$

(18) gives me α in terms of variables for which, as discussed in the Appendix, estimates are available. In the case of a research monopoly, I substitute (17) and (5) into (3) to obtain

$$\alpha = \frac{\frac{z\pi(1-c)}{r} - 2\left(Y\alpha\right) \pm \left[\frac{z\pi(1-c)}{r}\left(\frac{z\pi(1-c)}{r} - 4\left(Y\alpha\right)\right)\right]^{1/2}}{2r}, \tag{19}$$

as described more fully in Section 8.9 of the Appendix.

My Estimates Imply That the Innovation Race Is Not Monopolized It is evident from (19) that in monopoly α is only defined for $\frac{z\pi(1-c)}{4r} > Y\alpha$. Under monopoly, there is a ceiling on the amount of cost a firm will take on as a fraction of its total quasi-profits. When I plug my parameter estimates in Table 1 into (19), I find that the condition $\frac{z\pi(1-c)}{4r} > Y\alpha$ is not satisfied. This may be interpreted in two ways: (1) if there is in fact monopoly in drug innovation, then my model or estimates are wrong, or (2) there is not in fact monopoly in drug innovation. Adopting (2) accords with the result of Grabowski et al. (2002) that the internal rate of return for drug innovators is only slightly above cost of capital. I therefore assume that innovation in the drug market is competitive.

4.2 Other Considerations in Mapping the Drug Development Process onto the Model

I treat the entire drug development process, from basic R&D, through clinical trials, up to FDA approval, including post-approval costs of drug improvement, as part of the patent race. DiMasi et al. (2003) show that R&D expenditure is typically made over time, with more promising compounds, particularly those that have survived early trials, receiving more financing in later periods. I nonetheless treat all R&D investment as made at the beginning of the development process (i.e., at $\tau = 0$).

I treat the date of FDA approval as the date of invention (i.e., as $\tau = \mathcal{T}$ or, equivalently, as $t = 0$) and assume that litigation or settlement are resolved instantaneously at the end of the prechallenge period (i.e., at $\tau = \mathcal{T} + T_{min}$ or, equivalently, at $t = T_{min}$). I assume that all litigation costs are expended at that moment as well. In practice, litigation costs are expended over time and commence before the expiration of the prechallenge period because that period probably includes part of the 30 month stay, during which the parties are locked in suit.

My model accounts for the 180-day first-filer exclusivity period by assuming that profits do not fall immediately to zero upon entry (i.e., at $\tau = \mathcal{T} + T_E$ or $t = T_E$). I note, however, that the data on post-entry prices in Olson and Wendling (2013), upon which I rely in determining the duopoly profit rate, β, stretches out only 18 months. I nonetheless assume that duopoly profit remains the same all the way until patent expiry in the capped case. In the fixed case I assume that it ends after half a year (approximately 180 days). I ignore post-entry profits Brand may obtain by selling its own generic version of its drug at the generic price (i.e., I ignore the "authorized generics" described by Appelt (2015)).

Although I assume the existence of a single patent covering a drug in my model, my calibration applies to drugs covered by multiple patents. I use for T the average period from FDA approval until first Generic entry identified by Grabowski et al. (2014). This duration should be the average term of the youngest patents covering drugs. The existence of multiple

patents poses a problem for my assumption that immediate entry or entry at patent expiry are the only possible outcomes of litigation. If Brand staggers its patent applications and some patents hold up in court, whereas others are struck down, then entry dates under litigation intermediate between immediate entry and entry at expiration of the term of the youngest patent are possible. Intermediate dates are excluded, however, if Brand tends to obtain all its patents at around the same time. My results apply subject to this condition.

5 Summary of Estimates

I summarize the estimates I use for the variables in the model in Tables 1 and 2 and describe them in detail in Section 8.12 in the Appendix. I adopt a prechallenge period estimate of 6.9 years and a patent term estimate of 12.9 years, which leave 6 years of challengeable exclusivity.

6 Results

6.1 Capped Duopoly Period

I summarize my capped duopoly results in Tables 3 and 5. Permissible delay is 15.06 months and constitutes an upper bound on the delay threshold for patents in the strength range $[19.4\%, 94.6\%]$. However, at p^{max} Generic is not willing to delay at all. Because of discounting of future value, Generic will only settle for a *hastening* of entry of 2.95 months. Consumer harm is relatively low: \$1.83 million for delay from $p^{max}T$ until patent expiry. The results are, however, sensitive to changes in my parameter estimates. The maximum delay to which Generic is willing to agree at p^{max}, over the bounds on my parameter estimates listed in Table 1, is delay that is 2.3% in excess of permissible delay.

The results in the first row of Table 3 are sensitive to changes in my estimate for the price elasticity of demand for pharmaceuticals. As described in Section 8.12 in the Appendix,

elasticity drives up CS_M as it approaches zero. To give a measure of the sensitivity of my results to elasticity and other variables, I report in the second row of Table 3 the greatest permissible delay over the variable bounds listed in Table 1. I chose these bounds to limit the size of permissible delay and prevent relevance strength from falling too low. For small increases in elasticity, permissible delay jumps by about a year, even after restrictions designed to limit permissible delay are imposed on other variable bounds. An important limit on the size of the bounds that I chose is the tendency of CS_M to increase p^{max} and thereby to make delay until patent expiry permissible. I discuss the mechanism by which CS_M tends to do this in Section 3.2. My results are also sensitive to price after generic entry as a share of preentry price, which determines β, and the net present value of the drug as a share of R&D cost, which determines π.

Optimal Patent Strength and Term Optimal patent strength is p^{max}, the strength that maximizes consumer welfare under litigation, ${}^a\mathcal{V}_c^{*L}$. The estimate in Table 3 for this strength implies that consumers do best when Brand has only a 19.4% chance of success in patent litigation. This corresponds to a litigation entry date of 8.06 years after the branded drug is first marketed. Optimal settlement strength (not shown in the table) maximizes ${}^a\mathcal{V}_c^{*S}$; it is 18.2%, which corresponds to 7.99 years.

The optimal patent term (also not shown in the table) is the T_E that maximizes ${}^a\mathcal{V}_c^{*S}$, assuming that price falls to competitive levels immediately after entry. I find it to be 4.63 years, which is less than $T_{min} = 6.9$ years. This suggests that if all patents were to have 100% strength, be properly granted only to innovations that increase consumer value, and never be litigated, marketing exclusivity should be limited to 4.63 years. When generic entry is accompanied by a duopoly period, the cost of delay to consumers is reduced because monopoly does not replace competition. This explains why optimal patent *strength* is not zero even though T_{min} exceeds the optimal patent *term*.

6.2 Fixed Duopoly Period

I obtain the results summarized in Tables 4 and 5. Permissible delay falls to 0.23 months at p^{max}. Although, unlike in the capped case, I do not prove that the delay threshold is falling as strength rises, numerical optimization reveals that for my parameter estimates it is. So 0.23 months is an upper bound on the delay threshold for patents in the strength range $[0, 99.86\%]$. It is an expression of the extent to which litigation cost savings provide no justification for delay in this case, even after accounting for innovation, that for 98.88% of patent strengths consumers are better off under litigation. In this case Generic is willing to agree to enough delay to harm consumers, even without a reverse payment (6.68 months $>$ 0.23 months) and that level of delay causes $134 million in harm to consumers. Delay until patent expiry causes $1.3 billion in harm, before the additional $34.60 million dynamic cost of a reverse payment is taken into account. The sensitivity of the results is also substantially reduced. I achieve nearly the same greatest permissible delay (25.55 months) as in the capped case over the substantially broader bounds for elasticity and other variables described in Table 1. Generic's maximum willingness to delay without a reverse payment over my sensitivity bounds is delay until patent expiry.

Table 6 gives results for reverse payment magnitudes and harm. The $1.2 billion payment in the capped case is implausibly high given the reverse payment amounts reported by Hemphill (2009). The payment estimate achieved using the parameters that maximize delay over my bounds in the capped case is $354 million, which is more plausible. Generic is willing to exceed permissible delay by 2.3% under those parameters, even without a reverse payment. If those parameters are more plausible, then consumer harm without a reverse payment is also more plausible in the capped case. In the fixed case, the reverse payment is $160 million, which, as an estimate for the average drug, fits the observations of Hemphill (2009) much better.

7 Conclusion

I estimate conditions on the period of time for which drug companies may delay entry through settlement without harming consumers. In the case of a period of duopoly capped at patent expiry, I find that for any patent with strength between 19.4% and 94.69% delay in excess of 15 months harms consumers. However, I also find that if patent law optimizes the litigation outcome, then a generic drug maker is not willing to agree to any delay. These findings are sensitive to my parameter estimates and I find that under plausible alternative estimates a generic maker is willing to agree to harmful levels of delay without a reverse payment. In the case of fixed duopoly, I find that a settlement that delays entry by more than 0.23 months harms consumers for all but the strongest patents. I also find that Generic is willing to agree to enough delay to cause $134 million in harm to consumers in this case. These results suggest that gains from innovation to not eliminate the possibility of consumer harm from settlements of patent litigation that delay generic entry into a drug market. They also suggest that entry settlements generally, and not just settlements that involve a reverse payment, can cause substantial harm.

References

Appelt, S. "Authorized Generic Entry Prior to Patent Expiry: Reassessing Incentives for Independent Generic Entry." *Review of Economics and Statistics*, Vol. 97 (2015), pp. 654–666.

Denicolò, V. "Do Patents Over-Compensate Innovators?" *Economic Policy*, Vol. 22 (2007), pp. 679–729.

DiMasi, J.A. and Grabowski, H.G. "The Cost of Biopharmaceutical R&D: Is Biotech Different?" *Managerial and Decision Economics*, Vol. 28 (2007), pp. 469–479.

DiMasi, J.A., Hansen, R.W., and Grabowski, H.G. "The Price of Innovation: New Estimates of Drug Development Costs." *Journal of Health Economics*, Vol. 22 (2003), pp. 151–185.

DiMasi, J.A., Hansen, R.W., Grabowski, H.G., and Lasagna, L. "Cost of Innovation in the Pharmaceutical Industry." *Journal of Health Economics*, Vol. 10 (1991), pp. 107–142.

Duflos, G. and Lichtenberg, F.R. "Does Competition Stimulate Drug Utilization? The Impact of Changes in Market Structure on Us Drug Prices, Marketing and Utilization." *International Review of Law and Economics*, Vol. 32 (2012), pp. 95–109.

Edlin, A.S., Hemphill, C.S., Hovenkamp, H.J., and Shapiro, C. "The Actavis Inference: Theory and Practice." *Rutgers University Law Review*, Vol. 67 (2015), pp. 585–635.

Elhauge, E. and Krueger, A. "Solving the Patent Settlement Puzzle." *Texas Law Review*, Vol. 91 (2012), pp. 283–330.

Federal Trade Commission. "Pay for Delay: How Drug Company Payoffs Cost Consumers Billions." Staff Study, Federal Trade Commission (2010).

Frank, R.G. and Salkever, D.S. "Generic Entry and the Pricing of Pharmaceuticals." *Journal of Economics & Management Strategy*, Vol. 6 (1997), pp. 75–90.

Gilbert, R. and Shapiro, C. "Optimal Patent Length and Breadth." *RAND Journal of Economics*, Vol. 34 (1990), pp. 106–112.

Grabowski, H., Long, G., and Mortimer, R. "Implementation of the Biosimilar Pathway: Economic and Policy Issues." *Seton Hall Law Review*, Vol. 41 (2011), pp. 511–557.

—. "Recent Trends in Brand-Name and Generic Drug Competition." *Journal of Medical Economics*, Vol. 17 (2014), pp. 207–214.

Grabowski, H., Vernon, J., and DiMasi, J.A. "Returns on Research and Development for 1990s New Drug Introductions." *Pharmacoeconomics*, Vol. 20 (2002), pp. 11–29.

Grabowski, H.G. and Kyle, M. "Generic Competition and Market Exclusivity Periods in Pharmaceuticals." *Managerial and Decision Economics*, Vol. 28 (2007), pp. 491–502.

Grabowski, H.G. and Vernon, J.M. "Returns to R&D on New Drug Introductions in the 1980s." *Journal of Health Economics*, Vol. 13 (1994), pp. 383–406.

Greene, A. and Steadman, D.D. "Analyzing Litigation Success Rates." *Pharmaceuticals*, Vol. 212 (2010), pp. 1–24.

Harris, B.C., Murphy, K.M., Willig, R.D., and Wright, M.B. "Activating Actavis: A More Complete Story." *Antitrust*, Vol. 28 (2014), pp. 83–89.

Hemphill, C.S. "Paying for Delay: Pharmaceutical Patent Settlement as a Regulatory Design Problem." *N.Y.U. Law Review*, Vol. 81 (2006), pp. 1553–1623.

—. "An Aggregate Approach to Antitrust: Using New Data and Rulemaking to Preserve Drug Competition." *Columbia Law Review*, Vol. 109 (2009), pp. 629–688.

Kelly, C. "The Balance between Innovation and Competition: The Hatch-Waxman Act, the 2003 Amendments, and Beyond." *Food and Drug Law Journal*, Vol. 66 (2011), pp. 417–78.

Kobayashi, B.H., Wright, J.D., Ginsburg, D.H., and Tsai, J. "Actavis and Multiple ANDA Entrants: Beyond the Temporary Duopoly." *Antitrust*, Vol. 29 (2015), pp. 89–97.

Langenfeld, J. and Li, W. "Intellectual Property and Agreements to Settle Patent Disputes: The Case of Settlement Agreements with Payments from Branded to Generic Drug Manufacturers." *Antitrust Law Journal*, Vol. 70 (2003), pp. 777–818.

Lemley, M.A. "Rational Ignorance at the Patent Office." *Northwestern University Law Review*, Vol. 95 (2001), pp. 1495–1532.

Liu, S. and Chollet, D. "Price and Income Elasticity of the Demand for Health Insurance and Health Care Services: A Critical Review of the Literature." Report Reference No. 6203-042, Mathematica Policy Research (2006). Available at `www.mathematica-mpr.com/~/media/publications/PDFs/priceincome.pdf`.

Morgan, S., Grootendorst, P., Lexchin, J., Cunningham, C., and Greyson, D. "The Cost of Drug Development: A Systematic Review." *Health Policy*, Vol. 100 (2011), pp. 4–17.

Nordhaus, W.D. *Invention Growth, and Welfare: A Theoretical Treatment of Technological Change*. 1969.

—. "Schumpeterian Profits in the American Economy: Theory and Measurement." Working Paper No. 10433, National Bureau of Economic Research (2004). Available at `http://www.nber.org/papers/w10433`.

Olson, L. and Wendling, B.W. "Estimating the Effect of Entry on Generic Drug Prices Using Hatch-Waxman Exclusivity." Working Paper No. 317, Federal Trade Commission Bureau of Economics (2013). Available at `https://www.ftc.gov/sites/default/files/documents/reports/`

estimating-effect-entry-generic-drug-prices-using-hatch-waxman-exclusivity/wp317.pdf.

Saha, A., Grabowski, H., Birnbaum, H., Greenberg, P., and Bizan, O. "Generic Competition in the US Pharmaceutical Industry." *International Journal of the Economics of Business*, Vol. 13 (2006), pp. 15–38.

Salop, S.C. "Question: What Is the Real and Proper Antitrust Welfare Standard? Answer: The True Consumer Welfare Standard." *Loyola Consumer Law Review*, Vol. 22 (2009), pp. 336–353.

Scherer, F.M. "Research and Development Resource Allocation Under Rivalry." *Quarterly Journal of Economics*, Vol. 81 (1967), pp. 359–394.

Shapiro, C. "Antitrust Limits to Patent Settlements." *RAND Journal of Economics*, Vol. 34 (2003), pp. 391–411.

U.S. Supreme Court. "Federal Trade Commission v. Actavis." *Supreme Court Reporter*, Vol. 133 (2013), pp. 2223–2247.

Willig, R.D. and Bigelow, J.P. "Antitrust Policy Toward Agreements That Settle Patent Litigation." *Antitrust Bulletin*, Vol. 49 (2004), pp. 655–698.

Woodcock, R.A. "Uncertainty and Reverse Payments." Working Paper (2016). Available at ssrn.com/abstract=2763601.

8 Appendix

8.1 There are not always gains from trade from patent settlement

Shapiro (2003) argues that there are always gains from trade associated with patent settlement because settlement permits compromises that are unavailable under litigation. Here is a counterexample.

If there are gains from trade, then firms can do better in settlement without making consumers worse off. Let $\chi_s \in [0, \chi]$ be some action, $\chi_s \pi$ be its welfare to producers and $(\chi - \chi_s)CS$ be its welfare to consumers, where π, $CS > 0$.[9] Assume that there are only two possible litigation outcomes, affirmance of an intellectual property right, $\chi_s = \chi$, or rejection, $\chi_s = 0$. Then expected welfare for consumers is

$$p(\chi - \chi)CS + (1-p)\chi CS = (1-p)\chi CS = (\chi - p\chi)CS, \tag{20}$$

and for firms it is

$$p\chi\pi + (1-p)(0)\pi = p\chi\pi, \tag{21}$$

where p is the probability of litigation success.

Suppose that firms agree on a settlement action χ_s. It is evident that any $\chi_s > p\chi$ will reduce consumer welfare below expected welfare $(\chi - p\chi)CS$ and therefore make consumers worse off. But any settlement that makes firms better off than under expected welfare $p\chi\pi$ must increase χ_s above $p\chi$. So notwithstanding the fact that $\chi_s = 0$ happens under one state and $\chi_s = \chi$ under another, there is no settlement that raises profits for firms without reducing consumer welfare.

In order for settlement to make firms better off without hurting consumers, consumer

[9] I mean χ in this section and Section 8.2 to refer to the activity variable called x by Shapiro (2003). I do not use x for the activity variable because I follow Denicolò (2007) in using x to refer to the dynamic factor in my entry settlement model.

welfare must be concave. When it is concave, consumer welfare falls faster in expectation than in settlement, ensuring the existence of gains from settlement that firms can arrogate to themselves without hurting consumers relative to litigation.

In my example, $(\chi - \chi_s)CS$ cannot be concave so long as CS is constant in χ_s. Suppose instead that CS is a function of χ_s and that consumer welfare,

$$(\chi - \chi_s)CS(\chi_s), \tag{22}$$

is concave. Expected welfare for consumers is now $(\chi - p\chi)CS(0)$. As shown in Figure 5, there now exists an $\chi_s > p\chi$ for which $(\chi - p\chi)CS(0) = (\chi - \chi_s)CS(\chi_s)$. Firms are better off but consumers no worse off; there are gains from settlement.

What Shapiro (2003) probably had in mind is the common situation in which the settlement variable open to the parties is price. As shown in Section 8.2 of this Appendix, when demand is downward sloping and the parties may settle on price, consumer welfare is concave and there are gains to be had from settlement. Increases in the profit of firms create progressively larger shares of deadweight loss, so consumer welfare at first falls more slowly than it would if deadweight loss were to accrue at a constant rate, as it does in expectation.[10]

There is no reason to think that the available range of actions for settling parties always includes some action χ_s to which $CS(\chi_s)$ responds in a way that produces the required concavity. Indeed, Shapiro (2003) himself acknowledges that entry settlements, for which the settling parties can agree only on the date of entry and not on the price, are one such example. Before entry there is one price and profit level and after it another and the date of entry determines their relative duration in the same way that the probability of success determines their relative weights in expectation. Deadweight loss steps between two levels in both expectation and settlement, and for this reason settlement can behave no differently from, and indeed no more concavely, than litigation. Thus consumer value in this case looks

[10]Where gains exist, there is no reason that they must accrue to producers as opposed to consumers. In the presence of concave consumer welfare, a settlement that does not change the settlement variable makes firms no worse off and consumers much better off.

like $(\chi - \chi_s)CS$.

8.2 For downward-sloping demand, there are always gains from price settlement

I presuppose Section 3. Suppose that I am at the date of invention instead of at the date at which the level of R&D is decided. From (8), static consumer welfare is

$$u + (z - z_E)k, \tag{23}$$

where I have left off the factor $\frac{1}{r}$, which is irrelevant here. (23) is the counterpart in my model of (20). The constant u has no effect on gains from trade and can be ignored. If the activity variable is the date of entry, then it is determined by z_E in (23). The counterpart of CS in (20) here is then k, which does not vary with z_E. Thus consumer welfare here is linear, not concave, and there are no gains from trade.

Suppose, instead, that the activity variable is price and that the patentholder and challenger choose β to settle on price. β is not itself price, but as the share of monopoly profit earned at a given price, it can serve as a proxy for it. I mean by a settlement on price that the parties agree on a uniform price throughout the patent period, so I must have $z_E = 0$.

β enters into consumer welfare only through k. Expanding k, relying on $d(\beta) \equiv D - D(\beta)$, and setting $z_E = 0$, I obtain

$$u + (1 - \beta)\left[1 + \left[\frac{1 - \frac{D(\beta)}{D}}{1 - \beta}\right]\left(\frac{D}{\pi}\right)\right] z\pi, \tag{24}$$

which is the counterpart in my model of (22). In (24), the counterpart of χ_s is β, that of $CS(\chi_s)$ is $\left[1 + \left[\frac{1-\frac{D(\beta)}{D}}{1-\beta}\right]\left(\frac{D}{\pi}\right)\right] z\pi$, and that of $\chi - \chi_s$ is $1 - \beta$.

Because $\left[1 + \left[\frac{1-\frac{D(\beta)}{D}}{1-\beta}\right]\left(\frac{D}{\pi}\right)\right] z\pi$ is a function of β, it is possible for consumer welfare here to be concave, as required for gains from trade. For concavity to obtain in (22) (i.e., for the

second derivative to be negative), I must have

$$-2CS'(\chi_s) + (\chi - \chi_s)CS''(\chi_s) < 0. \tag{25}$$

Finding the first and second derivatives of $\left[1 + \left[\frac{1-\frac{D(\beta)}{D}}{1-\beta}\right]\left(\frac{D}{\pi}\right)\right] z\pi$, plugging them into (25), and simplifying, I have

$$\left[-D''(\beta) - \frac{4D'(\beta)}{1-\beta}\right] z. \tag{26}$$

As discussed by Shapiro (2003), if demand is downward sloping, then deadweight loss, $D(\beta)$, must rise in β at an increasing rate, so $D'(\beta) > 0$ and $D''(\beta) > 0$. One might say, with some abuse of the interpretation of the demand function, that if demand is falling then each volume decline associated with a higher price excludes a higher value consumer from the market, thereby increasing deadweight loss at an increasing rate. This implies that (26) is negative, as required for concavity.

So price settlements make consumer welfare concave, thereby ensuring that gains from trade emerge from the mere fact of settlement itself. A comparison of this result with (23), which is linear in its settlement variable, shows that it is not the mere fact of settlement, or the compromise it makes possible, that is responsible for these gains, but rather the concavity of consumer welfare in the settlement variable.

8.3 Dynamic value is increasing in private value and the entry date

8.3.1 Competitive patent race

Substituting v^p for $E[v^p]$ in (2) to reflect the case in which T_E is deterministic, and substituting x for $E[f(w)|Y]$, I have

$$x\left(v^p - L\right) - \alpha Y = 0. \tag{27}$$

Implicitly differentiating (27) with respect to Y and T_E, I obtain

$$\frac{dY}{dT_E} = \frac{-x\frac{dv^p}{dT_E}}{(v^p - L)\frac{dx}{dY} - \alpha}. \tag{28}$$

Multiplying through by $\frac{dx}{dY}$, I obtain

$$\frac{dx}{dT_E} = \frac{-\frac{dx}{dY}x\frac{dv^p}{dT_E}}{(v^p - L)\frac{dx}{dY} - \alpha}. \tag{29}$$

Because v^p is increasing in T_E and x is increasing in Y, the numerator must be nonpositive and indeed must be negative for $Y > 0$. Because x increases in Y at a decreasing rate, at the nonnegative Y that solves (27), $\frac{x(Y)}{Y} > \frac{dx}{dY}$ if $v^p > L$, and, because (27) implies that $\alpha = \frac{x}{Y}(v^p - L)$, I must have $(v^p - L)\frac{dx}{dY} < \alpha$. So the denominator is always negative. Thus $\frac{dx}{dT_E} \geq 0$ and indeed $\frac{dx}{dT_E} > 0$ for $Y > 0$. Implicitly differentiating (27) with respect to Y and v^p and multiplying by $\frac{dx}{dY}$, I obtain an expression for $\frac{dx}{dv^p}$ that must be nonnegative and indeed positive for $Y > 0$ by the same argument. This latter result implies that x is increasing as well in $E[v^p]$.

8.3.2 Monopoly

Substituting v^p for $E[v^p]$ and x for $E[f(w)|Y]$ in (3), I have

$$\frac{dx}{dY}(v^p - L) = \alpha \tag{30}$$

Implicitly differentiating (30) with respect to Y and T_E, and multiplying through by $\frac{dx}{dY}$, I obtain

$$\frac{dx}{dT_E} = \frac{-\frac{dx}{dY}\frac{dv^p}{dT_E}}{\frac{d^2x}{dY^2}(v^p - L)}.$$

This must be positive so long as $v^p > L$. Implicitly differentiating (30) with respect to Y and v^p and multiplying by $\frac{dx}{dY}$, I obtain an expression for $\frac{dx}{dv^p}$ that must be greater than zero by the same argument. This latter result implies that x is increasing as well in $E[v^p]$.

8.4 Litigation costs reduce expected dynamic value

For the competitive patent race case, I implicitly differentiate (27) with respect to Y and L, and multiply by $\frac{dx}{dY}$ to obtain

$$\frac{dx}{dL} = \frac{x\frac{dx}{dY}}{(v^p - L)\frac{dx}{dY} - \alpha},$$

which must be negative if $v^p > L$. For the monopoly case, I undertake the same operations on (30), arriving at

$$\frac{dx}{dL} = \frac{\left[\frac{dx}{dY}\right]^2}{\frac{d^2x}{dY^2}(v^p - L)},$$

which again must be negative if $v^p > L$ because x is increasing in Y at a decreasing rate. These results hold if $E[v^p]$ is substituted for v^p. These results hold as well if litigation costs are discounted by some factor $rZ_{min} > 0$.

8.5 The model holds if Brand has perfect knowledge regarding settlement and challenge occurrence

I want to show that if delay harms consumers in my model, which is to say, if ${}^a\mathcal{V}_c^{*S} < {}^a\mathcal{V}_c^{*L}$, it also harms them if I take into account consumer expectations regarding whether settlement will occur. I therefore wish to know whether ${}^a\mathcal{V}_c^{*S} <^a \mathcal{V}_c^{*L}$ implies that

$$E[\text{value with no ban on settlement}] < E[\text{value with a ban on settlement}],$$

where uncertainty regarding whether settlement will occur has been incorporated into the expectations.

I assume that Brand has perfect information about whether there will be a settlement. Under this assumption, consumers face uncertainty about whether they will enjoy settlement welfare or litigation welfare but need not worry that Brand adjusts its R&D effort in either scenario based on uncertainty about whether that scenario will obtain. I assume further that the probability of settlement occurrence, p_S, is independent of the other random variables in my model.

I wish to know whether

$$\underbrace{p_S {}^a\mathcal{V}_c^{*S} + (1-p_S) {}^a\mathcal{V}_c^{*L}}_{\text{expected value without a settlement ban}} < \underbrace{{}^a\mathcal{V}_c^{*L}}_{\text{expected value with a ban (i.e., } p_S = 0)} . \tag{31}$$

This must hold whenever $\mathcal{V}_c^{*S} < \mathcal{V}_c^{*L}$, so long as the probability of settlement is greater than zero. Thus whenever I have impermissible delay under my model, I have impermissible delay in the presence of uncertainty about settlement occurrence, regardless what probabilities consumers might wish to place on settlement occurrence.

8.6 The value-maximizing patent strength under litigation

Let t be $n - rL$ and d be $u + zk$. I set the derivative of ${}^a\mathcal{V}_c^{*L}$ with respect to pz equal to zero:

$$\frac{d\left(\frac{1}{r}\left[\left[1-\left(\frac{\alpha r^2}{p^{max}zm+t}\right)\right](d-kp^{max}z)\right]\right)}{dpz} = 0.$$

Carrying out the differentiation and rearranging, I obtain

$$-km^2\left(p^{max}z\right)^2 - 2ktmp^{max}z + \alpha r^2 md + \alpha r^2 kt - kt^2 = 0. \tag{32}$$

(32) may be solved for $p^{max}z$ using the quadratic formula.

8.7 Permissible delay

I presuppose Section 3. Let l be $u + k(z - E[z_E^{lit}])$, q be $E[z_E^{lit}]\pi(1 - s\beta) + zs\beta\pi$, and x_L be $1 - \left(\frac{\alpha r}{\frac{1}{r}\left[(E[z_E^{lit}])m+n\right]-Z_{min}L}\right)$. From (11) I have

$$\frac{1}{r}\left[\left[1 - \left(\frac{\alpha r}{\frac{1}{r}\left[(E[z_E^{lit}] + \Delta)m + n\right]}\right)\right]\left(u + (z - (E[z_E^{lit}] + \Delta))k\right)\right] - \frac{1}{r}\left[x_L l\right] = 0.$$

Rearranging, I obtain

$$-km\Delta^2 + \left[lm - k(q - \alpha r^2) - mx_L l\right]\Delta + (q - \alpha r^2)l - x_L lq = 0, \tag{33}$$

which can be solved for Δ using the quadratic formula.

8.8 Permissible delay in competition is decreasing in $E[T_E^{lit}]$ for my purposes

Implicitly differentiating (33), I obtain

$$\frac{d\Delta}{dE[z_E^{lit}]} = \frac{\left(ml\frac{\alpha r^2}{(q - rZ_{min}L)^2} - kx_L\right)q_\Delta}{\left(\frac{ml}{q_{2\Delta}}(\frac{\alpha r^2}{q - rZ_{min}L}) - k(1 - \frac{\alpha r^2}{q_{2\Delta}})\right)q_{2\Delta}} - 1, \tag{34}$$

where I have let q_Δ be $\left(E[z_E^{lit}] + \Delta\right)m + n$ and $q_{2\Delta}$ be $\left(E[z_E^{lit}] + 2\Delta\right)m + n$. Differentiating $\Delta_T \equiv -\frac{1}{r}ln(1 - z_E^{set*}) - E[T_E^{lit}]$ with respect to $E[T_E^{lit}]$, I have $\frac{1}{r}\left[\frac{1}{1 - z_E^{set*}}\right]\left[\frac{dz_E^{set*}}{dE[T_E^{lit}]}\right] - 1$. Substituting $z_E^{set*} = \Delta - E[z_E^{lit}]$ and $E[z_E^{lit}] = \frac{E[T_E^{lit}] - T_{min}}{T - T_{min}}(z - z_{min}) + z_{min}$ into this and solving, I obtain

$$\frac{d\Delta_T}{dE[T_E^{lit}]} = \left(\frac{1}{r}\left[\frac{1}{1 - z_E^{set*}}\right]\left[\frac{z - z_{min}}{T - T_{min}}\right]\left[\frac{d\Delta}{dE[z_E^{lit}]} + 1\right]\right) - 1, \tag{35}$$

Using a numerical optimization algorithm, it can be verified that (35) is negative over the modified bounds described in Table 1.

8.9 Determining α in the monopoly case

Into the monopoly condition defined by (4) and (5), I substitute (17) and (5) into (4), and $\frac{Y\alpha}{\alpha}$ for Y, to obtain $\frac{r}{\left(\frac{Y\alpha}{\alpha}+r\right)^2}v^p(1-c)-\alpha=0$. Rearranging, I have $r^2\alpha^2+(2r\left(Y\alpha\right)-rv^p(1-c))\,\alpha+\left(Y\alpha\right)^2=0$, from which I obtain $\alpha=\frac{v^p(1-c)-2(Y\alpha)\pm[v^p(1-c)(v^p(1-c)-4(Y\alpha))]^{1/2}}{2r}$ using the quadratic formula.

8.10 Determining relevance strength

To obtain the relevance strength p^R, I must find the $E[z_E^{lit}]$ for which litigation value equals settlement value at patent expiry. I call it z_E^R. I can then obtain p^R from the equation $p^Rz=z_E^R$. From (13), in competition I have

$$\frac{u}{r}\left[1-\left(\frac{\alpha r^2}{z\pi}\right)\right]=\frac{1}{r}\left[\left[1-\left(\frac{\alpha r}{\frac{1}{r}\left[z_E^Rm+n\right]-L}\right)\right]\left(u+(z-z_E^R)k\right)\right],$$

which can be rearranged to obtain

$$-km\left(z_E^R\right)^2+\left[md-k(t-\alpha r^2)-mu\left[1-\left(\frac{\alpha r^2}{z\pi}\right)\right]\right]z_E^R+d(t-\alpha r^2)-tu\left[1-\left(\frac{\alpha r^2}{z\pi}\right)\right]=0, \tag{36}$$

where I have employed the substitutions d and t defined in Section 8.6. (36) may be solved for z_E^R using the quadratic formula.

8.11 Fixed duopoly

I refer to duopoly that lasts until patent expiry T as "capped" and duopoly that lasts for only a fixed period H as "fixed". I consider only the competitive case (i.e., $n=1$ in this Section). I assume that there is a prechallenge period T_{min} (see Section 4.1.1 for more details). Results without such a period may be obtained by setting $T_{min}=0$.

To obtain expected consumer welfare in the fixed competitive dependence case, I first

obtain static consumer and Brand value over the domain $[0, \mathcal{R}_H]$, where $\mathcal{R}_H \equiv \frac{z-z_H}{1-z_H}$, which corresponds to entry dates over the period $[0, T-H]$. Changing the limits of integration in (7) for each of the three integrals to $[0, T_E]$, $[T_E, T_E+H]$, and $[T_E+H, \infty]$, respectively, and integrating, I obtain

$$V_c = \frac{1}{r}\left(CS_M + (1-z_E)\left[z_H k + (1-z_H)(\pi + D)\right]\right) \equiv V_c^{early}, \tag{37}$$

where $z_H \equiv 1 - e^{-rH}$. I have labeled V_c as V_c^{early} to indicate that (37) represents static consumer value for entry prior to $T-H$, and I recall that $k \equiv (1-\beta)\pi + d(\beta)$. Similarly, static Brand value (9) becomes

$$v^p = \frac{1}{r}\left[z_E \pi (1 - z_H s\beta) + z_H s\beta\pi\right] \equiv v^p_{early}, \tag{38}$$

and static Generic value (14) becomes

$$\frac{1}{r}\left[z_H(1-z_E)(1-s)\beta\pi\right]. \tag{39}$$

Over the domain $[\mathcal{R}_H, z]$, static settlement value is identical to that originally described by equations (8), (9), and (14).

Litigation can lead only to entry at $T_E = 0$ and $T_E = T$, so expected static value under litigation interpolates between early value at $T_E = 0$ and late value at $T_E = T$. Static consumer, rightsholder, and challenger value in litigation are

$$\frac{1}{r}\left[-z_E\left(\mathcal{R}_{min}k' - \mathcal{R}(\pi + D)\right) + z\mathcal{R}_{min}k' - z_{min}\mathcal{R}(\pi + D) + CS_M\right], \tag{40}$$

$$\frac{1}{r}\left[z_E(1 - \mathcal{R}_{min}z_H s\beta)\pi + z\mathcal{R}_{min}z_H s\beta\pi\right] - Z_{min}L, \tag{41}$$

and

$$\frac{1}{r}\left[(z-z_E)z_H\mathcal{R}_{min}(1-s)\beta\pi\right]-Z_{min}L_G, \tag{42}$$

respectively, where $k' \equiv z_H k + (1-z_H)(\pi+D)$, $\mathcal{R} \equiv \frac{1-z}{z-z_{min}}$, $\mathcal{R}_{min} \equiv \frac{1-z_{min}}{z-z_{min}}$ and L_G is the challenger's litigation cost.

Substituting (38) for v^p in (6) and (37) for V_c in (4) and setting $L=0$, I obtain

$$\frac{1}{r}\left[1-\frac{\alpha r}{\frac{1}{r}\left[E[z_E]m'+n'\right]}\right]\left[CS_M+(1-E[z_E])k'\right] \tag{43}$$

for expected consumer value for $E[z_E] \in [0, \mathcal{R}_H]$, $m' \equiv \pi(1-z_H s\beta)$, and $n' \equiv z_H s\beta\pi$. (43) intersects capped expected settlement value ${}^a\mathcal{V}_c^{*S}$ at $E[z_E]=\mathcal{R}_H$ and ${}^a\mathcal{V}_c^{*S}$ continues to hold for $E[z_E] \in [\mathcal{R}_H, z]$. I obtain expected consumer value under litigation over the entire domain $[0,z]$ by substituting (41) and (40) for v^p in (6) and V_c in (4) to obtain

$$\frac{1}{r}\left[1-\frac{\alpha r}{\frac{1}{r}\left[E[z_E]m'_L+n'_L\right]-Z_{min}L}\right]\left[-E[z_E]k'_L+k''_L+CS_M\right], \tag{44}$$

where $m'_L \equiv (1-\mathcal{R}_{min}z_H s\beta)\pi$, $n'_L \equiv z\mathcal{R}_{min}z_H s\beta\pi$, $k'_L \equiv \mathcal{R}_{min}k' - \mathcal{R}(\pi+D)$, and $k''_L \equiv z\mathcal{R}_{min}k' - z_{min}\mathcal{R}(\pi+D)$.

When $L=0$, expected litigation value meets expected settlement value at z_{min}, z, and possibly at a third point, as explained below. When expected value under litigation exceeds expected settlement value over (z_{min}, z) and $L=0$, $L>0$ implies that expected litigation value will intersect expected settlement value at two points, $z_l > z_{min}$ and $z_h < z$, for which $z_l < z_h$. For any $E[z_E] \in (z_l, z_h)$, consumers are better off under litigation than settlement. Assuming that $z_l \in (z_{min}, \mathcal{R}_H]$ and $z_h \in [\mathcal{R}_H, z)$, z_l may be determined by equating (44) and (43) and solving for $E[z_E]$, while z_h may be determined by equating (44) and ${}^a\mathcal{V}_c^{*L}$ and solving for $E[z_E]$. In both cases, I obtain a cubic equation in $E[z_E]$, which may be solved

using the cubic formula. A similar exercise may be carried out to determine intersection points in $[T_{min}, T]$; however, the solution must eventually be obtained numerically.

Δ for $E[z_E] \in [z_{min}, \mathcal{R}_H]$ may be obtained by setting

$$\frac{1}{r}\left[1 - \frac{\alpha r}{\frac{1}{r}\left[(E[z_E]+\Delta)m' + n'\right]}\right]\left[CS_M + (1-(E[z_E]+\Delta))\,k'\right] \tag{45}$$

equal to (44) and solving. I obtain a quadratic equation, $a\Delta^2 + b\Delta + c = 0$, where $a = -k'm'$, $b = (m'\left((1-E[z_E])k' + CS_M\right)) - (E[z_E]m' + n' - \alpha r^2)\,k' - m'v$, v is expected consumer value under litigation as defined in (44), and $c = \left((E[z_E]m' + n' - \alpha r^2)\left((1-E[z_E])k' + CS_M\right)\right) - (E[z_E]m' + n')v$. The equation can be solved using the quadratic formula. I ignore Δ for $E[z_E] \in (\mathcal{R}_H, z]$ because over the drug market variable bounds listed in Table 1 to which I apply my model, the litigation value maximizing $E[z_E]$ is always in $[z_{min}, \mathcal{R}_H]$ and for Δ^{max} over these bounds $E[z_E^{lit}] + \Delta^{max} < \mathcal{R}_H$.[11]

The patent strength that maximizes expected value under litigation, p^{max}, may be obtained by differentiating (44) with respect to $E[z_E] = pz$ and setting it equal to zero. Doing so and rearranging, I obtain another quadratic equation with coefficients $a = -(m'_L)^2 k'_L$, $b = -\left(2k'_L(n'_L - rZ_{min}L)m'_L\right)$, and $c = \alpha r^2 m'_L(k''_L + CS_M) + \alpha r^2 k'_L(n'_L - rZ_{min}L) + k'_L(n'_L - rZ_{min}L)^2$, which can be solved using the quadratic formula.

I now show that fixed settlement value exceeds capped settlement value in identical expectations and competition for at least some values of $E[z_E]$ immediately below $\mathcal{R}_H$, if $s < \frac{\pi}{\pi+D}$. This result holds for any x that is increasing in v^p; the Poisson innovation production function assumption that gives rise to (5) is not required. I first remark that (2) with v^p defined either as in (9) or v^p_{early} creates a one to one correspondence between $E[z_E]$ and Y for any given dynamic value $x = E[f(w)|Y]$. To show this more clearly, I rely on the

[11] When optimizing over all $E[z_E] \in [z_{min}, z]$ to identify the maximum permissible strengthening identified in Table 4, I checked that my result exceeds $\frac{z - z_H}{z - z_{min}}$ to ensure that the maximum I identify is a maximum over the entire domain.

linearity of z_E in (9) and v^p_{early}, to write $E[v^p] = v^p(E[z_E])$. I substitute x into (2) to obtain

$$v^p(E[z_E]) = \frac{\alpha Y}{x}, \tag{46}$$

which reflects the unique correspondence between $E[z_E]$ and Y. I proceed by first comparing fixed and capped expected consumer values as a function of Y, concluding that fixed must exceed capped for all values of Y corresponding to $E[z_E] \in [z_{min}, \mathcal{R}_H)$. I then consider the consequences of my results in Y-space for expected consumer value in $E[z_E]$-space.

Using v^p_{early} for v^p in (46), I obtain

$$E[z_E] = \frac{1}{\pi(1 - z_H s\beta)} \left[\frac{r\alpha Y}{x} - z_H s\beta\pi \right]. \tag{47}$$

Substituting this result and V_c^{early} into $\mathcal{V}_c^*$, I obtain for expected consumer value under fixed duopoly:

$$\frac{1}{r}x \left[CS_M + \frac{k'}{1 - z_H s\beta} \left(1 - \frac{r\alpha Y}{\pi x} \right) \right]. \tag{48}$$

Using (9) for v^p in (46), I obtain

$$E[z_E] = \frac{1}{m} \left[\frac{r\alpha Y}{x} - n \right]. \tag{49}$$

Substituting (49) and (8) into $\mathcal{V}_c^*$, I obtain

$$\frac{1}{r}x \left[CS_M + (1 - z)(\pi + D) + \frac{k}{1 - s\beta} \left(z - \frac{r\alpha Y}{\pi x} \right) \right] \tag{50}$$

for the capped duopoly case. Because $\frac{1}{r}x$ appears in both (48) and (50), the relative sizes of (48) and (50) are determined by the expressions in the square brackets in those equations, both of which are linear in $\frac{r\alpha Y}{\pi x}$. I note that $\frac{r\alpha Y}{\pi x}$ is increasing in Y and therefore in $E[z_E]$.

I am concerned only with $E[z_E] \in [z_{min}, \mathcal{R}_H]$. At $E[z_E] = \mathcal{R}_H$, fixed and capped value

coincide in both Y- and $E[z_E]$-space because the static values coincide at that point (i.e., $E[V_c^{early}] = E[V_c]$ and $E[v^p_{early}] = E[v^p]$ at that point). Because two lines only cross once, I need only show that for some $\frac{r\alpha Y}{\pi x}$ below that corresponding to $\mathcal{R}_H$, fixed value exceeds capped value in order to conclude that this holds over the entire domain $[z_{min}, \mathcal{R}_H)$. I choose the $\frac{r\alpha Y}{\pi x}$ corresponding to z_{min} in the fixed case. Setting $E[z_E] = z_{min}$ in (47) and rearranging, I have

$$\frac{r\alpha Y}{\pi x} = z_{min}(1 - z_H s\beta) + z_H s\beta. \tag{51}$$

Substituting this into (48), simplifying, and ignoring the CS_M term because it appears in both fixed and capped value, I obtain $(1 - z_{min})k'$. Substituting (51) into (50), and here again ignoring CS_M, I obtain $\frac{k}{1-s\beta}\left[(z - z_H s\beta) - z_{min}(1 - z_H s\beta)\right] + (1 - z)(\pi + D)$. Solving for the z_{min} for which the former exceeds the latter, I obtain

$$z_{min} < 1 + \frac{(1 - s\beta)(1 - z)(\pi + D) - k(1 - z)}{k(1 - z_H s\beta) - (1 - s\beta)k'},$$

where I have assumed that $\frac{k(1-z_H s\beta)}{1-s\beta} - k' < 0$. Remarking that $k(1 - z_H s\beta) - (1 - s\beta)k' = (1 - z_H)(k - (1 - s\beta)(\pi + D))$, I obtain $z_{min} < \mathcal{R}_H$. Thus assuming that $\frac{k(1-z_H s\beta)}{1-s\beta} - k' < 0$, fixed will exceed capped if, as seems reasonable to assume when H is small, as in the drug market, z_{min} is less than the point at which fixed and capped value meet.

$\frac{k(1-z_H s\beta)}{1-s\beta} - k' < 0$ holds if

$$k - (1 - s\beta)(\pi + D) < 0. \tag{52}$$

Because $k \equiv (1 - \beta)\pi + d(\beta)$, the behavior of $d(\beta)$ is key. If there is no deadweight loss, or deadweight loss increases linearly in β (i.e., $d(\beta) = (1 - \beta)D$), then (52) will be satisfied. If deadweight loss increases nonlinearly in β, then it may be violated. But the maximum size of deadweight loss, D, also plays a role. If deadweight loss is small, then the π terms

will be determinative, and for these the inequality is always satisfied. Specifically, because $d(\beta) \leq D$ if D is such that $(1-\beta)\pi + D < (1-s\beta)(\pi + D)$, then the inequality must always hold. $(1-\beta)\pi + D < (1-s\beta)(\pi + D)$ implies

$$s < \frac{\pi}{\pi + D}, \tag{53}$$

which constitutes a sufficient condition for (52) to hold.

Having shown the conditions for which fixed always exceeds capped for all Y that correspond to $E[z_E] \in [z_{min}, \mathcal{R}_H)$, it remains to determine what this means in $E[z_E]$-space. Because (38) is less than (9) over $[z_{min}, \mathcal{R}_H)$, it is evident from (46) that the $E[z_E]$ that achieves a given $\frac{r\alpha Y}{\pi x}$, and therefore a given Y, must be greater in fixed than in capped. It follows that the graph of fixed shifts to the right relative to capped when I move from Y-space to $E[z_E]$ space. However, because static values coincide at $\mathcal{R}_H$, fixed and capped continue to coincide at $\mathcal{R}_H$ in both spaces. Thus fixed value is pushed up against $\mathcal{R}_H$ in the transition from Y-space to $E[z_E]$-space. Because it is possible that this might cause fixed value to intersect with capped value for smaller $E[z_E]$s, I have not shown that fixed value exceeds capped value for all $E[z_E] \in [z_{min}, \mathcal{R}_H)$. However, because fixed remains a function in $E[z_E]$-space and the shift does not change the intersection point, $\mathcal{R}_H$, the shift does not push fixed past $\mathcal{R}_H$. Thus fixed always intersects capped value from above as $E[z_E]$ increases toward $\mathcal{R}_H$, and therefore fixed must always exceed capped over some portion of $[z_{min}, \mathcal{R}_H)$ immediately below $\mathcal{R}_H$. Thus at least for stronger patents (that are not so strong as to correspond to discount factors in excess of $\mathcal{R}_H$), consumers are always rendered better off in settlement by a fixing of duopoly. This result holds as well when the substitution $1 - e^{-rT_E}$ is made for $E[z_E]$ because it is a monotonic transformation.

I now show that in the absence of litigation costs, fixed value in litigation exceeds fixed settlement value for at least some $E[z_E]$ immediately below z (i.e., the graph of the former crosses that of the latter from above). This result again holds for any scaling factor x

that is increasing in v^p. I proceed again by first showing that for any Y corresponding to $E[z_E] \in (z_{min}, z)$ litigation exceeds settlement and then discussing the consequences of this result in $E[z_E]$-space. I consider first the case of $E[z_E] \in [z_{min}, \mathcal{R}_H]$ and then the case of $E[z_E] \in [\mathcal{R}_H, z]$.

In the case of $E[z_E] \in [z_{min}, \mathcal{R}_H]$, my expression for fixed settlement value remains (48). Using (41) for static producer value and (40) for static consumer value, with $L = 0$, I obtain litigation value of

$$-\frac{\mathcal{R}_{min}k' - \mathcal{R}(\pi + D)}{1 - \mathcal{R}_{min}z_H s\beta}\left[\frac{r\alpha Y}{\pi x} - z\mathcal{R}_{min}z_H s\beta\right] + z\mathcal{R}_{min}k' - z_{min}\mathcal{R}(\pi + D), \tag{54}$$

where I have left off the $\frac{1}{r}x$ and CS_M terms because here again they are shared by fixed value and therefore may be ignored in an analysis of relative sizes. Here again both values are linear in $\frac{r\alpha Y}{\pi x}$. Because they intersect at z_{min}, showing that litigation exceeds settlement at a value for $\frac{r\alpha Y}{\pi x}$ greater than that corresponding to z_{min} will suffice to show that litigation exceeds settlement over the entire segment. I choose $\mathcal{R}_H$ as my value for $\frac{r\alpha Y}{\pi x}$.

Fixed value is (37) evaluated at $E[z_E] = \mathcal{R}_H$, after x has been discarded because it appears in both expressions under comparison. Discarding $\frac{1}{r}$ and CS_M for the same reason, I obtain $(1 - \mathcal{R}_H)k'$. From (46), I have

$$\frac{z - z_H}{1 - z_H}(1 - z_H s\beta) + z_H s\beta = \frac{r\alpha Y}{\pi x(Y)}. \tag{55}$$

Plugging (55) into (54) and comparing with $(1 - \mathcal{R}_H)k'$, I obtain the following necessary condition for litigation to exceed settlement:

$$(z_{min} - \mathcal{R}_H)\frac{k'}{\pi + D} - (1 - z_H s\beta)(z_{min} - \mathcal{R}_H) > 0.$$

For $z_{min} < \mathcal{R}_H$, this becomes $\frac{k'}{\pi + D} - (1 - z_H s\beta) < 0$, which simplifies to (52).

Turning now to $E[z_E] \in [\mathcal{R}_H, z]$, I remark that (48) and (54) are again both linear in

$\frac{r\alpha Y}{\pi x}$ and meet in both $E[z_E]$- and Y-space at $E[z_E] = z$. I therefore need to show that litigation exceeds fixed at some $\frac{r\alpha Y}{\pi x}$ smaller than the intersection point. I again choose $\mathcal{R}_H$. Capped and fixed values intersect at this point, so the conditions just described under which litigation exceeds capped at this point apply as well to define when litigation exceeds fixed at this point.

I have therefore shown that in Y-space expected litigation value exceeds expected settlement value over (z_{min}, z). Because static private value in litigation is less than static private value in settlement over this entire domain, but coincides with settlement value at the endpoints in both Y- and $E[z_E]$-space, expected litigation value will be pulled to the right along the domain while maintaining the same endpoints in the transformation from Y-space to $E[z_E]$-space. This means that it is possible that expected litigation value may fall below expected settlement value for some lower values of $E[z_E]$. However, because the intersection point of capped expected settlement value and litigation value at $E[z_E] = z$ does not change and litigation value remains a function of $E[z_E]$, a portion of expected litigation value immediately below z must always exceed expected settlement value. The transformation from $E[z_E]$-space to $E[T_E]$-space similarly pulls expected litigation value to the right because $E[z_E]$ maps onto $1 - e^{-rT_E}$ in settlement but maps onto $p(T - T_{min}) + T_{min}$ in litigation. Therefore, by the same argument, a portion of expected litigation value immediately below T must exceed expected settlement value. For sufficiently large litigation cost, L, however, this will not hold.

8.12 Estimates

Here I describe how I arrive at estimates for the model's variables in the drug market context.

8.12.1 R&D cost growth, the tax rate, inflation, and αY

The debate over cost Morgan et al. (2011) observe that there is no "gold standard" study of the average cost of developing a new drug and nearly all major studies are based

at least in part on data sets that have not been made publicly available for audit. They find that published cost estimates from the 2000s vary fourfold from $422 million to $1800 million in 2009 dollars.

I use the estimate of DiMasi et al. (2003), which is closest to the mean estimate in this range. Morgan et al. (2011) point out that DiMasi et al. (2003) report cost for new compounds that have never been approved before for any use and that were developed entirely by the firms that provide data for their study. This means their estimate does not account for the use of old compounds in new combinations with other drugs, new dosages, or in treating different ailments.

Because the cost of reusing compounds may be significantly less than the cost of developing an entirely new compound, the cost estimate of DiMasi et al. (2003) is likely to be too high as an estimate for new drugs generally. The effect of this overstatement on my conclusions could be minimal if my benefits estimates are also similarly biased upward. However, if my benefits estimates are not also biased upward, then the effect will be to understate the amount of permissible delay. This is because the benefits of drugs relative to costs would be higher than originally thought, making patent strengthening through settlement more desirable.[12]

Deriving cost DiMasi et al. (2003) calculate the present value of expected drug development costs for the average drug using cost data for the years 1980 to 2000. They arrive at $897 million in pretax 2000 dollars, including both pre and post approval costs. By comparing this result with DiMasi et al. (1991), DiMasi et al. (2003) conclude that there is a 7.4% annual growth rate in capitalized R&D costs. Grabowski et al. (2002) remark that the mean drug introduction date in DiMasi et al. (2003) is 1997. Accordingly, I use the 7.4%

[12]Higher benefits relative to costs might suggest monopoly in innovation. In monopoly, the value ratio $\frac{z\pi(1-c)}{r\alpha Y}$ discussed in Section 8.12.3 exceeds 300% (see footnote 14). I back benefit $\frac{z\pi(1-c)}{r}$ out of (56) using the cost estimate (αY) of DiMasi et al. (2003) and then plug the low 2000s αY in Morgan et al. (2011) (exclusive of the single-drug Global Alliance estimate) into $\frac{z\pi(1-c)}{r\alpha Y}$. The result is below 300%, rejecting monopoly.

growth rate over 18 years from 1997 to 2015, as well as GDP Implicit Price Deflator data for inflation from 2000 to 2015, which yields inflation over the period of 33.84%, to obtain the following projection of 2015 costs: $4548.46 million.[13]

Because R&D costs are tax deductible, I follow Grabowski et al. (2002) in multiplying the pretax value by 70% (i.e., assuming a 30% tax rate) to obtain $3183.92 as my estimate of expected after tax R&D costs, αY.

8.12.2 *c*

Grabowski et al. (2002) employ an average contribution margin of 45% and non-R&D capital costs of 3.3% of sales. I therefore have $c = 58.3\%$.

8.12.3 Value ratio and π

I obtain π from αY by assuming that drug revenues as a share of R&D costs are stable over time and observing, as I did in calculating α, that my R&D cost estimate is based on data from a period when most entry took place at T and without litigation. I then find the constant revenue stream π for which the present value of revenues yields the required share of R&D costs. I assume that the revenue stream lasts my estimated duration of exclusivity, which I give below as 12.9 years.

Grabowski and Vernon (1994) report 1990 expected net present value of drug development of $22.2 million against costs of $201.9 million, or 11% of costs after rounding. Grabowski et al. (2002) report 2000 expected net present value of drug development of $45 million against costs of $480.3 million, or 9% of costs after rounding. I assume that the ratio in 2015 was 10%.

Assuming entry at patent expiry, the present value of revenues is $\frac{1}{r}z\pi$, so I have

$$\frac{1}{r}z\pi(1-c) = (1.10)\alpha Y, \tag{56}$$

[13] In November 2014, the authors of DiMasi et al. (2003) released an update incorporating data up to 2013. Their updated results are close to my estimate of cost once the new study period limit is taken into account.

where I have discounted the present value of revenues by $1 - c$ because the revenue numbers in Grabowski et al. (2002) are stated after removal of all costs other than R&D costs. Solving for π, I obtain \$1249.25 million.[14]

Grabowski et al. (2002) remark that neither revenues nor costs are incurred evenly over the life of a drug. The profit estimates of Grabowski et al. (2002) on which I rely include profits earned after patent expiry and the entrance of generics and during an initial period in which sales are low because marketing has not yet created maximum demand. The constant π that I derive is therefore necessarily only a rough estimate.

8.12.4 ϵ

A literature based on changes in drug copayments for insurance plans, summarized by Liu and Chollet (2006), suggests that the price elasticity of demand for pharmaceuticals ranges from -0.6 to -0.1. However, using data on generic entry, Duflos and Lichtenberg (2012) conclude that the volume-expanding effects of lower prices are fully counteracted by the volume-reducing effects of reduced marketing, with the result that volume does not change in response to generic entry.

Translating the conclusion of Duflos and Lichtenberg (2012) into my model requires some finesse because the model does not take the demand expanding effects of marketing into account. Simply treating the result in Duflos and Lichtenberg (2012) as implying an elasticity of zero is problematic because the constancy of sales volume is due to a shifting of the demand curve and not to the shape of the curve. An elasticity of zero implies that consumer value increases dollar for dollar as price and profit fall. But if volume is constant because of a reduction in demand, then consumer value may not change at all as a result of a price decline.

Despite these difficulties, I fit the result of Duflos and Lichtenberg (2012) into my model

[14]My use of (56) to obtain π means that the existence condition for monopoly in innovation, $\frac{z\pi(1-c)}{4r} > Y\alpha$ (as described in Section 4.1.3), depends entirely on my guess for the value ratio. The condition is satisfied only if the value ratio is greater than 300%. My guess for the value ratio is 10%.

by interpreting their work to imply an elasticity of zero. I draw some comfort in this choice from the observation that I may think of the demand associated with a large amount of marketing as in some sense a measure of the true value of the drug to consumers. In this account, marketing allows consumers to become aware of the value that they really place on the product or, alternatively, reduces their power to hold out for lower prices by insisting that they be charged based on an artificial, low demand line. As a result, the decline in demand associated with a reduction in marketing is not a true decline, and therefore the lower prices enjoyed by consumers after generic entry can be thought to correspond to higher value for consumers.

I therefore assume a zero elasticity with respect to a reduction in price from the monopoly level as a result of Generic entry. But I do not make the same assumption for an increase in price. If elasticity were zero for price increases, then Brand would charge an infinite price during the exclusivity term. I therefore assume that above Brand's monopoly price demand is somewhat elastic and the copayment elasticities summarized in Liu and Chollet (2006) apply to such price increases. I choose the average of the endpoints of the range of elasticities reported by Liu and Chollet (2006), or -0.35, as my elasticity guess.

8.12.5 D and $d(\beta)$

A consequence of assuming that elasticity for price drops is zero is that there is no deadweight loss, D, and consequently no share of deadweight loss, $d(\beta)$, that will be returned to consumers when price falls. I therefore set both of these equal to zero.

8.12.6 CS_M

I assume that, above the price charged by Brand when alone in the market, demand is linear, downward-sloping, and not perfectly inelastic. Consumer surplus is the area of a triangle

with base q and height $-\left(\frac{dp}{dq}\right)q$. So I have

$$CS = -\frac{1}{2}\left(\frac{dp}{dq}\right)q^2. \tag{57}$$

From the definition of elasticity, ϵ, I have $\frac{1}{\epsilon} = \frac{dp}{dq}\frac{q}{p}$ and therefore $\frac{p}{\epsilon q} = \frac{dp}{dq}$. Substituting this into (57), I obtain

$$CS = -\frac{pq}{2\epsilon}. \tag{58}$$

Although my model defines value in terms of consumer value at the monopoly price, CS_M, and monopoly revenue, π, it does not require that these be consumer and firm value under monopoly conditions (i.e., when firm value is maximized for a given demand schedule). They need only represent consumer and firm value when the rightsholder is alone in the market (i.e., the lowest static consumer value and greatest static firm value available). Because the estimates for ϵ that I use are less than one in absolute value, revenue cannot be at a maximum for these estimates (at the maximum $|\epsilon| = 1$). Revenue is lower and consumer value is higher than under a pure monopoly scenario. I therefore define CS_M and π in my model to be CS and pq, respectively, in (58), and obtain, from (58),

$$CS_M = -\frac{\pi}{2\epsilon}. \tag{59}$$

This yields a guess for consumer surplus at the monopoly price of $1784.64 million.

This guess gives a ratio of firm value (assuming exclusivity until patent expiry and no litigation) to total innovation value, $\frac{1}{r}z\pi/\frac{1}{r}(CS_M+\pi)$, of 32%. Nordhaus (2004) estimates that only 2.2% of innovation value is appropriated by firms. My estimate for CS_M is therefore too small or there are massive spillovers outside of the drug market itself. Either way, if innovation value really is much larger than it appears to be in my model of the drug market, any increase in profit to producers and thereby to R&D investment will be much more

valuable than it is pictured to be here. This in turn will increase permissible strengthening in the drug market relative to my estimates. I do not attempt to adjust consumer value based on this observation; my estimates are open to critique on this ground.

8.12.7 β, M, N, and s_{US}

β is the ratio of revenues earned by both Brand and Generic together after entry to the revenues earned by Brand before entry. The calculation of β is somewhat involved because I assume that generic entry following U.S. patent litigation affects only Brand's U.S. profits.

π is global revenue because the figures of Grabowski et al. (2002) from which I derive it are based on international sales data. Grabowski et al. (2002) reference an international sales multiplier of 2.19. Let M be that multiplier, Brand's U.S. market share after Generic entry be s_{US}, and the fraction of preentry price constituted by Generic's price be N. I have

$$\begin{aligned}\beta &= \frac{\pi\frac{M-1}{M} + \frac{\pi}{M}(s_{US} + (1 - s_{US})N)}{\pi} \\ &= \frac{1}{M}\left[M - 1 + (s_{US} + (1 - s_{US})N)\right].\end{aligned} \tag{60}$$

In (60), I obtain U.S. revenue by dividing total (international) revenue, π, by the multiplier, M. I discount the revenue earned by Generic in her share of the U.S. market, $1 - s_{US}$, by the postentry price ratio for generic drugs, N. I do not discount Brand's share, s_{US}, by this ratio, however, because, as I discuss more below, Brand's price does not drop after entry. I assume that Brand's price in non-U.S. markets does not fall either, and so I have $\pi\frac{M-1}{M}$ for non-U.S. postentry profits.[15]

(60) yields a guess of 90% for β. In calculating β, I employ the following estimates for s_{US} and N.

s_{US} Grabowski et al. (2014) estimate Brand market share after generic entry for the U.S.

[15] By assuming that profit is zero after the expiration of duopoly (e.g., after T in the capped case), I assume that the expiration of duopoly in the U.S. leads to competition and zero profits globally.

market only to be 16% for new molecular entities facing first generic entry from 2011 to 2012. Grabowski et al. (2014) use aggregate data for all generic entries and not exclusively for those pursuant to a Paragraph IV challenge. If generic entry after patent expiry drives share lower than entry pursuant to a Paragraph IV challenge, then this aggregate number will overstate the drop. Still, I use 16% as my guess.

N Frank and Salkever (1997) and Saha et al. (2006) find no or little reduction, respectively, in Brand price as a result of entry. Olson and Wendling (2013) find that the ratio of postentry price for generic drugs to the preentry price for the branded drug averaged 73.5% for Paragraph IV-based generic entrants in a sample of all oral solid medications sold in the U.S. from April 2003 to December 2010. Let N be the ratio of preentry price to postentry price for generic firms. I assume a postentry price discount of $N = 0.735$ for Generic and no price reduction for Brand.

8.12.8 *s*

s is the share of total postentry revenue of both Brand and Generic that is enjoyed by Brand. To obtain it, I remove Generic's postentry profits, $\frac{\pi}{M}(1 - s_{US})N$ from total postentry profits (the numerator in (60)) and find the ratio of the result to total postentry profits. I have $s = \frac{M-1+s_{US}}{M-1+(s_{US}+(1-s_{US})N)}$. This yields a guess of 68.62% for s. Brand's total share is much higher than its U.S. share of (s_{US} =) 16% because I assume that Generic does not enter internationally.

8.12.9 *r*

I follow DiMasi and Grabowski (2007) in using a real cost of capital for the drug industry of 11.5%.

8.12.10 T

T gives the exclusivity period for a drug that would apply were a court to reject a challenge. In this sense it is the official exclusivity period. Grabowski et al. (2014) find an average length of exclusivity of 12.9 years for all new molecular entities for the years 2011 to 2012. These include drugs for which entry occurred pursuant to a successful Paragraph IV challenge or a challenge settled for early entry, so actual exclusivity in the absence of a challenge is likely to be longer. I nevertheless use 12.9 years as my estimate for T.

8.12.11 T_{min}

Grabowski et al. (2014) find that for new molecular entities experiencing first generic entry in 2012 the average time between launch of the drug and Paragraph IV challenge was 6.9 years and that the trend is downward. I use 6.9 years as my estimate. This exceeds the five years of required exclusivity for new chemical entities under the Hatch-Waxman Act. One explanation may be that not all challenges are filed as early as the act permits. Another may be that the act permits imposition of a 30 month stay on entry while the parties litigate the patent. However, the estimate is shorter than the 7.5 years of exclusivity that includes the full 30 month stay. This may be because some challenges are resolved before the end of the stay and some drugs emerge from new chemical entity exclusivity without any remaining patent protection.

8.12.12 L

Elhauge and Krueger (2012) estimate average litigation costs of $10 million in 2011, which they consider a "good high-end estimate". I use the GDP Implicit Price Deflator over the years 2011 to 2015 to obtain a 2015 estimate for L of $10.61 million.

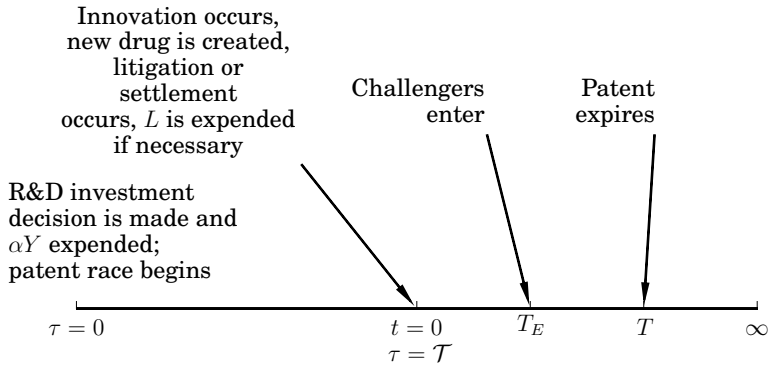
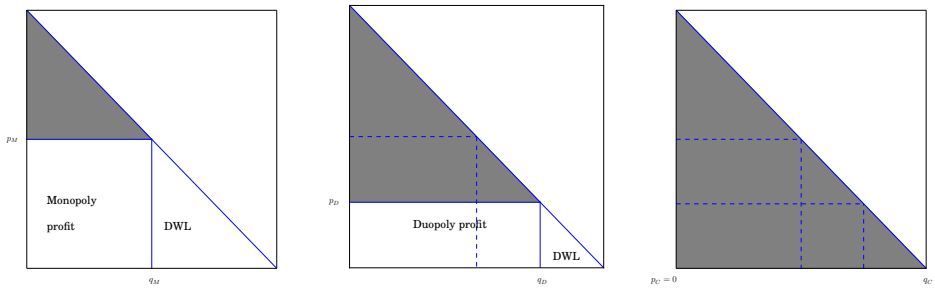


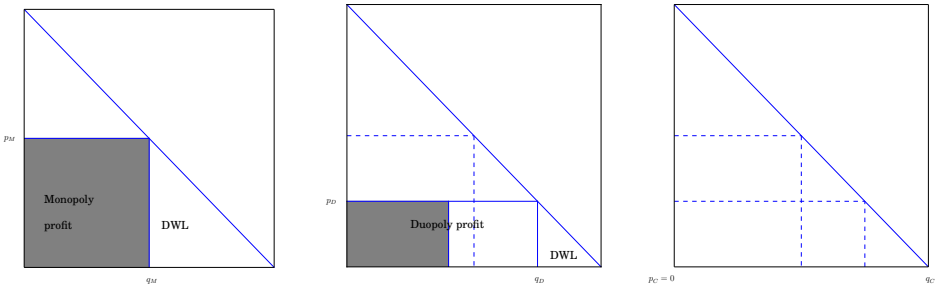
Figure 1

Consumer value



(a) (b) (c)

Firm value



(d) (e) (f)

Figure 2

Caption: The shaded areas in Figures 2(a) through 2(c) give static consumer welfare per unit time under monopoly, duopoly, and competition, respectively. The shaded areas in Figures 2(d) through 2(f) give Brand's static value per unit time under those three regimes.

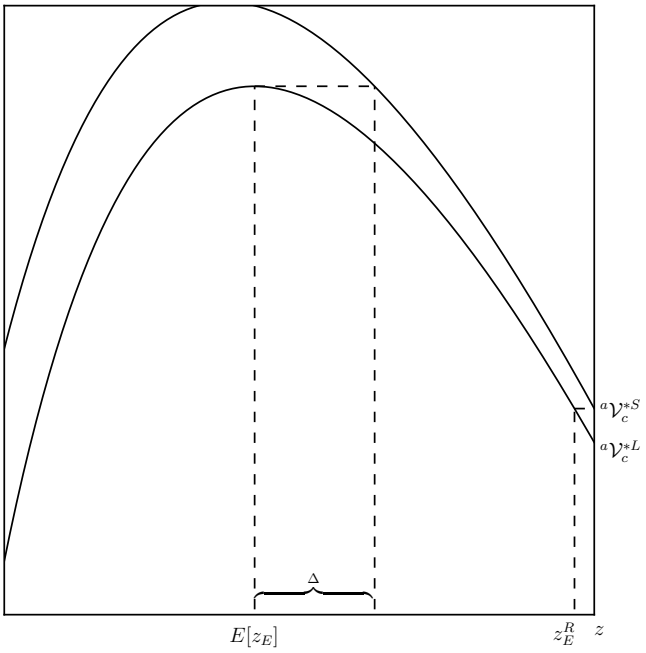


Figure 3

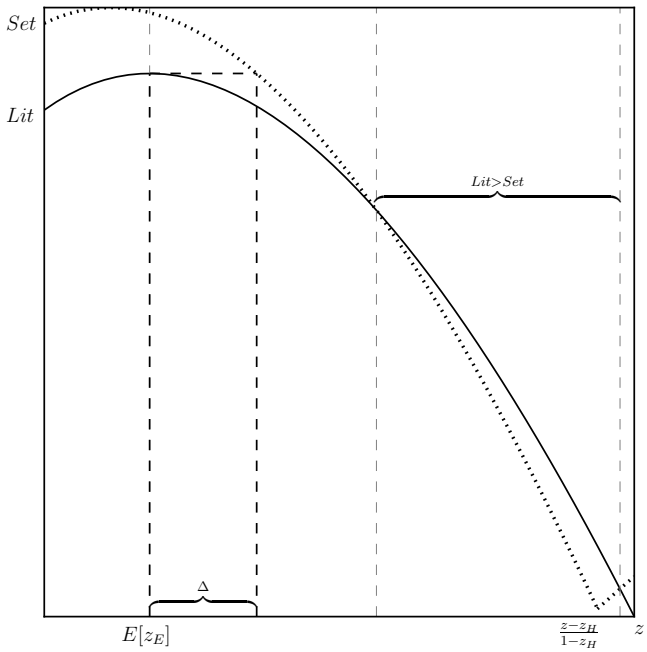


Figure 4

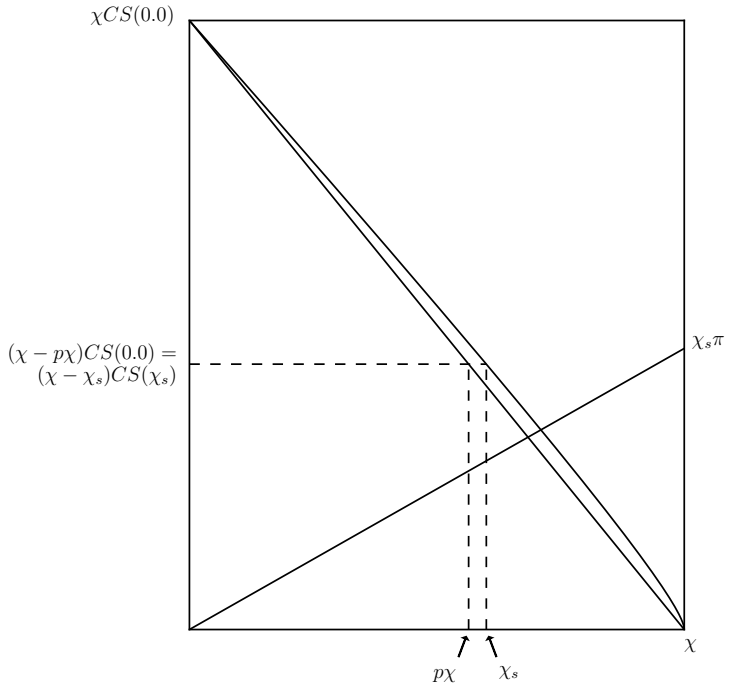


Figure 5

Table 1: Parameter estimates and bounds ($ in millions)

Variable	Estimate	Capped duopoly bounds	Fixed duopoly bounds
R&D cost (pretax)	$897.0	($880.0, $1000.0)	($700.0, $1000.0)
Inflation 2000-2015	33.84%	(30%, 34%)	(20%, 40%)
R&D cost annual growth rate	7.4%	(7%, 10%)	(6%, 10%)
Effective tax rate	30%	(29%, 33%)	(20%, 40%)
Net present value of drug as share of R&D cost	10%	(5%, 10.25%)	(1%, 12%)
Real interest rate (r)	11.5%	(10.9%, 15%)	(10%, 15%)
Share of sales not spent on costs other than R&D costs ($1-c$)	41.7%	(30%, 80%)	(30%, 80%)
Price after Generic entry as share of preentry price (N)	73.5%	(50%, 73.5%)	(50%, 90%)
Global sales multiplier (global sales as share of U.S. sales) (M)	2.19	(1.8, 2.3)	(1.8, 2.3)
Brand's share of U.S. market after Generic entry (s_{US})	16%	(10%, 25%)	(10%, 50%)
Litigation cost	$10.0	($5.0, $12.69)	($0, $30)
Price elasticity of demand for pharmaceuticals (ϵ)	-0.35	(-0.6, -0.34)	(-0.6, -0.19)
Duration of exclusivity (T)	12.9 years		
Duration of exclusivity prior to challenge (T_{min})	6.9 years		
Deadweight loss (D)	$0.0		

Table 2: Implied estimates ($ in millions)

Variable	Guess
αY	$3183.92
π	$1249.25
β	89.84%
s	68.62%
CS_M	$1784.64
α	$2768.63
$d(\beta)$	$0.0
L	$10.61

Table 3: Capped duopoly results and sensitivity

Permissible delay (Δ_T) at p^{max}	Optimal patent strength (p^{max})	Patent strength required for delay until patent expiry to benefit consumers (relevance strength)	Consumer harm from delay until patent expiry (evaluated at p^{max})
15.06 months	19.4%	94.6%	\$1.83 million
Extrema over the bounds in Table 1			
Maximum permissible delay	Maximum optimal patent strength	Minimum relevance strength	Minimum consumer harm from delay until patent expiry (evaluated at p^{max})
27.42 months (2.29 years)	63.7%	80.3%	\$0.14 million

Table 4: Fixed duopoly results and sensitivity

Permissible delay (Δ_T) at p^{max}	Optimal patent strength (p^{max})	Size of zone for which litigation value exceeds settlement value (as a share of challengeable patent term)	Patent strength required for delay until patent expiry to benefit consumers (relevance strength)	Consumer harm from delay until patent expiry (at greatest delay acceptable to Generic without a payment)
0.23 months	0.0%	5.93 years (98.88%)	99.86%	\$1313.61 million (\$134.61 million)
Extrema over the modified bounds in Table 1				
Maximum permissible delay	Maximum optimal patent strength	Minimum size of zone described above	Minimum relevance strength	Minimum consumer harm from entry at patent expiry
25.55 months (2.23 years)	0.0%	3.59 years (59.83%)	96.49%	\$108.49 million

Table 5: Generic's willingness to delay

Duopoly capped at patent expiry	
Willingness to delay (evaluated at p^{max})	Maximum share of impermissible delay ($\frac{\Delta_G^T - \Delta_T}{T-(T_E+\Delta_T)}$) over Table 1 bounds (consumer welfare loss from delay until patent expiry)
-2.95 months	2.3% ($461,000)
Duopoly fixed at half a year	
Willingness to delay (evaluated at p^{max})	Maximum share of impermissible strengthening ($\frac{\Delta_G^T - \Delta_T}{T-(T_E+\Delta_T)}$) over Table 1 modified bounds (consumer welfare loss from delay until patent expiry)
6.68 months	100% ($647.94 million)

Note: The consumer harm estimates that I report in parentheses are for the parameters that create the associated maximum. The negative sign on permissible delay in the case of duopoly until patent expiry means that a hastening is required.

Table 6: Reverse payments

Duopoly capped at patent expiry		
	Reverse payment required to induce delay until patent expiry	Additional consumer harm from delay until patent expiry that is due to reverse payment
Using estimates in Table 1	\$1220.18 million	\$35.16 million
At maximum in Table 5	\$354.24 million	\$1.64 million
Duopoly fixed at half a year		
Using estimates in Table 1	\$160.52 million	\$34.60 million
At maximum in Table 5	\$0 (i.e., Generic is willing to delay until patent expiry without a payment)	\$0

Note: The payments I report are the minimum payments that Brand must make in order to make Generic willing to agree to delay until patent expiry. If Generic drives a hard bargain, Brand might need to pay more.