Reducing Drug Prices without Depressing Innovation *

Stephen W. Salant
William James Adams

January 15, 2019

Abstract

Prices of biopharmaceuticals in the United States exceed the prices of the same drugs negotiated by foreign governments, which, in turn, exceed their marginal costs of production. We present a model that accounts for these stylized facts and use it to predict the consequences of three policies proposed to reduce domestic drug prices: (1) facilitating drug imports from Canada and Western Europe; (2) requiring that Medicare pay the same prices for drugs as foreign governments; and (3) reducing the profit of downstream channel players (wholesalers, insurance companies, pharmacy benefit managers, and pharmacies) by promoting competition downstream. If not offset, all but the last of these price-reducing policies would eventually depress drug innovation. We conclude by discussing the least expensive way of restoring innovation while maintaining lower domestic prices. Although the model described here is conceptual and its results qualitative, it is the centerpiece of a calibrated simulation model (code to be made available on request) that can be used to quantify the effects of the alternative policies. We will describe the simulation model in a companion paper.

*We are indebted to Yuan Chen, Yichuan Wang, Haozhu Wang, and Mingyuan Zhang for their invaluable research assistance and to Anna Lauren Schmidt for her comments on an earlier draft. We gratefully acknowledge financial support of the Michigan Institute of Teaching and Research in Economics (MITRE).

1 Introduction

How to lower biopharmaceutical prices in the United States without deterring innovation of new drugs (CEA 2018) constitutes a major policy dilemma. Before sensible policy can be devised to resolve this dilemma, it is necessary to understand the process by which promising molecules are discovered, developed into marketable drugs, and sold in the United States and abroad. We have developed a theoretically based MATLAB model to describe this process in some detail. It can be used to *quantify* the effects on drug innovation of interventions in the product market. In this preliminary paper, we confine attention to the core of that model: our theoretical characterization of drug pricing in the United States and abroad. For concreteness, we focus on drugs to treat the hepatitis C virus (HCV).

The market in pharmaceuticals in the United States and abroad is undeniably complicated. Foreign governments bargain with each drug manufacturer, and the bargaining concerns not only the price of a new drug but also the date when it will be made available to that particular buyer. Capturing these interactions would require a dynamic game with a complex state space.[1] Although such a dynamic game might be realistic, it would be completely intractable.

A more tractable conceptual model is needed if we are to anticipate how alternative government policies will affect domestic drug prices and innovation. Berndt (2007) has suggested the use of the standard model of third-degree price discrimination. But that model assumes the foreign price is set by the manufacturer without any pressure from foreign governments. Egan and Philipson (2013) propose the use of a model of public goods. But in their model the formation of price in the United States and abroad is treated symmetrically. Both models achieve their tractability by abstracting from the lags in the introduction of new drugs (Danzon and Furukawa 2008) and by focusing instead only on prices. But because these traditional models were developed decades before the current debate about drug pricing, they omit institutional features of the drug market too important to ignore.

Our approach has been to build a new model that takes account of the asymmetry between the way prices are determined in the United States and abroad but is nonetheless tractable. Our model allows for the possibility that drug prices in Canada and Europe may be unconnected to prices in the United States but also for the very real possibility that they are connected through the threat of arbitrage. As internet shopping expands, this threat that cheaper medicines will be purchased from abroad can only grow in importance. Not only does our model illuminate the effects of policies currently under discussion, but it is also so tractable that the analysis is conducted graphically.

Any model of the international pharmaceutical market must explain two puzzling stylized facts: (1) Americans pay much more than Canadians and Europeans for the same drugs, and (2) even the lower European price vastly exceeds the marginal cost of production. To be concrete, Americans pay at least \$65,000 for the same HCV cure that Europeans buy for \$40,000 even though the marginal cost of producing this cure is estimated to be less than \$140.[2] The standard answer is that in Europe and Canada, governments use their considerable bargaining power to get the lowest price from manufacturers, whereas no comparable bargaining occurs in the United States

According to the Council of Economic Advisers, the price in Western Europe is currently bargained down to the marginal cost of production: "Most OECD nations employ price controls in an attempt to constrain the cost of novel biopharmaceutical products, e.g. through cost-effectiveness or reference pricing policies. In essence, in price negotiations with manufacturers, foreign governments with centralized pricing exploit the fact that once a drug is already produced, the firm is always better off selling at a price above the marginal cost of production and making a profit, regardless of how small, than not selling at all. *Thus, the foreign government can insist on a price that covers the marginal production cost—but not the far greater sunk costs from years of research and development—and firms will continue to sell to that country.*" (CEA 2018, 15; emphasis added).

[1] Since the agreement between one bargaining pair depends in part on the previous agreements by other bargaining pairs, the state variable would have to keep track of the identities of the bargaining pairs that have reached agreement and the agreed-upon price and date of introduction of the particular drug.

[2] We accept here as a stylized fact the widespread view that the prices of drugs in the United States exceed the prices of those same drugs abroad. However, it must be recognized that the *magnitude* of this excess is impossible to quantify, since the rebates and discounts manufacturers routinely offer their customers in the United States and abroad are shrouded in secrecy. If American buyers receive large rebates, net prices received by manufacturers might not be so much higher in the United States than they are abroad.

The academic literature (Grossman and Lai 2008; p. 386 and Figure 1) also predicts that when re-imports are illegal, governments imposing price controls will bargain down to the marginal cost of production under the plausible assumption that these countries are not too sizable compared with the region that innovates.

And yet, contrary to these predictions, the prices negotiated by Canada and the governments in Western Europe are sometimes many hundreds of times larger than the marginal costs of production. For example, no price in Western Europe for a 12-week course of the HCV drug Sovaldi is below $40,000. And yet "a recent study estimated the cost of production of sofosbuvir [Sovaldi] to be US $68-$136 for a 12-week course of treatment based on the same manufacturing methods used in the large-scale generic production of HIV/AIDS medicines (Hill et al., 2014), and its findings have not been challenged" (Iyengar et al. 2016.) Nor is Sovaldi unique in this regard. "Predicted manufacturing costs (US dollars) for 12-week courses of HCV DAAs [direct-acting antivirals] were $21-$63 for riba-virin, $10-$30 for daclatasvir, $68-$136 for sofosbuvir, $100-$210 for faldaprevir, and $130-$270 for simeprevir" (Hill et al. 2014).

The real question is not why prices in Western Europe are so low but why they are so high. They are perhaps low relative to US prices (Sovaldi lists for $84,000 in the United States), but they are high relative to their marginal costs of production.

The literature's conclusion that foreign governments will bargain prices down to the marginal cost of production is very plausible if negotiated prices are "unconnected" to prices in the United States[3] That prices in Canada and Western Europe greatly exceed the marginal cost of production, however, convinces us that the markets *are* connected and that the standard literature mischaracterizes the situation.

This is no mere academic quibble. Current policy proposals rely on this mischaracterization. If the United States and foreign markets are unconnected, then any plan to force foreign governments somehow to pay more for their drugs not only will raise drug company profits and hence increase innovation but will do so without raising US prices at the same time. But if, as the evidence seems to suggest, the two markets are connected, then such a policy will *raise* US prices when the goal of US policy is to lower them.

Logically, negotiated prices exceed marginal costs for one of two reasons: either (1) manufacturers would reject demands for prices closer to marginal cost or (2) manufacturers would accept such demands but negotiators have no desire to bargain so aggressively. Egan and Philipson (2013) make the latter argument. They contend that governments refrain from bargaining for even lower prices out of fear of depressing future innovation (innovation costs for current drugs being sunk). Given that the discovery of promising molecules and their development into drugs takes more than a decade and is fraught with uncertainty, we are skeptical that foreign governments desire no lower prices on this account.

We think a more plausible explanation for why Canadians and Europeans do not demand prices closer to marginal cost is that they anticipate that drug manufacturers would reject such demands out of fear that arbitrageurs would resell drugs in the United States if manufacturers lowered prices any further. For example, while current price differentials are insufficient to induce massive imports from Canada, the more desperate among us have long had their prescriptions filled in Canada. In response, drug company representatives have given some Canadian drugstores an ultimatum: pay a price closer to the US price or receive no new drugs. According to the New York Times (Simon 2003), "One drug industry executive in the United States said that the gap in American and Canadian medicine prices might discourage manufacturers from releasing some new drugs in Canada. 'From now on, if the Canadians don't give us a price close to our United States price, I'm not selling it there,' he said. 'I would rather not have people in Congress see us launch a new product in the United States with a price a lot higher than our Canadian price.' "

In a valuable article on parallel trade, Grossman and Lai (2008) consider two cases. In the first, arbitrage between markets is costless and the markets are connected. In the second, it is illegal to resell drugs sold at a low price in the higher-price market, and the markets are unconnected. In our view, however, this neglects an intermediate case of importance where reselling drugs is illegal but nonetheless the markets are connected. Banning pharmaceutical imports does not eliminate importation; it merely makes engaging in it more costly, since one is penalized if caught. Massive arbitrage would still occur if the capital gain from buying low and selling high exceeded the expected penalty the arbitrageurs anticipate.[4]

[3]Of course, like any other theoretical prediction, this one rests on assumptions. In particular, foreign governments are assumed to propose prices on a take-it-or-leave-it basis, and information is assumed to be complete.

[4]In reality, limited arbitrage occurs, since some inframarginal importers expect lower penalties.

Foreign governments stop just short of triggering massive arbitrage when bargaining prices down, recognizing that manufacturers would resist any further concessions. In this situation, the *threat* of arbitrage still connects the low- and high-price markets.

The evidence that manufacturers recognize that massive arbitrage would endanger their profits is the huge sums they spend to prevent it. In the United States, where importing drugs is illegal, manufacturers and the nonprofit "pro-consumer" organizations that developers have been caught funding surreptitiously (Kopp and Bluth 2017) lobby Congress to preserve the import ban using the pretext that imports from Canada or Western Europe are "unsafe." However, a private firm, PharmacyChecker.com, has developed extensive methods to determine which online foreign pharmacies are safe (Honest Apothecary 2013). Sampling from pharmacies certified safe by PharmacyChecker.com has demonstrated convincingly (Bate et al. 2013) that drugs purchased from these certified online foreign pharmacies are as safe as drugs purchased in domestic, brick-and-mortar pharmacies.[5] Since parallel trade within the European Union is legal, these same companies, at considerable cost, have had to devise other strategies to limit the damage massive parallel trade would do to their profits.[6]

In our model, we recognize that importing drugs sold initially in Canada or Western Europe is illegal but treat the expected penalty (Δ) from the activity as exogenous. If the expected penalty is sufficiently high ($\Delta \geq \Delta^*$), the markets are unconnected and the negotiated foreign price equals the marginal cost of production while the expected price in the United States is the oligopoly price. At the other extreme of a zero expected penalty ($\Delta = 0$), behavior is similar to the case Grossman and Lai (2008) considered where arbitrage is legal. Our formulation, however, permits consideration of the *intermediate* case ($0 < \Delta < \Delta^*$) where the expected penalty is small enough that the threat of arbitrage induces manufacturers to reject prices in Canada and Western Europe any closer to the marginal cost of production. In such a case, any plan to force foreign governments to raise their prices would result in higher US prices as well.

Having considered why a negotiated price of \$40,000 exceeds a much smaller marginal cost of production, we now ask why a price in the United States might be at least \$25,000 higher than the negotiated price in Europe. It is conceivable that the mass of potential arbitrageurs expect a penalty of \$25,000 per cure if they buy low and sell high. But that seems to us implausible. So we consider a complementary explanation: that the threat of arbitrage governs the price manufacturers charge wholesalers but that these wholesalers and the other channel players downstream then mark up the price a *second* time in generating the retail price in the US market. In a careful study, Sood et al. (2017) show that when wholesalers pay manufacturers \$54 for a drug to cover production costs plus the manufacturer markup, the price consumers pay is not \$54 but \$100! The \$46 added to the wholesale price covers the costs and markup of the four channel players: (1) wholesalers, (2) pharmacies, (3) pharmaceutical benefit managers (PBMs), and (4) insurers.[7] Although the medications go in a straightforward fashion to the wholesalers and then to the pharmacies, financial flows also involve the insurers and the PBMs. If we were to ignore this \$46 payment markup that goes to the four types of channel players, our predictions about the price consumers pay would be much too low. On the other hand, a paper like ours that takes an international perspective cannot get bogged down in modeling the complex financial flows among the domestic channel players. Our approach is to consolidate the four types of channel players into a *four-function* player: the wholesaler-pharmacy-insurer-PBM. We posit that there are M (exogenous) four-function channel players and assume for simplicity that they have zero costs. In effect, we treat the \$46 of Sood and colleagues as a second

[5]Firm profit, not consumer safety, motivates these lobbying expenditures. As Kesselheim and Choudhry (2008) emphasize, "Concerns about the integrity of imported brand-name and generic drugs from these markets [Canada and Europe] are often exaggerated, and US regulators should be able to readily ensure the safety of imported products." According to Outterson (2005), "The most thorough recent analysis . . . concludes that Canadian drug supply is actually safer on balance than that of the United States. . . . The EU has many years of experience with parallel trade in pharmaceuticals, without significant safety issues." Outterson (2005) points out that the behavior of manufacturers itself reflects a disregard for consumer safety."By cutting off direct supplies to exporting pharmacies, the pharmaceutical companies force additional intermediaries into the supply chain, which increases safety and handling problems, increases inefficiencies and increases the opportunity for spoilage and introduction of counterfeits. If the concern is truly patient safety, supply restrictions are a crude and counterproductive tool."

[6]Pharmaceutical companies have employed a variety of strategies to prevent international arbitrage inside the European Union. These include strategic use of marketing authorizations, patents, trademarks, vertical restraints, launch timing, and refusals to supply.

[7]It is entirely possible that Sood et al. (2017) have overstated the markup of the manufacturer and understated the markup of the channel players. This would occur if they failed to identify some of the hidden rebates that PBMs secure from manufacturers by threatening to remove their products from the formulary list.

markup secured by these players. The size of that markup depends on the number of multifunction channel players (M) and on domestic demand at high prices.

The goal is to understand the long-run effects on innovation of proposed policy interventions in the product market.[8] We consider various interventions to lower the prices that US consumers pay: (1) reducing the expected penalty of importing non-counterfeit drugs from Europe and Canada; (2) allowing Medicare either to pay the price negotiated by foreign governments or to negotiate drug prices itself; and (3) reducing the second markup by increasing competition in the channel.[9]

Typically, a policy anticipated to lower prices in the US market will depress innovation and the expected number of drugs produced.[10] Most biopharmaceutical research is conducted in universities and independent laboratories rather than inside big pharmaceutical companies. According to Shepherd (forthcoming), "Approximately three-fourths of new drugs are externally sourced. Internal R&D is no longer the primary source, or even an important source, of drug innovation in large pharmaceutical companies." The role of the large pharmaceutical companies is to acquire promising molecules from the academics, surmount the remaining FDA hurdles, and bring the drugs to market. Manufacturers anticipating lower profits because of government intervention would pay academic researchers less for the promising molecules they discover and, expecting lower reward for their discoveries, those researchers with the lowest probabilities of finding a promising molecule would cease to search for one.[11] As a result, there would be less innovation.

To offset or "sterilize" these depressing effects on innovation, a second policy instrument is required. The government has a choice: it can replace the money the drug companies cease paying academics who succeed in finding promising molecules, so that the academic who was just indifferent between searching for a molecule and abandoning the search continues to be indifferent. Or the government can pay *everyone* who commits to search for a molecule *prior* to the outcome of their research gambles just enough that the marginal academic remains indifferent. Both strategies would restore innovation to its previous level, but one always turns out to be less expensive for the government. It is always cheaper for the government to pay everyone *before* discoveries are made, even though many of the people compensated ultimately discover nothing useful. We refer to this as the "paradox of sterilization."

We proceed as follows. In Section 2, we describe the product-market game. In Section 3, we characterize its equilibrium. In Section 4, we show how various policies affect prices consumers pay in the United States and abroad and the profits of drug manufacturers. In Section 5, we show how to adapt the equilibrium analysis and its comparative statics when the intervention of the channel players results in double marginalization. In Section 6, we explain the paradox of sterilization and provide intuition. Section 7 concludes the paper.

2 Description of the Game

We envision the following game. A single negotiator specifies a discounted price p^N per cure at which to purchase medication for each of the (exogenous) Q^N HCV sufferers he represents. The negotiator proposes this price sequentially to each of the n drug manufacturers (hereafter referred to as "developers" in recognition of the many activities these firms do to develop a promising molecule into a marketed

[8] The immediate effect of the proposed policy interventions benefits domestic consumers by reducing drug prices. But two of the three policies we consider would depress innovation and therefore would harm future generations of consumers. A model developed by researchers at RAND (Lakdawalla et al. 2009) focuses on this intergenerational trade-off. The RAND team's approach relies on hazard functions and exploits historical data. Our model abstracts from these transitory, intergenerational effects and focuses on the long-run, steady-state consequences that price-reducing policies would have on innovation in the absence of sterilization policies. Moreover, we assume that the government sterilizes the price-reducing policies so that they do not depress innovation. The RAND model is heavily empirical and focuses on the transition to the long-run, steady-state equilibrium if price-reducing policies are allowed to depress subsequent innovation. Our model abstracts from transitory effects and shows how the government can ensure at least cost that future innovation does not fall when price-reducing policies are imposed and maintained. Hence, in our view, the two approaches nicely complement each other.

[9] Another useful reform would be for each developer to compensate directly the wholesalers and pharmacies involved in distributing its product if and only if that player imposes no second markup over the manufactuers price.

[10] Important exceptions are policies that reduce the second markup. Such policies can reduce US prices and simultaneously increase manufacturer profits and hence innovation.

[11] It is important to note, however, that those least likely to succeed are the ones who abandon the search. The lower their success probabilities relative to the academics who continue to search, the less their departure will depress innovation.

drug). Each developer either accepts or rejects his proposal and announces its decision publicly. Those rejecting the proposal then produce and sell only in the unnegotiated (US) market. Those accepting it not only sell in the domestic market but also share equally the Q^N additional sales at price p^N per cure in foreign markets. The same equilibrium arises if the n developers accept or reject the negotiator's proposal simultaneously instead of sequentially.[12]

Given the extremely low marginal costs of production for HCV cures (Hill et al. 2014) reported in Section 1, we assume that producing additional units is costless. We assume that, for every dollar paid for the medication, developers receive the share χ and the multifunction channel players receive the fraction $1-\chi$. We alter this assumption in Section 5, where we take account of double marginalization.

Developers benefit in two ways if they accept the negotiator's proposal. First, each developer sells more of a drug that is costless to produce. Second, revenue from these foreign sales does not have to be shared with domestic channel players.

But there is also a cost. With Q^N more drugs in circulation, there is also a threat that supplies sold to Canada and Western Europe would flow into the United States if the price differential between the two regions exceeds an exogenous expected penalty for smuggling of Δ dollars per cure. The threat of arbitrage ensures that the price in the US market will not exceed the price in the rest of the world by more than Δ. The consequences of any developer accepting the negotiator's proposal is thus a price for every cure sold in the US market of at most Δ more than the price abroad.

We assume that the drugs in this therapeutic class are perfect substitutes and therefore sell at the same price. In fact, the new cures for HCV do appear to be very close substitutes.[13] We assume that consumers are covered by insurance, so their demand ($D(p)$) is less elastic than if they paid everything out of pocket.[14] Throughout, we make assumptions on the domestic demand function ($D(p)$) sufficient for (1) the total revenue function ($pD(p)$) to be concave, (2) the expected penalty from smuggling to be smaller than the revenue-maximizing monopoly price ($\Delta < argmax_{p\geq 0} pD(p)$), and (3) a unique Cournot equilibrium to exist.

3 Equilibrium and Determination of the First Markup

The negotiator approaches each developer in sequence and proposes to pay p^N per cure for $\frac{Q^N}{l}$ cures, where $l = 1, \ldots, n$ is the number of developers that accept. Each developer accepts or rejects the proposal, and the negotiator moves on to the next developer. To determine the subgame-perfect equilibrium, we first determine the payoffs in the various subgames that can arise.

If no developer accepts the negotiator's proposal, then each of the n developers simultaneously decides how much to produce and sell in the US market. In the equilibrium of this subgame, every developer acts like a symmetric Cournot oligopolist selling a perfect substitute. Developers receive an exogenous fraction χ of the Cournot profits generated.

If instead one or more developers *accept* the negotiator's proposal but it is so high that $p^N + \Delta \geq p^{Cournot}$, then the price in the US market remains $p^{Cournot}$. In these subgames, developers rejecting the proposal would earn Cournot profits while the l firms accepting it would each earn an additional $p^N Q^N / l$.

If, however, one or more developers accept the negotiator's proposal and $p^N + \Delta < p^{Cournot}$, then every developer would realize that smuggled drugs would flood the US market if the domestic price strictly exceeded $p^N + \Delta$. In these subgames, limit-pricing occurs. Each developer sells more than its Cournot output ($D(p^N + \Delta)/n > D(p^{Cournot})/n$.) No developer would unilaterally sell less than this since doing so would lower its sales without raising the price per cure ($p^N + \Delta$). Nor would any developer unilaterally sell more since, with every firm producing an output exceeding the Cournot level, the (right) marginal revenue is strictly negative. Hence, in the equilibrium of subgames that follow acceptance of any proposed $p^N < p^{Cournot} - \Delta$, the price in the US market would be $p^N + \Delta$, but no smuggling would occur.

[12] However, in the simultaneous-move version, there is also a spurious equilibrium where every firm accepts the negotiator's proposed price even if he offers only a penny.

[13] According to Newsweek (Wapner 2017), "A curative drug [for hepatitis C] was approved a few years ago but was incredibly expensive. When a second curative treatment [for hepatitis C] emerged, Express Scripts told the first manufacturer that it would not put its drug on Express Scripts formulary unless the company lowered the price to that of the second drug."

[14] For a qualitative discussion of such demand curves, see Kina and Wosinka (2009, 495); for a derivation of the demand for drugs by consumers with constant coinsurance rates, see Berndt and Newhouse (2010, 34-35).

We now consider how each developer in the sequence would respond to any proposed p^N if the sales behavior described above was anticipated. Each developer in the sequence will find itself in one of three situations: (1) some previous developer has accepted the negotiator's proposal; (2) no previous developer has accepted the proposal, but it is nonetheless still in the interest of the last developer to accept the negotiator's proposal; or (3) no previous developer has accepted the proposal, and it is also in the interest of the last developer to *reject* the proposal.

In situations (1) and (2), the developer anticipates that no matter what it does, the price in the unnegotiated market will be $p^N + \Delta$. Since in either situation accepting the proposal results in additional sales at no cost, the developer will always accept the proposal.

In situation (3), the developer always rejects the negotiator's proposal. For, the developer anticipates that if it is rational for the last developer to reject the proposal, then it must also be rational for every prior developer to reject that proposal since, unlike the last developer, prior developers would have to divide up the Q^N additional sales among themselves and hence the negotiator's offer is less valuable to them. Thus, in situation (3), each developer is pivotal: if the developer accepts the proposal, this induces everyone subsequent to this developer to accept it also.

Note that each developer in the sequence is a copycat: it makes the same decision as the one it anticipates the last firm will make. Anticipating this response the negotiator will choose the lowest price that the final developer in the sequence will accept. Denote this price as $\hat{p}^N$. This price is defined as the smallest solution to the following equation, assuming it is nonnegative.[15]

$$p^N Q^N + \frac{\chi(p^N + \Delta)D(p^N + \Delta)}{n} = \frac{\chi\pi^{Cournot}(n)}{n}. \tag{3.1}$$

Since at this price every developer will accept the proposal, each firm will receive $1/n^{th}$ of the additional Q^N sales. The n developers produce in aggregate $Q^N + D(\hat{p}^N + \Delta)$. They sell Q^N units in the negotiated market and $D(\hat{p}^N + \Delta)$ in the unnegotiated market. Denote the revenue received by each of the n developers as $R(n)$. Each developer earns revenue

$$\begin{aligned} R(n) &= \frac{\hat{p}^N Q^N}{n} + \frac{\chi(\hat{p}^N + \Delta)D(\hat{p}^N + \Delta)}{n}. & (3.2) \\ &= \frac{\chi\pi^{Cournot}(n) - (n-1)\hat{p}^N Q^N}{n}, & (3.3) \end{aligned}$$

where the last line is obtained by substituting into (3.2) the solution to (3.1).

It is helpful to rearrange equation (3.1) as follows:

$$(p^N + \Delta)D(p^N + \Delta) = \pi^{Cournot}(n) - \frac{p^N n Q^N}{\chi}. \tag{3.4}$$

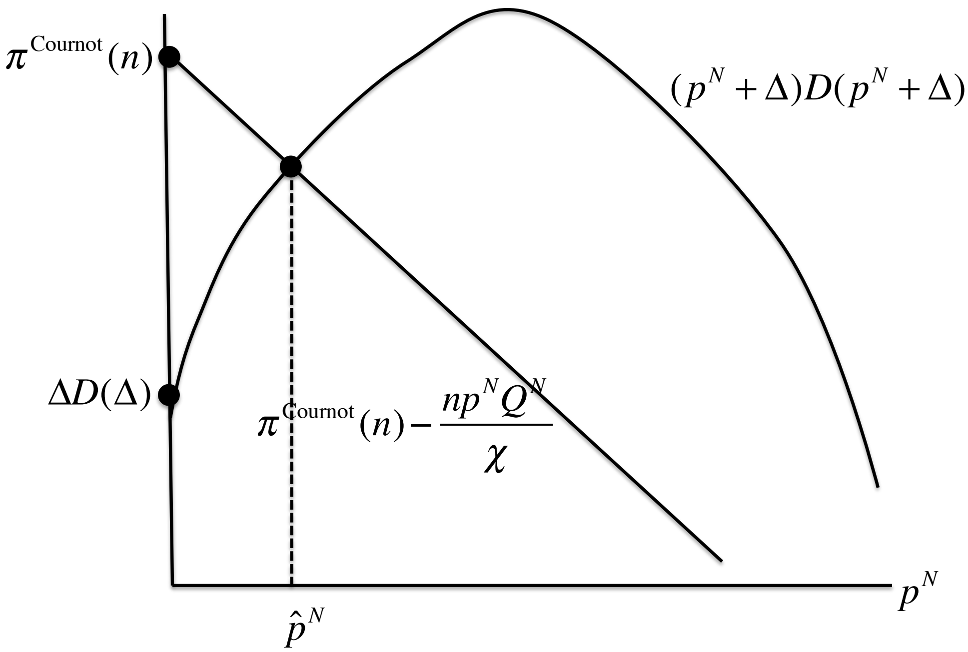
The right-hand side is a decreasing linear function of p^N with vertical intercept $\pi^{\text{Cournot}}(n)$ and slope $-nQ^N/\chi < 0$. The left-hand side is a strictly concave function with vertical intercept $\Delta D(\Delta) \geq 0$. Given our assumption that $0 \leq \Delta < argmax_{p \geq 0} pD(p)$, the single-peaked function $\big(p^N + \Delta\big)\big(D(p^N + \Delta)\big) > 0$ is strictly increasing at its vertical intercept.

Since Cournot profit is strictly smaller than monopoly profit (for $n = 2, \ldots$), the vertical intercept of the line is strictly smaller than the peak of the concave profit function. There are two possible cases. To distinguish them, it is helpful to denote the expected penalty equal to the Cournot price as $\Delta^* (= p^{Cournot})$. In the first case ($\Delta < \Delta^*$), the domestic and foreign markets are connected; in the second case ($\Delta \geq \Delta^*$), the two markets are unconnected. The first case (respectively, the second case) arises if the vertical intercept of the single-peaked function lies below (resp. above) the vertical intercept of the downward-sloping line.

In the connected case, the horizontal component of the point of intersection is the negotiated price ($\hat{p}^N$), and the vertical component is the total revenue in the domestic market, which gets divided between the developers and the channel players. In the unconnected case, the negotiated price equals the marginal production cost (assumed, for simplicity, to be zero), and the price in the US market is the Cournot price. We depict the solution in the connected case in Figure (3.1):

[15] The smallest solution will be negative if the expected penalty is larger than the Cournot price: $\Delta > p^{Cournot}$, which implies

Figure 3.1: Determination of the Negotiated Price



4 Comparative Statics

In this section, we consider four policies that would reduce the domestic price of prescription drugs: (1) ceasing to discourage imports from online pharmacies certified safe by PharmacyChecker.com ($\Delta \downarrow$); (2) allowing Medicare to pay the price negotiated by foreign governments ($Q^N \uparrow, D(p) \downarrow$); (3) increasing competition among developers ($n \uparrow$); and (4) increasing competition among downstream channel players ($\chi \uparrow$). We also show how these policies would affect the price negotiated by foreign governments and the profit of each developer. Under three of the four policies, developer profit falls when the domestic price falls. As a result, innovation would fall unless another policy instrument is used to offset the effect. Increasing competition among downstream channel players is unique in reducing the domestic price while *at the same time* stimulating innovation.

In analyzing the effects of a change in each exogenous parameter, we first consider the case where the two markets are connected and then the case where they are unconnected. The former case can easily be deduced from Figure 3.1. Results for both cases are summarized in Table 1.

4.1 Reducing the Expected Penalty for Re-importing Prescription Drugs

An exogenous reduction in Δ will raise the foreign negotiated price. For if the negotiated price did not change, developers would earn strictly more by selling exclusively in the domestic market and would reject the negotiator's proposed price (see equation (3.1)). To acquire any drugs, therefore, the negotiator

$\Delta D(\Delta) < \pi^{Cournot}$. In this "corner" case, $\hat{p}^N = 0$, and the US price will be $p^{Cournot}$. Since the price differential between the two regions is strictly smaller than Δ, no one will be tempted to smuggle. Each developer in this case earns $R(n)/n = \chi\pi^{Cournot}(n)/n$.

would have to propose a higher price. The domestic price, however, must strictly fall. Otherwise, the left-hand side of (3.1) would strictly exceed the right-hand side.

In terms of Figure (3.1), an exogenous decrease in Δ will shift the single-peaked function down in the neighborhood of the equilibrium. To see this, note that at a fixed Δ, $(p^N + \Delta)D(p^N + \Delta)$ is strictly increasing in p^N where it intersects the downward-sloping line, and hence at a fixed p^N, this function must be strictly increasing in Δ. But if Δ decreases, it will shift the curve downward to the left of its peak (and upward to the right of its peak), and consequently the intersection with the unchanged downward-sloping line will occur at a higher $\hat{p}^N$. Equation (3.3) implies that a reduction in Δ will cause R to fall. Each developer loses more from the price decrease in the US market than it gains from the increase in the negotiated price.

If Δ is so large that the two markets are unconnected ($\Delta \geq \Delta^*$), then a reduction in Δ within this region will affect neither the two prices nor any developer's profit.

4.2 Switching HCV Sufferers from the Market-Determined Price to the Negotiated Price

If Medicare is allowed to pay the price that foreign governments negotiate with developers (or to negotiate the way they do), some HCV sufferers in the US market would switch to paying the negotiated price. We assume that at any price, $D(p)$ shifts leftward, reflecting the loss of the demands of these domestic HCV sufferers on Medicare. Since the total number of infecteds globally (W) is unchanged, $Q^N + D(0) = W$. As a result, when Q^N increases, $D(p)$ shifts leftward by an equal amount.

If the markets are connected, the negotiated and market-determined prices either both rise or both fall, as does the profit of each developer. It turns out that both prices must fall provided relatively weak assumptions on demand are satisfied. If the two markets are unconnected,[16] an exogenous increase in the number paying the negotiated price and a simultaneous decrease in the number paying the market-determined price (which equals the Cournot price) will leave the negotiated price unchanged and will depress the market-determined price. The revenue of each developer will decline. These claims are established in the Appendix.

When campaigning, candidate Donald Trump advocated that Medicare Part D sufferers cease to pay the market-determined price and begin to pay a negotiated price. As president, he has proposed that Medicare pay developers the price negotiated by foreign governments. While this recent proposal was limited to drugs administered in doctors' offices or hospitals, we ask what would happen if the policy applied to all HCV drugs (or whatever biopharmaceutical is under consideration). Since Q^N would then increase, this policy would lower the US price. But it would also lower developer profit, depressing innovation.

4.3 Increasing the Number of Developers

An exogenous increase in the number (n) of developers will cause the negotiated price to fall. For if the negotiated price did not change, the developers would strictly prefer to sell in both markets rather than to sell exclusively in the US market (see equation (3.1)), and the negotiator would seize the opportunity to propose a lower price. For the price differential to remain unchanged, the domestic price must fall by the same amount. Each developer's profit would also fall, since total revenue in each market falls and must be divided among a larger number of developers. Graphically, an increase in n does not affect the domestic industry revenue curve in Figure (3.1) but shifts the intercept of the line down since, in a symmetric Cournot equilibrium, industry profits decline as the number of competitors increases. The increase in the number of developers also causes the line to steepen. As a result, the intersection point has a smaller horizontal component (the negotiated price) and a smaller vertical component (domestic industry profit).

If Δ is so large that the two markets are unconnected ($\Delta \geq \Delta^*$), an exogenous increase in the number of developers will leave the negotiated price at the marginal cost of production and will reduce the domestic price because there would be more Cournot competitors. In the foreign market, profits would continue to

[16] The markets are connected if $Q^N \in [0, W - m(n+1)\Delta)$ and unconnected if $Q^N \in [W - m(n+1)\Delta, W - 2m\Delta)$.

Table 1: Comparative Statics

	p^N	p^U	R
Q^N	$-$ (NC)	$-$ $(-)$	$-$ $(-)$
n	$-$ (NC)	$-$ $(-)$	$-$ $(-)$
Δ	$-$ (NC)	$+$ (NC)	$+$ (NC)
χ	$+$ (NC)	$+$ (NC)	$+$ $(+)$

Results for the case of unconnected markets are in parentheses. NC denotes "No Change"

be zero while in the domestic market, the reduced industry revenue divided among a larger number of developers would result in lower profits per developer.

We summarize these comparative-statics results in Table 1.

4.4 Reducing the Cut Taken by Channel Players

An exogenous reduction in the share of domestic profits siphoned off by the channel players that inhabit the domestic market (an increase in χ), raises $\hat{p}^N$. For if the negotiated price did not change, developers would earn strictly more by selling exclusively in the domestic market and would reject the proposed price of the negotiator (see equation (3.1)). To acquire any drugs, therefore, the negotiator would have to propose a higher price. Developers earn higher revenue from sales in each market. In the foreign market, they sell the same volume at a higher price; in the domestic market, they receive a larger share of an increased total revenue. Graphically, an increase in χ makes the slope of the downward-sloping line in Figure 3.1 flatter without changing its intercept or the single-peaked, domestic total revenue function. Consequently, both the horizontal component (the negotiated price) and the vertical component (the total revenue in the domestic market) rise.

If Δ is so large that the two markets are unconnected ($\Delta \geq \Delta^*$), then a reduction in the channel players' share of domestic profits (an increase in χ) will no longer have any effect on the price in either market. The domestic price remains at the Cournot level ($p^U = p^{Cournot}$), and the foreign price remains equal to marginal production cost ($p^N = 0$). But since each developer will get a larger share of the unchanged Cournot industry profits, each developer earns larger profits.

4.5 Functional Forms

We assume the aggregate demand by insured HCV sufferers in the US market is $D = a - mp$, where $a > 0$ and $m > 0$ are exogenous parameters. Our assumption that Δ is smaller than the monopoly price reduces to $a - 2m\Delta > 0$ and is sufficient to ensure that the product $(p^N + \Delta)(a - m(p^N + \Delta)$ as a function of p^N has a strictly positive value and a strictly positive slope at its vertical intercept ($p^N = 0$).

Let Q^N denote the number of HCV sufferers paying the negotiated price. If the n developers choose to sell only in the US market (and nothing abroad), a Cournot equilibrium results. As is well known, in this case the price, aggregate sales, and aggregate profits are, respectively,

$$p^{Cournot} = \frac{a}{m(n+1)} \tag{4.1}$$

$$Q^{Cournot} = \frac{an}{n+1} \tag{4.2}$$

$$\pi^{Cournot} = \frac{na^2}{m(n+1)^2}. \tag{4.3}$$

In the connected case, the negotiator will bargain the foreign price down to the point where the final developer in the sequence is indifferent between accepting and rejecting his proposal:

$$p^N Q^N (\frac{1}{\chi}) + \frac{(p^N + \Delta)(a - m[p^N + \Delta])}{n} = \frac{na^2}{nm(n+1)^2}. \tag{4.4}$$

This is a quadratic in p^N:

$$m(p^N)^2 + (2m\Delta - \frac{nQ^N}{\chi} - a)p^N + (\pi^{Cournot}(n) + m\Delta^2 - a\Delta) = 0, \tag{4.5}$$

Two values will solve it. Assuming it is nonnegative, the smaller root is the relevant one; otherwise, $\hat{p}^N = 0$. Therefore, in the linear demand case,

$$\hat{p}^N = \max\left(0, \frac{-(2m\Delta - [nQ^N/\chi] - a) - \sqrt{(2m\Delta - [nQ^N/\chi] - a)^2 - 4m(\pi^{Cournot}(n) + m\Delta^2 - a\Delta)}}{2m}\right). \tag{4.6}$$

Since Figure 3.1 determining the negotiated price ($\hat{p}^N$) and Table 1 describing the comparative statics hold for general demand curves, they hold for linear ones. As we will see, this figure and table also prove useful when discussing the second markup. For future use, we summarize the function defined in equation (4.6) as $\hat{p}^N(\chi, \Delta, n, Q^N)$.

5 The Second Markup

5.1 No Threat of Arbitrage

While developers add to the marginal cost of production (assumed negligible) a markup of $p^N + \Delta$, consumers do not pay this price, according to Sood et al. (2017). Four types of channel players (wholesalers, pharmacies, PBMs, and insurers) are responsible for a second markup. To simplify, we assume that there are M channel players. Each player is assumed to perform all four functions and is designated a multifunction channel player. We assume that each multifunction channel player exercises market power in the product market.[17] This seems to us appropriate since, according to Sood et al. (2017), the top three PBMs have a market share of more than two-thirds, the top three wholesalers have a market share of more than four-fifths, and the top three pharmacies have a market share of roughly one-half. Moreover, CVS and Express Scripts are not only the top two pharmacies but are also the top two PBMs.

To describe the strategic interaction between the upstream drug developers and the downstream multifunction channel players, we adapt the model of Greenhut and Ohta (1979). The M multifunction channel players, in their capacity as wholesalers, buy cures from the n independent developers and then, in their capacity as pharmacies, sell them to customers at a markup over the wholesale price.

Assume consumer willingnesses to pay when insured generate the given linear inverse demand curve:

$$p^R(Q) = (a - Q)/m, \tag{5.1}$$

where p^R denotes the retail price.

[17] PBMs often claim that they exercise oligopsony power. In the absence of unambiguous evidence to support such claims, we assume that they take the developer's wholesale price as given.

Channel players buy from developers at the wholesale price (p^w) and then sell to consumers at the higher retail price (p^R). This downstream stage is exactly like a traditional Cournot oligopoly model except that the acquisition cost p^w, which channel players take as given, replaces the constant marginal cost of production. Channel player i chooses how much (q_i) to purchase from developers at the given price p^w and resell to consumers to maximize

$$\left(\frac{a - q_i - Q_{-i}}{m}\right) q_i - p^w q_i.$$

In the symmetric Nash equilibrium, each channel player (in its capacity as a pharmacy) sells to consumers the same amount that it has purchased (as a wholesaler) from the developer, an amount that equates perceived marginal revenue to the cost of acquiring additional units:

$$\frac{a - Q}{m} - \frac{Q}{Mm} = p^w. \tag{5.2}$$

In the upstream stage, developers recognize that if they produce a larger quantity, channel players will increase their purchases only if the wholesale price is reduced. That is, they regard (5.2) as an inverse demand curve with slope $\frac{dp^w}{dQ} = -\frac{M+1}{Mm} < 0$.

The n developers decide independently how much to sell to the channel players. In the symmetric Nash equilibrium, each developer manufactures and sells an amount such that its perceived marginal revenue from additional sales equals the marginal cost of production (assumed by us to be zero).[18]

$$\left(\frac{a - Q}{m} - \frac{Q}{Mm}\right) - \frac{Q}{n}\left(\frac{M+1}{Mm}\right) = 0. \tag{5.3}$$

From equation (5.3), we conclude:

$$\hat{Q} = \frac{M}{(M+1)}\frac{an}{(n+1)} = \frac{M}{(M+1)} Q^{Cournot}, \tag{5.4}$$

where the "hat" denotes variables determined in the subgame-perfect equilibrium. Aggregate sales are smaller in this two-stage oligopoly than under one-stage oligopoly because of the second markup. In the case of linear inverse demand, aggregate output is smaller by the factor $\frac{M}{M+1}$. Evaluating $p^w(Q)$ and $p^R(Q)$ at $\hat{Q}$, we conclude:

$$\hat{p}^w = \frac{a}{m(n+1)} = p^{Cournot} \tag{5.5}$$

$$\hat{p}^R = \frac{a(M+n+1)}{m(M+1)(n+1)}. \tag{5.6}$$

Note that, in the absence of a threat of arbitrage, developers would charge channel players a wholesale price $\hat{p}^w$ equal to what they would have charged final consumers if they had dealt with them directly; the channel players then add on their own markup, reducing the quantity sold to $\frac{M}{M+1}$ of its former level. Intuitively, the wholesale price equals the Cournot price because the quantity demanded of the developers is $\frac{M}{M+1}$ as large at every price. Since this is like a change in the units used to measure quantity, it does not affect the equilibrium price per cure. It does, of course, reduce the profit per developer by $\frac{M}{M+1}$:

$$\hat{R} = \left(\frac{M}{M+1}\right)\frac{a^2}{m(n+1)^2} = \left(\frac{M}{(M+1)}\right)\frac{\pi^{Cournot}}{n}. \tag{5.7}$$

[18] Although the cost of an additional course of a drug (a "cure") is negligible, the costs of developing that drug are substantial (DiMasi et al. 2003; DiMasi and Grabowski 2007). These costs do not figure into our calculations, however, because they are sunk before the strategic interactions modeled here begin.

5.2 Threat of Arbitrage

Suppose risk-neutral, price-taking arbitrageurs can purchase cures at price p^N and can attempt to resell them to channel players, in their capacity as wholesalers, at a profit.[19]. Since such arbitrage is illegal, we assume arbitrageurs expect to be fined Δ per unit resold if they are caught. Arbitrage would be deterred if the channel players could purchase all the cures they wanted from the developers at $\hat{p}^w \leq p^N + \Delta$.

But suppose $\hat{p}^w > p^N + \Delta$. Then the channel players, as wholesalers, can buy the same drugs from arbitrageurs more cheaply than from the developers. If developers continued to restrict supply to $\hat{Q}$, they would no longer be able to drive the wholesale price above $p^N + \Delta$. Wholesalers would pay no more than that. They would simply buy the additional products they would demand at that reduced price from the arbitrageurs. This would not be optimal for the developers, however. Since their marginal production cost is negligible, they would expand their sales at the wholesale price of $p^N + \Delta$ to fill the entire aggregate demand of the multifunction channel players at that price. Hence,

$$p^w = p^N + \Delta.$$

Recall that equation (5.2) is the inverse demand curve of the developers. At a wholesale price of $p^N + \Delta$, they would purchase $\tilde{Q}$ units, solving

$$p^N + \Delta = \frac{a - \tilde{Q}}{m} - \frac{\tilde{Q}}{Mm}. \tag{5.8}$$

At that wholesale price, the channel players would therefore demand

$$\tilde{Q} = \left(\frac{M}{M+1}\right)\left(a - m(p^N + \Delta)\right). \tag{5.9}$$

Just as when the developers sold directly to consumers, the *threat* of arbitrage reduces to $p^N + \Delta$ what the developers charge the channel players in their capacity as wholesalers. But with channel players adding a second markup, consumers pay a higher retail price and purchase only $\frac{M}{M+1}$ times as much.

5.3 Determination of p^N When Wholesalers Can Purchase from Arbitrageurs as Well as Developers

As before, the negotiator approaches developers sequentially and offers to purchase a total of Q^N additional cures, equally divided among the developers accepting his offer, at the negotiated price p^N. By accepting the proposal, each developer sells more cures, but as in Section 3, the developer's profits in the domestic market are smaller than they would have been if smugglers had no supplies to resell to the channel players. To show this, we use equation (5.2) to express the aggregate profit of the developers as

$$R^{ag} = Q\left(\frac{a - Q}{m}\right) - \frac{Q^2}{Mm}. \tag{5.10}$$

Differentiating twice, it is straightforward to show: (1) R^{ag} is globally concave ($\frac{d^2R^{ag}}{dQ^2} < 0$) and (2) since $\frac{dR^{ag}}{dQ} = \frac{aM - 2Q(M+1)}{Mm}$, this first derivative is negative when evaluated at $\hat{Q}$. Together, these two observations imply that when aggregate sales increase, the profit of every developer declines.

The arguments in Section 3 for determining the negotiated price still apply. Denote that price $\tilde{p}^N$. The lowest negotiated price that would be acceptable to the n developers satisfies the following equation:

$$\tilde{p}^N Q^N + \left(\frac{M}{M+1}\right)\frac{(a - m[\tilde{p}^N + \Delta])}{n}(\tilde{p}^N + \Delta) = \left(\frac{M}{M+1}\right)\left(\frac{a^2}{m(n+1)^2}\right). \tag{5.11}$$

[19] It has also been suggested that imported products could be acquired by channel players in their capacity as pharmacies (Outterson 2005).

Multiplying every term by $n(\frac{M+1}{M})$, we conclude:

$$\tilde{p}^N n Q^N \left(\frac{M+1}{M}\right) + (a - m[\tilde{p}^N + \Delta])(\tilde{p}^N + \Delta) = \frac{na^2}{m(n+1)^2}. \tag{5.12}$$

Compare this equation with equation (4.4). Note that it is identical except that $\frac{M}{M+1}$ replaces χ. Therefore, $\tilde{p}^N = \hat{p}^N(\frac{M}{M+1}, \Delta, n, Q^N)$, where $\hat{p}^N(\cdot) was defined at the end of Section 4$. Figure 3.1 can be regarded as determining $\tilde{p}^N$ instead of $\hat{p}^N$.

5.4 A Comparison

Suppose that in the model without a second markup we set $\chi < 1$ and the other exogenous variables in any admissible way. Can we *duplicate* the results in the model with the second markup? We claim that we can. Set $M = \frac{\chi}{1-\chi}$, in the model with the second markup. Then $\chi = \frac{M}{M+1}$. First consider the "connected" case.

1. In this case, the foreign price in the model with the second markup will equal the foreign price in the model without the second markup.
2. The wholesale price ($\tilde{p}^w$) the developers charge the channel players equals what they would have charged consumers (p^U) in the model without a second markup.
3. The revenue each developer earns from sales to the channel players in their capacity as wholesalers in the model with a second markup is $R = \frac{\hat{p}^N Q^N}{n} + \frac{M}{M+1}\frac{(a-m\tilde{p}^w)\tilde{p}^w}{n}$. The factor $\frac{M}{M+1}$ enters because the second markup reduces demand in the second market. But since $\chi = \frac{M}{M+1}$ and $p^U = \tilde{p}^w$, this formula reduces to $R = \frac{\hat{p}^N Q^N}{n} + \chi\frac{(a-mp^U)p^U}{n}$, the revenue developers would earn in the model without a second markup. Without the second markup, the quantity sold in the domestic market is larger, but since developers receive the same price per unit and are taxed at rate $\chi = \frac{M}{M+1}$, their profit is the same as when there is a second markup.

The models give the same results when each is in the connected case.

If each is in the unconnected case, then the foreign price in the model with a second markup is zero (the negligible marginal production cost), and that is also the foreign price in the model without a second markup.

Developers charge channel players $\tilde{p}^w = p^{Cournot}$, exactly what they would charge consumers in the model without a second markup.

Since the foreign price is bargained down to the negligble marginal cost of production, the revenue of each developer in the model with a second markup comes entirely from the domestic market: $\frac{M}{M+1}\frac{\pi^{Cournot}}{n}$. Here again, the second markup causes the quantity sold to fall by the factor $\frac{M}{M+1}$. In the model with no second markup, developers sell more at the same price of $p^{Cournot}$ per unit. But since $\chi = \frac{M}{M+1}$, this too is exactly what each developer earns in the model without a second markup. The revenue in this case is diminished by χ, which is like a profits tax.

It remains to prove that the markets are connected in the model without a second markup if and only if the markets are connected in the model with a second markup. In both models, the markets are connected if $\Delta < p^{Cournot}$ and unconnected if $\Delta > p^{Cournot}$.

How then do the models differ? The model with a second markup differs from the model without a second markup in two ways: (1) with a second markup, developers sell less (and fewer consumers are served); (2) moreover, there is an additional equation defining the retail price. We now show that the retail price equals a weighted average of (1) the choke price of the linear inverse demand curve (a/m) and (2) the wholesale price.[20]

[20] With linear inverse demand, the price in Cournot equilibrium is always the weighted average of marginal cost and the choke price. In the model without a second markup, the marginal cost of each of n developers was assumed to be zero. Hence, $p^{Cournot} = \left(\frac{1}{n+1}\right)\frac{a}{m} + \left(\frac{n}{n+1}\right)0 = \frac{a}{m(n+1)}$.

Table 2: Comparative Statics with Second Markup

	p^N	p^w	R	p^R
Q^N	$-$ (NC)	$-$ $(-)$	$-$ $(-)$	$-$ $(-)$
n	$-$ (NC)	$-$ $(-)$	$-$ $(-)$	$-$ $(-)$
Δ	$-$ (NC)	+ (NC)	+ (NC)	+ (NC)
M	+ (NC)	+ (NC)	+ (+)	$-$ $(-)$

$$p^R = \frac{a-\tilde{Q}}{m} \tag{5.13}$$

$$= \frac{a}{m} - \left(\frac{M}{M+1}\right)\frac{(a-m\tilde{p}^w)}{m} \tag{5.14}$$

$$= \left(\frac{1}{M+1}\right)\frac{a}{m} + \left(\frac{M}{M+1}\right)\tilde{p}^w. \tag{5.15}$$

5.5 The Response of the Retail Price to Policy Changes

In this subsection, we deduce the effects of different policies on the retail price in the model with a second markup.

If more people switch from paying the US price to the negotiated price, then the retail price falls. To see this, note that the retail price is the weighted average of the choke price and the wholesale price. An increase in Q^N with an offsetting reduction in a reduces both the choke price and (from Table 1) the wholesale price. Since the weights do not change, the retail price must fall in both the connected and unconnected cases. If the number of developers (n) is larger, the retail price falls. This follows since the choke price is unaffected, as are the weights. But the wholesale price decreases in both the connected and unconnected cases.

If the expected penalty (Δ) is larger, then the retail price rises in the connected case but does not change in the unconnected case. In the connected case, the increase in Δ increases the wholesale price without affecting the choke price or the weights. So the retail price increases. In the unconnected case, the wholesale price remains $p^{Cournot}$ and hence the retail price is unaffected.

If the number of multifunction channel players increases, less weight is put on the choke price and more on the wholesale price, which is smaller. As a result, the retail price falls as M increases. In the unconnected case, the wholesale price is constant and the retail price converges to it monotonically. In the connected case, the wholesale price increases in M but approaches a limit since it is bounded above by $p^{Cournot}$. The retail price declines monotonically. The difference $p^R - p^w$ declines monotonically to zero.

In the unconnected case, a Trumpian price floor binds whenever it exceeds the marginal cost of production. Since we assume this cost to be zero, the floor binds for any positive floor. Since the wholesale price remains $p^{Cournot}$, the retail price is unaffected if the floor is raised. In the connected case, however, a unit increase in the floor increases the wholesale price by one unit and the retail price by $\frac{M}{M+1}$ units.

We summarize the comparative statics results for the model with the second markup in Table 2:

We have simulated the model with the second markup and depict the results in Figures 5.1 to 5.6.

As Figure 5.1 reflects, a reduction in the penalty expected for re-importing drugs will lower the retail price in the importing (domestic) market while raising the price in the exporting (foreign) market if the markets are connected. Developer revenue will fall and, as a result, innovation will be depressed. If

Figure 5.1: Effects of Changing Expected Penalty for Arbitrage

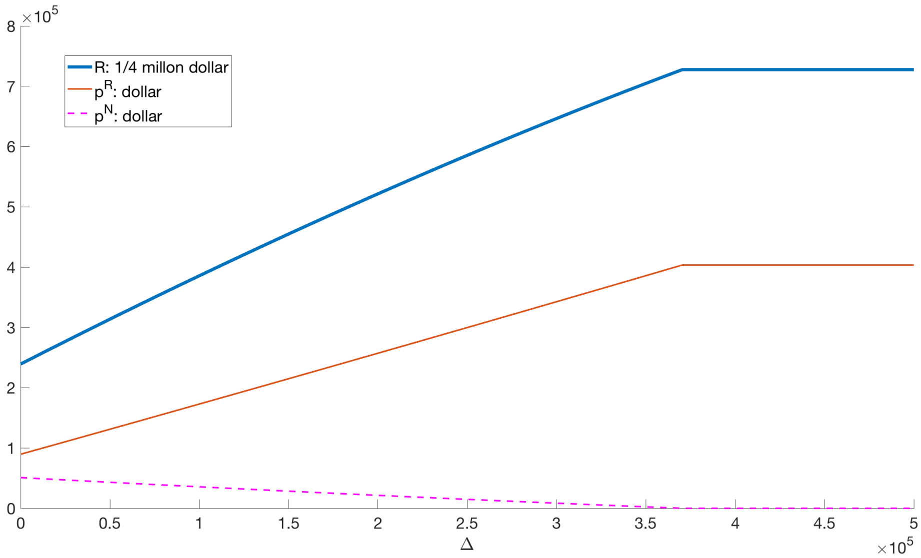


Figure 5.2: Effects of Switching People from the Unnegotiated Market to the Negotiated Market

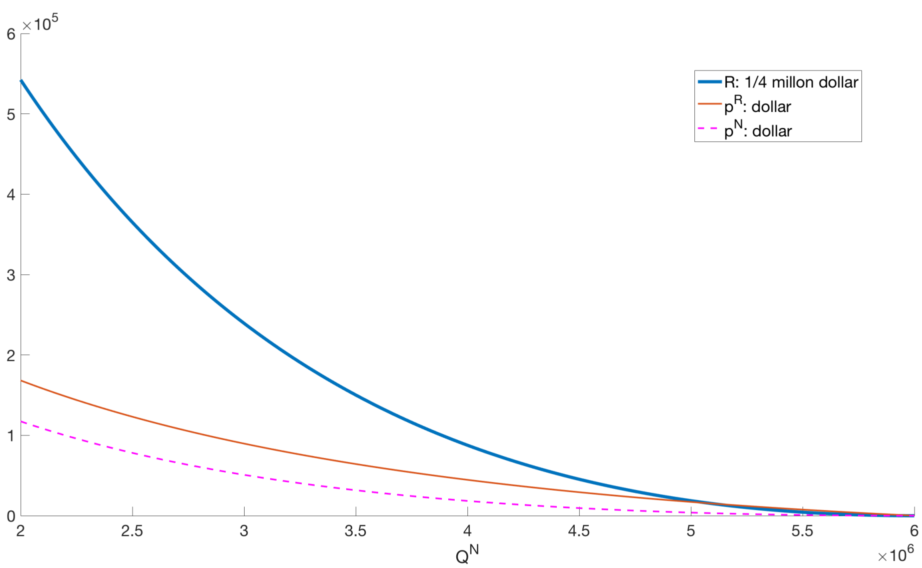


Figure 5.3: Increasing Competition among Channel Players When Markets Are Connected

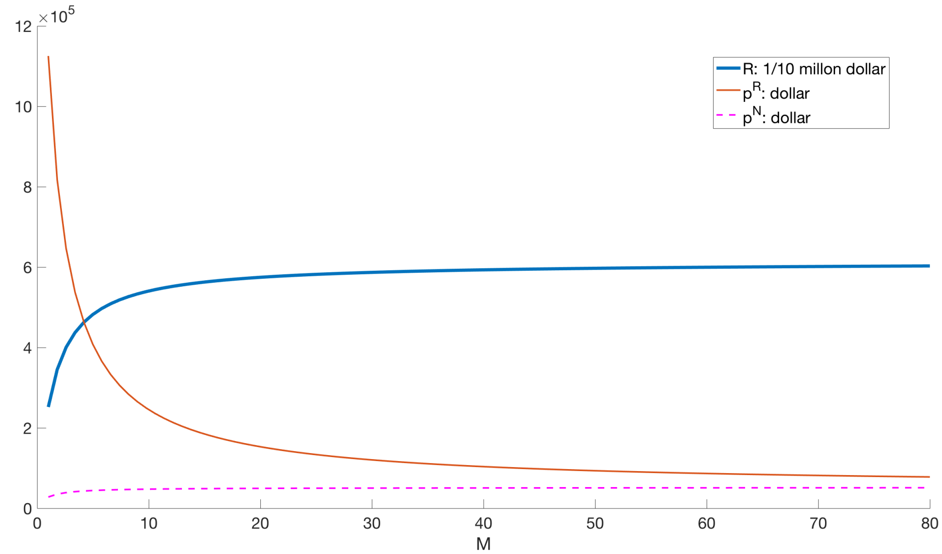


Figure 5.4: Increasing Competition among Channel Players When Markets Are Unconnected

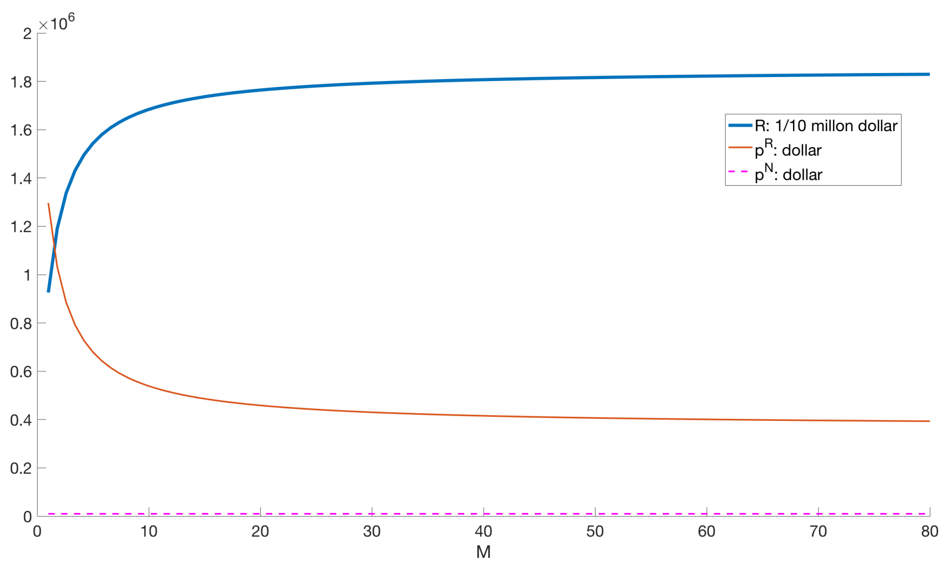


Figure 5.5: Effects of Raising the Floor on the Foreign Price When Markets Are Connected

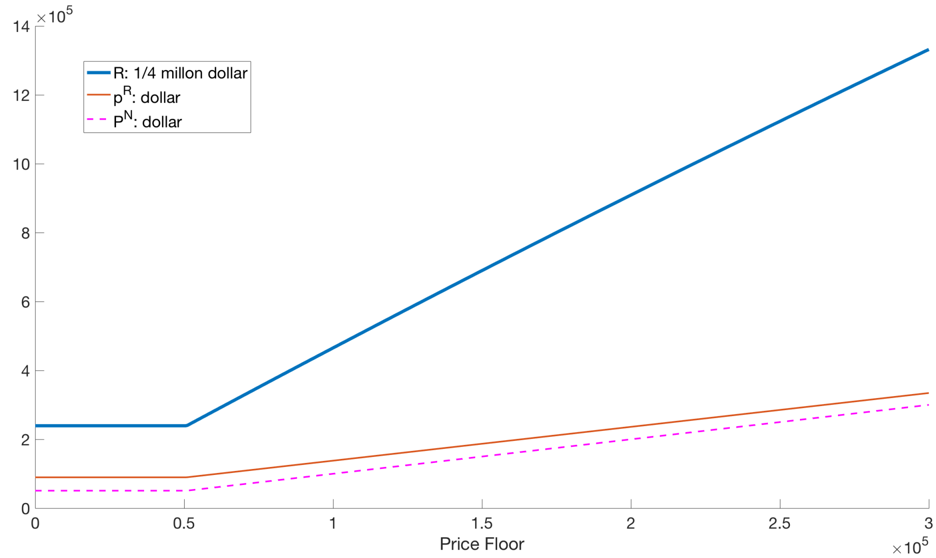
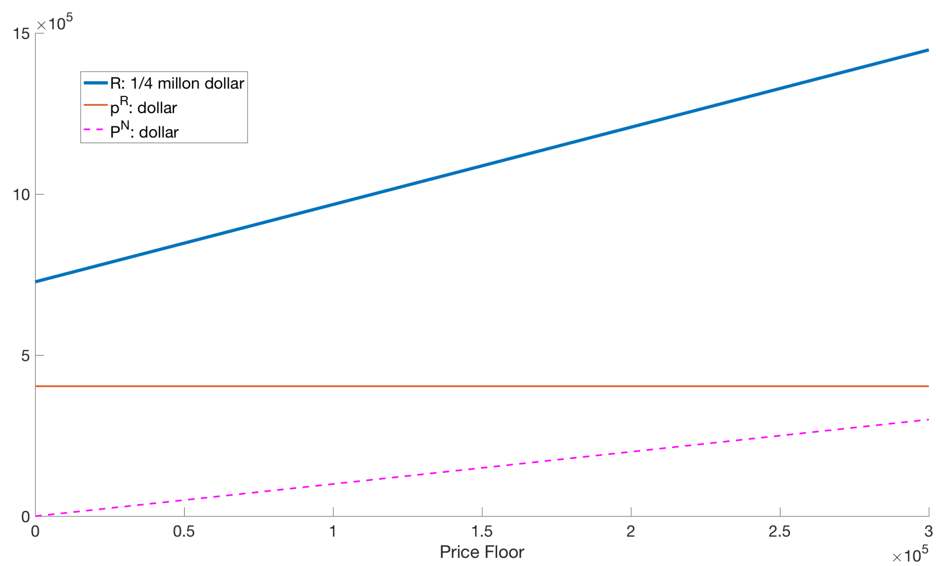


Figure 5.6: Effects of Raising the Floor on the Foreign Price When Markets Are Unconnected



the markets are unconnected, however, reducing the expected penalty will have no effect. The boundary between the two regions occurs when $\Delta = p^{Cournot}$.

As Figure 5.2 reflects, if HCV sufferers on Medicare begin to pay the negotiated price, both the domestic price and the foreign price will fall, as will developer revenue. Thus, the policy will depress innovation.

As Figures 5.3 and 5.4 reflect, increasing competition among channel players will reduce the domestic price (in the limit, to the wholesale price). Unlike the other policies, this policy increases developer revenue and innovation at the same time that it reduces the domestic price. In the connected case, the foreign price will rise, while in the unconnected case, it will remain at marginal cost.

Finally, Figures 5.5 and 5.6 depict the consequences of the Trumpian policy of a price floor on the foreign price. As Figure 5.6 reflects, in the unconnected case, any floor above marginal cost will bind, and it will increase developer revenue and innovation without raising the domestic price. However, as Figure 5.5 reflects, if the markets are connected, a binding price floor will also raise the domestic price as well as developer revenue and innovation.

5.6 Caveats about Linear Demand

To construct Figures 5.1 to 5.6, we set the exogenous parameters (those not varied in a figure) as follows: $\Delta = \$100; W = 6 * 10^6; m = 1.35; Q^N = 3 * 10^6; n = 5; M = 55$. These parameters imply a foreign price of $50.8 thousand and a domestic retail price of $89.7 thousand. The most unrealistic of these numbers is clearly the number of channel players (M). If $M = 4$ (instead of 55), the foreign price is predicted to be $42.8 thousand and the domestic retail price $478.7 thousand (see Figure 5.3). Clearly, no insurance company would reimburse for a 12-week cure costing nearly half a million dollars. This makes clear that the assumption of linear inverse demand at high prices is unrealistic. That is, we have omitted an important check on the market power of the channel players: the refusal of consumers or the companies that furnish their insurance to pay such exorbitant prices. If we had assumed lower demand at high prices than linear demand implies, the retail price would have fallen to something more realistic even with only four channel players.[21]

6 The Adverse Impact on Innovation of Price-Reducing Policies

Most biopharmaceutical innovation is done by academics (Shepherd, forthcoming), some more capable of searching for a promising molecule than others. Assume there are N academics with distinct, strictly positive probabilities of finding a promising molecule if they commit to looking for one. Let $p_i > 0$ denote the success probability of academic i, where $p_1 > p_2 \ldots > p_N$. We assume that these academics do not differ in other respects. In that case, the number of academics willing to search for a molecule is a strictly increasing step function of the payoff they expect to receive if they find a promising molecule. For any given expected payoff (denoted V), academics with sufficiently high success probabilities will gear up to search for a promising molecule, while those with insufficiently high probabilities will pass up the opportunity. We assume each academic takes V as given.[22]

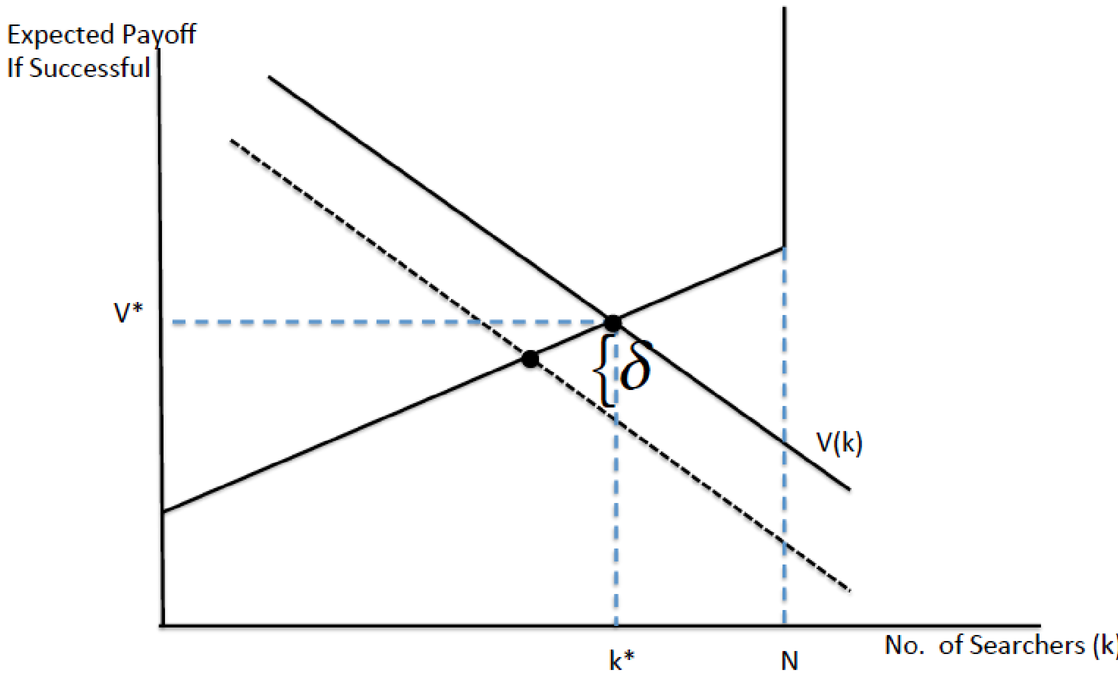
We assume that a developer who expects to receive more revenue from sales of a drug it develops from a promising molecule is willing to pay academics more for their promising molecules. Hence, we assume that V is a strictly increasing function of R.[23] The results in Table 1 and Table 2, therefore, can be reinterpreted as implying that an increase in downstream competition (an increase in M or χ) raises V while promotion of imports from Canada or Western Europe or the requirement that Medicare pay the

[21] To see this, perturb the initial equilibrium by assuming that the inverse demand curve is kinked at the equilibrium value $\tilde{p}^w$ and is then linear with a flatter slope ($\bar{m}$) and a smaller choke price ($\frac{\bar{a}}{\bar{m}}$). Thus, the inverse demand curve is piecewise linear and concave. The retail price will then be a weighted average: $\frac{1}{5}\frac{\bar{a}}{\bar{m}} + \frac{4}{5}\tilde{p}^w$. The flatter the new segment of the inverse demand curve, the closer the new choke price will be to the kink at the height of the old wholesale price ($50.7 thousand). The retail price will be a weighted average of this smaller choke price and the new wholesale price, which must be smaller than the new choke price. As a result, even with four channel players, the retail price could easily be reduced to something more realistic.

[22] How sensitive the supply of searchers is to the payoff they expect if they are successful depends on the vector of success probabilities. For simplicity, we assume that each success probability is distinct. As a result, the supply curve will have N steps and will steepen as the difference between the success probabilities grows. If instead we had assumed that all N academics had the same success probability, then all N of them would have the same asking price.

[23] This assumption is not only plausible but consistent with our more complex MATLAB simulation model.

Figure 6.1: Effect of Reduction in Expected Payoff if Successful Reduces the Number of Academics Searching for a Molecule



price that foreign governments negotiate lowers V. That is, $V_M > 0, V_\Delta > 0, V_{Q_N} < 0$. Recall that increasing the number of developers (n) competing in the product market lowers R and hence V. We assume that when more academics search for promising molecules, more are discovered and eventually more firms sell whatever drugs can be developed.[24] Consequently V is a decreasing function of k.

We summarize this discussion in Figure 6.1. On the horizontal axis, we plot the number of academics who search for a promising molecule ($k \leq N$). The vertical axis is in dollars. The upward-sloping line depicts the number of academic searchers (k) as a function of the payoff they expect to receive if they are successful; this is a continuous approximation of the step function. The downward-sloping line is $V(k)$.[25]

Denote the intersection of the two curves as (k^*, V^*). This intersection point corresponds to the unique equilibrium: exactly k^* academics voluntarily search for molecules because they expect to receive V^* dollars if successful, and developers voluntarily pay each successful academic V^* dollars because of the revenue they anticipate receiving in the product market when k^* academics search for molecules.

$V(k)$ will shift if the government intervenes in the product market in order to lower the domestic price of pharmaceuticals. Increasing competition downstream would lower the domestic price and would increase the profits of each developer selling a drug. Hence, $V(k)$ would shift up and, in the long run, innovation would increase. Since there is no adverse effect on innovation, there is no "policy tradeoff"; hence, there is nothing further to discuss.

On the other hand, reducing the domestic price by promoting the importation of biopharmaceuticals from Canada or Western Europe or by requiring Medicare to pay the same price that foreign governments have negotiated shifts down $V(k)$. In the long run, innovation falls. Note that the academics who cease to search for a promising molecule will be those with the *lowest* probabilities of finding one.

[24]This assumption is not only plausible but consistent with our more complex MATLAB simulation model.

[25]Our results would not change if $V(k)$ were horizontal or even upward-sloping, provided it crosses the supply curve from above.

6.1 When to Sterilize to Offset the Adverse Impacts

Suppose that in response to reduced revenue in the drug market, developers reduce payments for promising molecules. Innovation falls in the long run unless a second policy instrument is used. In particular, suppose that if k^* academics continued to search, they would receive $\delta > 0$ less than before the government's product-market intervention—only $V(k^*) - \delta$. The government can restore innovation by paying each academic who has discovered such a molecule δ to replace what the developers cease to pay. Alternatively, the government can restore innovation by paying each academic *before* he knows whether he has succeeded or failed an amount s just sufficient to restore innovation to its previous level. Which strategy is less expensive for the government?

If each of the k^* academics committed to searching is paid $s = p_{k^*}\delta$, the marginal researcher will continue to search and so will the $k^* - 1$ other researchers since they have an even better chance of finding a promising molecule. Multiplying by k^*, we conclude:

$$k^* s = k^* p_{k^*} \delta < (p_1 + p_2 + \ldots + p_{k^*})\delta. \tag{6.1}$$

The inequality follows because there are k^* terms in the parentheses on the right-hand side, strictly larger than p_{k^*}.[26] The left-hand side is the cost to the government of paying each of the k^* researchers s *before* they learn whether their research has succeeded or failed. The right-hand side, which is strictly larger, is the expected cost to the government of paying δ to each of the k^* researchers lucky enough to find a promising molecule.

The marginal researcher is indifferent whether the government pays him s before the outcome of his research gamble is known or δ if he is successful. But *every* other researcher strictly prefers to receive δ if successful. In fact, since $s - p_i\delta < 0$ for $i = 1, \ldots k^* - 1$, the higher the success probability of the inframarginal researcher, the more he loses if the government sterilizes before, rather than after, the outcomes of the research gambles are known. Sterilizing before the research outcomes are known redistributes inframarginal rents from researchers to the government, with those with the highest success probabilities paying the most. Nonetheless, as long as the marginal researcher continues to search for a promising molecule, so will the researchers with higher success probabilities.

7 Conclusion

The purpose of this paper is to explain the basic stylized facts in the international biopharmaceutical industry and, assuming that explanation is correct, to deduce from it the effects on prices and future innovation of imposing product-market policies that have been frequently discussed: (1) promoting foreign imports from online pharmacies in Canada and Western Europe certified safe by PharmacyChecker.com; (2) requiring that Medicare pay no more than the prices that foreign governments have negotiated; and (3) increasing competition among players downstream in the distribution chain (wholesalers, insurers, pharmacy benefit managers, and pharmacies). While all three of these policies will reduce domestic prices, only the last at the same time stimulates future innovation by academic researchers. To prevent the other two policies from depressing future innovation, a second instrument is required. One possibility is for the government to reward research success by exactly as much as developers reduce their rewards when the product market becomes less profitable. We show, however, that a cheaper way to restore innovation is to reward each researcher looking for a molecule *before* the outcome of his molecule search is known. The latter policy is cheaper because the percentage reduction in the subsidy paid is always larger than the percentage increase in the number of people who receive the subsidy. Subsidizing ex ante has notable redistributive effects. The marginal researcher is indifferent whether he is subsidized the smaller amount before the outcome of his research is known or the larger amount if and only if he is successful. But the higher the success probability of a researcher, the more he loses if the government subsidizes *ex ante*.

[26] The inequality would also hold if at least one of the $k^* - 1$ probabilities strictly exceeds p_{k^*}.

Appendix: Switching HCV Sufferers from the Market-Determined Price to the Negotiated Price

If some sufferers cease to pay the US price and instead pay the negotiated price, both prices must fall provided relatively weak assumptions on demand are satisfied. Let $D(p;a)$ denote domestic demand as a function of the endogenous domestic price p and the exogenous shift parameter a. Increases in a raise demand ($D_2(p;a) > 0$) and reduce the number of infecteds whom the negotiator represents ($Q'^N(a) < 0$). Assume that the Cournot price, $p^{Cournot}$, is differentiable and weakly increasing in a. In addition, we assume that $D_2(\check{p}, a) \geq D_2(p, a)$ for any $\check{p} \geq p > 0$. It is easy to verify that the linear demand function $D(p;a) = a - mp$ satisfies these restrictions, which are sufficient but not necessary for both the domestic and foreign prices to fall when HCV sufferers in the United States switch to the negotiated price.

Rewrite (3.4) as the following function of endogenous variable p^N and the exogenous variable a.

$$p^{Cournot}(a)D\left(p^{Cournot}(a);a\right) - (p^N + \Delta)D(p^N + \Delta;a) - \frac{np^N Q^N(a)}{\chi} = 0.$$

The left-hand side is the difference between the downward-sloping linear function of p^N and the upward-sloping segment of the single-peaked function of p^N. Hence, it is a strictly decreasing function of p^N in the neighborhood of the solution. We now show that this function shifts up if a increases exogenously. Hence, $\frac{dp^N}{da} > 0$ or, equivalently, $\frac{dp^N}{dQ^N} = \frac{dp^N}{da}\frac{da}{dQ^N} < 0$.

Holding p^N at its equilibrium value, we partially differentiate the left-hand side with respect to a and simplify this partial derivative to obtain:

$$\begin{aligned} &p'^{Cournot}(a)\left\{D\left(p^{Cournot}(a);a\right) + p^{Cournot}(a)D_1\left(p^{Cournot}(a);a\right)\right\} \quad + \\ &\qquad D_2(p^N+\Delta;a)\left\{p^{Cournot}(a)\frac{D_2(p^{Cournot}(a);a)}{D_2(p^N+\Delta;a)} - (p^N+\Delta)\right\} \quad + \quad \frac{-Q'^N(a)np^N}{\chi}. \end{aligned}$$

The partial derivative with respect to a is, therefore, the sum of three terms. The first term is the product of two positive factors; the second factor is positive since the Cournot price is smaller than the monopoly price and the revenue function is increasing below the monopoly price. The second term is positive because it is a product of two factors each of which is positive; the second of these factors is positive because the domestic price is smaller than the Cournot price and the Cournot price is multiplied by a fraction weakly exceeding one. The third term is positive since $Q'^N(a) < 0$. An increase in a (or equivalently, a decrease in Q^N) increases p^N, and since the difference in the two prices must remain unchanged (Δ), the domestic price also strictly increases. An increase in Q^N would drive both prices down. Since the negotiated price falls although more sales occur at that price, the revenue generated from such sales ($p^N Q^N$) may increase or decrease. We consider each case in turn. If $p^N Q^N$ decreases, then each developer earns less revenue in each market, and hence developer profits fall. If, on the other hand, the revenue from sales at the negotiated price rises, this increase turns out to be insufficient to outweigh the revenue decrease in the domestic market. This follows from (3.3), which takes account not merely of the definition of $R(n)$ in (3.2) but also of the size of p^N in equilibrium. The numerator in the right-hand side of (3.3) consists of the difference of two terms. Since the first term falls and the second term rises when $p^N Q^N$ rises, the numerator falls. Therefore, when HCV sufferers formerly paying the market-determined price begin to pay the negotiated price, both prices fall and so does the profit of each developer.

If the two markets are unconnected,[27] an exogenous increase in the number paying the negotiated price and a simultaneous decrease in the number paying the market-determined price (which equals the Cournot price) will leave the negotiated price unchanged and will depress the market-determined price. The revenue of each developer will decline. To verify this, we show that developer revenue is strictly increasing in a. Differentiate the revenue function $p^{Cournot}(a)D\left(p^{Cournot}(a);a\right)$ to obtain:

$$p'^{Cournot}(a)\left\{D\left(p^{Cournot}(a);a\right) + p^{Cournot}(a)D_1\left(p^{Cournot}(a);a\right)\right\} + p^{Cournot}(a)D_2\left(p^{Cournot}(a);a\right).$$

[27] The markets are connected if $Q^N \in [0, W - m(n+1)\Delta)$ and unconnected if $Q^N \in [W - m(n+1)\Delta, W - 2m\Delta)$ for $n = 2, 3, \ldots$

The derivative is the sum of two terms. Each term is the product of two positive factors. The second factor of the first term is strictly positive because the Cournot price is strictly below the monopoly price.

8 References

Bate, R., G.Z. Jin, and A. Mathur. 2013. "In Whom We Trust: The Role of Certification Agencies in Online Drug Markets." NBER Working Paper No. 17955.

Berndt, E. 2007. "A Primer on the Economics of the Re-importation of Prescription Drugs." Managerial and Decision Economics 28: 415-435.

Berndt, E., and J. Newhouse. 2010. "Pricing and Reimbursement in U.S. Pharmaceutical Markets." NBER Working Paper 16297, National Bureau of Economic Research.

CEA (Council of Economic Advisers). 2018. "Reforming Biopharmaceutical Pricing at Home and Abroad." Washington, DC: Government Printing Office.

Danzon, P., and M. Furukawa. 2008. "International Prices and Availability of Pharmaceuticals in 2005."Health Affairs 27 (1): 221-233.

DiMasi, J.A., and H.G. Grabowski. 2007. "The Cost of Biopharmaceutical R&D: Is Biotech Different?" Managerial and Decision Economics 28 (4-5): 469-479.

DiMasi, J.A., R.W. Hansen, and H.G. Grabowski. 2003. "The Price of Innovation: New Estimates of Drug Development Costs." Journal of Health Economics 22 (2): 151-185.

Egan, M., and T. Philipson. 2013. "International Health Economics." NBER Working Paper 19280, National Bureau of Economic Research.

Greenhut, M.L., and H. Ohta. 1979."Vertical Integration of Successive Oligopolists. " American Economic Review 69 (1): 137-114

Grossman, Gene M., and Edwin L.-C. Lai. 2008. "Parallel Imports and Price Controls." RAND Journal of Economics 39 (2): 378-402.

Hill, A., S. Khoo, J. Fortunak, B. Simmons, and N. Ford. 2014. "Minimum Costs for Producing Hepatitis C Direct-Acting Antivirals for Use in Large-Scale Treatment Access Programs in Developing Countries." Clinical Infectious Diseases 58 (7): 928-936.

Honest Apothecary. 2013. "Buying Prescriptions Online & Pharmacy Checker.com." June 13. http://www.thehonestapothecary.com/2013/06/13/buying-prescriptions-online-pharmacychecker-com/.

Iyengar, S., K. Tay-Teo, S. Vogler, P. Beyer, S. Wiktor, K. de Joncheere, and S. Hill. 2016. "Prices, Costs, and Affordability of New Medicines for Hepatitis C in 30 Countries: An Economic Analysis." PLOS Medicine 13 (5): 1-22.

Kesselheim, A., and N. Choudhry. 2008. "The International Pharmaceutical Market as a Source of Low-Cost Prescription Drugs for U.S. Patients. " Annals of Internal Medicine 148 (8): 614-619.

Kina, S., and M. Wosinka. 2009. "Pharmaceutical Pricing. " In Handbook of Pricing Research in Marketing, edited by V.R. Rao, 488-511. Cheltenham, UK: Edward Elgar.

Kopp, E., and R. Bluth. 2017. "Nonprofit Working to Block Drug Imports Has Ties to Pharma Lobby " NPR Newscast, April 18.

Lakdawalla, D., D. Goldman, P-C Michaud, N. Sood, R. Lempert, Z. Cong, H. de Vries, and I. Gutierrez. 2009. "U.S. Pharmaceutical Policy in a Global Marketplace. " Health Affairs 28 (1): w138-w150.

Outterson, K. 2005. "Pharmaceutical Arbitrage: Balancing Access and Innovation in International Prescription Drug Markets." Yale Journal of Health Policy, Law, and Ethics 5 (1): article 4.

Shepherd, J. Forthcoming. "Consolidation and Innovation in the Pharmaceutical Industry: The Role of Mergers and Acquisitions in the Current Innovation Ecosystem." Journal of Health Care Law Policy.

Simon, Bernard. 2003. "Curtailing Medicines from Canada." New York Times, Late Edition (East Coast), Nov. 11: W1.

Sood, N., T. Shih, K. Van Nuys, and D. Goldman. 2017. "The Flow of Money through the Pharmaceutical Distribution System." Leonard Schaeffer Center for Health Policy and Economics, University of Southern California.

Wapner, Jessica. 2017. "Understanding the Hidden Villain of Big Pharma: Pharmacy Benefit Managers." Newsweek, March 17.