THE ORGANIZATION OF INNOVATION:

THEORY AND EVIDENCE FROM THE PHARMACEUTICAL INDUSTRY

By

Thomas Jungbauer
Johnson Graduate School of Management
Cornell University
401A Sage Hall
Ithaca, NY 14853
(607) 255-2501, jungbauer@cornell.edu

Sean Nicholson
Department of Policy Analysis and Management
Cornell University
2307 Martha Van Rensselaer Hall
Ithaca, NY 14853
(607) 254-6498, sn243@cornell.edu

June Pan
VISA
Washington, DC

and

Michael Waldman
Johnson Graduate School of Management
Cornell University
323 Sage Hall
Ithaca, NY 14853
(607) 255-8631, mw46@cornell.edu

February 2020

* We thank Catherine Barrera, Christopher Cairns, and George Jakubson for helpful comments.

ABSTRACT

A firm that develops a new product potentially cannibalizes sales of existing products in the firm's product portfolio, where such cannibalization is more costly the more profitable sales of the cannibalized products are. Thus, if the firm is currently producing a product for which its market power is substantial, it will want to control the research and development process in order to limit cannibalization. In this paper, we explore how this basic logic affects the organization of investments in research and development. We first build and analyze a theoretical model of the research and development process in which conducting R&D in-house provides the firm more control over the new product's location in product space. We then explore the model's testable predictions using data from the pharmaceutical industry concerning patents, patent expiration, and decisions concerning whether various stages of the research and development process are conducted in-house or outsourced. Our empirical findings support the model's testable predictions.

Keywords: research and development, outsourcing, vertical integration

I. INTRODUCTION

Research and development intensive firms face the question of whether to conduct R&D in-house or to outsource. Despite the popularity of outsourcing, partnering with outside firms is not without costs. For example, Peter Chambre, past chief executive of Cambridge Antibody Technology, once remarked that "he would prefer to delay licensing products to drug companies until they are at a later stage to generate more value from the company's technology" (Mills, 2004). More generally, various articles discuss pros and cons of outsourcing, where a common cited disadvantage is a loss of control by the party purchasing the service (Patel 2017; Raineri 2019). In this paper, we theoretically and empirically investigate a new idea concerning loss of control and outsourcing using a variant of the well known property rights approach to the theory of the firm put forth initially in Grossman and Hart (1986) and Hart and Moore (1990).

Consider a firm that is developing a new product, and chooses whether to conduct the R&D associated with the development of that new product in-house or by outsourcing. The new product will potentially cannibalize sales of existing products in the firm's product portfolio. For example, a drug company that develops a new product for the treatment of depression will hurt the sales of its existing products that can also be used to treat depression. Such cannibalization is more costly the more profitable are sales of the cannibalized products in the absence of the new product introduction. Thus, if the firm developing the new product currently produces a product for which its market power is substantial, it will want to control the research and development process in order to limit the degree of cannibalization. Our main argument is that, because of incomplete contracting problems that arise when R&D investments are outsourced, the firm's ability to limit cannibalization is higher when it conducts the R&D in-house. Thus, the firm will choose to conduct the R&D in-house when the potential costs of cannibalization are high.

We start by constructing and analyzing a theoretical model that formalizes this argument. In our model, a firm, call it the originator, develops a new product and chooses whether to conduct R&D in-house or to outsource. The firm that conducts the R&D cannot perfectly control the exact location of the new product in product space, but instead chooses a mean location and an investment level that determines the expected distance of the actual location from the mean location. A higher investment level translates into a smaller expected distance from the mean. We capture the cost of cannibalization by assuming that the originator also owns an existing patented product in the same product class as the new one. We consider a T-period model in which we vary the cost of cannibalization by varying the period in which the patent on the existing product expires. For each possible period of patent expiration, the analysis compares the investment in location precision when the R&D investment is conducted in-house with the investment when it is outsourced. We then use this comparison to characterize equilibrium behavior.

Our analysis generates three testable predictions. First, compared to firms that do not own existing patented products, a firm that does is more likely to choose in-house development for the development of a new product in the same product class as one of its existing patented products. Second, the probability of in-house development is positively related to the remaining patent duration of the existing patented product. Third, the probability of in-house development is also positively related to the expected market share of the existing patented product at the date the new product is expected to reach the market. In each case, the basic logic behind the prediction follows the same path. As discussed above, the incentive to choose in-house development is higher the higher is the profitability of the cannibalized product in the absence of the new product introduction.

We test these predictions using data from the pharmaceutical industry, where the patent system is a defining feature of the industry and a large number of firms employ both in-house and outsourced development. Using detailed compound-level data from the Pharmaprojects dataset, we find empirical support for all three testable predictions. First, we find that the probability of in-house development increases when the originating firm has one or more existing patented compounds in the same therapeutic class as the compound under development. Second, given the presence of existing patented compounds in the same therapeutic class, the probability of in-house development increases with the remaining patent lengths of these compounds. Third, by supplementing the main data source with IMS sales data,[1] we are able to show that the originator is more likely to develop the new compound in-house the higher is the market share the firm expects for its same-class patented drugs at the time the new compound is expected to reach the market.

As discussed in more detail in the next section, the main point of this paper is that the vertical integration decision can depend on factors not focused on in the main theories of vertical integration. Specifically, most of the current literature on vertical integration focuses either on the characteristics of the input and/or the characteristics of the product or products the input is used to produce. Our argument, however, is fundamentally different. In our model, whether or not to conduct R&D in-house depends on the characteristics of products in the firm's product portfolio for which the R&D is not an input. The logic, which is discussed above, is that even though the R&D is not an input for these "other" products, the specific nature of the R&D can affect the value of these other products which translates into a benefit of conducting the R&D in-house.[2]

[1] IMS Health is an American company that provides data and related services to the healthcare industry.

[2] One exception is Novak and Stern (2009) that finds empirical evidence in automobile production for complementarity concerning vertical integration decisions involving inputs closely related in the production process.

Another important aspect of our findings concerns how the competitiveness of the market affects the vertical integration decision. As discussed in more detail in the next section, a number of papers have investigated the relationship between product market competition and vertical integration, where some of the studies find a positive effect and some a negative effect. We find that in the case of R&D the relationship is more nuanced. Specifically, we find that more competition lowers the frequency of in-house R&D, but that an important avenue for this relationship is that more competition decreases the positive effect that having a patented product in the same product class has on the likelihood of in-house development.

The outline for the paper is as follows. Section II below reviews the relevant literature. Section III then presents the model and provides a preliminary analysis. Section IV presents a full equilibrium analysis and discusses testable predictions. Section V describes the data used in the empirical testing. Section VI presents the empirical analysis. Section VII discusses the extent to which our findings can be explained by alternative theories concerning the in-house versus outsourcing decision. Finally, Section VIII presents concluding remarks.

II. LITERATURE REVIEW

One body of work this paper is related to is the extensive literature concerning vertical integration. The two main theories in this literature are the transaction cost theory of the firm due initially to Williamson (1975,1979) and Klein, Crawford, and Alchian (1978), and the property rights theory of the firm put forth initially in Grossman and Hart (1986) and Hart and Moore (1990).[3] The property rights theory of the firm, which is of particular relevance to our study,

They provide two possible explanations for the result, neither of which, however, is related to our argument. See also Barrera and Waldman (2019) for a related analysis.

[3] See Gibbons (2005) and Lafontaine and Slade (2007) for surveys that discuss both theories. Williamson (2010) focuses on the transaction cost theory, while Hart (2017) focuses on the property rights approach.

employs a framework in which contracts are incomplete and vertical integration is used to reduce inefficiencies due to incomplete contracting. For example, in the original Grossman and Hart (1986) study, two parties make ex ante investments that are non-contractible and utility is non-transferable ex ante. They show that one party purchases the assets of the other when the former's investment is more important than the latter's.

A related study by Aghion and Tirole (1994) uses the incomplete contracting framework to examine the organization of innovation. In that analysis, the two key choices are research and development efforts and financing, where the organization of innovation determines incentives for R&D efforts and the costs of financing. Their theoretical analysis suggests that when R&D efforts are more important, then R&D is more likely to be conducted by an independent unit, while financing being more important yields the opposite. Lerner and Merges (1998) study the determinants of control rights in biotechnology alliances, and find results mostly consistent with the Aghion and Tirole theory.

Our paper falls into the property rights approach in that we also employ an incomplete contracting assumption and ex ante investments to analyze integration decisions, where our focus is whether R&D is conducted in-house or outsourced. Like in Grossman and Hart (1986), we assume that ex ante development decisions are non-contractible. In our model, the choice to vertically integrate means that the originating firm conducts R&D in-house and maintains control over investment decisions that influence the location of the new product in product space. The choice not to vertically integrate means R&D is outsourced, and the originating firm loses the ability to influence the new product's location in product space. We show that the originator chooses to vertically integrate and retain the ability to influence the new product's location in

product space, when limiting cannibalization of the firm's existing patented products in the same product class is more important than reducing the fixed costs of development.

One difference between our model and previous papers on the property rights theory of the firm concerns the nature of the difference in investment outcomes as a function of which party has control rights. In Grossman and Hart (1986) and Aghion and Tirole (1994), the two firms have different investment technologies and the firm with the superior technology purchases the assets of the other firm. In contrast, in our model the key element is that the originator has higher investment incentives because of the potential cannibalization of other products in the originator's product portfolio. Vertical integration, which is associated with higher investment levels, is thus chosen when the benefit associated with these higher levels exceeds any reduced fixed costs associated with outsourcing.[4]

As mentioned in the Introduction, this paper also contributes to the literature on the relationship between vertical integration and product market competition. Most of that literature is empirical. A number of early studies such as Tucker and Wilder (1977), Levy (1985), and Balakrishnan and Wernerfelt (1986) focus on US manufacturing and find a positive correlation between vertical integration and product market competition, while Aghion et al. (2006) finds a U-shaped relationship using UK manufacturing data. The more recent studies of Galdon-Sanchez et al. (2015) concerning services and Gil and Ruzzier (2018) focused on the Spanish television industry find a negative relationship which is in contrast to most of the earlier studies. We develop a theory that predicts a negative relationship between competition and in-house R&D, where the negative relationship is due to reduced incentives for in-house development when the originator owns an existing patented product in the same product class as the product under development.

[4] An implicit assumption, which is important for this result, is that the originator is not able to sell the related products in its product portfolio to an outsourcing firm. In Section IV, we provide a detailed discussion of this assumption.

We also provide empirical testing using pharmaceutical data that supports the predicted relationship.

This paper is also related to the literature concerning planned obsolescence and new product introductions. Papers in that literature, such as Waldman (1993,1996), Choi (1994), and Nahm (2004), focus on settings in which a durable goods seller with market power introduces a new product that makes used units obsolete. One of the main results in that literature is that, if the firm sells its output, then it faces a time inconsistency problem concerning new product introductions which reduces profitability. By renting, however, the firm avoids the time inconsistency problem because it retains ownership of the used units, and thus internalizes the effect of a new product introduction on the value of used units. As a consequence, a firm in such a setting will want to rent its output, if feasible, because renting avoids time inconsistency which means the firm has efficient (private) incentives for the development of new products.

The main message of our analysis is similar. Like in the literature on planned obsolescence and new product introductions, in our analysis R&D investments are privately efficient when the firm choosing the investments for the development of the new product is also the firm that owns the existing asset – in this case the profit stream associated with sales of the existing patented product – that will lose value when the new product is introduced. So, just like renting being chosen by the durable goods seller in the planned obsolescence literature, the seller of the existing patented product develops the new product in-house as long as any higher costs associated with in-house development are not too large.

This paper also contributes to the literature on cannibalization, and, in particular, what firms can do to limit cannibalization when introducing a new product. For example, Moorthy and Png (1992) show that a monopolist selling a product line can sometimes increase its profitability

by delaying the introduction of a lower quality product which allows the firm to increase the high quality price. This is optimal when cannibalization is an issue and consumers are more impatient than the seller. More recently, Siebert (2015) shows that a firm's optimal strategy in entering a new market in a duopoly setting with vertical differentiation is to introduce a single product. The result arises because avoiding cannibalization is more important in their setting than price discrimination. We contribute to this literature by showing how the desire to limit cannibalization can affect the internal organization of the firm. The logic is that limiting cannibalization depends on product location, where the ability to control product location is improved when the R&D process is in-house. So, if limiting cannibalization is an important consideration, then we should expect a higher frequency of in-house R&D.

Finally, most of the prior literature on the issue of firm boundaries in the pharmaceutical industry has focused either on why firms form alliances, or the outcomes of alliances. For example, Nicholson et al. (2005) shows that biotech companies send positive signals to investors by forming alliances with larger pharmaceutical firms, while Danzon et al. (2005) finds that success rates of complex phase 2 and phase 3 trials are higher for products developed in an alliance. Few papers focus on the characteristics of R&D projects that tend to result in in-house development rather than outsourcing. One such paper is Azoulay (2004) which finds that pharmaceutical firms are more likely to outsource data-intensive clinical trials while knowledge-intensive trials are typically conducted in-house. Our paper is the first to offer a patent protection perspective on the choice of pharmaceutical alliance decisions at the project level.

II. MODEL AND PRELIMINARY ANALYSIS

In this section, we present a T-period model of a firm's decision to conduct R&D either in-house or through outsourcing. We then provide preliminary results concerning R&D decisions made given the firm chooses in-house development, and when it chooses to outsource. In the next section, we provide a full equilibrium analysis, and also derive testable predictions.

A) The Model

In our model, there is a single risk neutral firm that owns an existing patented product, and has decided to develop a new product in the same product class. We call this firm the originator. The originator has a marginal cost c_1 for producing units of the existing patented product and no fixed costs. There are also generic producers that can produce the existing patented product at marginal cost c_1 after patent expiration.

In addition to the originator and the generic producers, there is a pool of identical risk neutral in-licensing firms that we refer to as the licensees. The licensees have a potential cost advantage in developing the new product in comparison to the originator. In particular, the originator has a fixed cost of development F_O which is a random draw from the probability density function f(.) with support (F_{min},∞), while the licensees incur a fixed cost of development F_L, $F_{min} \leq F_L < \infty$. We use Δ to refer to the difference in fixed costs, i.e., $F_O - F_L = \Delta$. We also assume that the marginal cost of production for the new product is lower for the firm that develops the product. Specifically, the developer has a marginal cost of production for the new product equal to c_2, while the marginal cost of production for a firm that did not develop the new product is $c_2^+ > c_2$. In other words, there are economies of scope between developing and producing the new product.

We assume that there are T total periods, T≥4, and no discounting, where the new product is developed in period two and the patent for the new product lasts through period T. The patent on the existing product, on the other hand, expires at the end of period t_E, where t_E can take on any value between one and T. Much of our focus is how behavior changes as a function of t_E.

If the originator chooses in-house development, then the originator develops the new product, chooses the new product price each period, and produces and sells the new product each period.[5] If the originator chooses to outsource development, on the other hand, then there is a contract between the originator and the licensee. The contract specifies for each period who produces the product, who sells the product (the firm that sells the product is the firm that receives payments from the consumers), who chooses the new product price each period, and a payment each period from the originator to the licensee which can depend on that period's new product quantity (the payment can be negative).[6] We also assume the contract to be renegotiation proof, where this assumption is described in more detail below. Further, actions and outcomes associated with the development process itself are assumed to be non-verifiable and thus non-contractible. This means that, if the originator chooses in-house development, then the originator makes the choices associated with the development process. But if outsourcing is chosen, then the licensee makes these choices, and payments cannot be directly contingent on these choices.

Following Salop (1979), the product space is characterized by a unit circle, where the location of the new product on the unit circle depends on the non-contractible development

[5] We do not allow for a contract that would assign production, selling, and pricing decisions to another firm when the originator chooses in-house development. Given the originator is as or more efficient than other firms in these activities when the originator chooses in-house development, allowing for such a contract would not change the equilibrium outcome.

[6] Implicitly, we are assuming that the new product price is non-verifiable and thus not contractible. This assumption is not essential for our main results, but rather serves to simplify the analysis. We also assume that the payment from the originator to the licensee in any period t cannot depend on the new product quantity in a different period. This assumption is also not essential for our main qualitative results, but rather serves to simplify the description of equilibrium behavior.

decisions. That is, the firm developing the new product (either originator or licensee) makes choices that serve to determine the location of the new product relative to the existing patented product. Due to the stochastic nature of the development process, however, the developer does not directly control the location of the new product but instead chooses a means value for the location, l^M, and an investment level, k, that determines the expected absolute distance between the mean location and the realized location.

To be precise, the clockwise distance between the new product and the existing patented product on the unit circle is given by $l=l^M+\varepsilon$, where ε is a random draw from one of the following two uniform distributions: $U[-\alpha,\alpha]$ and $U[-\beta,\beta]$, $\alpha<\beta\leq¼$.[7] The higher the investment level, the more likely is the random draw from the uniform distribution with the smaller range. Let p(k) denote the probability that ε is drawn from $U[-\alpha,\alpha]$ given the investment level equals k. We assume $p(0)=0$, $p'(0)=\infty$, $p'(k)>0$ and $p''(k)<0$ for all $k\geq0$, and $p(\infty)<1$.

On the demand side, there is a continuum of consumers of unit mass uniformly distributed along the circumference of the circle. A consumer can buy any weakly positive number of one of the products, i.e., a consumer can buy units of the originator's existing patented product or units of the new product but we do not allow mixing.[8] To be precise, the valuation a consumer places on consuming unit q of a product is given by $V(q)=V^+-vq$, so the valuation function is characterized by decreasing marginal utility of consumption. A consumer also faces a distance cost for consuming a product not at the consumer's exact location in product space. The distance cost a consumer incurs from consuming a unit located a distance s from the consumer's location in

[7] The assumption $\beta\leq¼$ is imposed for tractability reasons. It ensures that, after the patent on the old product expires, some consumers on both "sides" of the new product continue to purchase the new product.

[8] The assumption that there is no mixing is consistent with typical demand behavior in the pharmaceutical industry. For example, individuals treated for depression seldom take different anti-depressant drugs at the same time due to concerns of possible unwanted drug interactions.

product space equals ds, d>0. We also assume V^+ to be sufficiently large such that the market is always fully covered in equilibrium. For any product price, P, a consumer who chooses to purchase that product purchases the amount that maximizes net utility from consumption, i.e., the consumer chooses the value for Q that maximizes $\int_0^Q (V^+-vq)dq-(P+ds)Q$. In turn, in facing prices for the two products, a consumer chooses to purchase the product that results in the highest net utility for the consumer given the quantity choices that maximize net utility.[9]

The timing of the game is as follows. At the beginning of the first period, the originator chooses a price for the existing patented product for the first period, consumers make purchase decisions, and the value for F_O is realized and publicly revealed.[10] In the first period, the originator also decides whether to develop the new product in-house or outsource the development to a licensee. If the originator chooses to outsource development, then the first period proceeds with a contracting stage. In particular, each firm in the pool of licensees makes a take-it or leave-it offer of a licensing contract to the originator and the originator chooses a licensee.[11]

At the beginning of the second period, the originator chooses a second period price for the existing patented product and consumers make purchase decisions thereafter. If the patent has not expired, then the originator sets the monopoly price, while it if has expired then competition with generic producers means the price equals marginal cost equal to c_1. The developer (either originator or licensee) also chooses a mean value for the new product's location in product space

[9] An alternative theoretical approach is to assume that each consumer purchases zero or one unit of the product that provides the consumer with the highest net utility, but that there are multiple consumers in each location whose valuations are uniformly distributed over $[0,V^+]$. This alternative specification is mathematically equivalent to the specification we analyze.

[10] The assumption that the realization of F_O is publicly revealed is not essential for our results. In particular, results would be the same without this assumption if we focused on Perfect Bayesian equilibrium rather than Subgame Perfect Nash equilibrium which is what we assume below.

[11] We assume that each firm in the pool of licensees either works as the developer of the originator's new product or develops no new product. We thus abstract away from the possibility of cross-subsidization which is the focus of Lerner and Malmendier (2010).

and an investment level in location precision, where these choices are the private information of the developer. After these choices, there is a realization concerning the uniform distribution from which the noise term is drawn from and then a random draw from this distribution. Thus, by the end of the second period the new product's location in product space is determined.[12] This location is publicly observable but not verifiable by the courts.

In the third period, the new product is brought to the market. If the patent on the existing product expires before the third period, then the price for this product is at marginal cost and the firm with control rights for the pricing of the new product takes this price as given in choosing a new-product price. If the patent on the existing patented product has not expired, then the originator chooses prices for both products if it has control rights for the pricing of the new product. If the patent on the existing patented product has not expired and the licensee has control rights concerning the pricing of the new product, then the two prices are determined by Bertrand competition between the two firms. In the following periods, prices are determined using the same rules as in the third period. Also, our focus throughout the paper is Subgame Perfect Nash equilibrium.

As mentioned above, we restrict the analysis to contracts that are renegotiation proof.[13] This means that the contract between the originator and the licensee when the originator outsources must be such that, in every period starting with period two, the parties do not have an incentive to renegotiate. To be specific, at the beginning of each period starting with period two, if the originator (licensee) makes a take-it or leave-it offer of a new contract to the licensee (originator),

[12] In the model, we abstract away from the possibility that the new product will fail in the development process. Adding such a probability would not affect the qualitative nature of the equilibrium or, in particular, the model's testable predictions.

[13] Focusing on renegotiation proof contracts is a standard approach employed in many contracting papers. For early papers that focus on how the possibility of renegotiation affects equilibrium contracting see, for example, Dewatripont (1988), Hart and Moore (1988), and Demougin (1989).

the equilibrium contract is such that the licensee (originator) has at least a weak incentive to turn down any offer that would make the originator (licensee) better off. In other words, the equilibrium contract is such that no pareto-improving renegotiation outcome exists.

B) Preliminary Results

We start with results concerning the nature of the equilibrium contract when the originator chooses to outsource. As captured in Lemma 1, production, sales, and pricing are all assigned to the licensee in every period after patent expiration of the existing product. In contrast, prior to patent expiration of the existing product, production is assigned to the licensee, but sales and control rights for pricing remain in the hands of the originator.

Lemma 1: Consider an equilibrium to the game in which the originator has chosen to outsource development of the new product. Then, the contract between the originator and the licensee is such that i) through iii) hold.

i) In any period t, $2<t\leq t_E$, the contract assigns production to the licensee, but sales and control rights for pricing of the new product remain with the originator. Also, the payment from the originator to the licensee is a fixed amount plus the number of new units sold that period multiplied by c_1.

ii) In any period t, $t>\max\{2,t_E\}$, the contract assigns production, sales, and pricing of the new product to the licensee. Also, the payment from the originator to the licensee is a fixed amount.

iii) The fixed payments from the originator to the licensees sum to the fixed amount that results in zero expected profits for the licensee.

The logic for part i) is as follows. First, the licensee is assigned production of the new product, because we assume that it is less costly for the developer to produce the new product. Second, consider a period prior to patent expiration of the existing patented product, and suppose that the contract does not specify that the originator sells the product and has control rights concerning pricing of the new product. Because the location of the new product is fixed at that point in time, the joint profits of the originator and licensee in that period are maximized by giving sales and the control rights over pricing to the originator, so that it can choose the prices that maximize the joint profits of the two products. Given that we assume the original contract is renegotiation proof, this means that sales and control rights for the pricing decision must be assigned to the originator in the contract. Also, having the payment from the originator to the licensee be a fixed amount plus the number of new units sold multiplied by c_1 means higher joint profits, because in choosing prices the originator will internalize all the returns associated with the pricing decisions.

Now consider a period after patent expiration of the existing product. If the contract assigns production, sales, and pricing to the licensee, the contract will not be renegotiated because the licensee can set the price just as effectively as the originator after patent expiration of the existing patented product. In turn, since assigning sales and the pricing decision to the licensee increases the licensee's investment incentives, this is the equilibrium outcome. Also, the payment from the originator to the licensee is a fixed amount, so that the licensee internalizes all of the effects of its pricing decision. Finally, iii) follows given competition across licensees ensures a zero profit condition on the part of the licensee.

The next step of the analysis is to consider decisions concerning new product location as a function of whether the originator chooses in-house development or outsourcing. Let $L(j,t_E)$

denote the mean distance between the new product and the existing patented product in product space as a function of whether development is in-house, j=I, or outsourced, j=O, and the period of patent expiration, t_E=2,…,T. Similarly, K(j,t_E) is the investment in location precision as a function of whether development is in-house or outsourced and the period of patent expiration.

Lemma 2: Holding all other parameters fixed, if the in-house versus outsource decision is taken as given rather than as endogenously determined, then i) through v) describe L(j,t_E) and K(j,t_E).

i) L(I,t_E)=L(O,t_E)=½ for all t_E, t_E=2,…,T.

ii) K(I,1)=K(I,2)=K(O,1)=K(O,2).

iii) K(I,t_E)>K(O,t_E) for all t_E>2 and K(O,T)=0.

iv) K(I,T)>K(I,T-1)>…>K(I,2)=K(I,1).

v) K(O,1)=K(O,2)>K(O,3)>…>K(O,T)=0.

Consider first what happens when the originator chooses in-house development. Clearly, for any value of t_E profits are maximized when the new product's location is exactly half way around the unit circle from the location of the existing patented product. So L(I,t_E)=½ for all t_E, t_E=1,2,…,T. Now consider the investment in location precision. The firm's return to having the new product's location close to the mean location is higher prior to patent expiration, because prior to that date being closer to the mean translates into higher profits for both new product sales and sales of the existing patented product. So the investment level increases the later is patent expiration of the existing patented product, i.e., K(I,T)>K(I,T-1)>…>K(I,2)=K(I,1).

Now suppose the originator outsources. If t_E=1 or 2, then the patent expires by the time the new product enters the market. In this case, if sales and control rights concerning pricing are

given to the licensee, joint surplus is not improved by shifting these back to the originator in a renegotiation. So sales and control rights concerning pricing are given to the licensee in the contract, and the licensee's investment incentives are the same as the originator given in-house development. As a result, $L(I,1)=L(I,2)=L(O,1)=L(O,2)=½$ and $K(I,1)=K(I,2)=K(O,1)=K(O,2)$.

Suppose the originator opts for outsourcing and $t_E=T$. In this case, in each of periods 3 through T which are after the new product's location has been determined, joint surplus is maximized if the originator receives the returns associated with new product sales and has control rights over the pricing decision. So that is what is specified in the contract given our focus on renegotiation proof contracts. In turn, this means that the licensee has no incentive to invest in location precision, so $L(I,T)=L(O,T)=½$ and $K(I,T)>K(O,T)=0$. Note that the mean location specified in the contact is still ½ since this improves joint surplus.

Finally, suppose that the originator chooses outsourcing and $2<t_E<T$. Because the contract must be renegotiation proof and the patent is still valid for sales of the existing patented product up through period t_E, sales and control rights for pricing the new product reside with the originator up through t_E. In contrast, after period t_E, the patent has expired with the result that sales and control rights for pricing the new product reside with the licensee. The result is that the licensee's incentive to invest is higher than when $t_E=T$, but lower than when $t_E=1$ or 2, and in this range the incentive to invest falls with t_E, i.e., $K(O,1)=K(O,2)>K(O,3)>\ldots>K(O,T)=0$. Also, the incentive to invest is less than under in-house development, i.e., $K(O,t_E)<K(I,t_E)$ given $2<t_E<T$, since with in-house development the developer in every period sells the product, has pricing control rights, and therefore internalizes all the returns associated with the pricing decisions. Further, similar to the other cases, $L(O,t_E)=½$ given $2<t_E<T$.

In summary, we have established a number of results. First, mean location is always at ½ which creates the maximum expected distance in product space between the new product and the existing patented product. Second, if the firm chooses in-house development, then its investment in location precision is higher the later the patent expiration of the existing patented product. Third, prior to patent expiration of the existing product, if development is outsourced, the originator sells the product, retains control rights for pricing, and payments are such that it internalizes all the returns associated with the pricing decisions. But after patent expiration, it is the licensee who sells the product, has pricing control rights, and internalizes the returns associated with the pricing decisions. Fourth, if outsourcing is chosen, the investment in location precision is less than the investment associated with in-house development except when t_E=1 or 2 in which case the investments are the same. Fifth, if outsourcing is chosen, the investment in location precision is lower the later is the patent expiration of the existing patented product.

IV. EQUILIBRIUM ANALYSIS AND TESTABLE PREDICTIONS

This section starts with a characterization of the in-house versus outsourcing decision and then presents testable predictions. It then proceeds to discuss additional considerations.

A) Equilibrium Analysis

The focus of our analysis is the originator's choice concerning whether to conduct development in-house or to choose outsourcing. The potential benefit to outsourcing is that the fixed cost of development is lower by the amount Δ. The cost of outsourcing is that, as shown in the previous section, the expected investment in location precision is lower if t_E>2, and this serves to lower originator profits because the expected distance in product space between the new product

and the existing patented product is smaller. A comparison of this benefit and cost determines whether the originator chooses in-house development or outsourcing.

In the analysis that follows, our focus is the originator's choice of in-house development versus outsourcing as a function of the difference in fixed costs associated with in-house development. Proposition 1 captures that there is a critical value for this difference, call it Δ^*, $\Delta^*>0$, such that the originator chooses in-house development when $\Delta\leq\Delta^*$ and outsourcing when $\Delta>\Delta^*$.[14] The straightforward logic for this result is that in-house development is chosen when this choice is associated with a cost advantage or small disadvantage, while outsourcing is chosen when there is a large disadvantage associated with in-house development. The proposition also captures additional results which we discuss below.

Proposition 1: Holding all other parameters fixed, there exists a value Δ^* such that the originator chooses in-house development when $\Delta\leq\Delta^*$ and chooses outsourcing when $\Delta>\Delta^*$, where Δ^* is a strictly increasing function of t_E for $t_E\geq 2$ and equals 0 if $t_E=1$ or 2. Also, equilibrium behavior satisfies results in Lemmas 1 and 2, where the equilibrium contract given $\Delta>\Delta^*$ is unique up to the timing of the payments described in Lemma 1.

Consider first $t_E=1$ or 2. In these cases, the patent on the existing product expires before the new product reaches the market. As found in the previous subsection, when this is the case there is no advantage in terms of investments in location precision from choosing in-house development. The reason is that, if the originator chooses to outsource, then the licensee sells the product and is given pricing control rights in each period after the new product is introduced. As

[14] To simplify the exposition, we assume that the originator chooses in-house development whenever it is indifferent between the two choices.

a result, the licensee internalizes all the returns associated with the the pricing decisions concerning the new product. This means the investment in location precision is independent of whether the originator chooses in-house development or outsources. So this choice depends solely on which of the two options has lower fixed costs associated with the development process, i.e., $\Delta^*=0$ in this case.

Now consider what happens when $t_E>2$. In each period in which the patent on the existing patented product has not yet expired, we know from the previous section that independent of the in-house versus outsourcing decision the originator sells the new product, has control rights concerning pricing, and internalizes the returns associated with the pricing decisions. On the other hand, after expiration of the patent on the existing patented product, the originator sells the new product, has control rights concerning pricing, and internalizes the returns associated with pricing given in-house development, but it is the licensee who sells the product, has control rights concerning pricing, and internalizes the returns associated with pricing given outsourcing. Suppose the originator chooses to outsource. Then in some periods in which the new good is sold, the licensee does not have control rights on pricing and does not internalize the returns associated with the pricing decisions. This means that the licensee's incentive to invest in location precision is less than the originator's incentive to invest given in-house development. Given this advantage associated with in-house development, the originator will only choose outsourcing if there is a sufficiently large reduction in the fixed cost of development associated with outsourcing, i.e., $\Delta^*>0$ if $t_E>2$.

Now consider two values for t_E, t' and $t'+1$, where $t'\geq 2$. The difference between these two values is that the patent on the existing patented product has expired in period $t'+1$ when $t_E=t'$, but is still valid when $t_E=t'+1$. If the originator chooses in-house development, then in period $t'+1$ the

originator has control rights on pricing and internalizes the returns associated with pricing decisions both when $t_E=t'$ and when it equals $t'+1$. The return to investing in location precision, however, is higher when $t_E=t'+1$ because the patent on the existing patented product has not yet expired in $t'+1$ which means the profit ramifications associated with pricing in $t'+1$ are higher when $t_E=t'+1$. So if the originator chooses in-house development, then it will invest more in location precision when $t_E=t'+1$.

Suppose instead the originator chooses outsourcing. Then in period $t'+1$ the licensee has pricing control rights and internalizes the returns associated with pricing when $t_E=t'$, but when $t_E=t'+1$ the originator has pricing control rights in period $t'+1$. This means that the licensee's incentive to invest in location precision is lower when $t_E=t'+1$ than when $t_E=t'$. So if the originator chooses outsourcing, the investment in location precision is lower when $t_E=t'+1$. Combining this result with the previous one yields that the expected loss due to a lower investment in location precision when the originator chooses outsourcing is higher when $t_E=t'+1$. So the reduction in fixed costs associated with outsourcing required for the originator to make that choice must be higher when $t_E=t'+1$, i.e., Δ^* increases with an increase in t_E.

Overall, there are different investment incentives in location precision between in-house development and outsourcing. In turn, these differences determine the reduction in the fixed costs of development associated with outsourcing needed for outsourcing to be chosen by the originator. When the originator chooses in-house development, then in each period the firm with control rights on pricing and which internalizes the returns associated with pricing decisions is also the firm that chooses the investment in location precision. In contrast, when outsourcing is chosen, this is not the case. As a result, if the patent expires after the introduction of the new product, then there is underinvestment in location precision given outsourcing, and for outsourcing to be chosen there

must be a corresponding reduction in the fixed costs of development. Further, the higher is the number of periods in which the patent on the existing patented product has not expired after the introduction of the new product, the larger is the underinvestment associated with outsourcing, and thus the larger is the required reduction in the fixed cost of development associated with outsourcing for outsourcing to be chosen.

B) Testable Predictions

We now discuss testable predictions. The first two testable predictions follow immediately from results stated in Proposition 1.

Testable Prediction 1: A firm developing a new product has a higher probability of choosing in-house development if it sells an existing patented product in the same product class, and the new product is expected to reach the market before this patent expires.

Testable Prediction 1 is basically the Proposition 1 result that $\Delta^*=0$ given $t_E=1$ or 2 and $\Delta^*>0$ for all $t_E>2$. Remember that Δ^* determines the probability that in-house development is chosen, where a higher value for Δ^* translates into a higher probability that the choice is in-house development. Proposition 1 says that when $t_E=1$ or 2, i.e., at the time the new product reaches the market the patent on the existing product will have expired, that $\Delta^*=0$. In other words, in this case the in-house versus outsource decision is determined solely by which choice results in lower costs. But if $t_E>2$, i.e., the patent on the existing patented product will be valid at the date the new product reaches the market, then $\Delta^*>0$ which means that outsourcing is only chosen if it is associated with a cost advantage. Or overall, there is a higher probability of in-house development when the patent

on the existing patented product will still be valid at the date the new product is expected to reach the market.

Testable Prediction 2: Consider a firm developing a new product that owns an existing patented product in the same product class. The longer this patent is expected to be valid after the new product reaches the market, the higher is the probability the firm chooses in-house development.

Testable Prediction 2 is the Proposition 1 result that Δ^* increases with an increase in t_E. As before, Δ^* determines the probability of in-house development, where a higher value for Δ^* means a higher probability of in-house development. The proposition states that an increase in t_E increases Δ^*, where the logic is that an increase in t_E raises investment incentives given in-house development, but does not given outsourcing. So the underinvestment given outsourcing rises with an increase in t_E, which means the fixed cost reduction associated with outsourcing needed for outsourcing to be chosen is higher. This is equivalent to saying that when the patent on the existing patented product is expected to be valid for a longer period of time after the new product reaches the market, i.e., t_E is higher, Δ^* rises which translates into a higher probability of in-house development.

The third prediction concerns market share. In our model, the originator is a monopolist in the product class. But suppose that, rather than being a monopolist, the originator was one of a small number of firms selling products in the product class. In this case, the return to the originator of increased location precision would be positively related to the market share of the firm's existing patented products at the date the new product would reach the market. If this share was low, then cannibalization would be mostly in terms of other firms' patented products and sales of products

not under patent protection, so the firm's incentive to control product location of the new product would be low. But if the share was high, then the firm's incentive to control product location of the new product would be high because the return to avoid cannibalization of its own patented products would be high. This logic leads to our third testable prediction.

Testable Prediction 3: Consider a firm developing a new product that sells existing patented products for which the patents are scheduled to expire after the new product reaches the market. The probability of in-house development will be higher the larger is the predicted market share of the firm's existing patented products at the date of the new product's introduction.

One way to think about this prediction is to focus on the two returns to location precision in our argument. One return is that by reducing the expected deviation between the realized location of the new product in product space and the optimal location, the firm increases the profitability of the new product The second is that by reducing this expected deviation, it also increases the profitability of its existing patented products prior to the expiration of those patents. Increasing the market share of the existing patented products in the product class makes the second factor more important, which means an increase in the returns to improved product location. So when that market share is higher, we should expect a higher probability of in-house development since in-house development increases investments in location precision.

C) Additional Theoretical Considerations

In this subsection, we consider two alternative strategies that firms might adopt to address the underinvestment problem concerning location precision associated with outsourcing. Consider

an originator that owns an existing patented product, with this patent still valid when the new product reaches the market. The first alternative strategy is that, if the originator chooses to outsource the development process, it also sells the patent on the existing patented product to the licensee. This would avoid the underinvestment problem due to outsourcing, because, if the licensee owns the patent on the existing patented product, then the contract between the originator and the licensee will be written in such a way that the licensee invests optimally in location precision.

An important problem with this alternative strategy, however, is that if the originator has private information concerning the value of the patent on the existing patented product, then adverse selection would make this strategy unattractive. This is basically an application of Akerlof's (1970) seminal adverse selection argument. If the originator has private information, then licensees will only be willing to offer an amount for the existing patent that reflects the expected value of the existing patent given that a sale actually takes place. Akerlof's insight was that in such a case the sale only takes place when the realized value of the existing patent is close to the minimum possible value. In other words, the originator will not sell the existing patent if the originator's private information indicates that the value of the patent is not close to this minimum value. So this alternative strategy might be used to avoid the underinvestment problem, but only in the rare cases in which the private information of the originator indicates to the originator that the value of the existing patent is very low.[15]

[15] A related alternative is that, instead of the originator selling the patent on the existing patented product to the licensee, the contract is such that the payment between the parties in each period prior to patent expiration depends on that period's price and quantity for the original patented product in a manner that is equivalent to the originator selling the patent to the licensee. This alternative has the same adverse selection drawback as the alternative of the originator selling the patent to the licensee.

The second alternative strategy is to make the payments to the licensee when outsourcing is chosen a function of either investment levels chosen by the licensee or the location in product space of the new product. We have assumed that investment levels and product location are not contractually verifiable, and thus that this type of strategy is not feasible. We feel this approach is realistic. The research and development process is quite complex. Trying to determine in a real world setting how much was invested in controlling the location of the new product in product space seems to us far beyond the ability of courts in contract enforcement. So we believe that our assumption that the investment in location precision is not contractible is realistic.

The idea that a contract might contain provisions related to the new product's location in product space seems more realistic. However, it is the case that courts would have difficulty enforcing such provisions because of the difficulties involved in verifying a product's exact location in product space. So, although contractual provisions based on product location are likely feasible to an extent, the ability of such provisions to completely avoid the underinvestment problem associated with outsourcing is likely limited. Our assumption that such provisions are not enforceable is a tractable way of capturing that these types of contract provisions are of limited use in real world settings.

V. DATA FROM THE PHARMACEUTICAL INDUSTRY

The pharmaceutical industry is an excellent candidate for testing our theory for a number of reasons. First, the industry spends a substantial amount on R&D for the development of new drugs each year. For example, as reported in Schulze et al. (2014), total global spending on R&D in the pharmaceutical industry was approximately 137 billion in US dollars in 2013. Second, patents are heavily used in this industry, which means that it is possible to measure variation in the

cost of cannibalization in this industry. Third, uncertainty in the pharmaceutical development process means that this industry matches the model's assumption that firms cannot perfectly control the exact location of new products in product space. Fourth, it is common practice at pharmaceutical firms to develop some new drugs in-house, while outsourcing the development of others. This suggests sufficient variability concerning the in-house versus outsource decision to make testing our theory possible using data from this industry.

A) Main Data Source and Descriptive Statistics

Our principal data source is the Pharmaprojects dataset. This dataset was assembled by Informa and contains information concerning the development of new pharmaceutical projects throughout the world. The dataset covers information for the time period 1989 through 2004. For each chemical compound under development, the dataset contains the name of the originator, the therapeutic class, the active ingredient, patent number and patent filing date, if any, whether development was outsourced and if so the names of the licensees, and the beginning and end dates of licenses and development stages.

A key issue for our empirical analysis is defining whether development is in-house or outsourced. In particular, if a compound is developed in-house or outsourced depends on whether the originator ever signs a licensing contract and, if it did, the stage of development at which the earliest licensing contract was signed.[16] Clearly, if there was never a licensing contract, then development was in-house which is how we categorize it. On the other hand, if there was a license

[16] Note that when two firms merge, our main dataset updates the company name of the originator of the compound to the name of the acquiring firm. For example, if Warner-Lambert was the originator of compound X in 1997 and was acquired by Pfizer in 1999, then in our main dataset it is possible that Warner-Lambert would not be identified as the originator which could create misclassification if there was a licensing contract between Warner-Lambert and Pfizer prior to 1999. We compiled a list of mergers and acquisitions and assigned compounds to the correct originating firms to avoid any statistical problems related to such misclassifications.

at some point in the development process, we categorize the development process as in-house if the first license is signed late in the development process, and outsourced if the first license is signed early in the development process. The basic idea is that the main decisions affecting the product's location in product space are typically made early in the development process.

Table 1 presents a summary of the development phases as described by the FDA. Pre-clinical consists mostly of tests on laboratory animals, while phase I focuses on safety and phase II on effectiveness and side effects. Both phase I and phase II are conducted on a relatively small scale typically, with the former recruiting around 20 to 80 subjects and the latter between a few dozen to about 300 subjects. Phase III continues testing on safety and effectiveness employing a much larger sample, usually ranging from several hundred to 3,000. Conceivably, a developer could still affect a new drug's location in product space through recruitment of specific population groups and testing of specific side effects. However, once phase II is completed and the FDA meets with the developer to discuss plans for phase III, it becomes quite difficult for the developer to make significant changes that would affect the new drug's expected location in product space.

With this in mind, we categorize a drug as being developed in-house if there was no license concerning the development process prior to the beginning of phase III. And we categorize the development of a drug as being outsourced if there was a license concerning the development process prior to the start of phase III. Restricting focus to compounds for which there was at least one drug development license, the initial license occurred during pre-clinical testing in 48.9 percent of the cases, during phase I in 10.0 percent of the cases, during phase II in 14.3 percent of the cases, during phase III in 14.9 percent of the cases, and in 11.9 percent of the cases the first license was only agreed to after the product was launched.

We construct three measures of whether the originator of a new product owns existing patented products in the same product class. The first measure, which we call EOP1, is an indicator variable that takes on a value of one if the originator owns one or more other patented products in the same product class and a value of zero otherwise.[17] The second measure, which we call EOP2, equals the number of other patented products owned by the originator which are in the same product class. With a third measure, called EOP3, we try to distinguish between existing patents in later development stages from those in earlier stages. Our goal is to construct a measure of the expected number of other patented compounds in the same product class owned by the originator that will eventually reach the market. Following Higgins and Rodriguez (2006), we construct EOP3 using a count of same-class same-firm patented compounds weighted by the the probability of becoming an approved drug conditional on the current stage of development. Based on existing research, the probabilities are 0.08 for pre-clinical, .020 for phase I, 0.28 for phase II, 0.58 for phase III, and 1.0 for launched drugs.

We also employ patent length variables. The first patent length variable, called LOP1, is the remaining length of the patent with the largest remaining length of all the patents owned by the same firm which are in the same class as the drug under development. The second patent length variable, LOP2, is the sum of the remaining patent lengths of all the drugs in the same product class owned by the same firm as the drug under development. The third patent length variable, LOP3, is the weighted sum of the remaining patent lengths for all the drugs in the same product class owned by the same firm as the drug under development, where the weights are based

[17] According to the Pharmaprojects Therapeutic Class Codes, there are 17 broadly defined categories. These categories are alimentary/metabolic products, blood and clotting products, cardiovascular products, dematological products, formulations, genitourinary products, hormonal products, immunological products, anti-infective products, anticancer products, musculoskeletal products, neurological products, anti-parasitic products, respiratory products, sensory products, biotechnology products, and miscellaneous products.

on the probabilities that drugs still under development reach the market as a fuction of the current stage of development.

One potential alternative explanation for how an originator chooses between in-house development and outsourcing is that it chooses to develop a drug in-house when it has more experience and expertise. In some of our empirical tests, we thus control for an originator's prior experience and expertise in developing drugs in the product class of the current drug under development. We define class-specific experience as the sum of the compound-year observations within a firm's particular therapeutic class up to the current year.[18] Including this variable in our tests allows us to separate the cost-side explanation for the in-house versus outsource decision from our explanation which concerns investment incentives.

In addition to class-specific experience, in some of our tests we also include a measure of economies of scope as another cost-side control. Following Danzon et al. (2005), we construct a Herfindahl Hirschman Index (HHI) for each firm's therapeutic scope by summing the squares of the percentage of compounds being developed for each therapeutic class within a firm in a given year. The bigger the value for HHI, the more concentrated is the firm's development portfolio in terms of therapeutic class. One might hypothesize that economies of scope arise when the development portfolios are less concentrated, and so lower HHI should be associated with a higher probability of in-house development.

We also construct two variables that are meant to capture market-level variability in a firm's incentive to avoid cannibalization. The first variable, PDM, is the number of patented drugs on the market that are in the same therapeutic class as the drug under development, but which are owned by firms other than the originator of the drug under development. The second variable,

[18] We have also conducted tests including class-specific experience by development phase measures. This did not change the qualitative nature of the results as they related to the testable predictions we focus on.

TDM, is the total number of drugs on the market that are in the same therapeutic class as the drug under development. The purpose of constructing these two variables is that including them in our tests allows us to control for the possibility that an increase in the number of competing drugs on the market, holding fixed the total number of drugs on the market, lowers the incentive for the originator to avoid cannibalization. Note that Table 2 provides a full list of our constructed variables along with their definitions.

Table 3 reports descriptive statistics for the development decision and patent data. The sample used in the main analysis contains 11,739 compounds originated by 610 firms between 1989 and 2004. On the compound-year level, 75.9 percent of the observations are in-house development. The data for the first patent existence measure, EOP1, shows that in 75.3 percent of the compound-year observations characterized by in-house development and 63.8 percent of the observations characterized by outsourcing, the originator owned at least one other patented compound in the same product class. In addition, compared to compound-year observations characterized by outsourcing, in-house observations have a higher number of other patented compounds in the same product class owned by the originators (EOP2), as well as a higher expected number of other patented compounds owned by the originators that are in the same product class and that are expected to reach the market (EOP3).

B) Secondary Data Source and Descriptive Statistics

To supplement the analysis, we also use the IMS dataset to create a market share measure based on drug sales. The IMS data includes a list of all drugs and their annual sales in the US between 1992 and 2004. The sales data are merged into the principal dataset from Pharmaprojects based on the name of the drug, its therapeutic class, as well as whether the drug is branded or not.

Because the IMS dataset classifies drugs according to the Anatomical Therapeutic Chemical (ATC) system which differs somewhat from the classification system used in the Pharmaprojects dataset, the tests that employ our market share measure only focus on the drug classes that are common to both classification systems.[19] We calculate for each drug under development in the principal dataset the summation of market shares for other patented drugs in the same class in each year. This variable, referred to as MSP, allows us to test how market share based on sales data affects a firm's incentive to choose in-house development.

VI. EMPIRICAL TESTS

In this section we empirically test the theoretical predictions derived in Section IV. We start with the two predictions concerning the role of other patents owned by the originator, and then consider the prediction involving market share. Robustness checks are presented at the end of the empirical analysis.[20]

A) Patent Existence

To investigate whether owning a patent in the same product class as the drug under development has an effect on the originator's decision whether to develop the product in-house or outsource, we estimate the following logit specification.

$$(1) \qquad \text{Prob}(Y_{ijkt}=1) = \Lambda(\alpha_0+\alpha_1 EOP_{ijkt}+\alpha_2 X_{ijkt}+\alpha_3 Z_{jkt}+\alpha_4 W_{jt}+C_k+T_t)$$

[19] Of the drugs that are launched in the Pharmaprojects dataset, we are able to match 57 percent to a listing in the IMS dataset. The 12 classes of drugs common to both datasets are alimentary/metabolic products, blood and clotting products, cardiovascular products, dematological products, genitourinary products, hormonal products, anti-infective products, musculoskeletal products, neurological products, anti-parasitic products, respiratory products, and sensory products.

[20] A more complete set of empirical tests can be found in Pan (2016).

$\Lambda(.)$ is the standard Logistic CDF. The subscripts i, j, k, and t index compound, firm, therapeutic class, and year, respectively. Y_{ijkt} is an indicator variable for in-house development. EOP_{ijkt} is a patent existence variable, where in some tests it is the indicator measure (EOP1), in other tests it is the indicator measure (EOP2), and in others it is the measure of the expected number of other patented products that will eventually reach the market (EOP3). X_{ijkt} is a vector of development phase indicator variables. Note that because of the way we define in-house development versus outsourcing, i.e., a drug under development is said to be outsourced when the originator signs a license prior to a phase III trial, only decisions made prior to phase III matter. Hence, in the analysis pre-clinical testing is the omitted comparison group, and controls for phase I and phase II trials are included. Z_{jkt} is the originator's development experience in the therapeutic class of the drug under development, while W_{jt} is the originator's therapeutic scope.

Equation (1) also includes therapeutic class fixed effects, C_k, to control for unobserved class characteristics that affect both patent existence and integration decisions in development. Year fixed effects, T_t, control for across-time differences in firms' preferences concerning the in-house versus outsource decision. From testable prediction 1 we expect α_1 to be positive, i.e., an originator that owns an existing patented product in the same product class as the product under development should be more likely to choose in-house development.[21]

Table 4 reports the results. Each patent existence measure (EOP1, EOP2, and EOP3) is estimated under two different specifications. In the first specification therapeutic experience and therapeutic scope are omitted. In the second specification experience and scope are included. All regressions employ robust standard errors to account for heteroskedacticity. Also, standard errors

[21] We do not include firm fixed effects because, in most cases, a firm has a limited number of products under development in a product class, so including firm fixed effects would eliminate most of the variability we use to estimate the coefficients of interest.

are clustered at the compound level to account for potential correlation concerning the in-house versus outsource decision for a particular compound across observations.

The first two columns report results for EOP1. Focusing on the coefficient of interest which is the coefficient on EOP1, we see that the coefficient is positive and statistically significant at the one percent level in each regression. The other results of interest are that in column 2 the coefficient on the experience variable is positive and statistically significant at the one percent level, and the coefficient on the scope variable is negative and statistically significant at the one percent level. The former result is consistent with therapeutic experience lowering the costs of in-house development and thus making in-house development more likely, while the latter result is consistent with therapeutic scope lowering costs which makes in-house development more likely (remember that our scope variable is such that a higher value means a less diversified portfolio of projects).

The other result to note is that the size of the coefficient on EOP1 falls as we move from column 1 to column 2. This result suggests a positive correlation between EOP1 and the experience variable and/or a negative correlation between EOP1 and the scope variable. This, in turn, suggests that part of the larger positive coefficient in column 1 may not be due a direct relationship between patent existence and in-house development, but to correlations between patent existence, experience, and scope, and correlations between experience, scope, and in-house development.

Columns 3 and 4 consider the same set of tests focusing on our patent count variable, and columns 5 and 6 consider the same tests focusing on the variable that measures the expected number of other patented products in the same class that will eventually reach the market. In each set of tests the results are similar to what we found in columns 1 and 2. For example, the coefficient

on the patent variable is always positive and statistically significant at the one percent level without controls for experience and scope. Also, in each set of tests the absolute value of the coefficient on the patent variable is higher in the odd numbered column which do not include controls for experience and scope.

There are, however, a few differences worth pointing out. First, in column 4 the coefficient on the experience variable is positive, but it is not statistically significant at standard confidence levels. In the analogous tests in columns 2 and 6 the experience coefficient is statistically significant at the one percent level. Second, in column 6 the coefficient on the patent count variable is positive but only statistically significant at the ten percent level – in the analogous tests in column 2 and 4, the coefficient on the patent variable is statistically significant at the one percent level.

Our preferred specification is the one which controls for experience and scope. Using this specification, we now report results concerning how much less licensing occurs prior to phase III when the originator owns patented products in the same product class. The baseline probability that a license for the development of a new drug is agreed to prior to phase III is 9.5 percent. Employing the coefficient on EOP1 in column 2, we have that licensing prior to phase III is 3.4 percent less likely per year when the originator owns at least one other patented compound in the same therapeutic class relative to the probability when the originator owns no such patented compound. This translates into a 20.2 percent cumulative decrease from the baseline in the probability of a licensing contract prior to phase III.

We can also do similar calculations employing analogous coefficients in columns 4 and 6. Using the coefficient on EOP2 in column 4, we estimate that increasing the number of other patented compounds in the same class owned by the originators by one standard deviation

decreases the probability of outsourced development by 7.4 percent per year, or cumulatively a 44.2 percent decrease from the baseline in the probability of a licensing contract prior to phase III. Using the coefficient on EOP3 in column 6, we estimate that increasing by a standard deviation the expected number of other same-class compounds owned by the originators that are expected to reach the market decreases the probability of outsourced development by 2.8 percent per year, or a cumulative decrease of 17.1 percent from the baseline in the probability of a licensing contract prior to phase III.

B) Patent Length

To investigate whether the length of patents owned by originators in the same product class as the drug under development has an effect on the originator's decision whether to develop the product in-house or outsource, we estimate the following logit specification.

(2) $$\text{Prob}(Y_{ijkt}=1) = \Lambda(B_0+B_1LOP_{ijkt}+B_2X_{ijkt}+B_3Z_{jkt}+B_4W_{jt}+C_k+T_t)$$

LOP_{ijkt} is a patent length variable, where in some tests it is the length of the longest patent of the drugs the originator owns in the same class as the drug under development (LOP1), in other tests it is the sum of the patent lengths of the drugs the originator owns in the same class as the drug under development (LOP2), and yet in other tests it is the weighted sum of the patent lengths of the drugs the originator owns in the same class as the drug under development (LOP3). The control variables for development phase (X_{ijt}), therapeutic experience (Z_{jkt}), scope (W_{jt}) and fixed effects for therapeutic category (C_k) and year (T_t) are the same as in equation (1). From testable prediction 2, we expect B_1 to be positive, i.e., an originator with longer patent life for drugs it owns in the same product class as the drug under development should be more likely to choose in-house development.

Table 5 reproduces the tests in Table 4 where we substitute our patent length variables for the patent existence variables. We start by discussing the results in columns 1 and 2 which employ the patent length variable LOP1. The results here are similar to what we saw for patent existence in Table 4. First, the coefficient on LOP1 is positive in both regressions, where it is statistically significant at the one percent level in each regression. Second, the coefficient on LOP1 in column 2 is smaller in absolute value than the column 1 coefficient. Third, in column 2 the coefficient on the experience variable is positive and statistically significant at the one percent level, while the coefficient on the scope variable is negative and statistically significant at the one percent level. The results in columns 3 and 4 for LOP2 and in columns 5 and 6 for LOP3 exhibit similar patterns.

As in the case of patent existence, our preferred specification is the one in which experience and scope variables are included. So we focus on that specification in reporting results concerning how much changes in patent length of patents of other products in the same class owned by originators affects the probability development is conducted in-house. Consider the coefficient in column 2 on LOP1. This coefficient tells us that a one standard deviation increase in the length of the largest patent on any other patent held by the originator in the same product class as the product under development is associated with a 1.1 percent decrease per year in the probability the originator agrees to a licensing contract prior to phase III. This translates into a cumulative decrease of 6.3 percent from the baseline in the probability of licensing prior to phase III.

We can also conduct similar exercises using results from columns 4 and 6. Employing the coefficient on LOP2 in column 5 yields that increasing the sum of patent lengths of patented drugs in the same product class owned by the originators by one standard deviation decreases the probability of outsourced development by 6.3 percent per year. This translates into a cumulative decrease of 37.9 percent from the baseline in the probability of a licensing contract prior to phase

III. Similar calculations using the coefficient on LOP3 in column 6 yield that increasing the weighted sum of patent lengths of patented drugs in the same product class owned by the originators by one standard deviation decreases outsourced development by 3.2 percent per year, which translates into an 18.9 percent cumulative decrease from the baseline in the probability of a licensing contract prior to phase III.

C) Market Share

This subsection considers tests related to the third testable prediction, which is that in-house development should be more common when the originator's market share in the product class is larger. We conduct two set of tests related to this prediction. The first uses the Pharmaprojects dataset to consider how the number of competing drugs owned by firms other than the originator affects the originator's incentive to choose in-house development. In the second we use IMS sales data to construct an expected market share measure for originators, and then directly test how expected market share affects the choice of in-house versus outsourced development.

As just indicated, we start with tests concerning the number of drugs in the product class owned by other firms. In the following logit specification, we develop a test by interacting the patent existence measure, EOP, with the number of other firms' patented drugs on the market that are in the same therapeutic class as the drug under development. The specific equation that we estimate takes the following form.

$$(3) \qquad \text{Prob}(Y_{ijkt}=1) = \Lambda(\gamma_0+\gamma_1 EOP_{ijkt}+\gamma_2 EOP_{ijkt} x PDM_{jkt}+\gamma_3 PDM_{jkt} + \gamma_4 TDM_{kt}+\gamma_5 X_{ijkt}+\gamma_6 Z_{jkt}+\gamma_7 W_{jt}+C_k+T_t)$$

TDM is the total number of drugs on the market that are in the same therapeutic class as the drug under development. Given we control for TDM, the effect that an existing patent in the same

product class owned by an originator should be smaller when there is a higher number of competing patents on the market owned by other firms, i.e., γ_2 is predicted to be negative. Note that other controls are the same as in equations (1) and (2) and our focus is our preferred specification which includes controls for experience and scope.

Table 6 reports results for estimating equation (3). Column 1 shows results when EOP1 is the patent existence variable. The main result here is that the coefficient on the patent existence variable is positive and statistically significant at the one percent level, while the coefficient on the interaction term is negative and statistically significant at the one percent level. Columns 2 and 3 show results for the same test, except that in column 2 EOP2 is the patent existence variable, while in column 3 it is EOP3. In each of columns 2 and 3 the pattern is the same, except that in column 2 the coefficient on the interaction term is statistically significant at the five percent level rather than the one percent level. Overall, the results in this table are consistent with testable prediction 3. That is, the effect of patent ownership by the originator of other products in the same product class as the product under development is reduced by an increase in the number of patented products in the same product class owned by other firms.

We now consider a similar set of tests, except our focus is the effect of patent length rather than patent existence on the in-house versus outsourcing decision. In particular, we estimate a logit specification similar to equation (3), except now the explanatory variable of interest is a measure of patent length rather than patent existence. The exact specification we consider is given in equation (4).

$$(4) \qquad \text{Prob}(Y_{ijkt}=1) = \Lambda(\delta_0+\delta_1 LOP_{ijkt}+\delta_2 LOP_{ijkt} x PDM_{jkt}+\delta_3 PDM_{jkt} + \delta_4 TDM_{kt} +\delta_5 X_{ijkt}+\delta_6 Z_{jkt}+\delta_7 W_{jt}+C_k+T_t)$$

We again focus on our preferred specification which includes controls for experience and scope.

Results are reported in Table 7. Column 1 reports results where LOP1 is the patent length measure. In this column the coefficient on LOP1 is positive and statistically significant at the one percent level, while the coefficient on the interaction term is negative and statistically significant at the one percent level. Columns 2 and 3 consider the same tests as in column 1, except that in column 2 LOP2 is the patent length variable and in column 3 is it LOP3. The pattern of results in columns 2 and 3 is the same as in column 1, except that in column 2 the coefficient on the interaction term is statistically significant at the ten percent level rather than the one percent level. Overall, we again find results consistent with the third testable prediction. That is, longer patent length for products owned by the originator in the same product class as the product under development increases the probabuility of in-house development, but the effect is weaker the higher the number of patented products in the product class owned by other firms.

We now consider a second approach for measuring how competition from other firms' patented products in the same product class affects the correlations we found in the previous subsections. In particular, rather than focusing on the number of other patented products owned by other firms, we focus on how a firm's expected market share of patented products in a product class affects the in-house versus outsource decision. Note that construction of our market share measures requires IMS data which only covers the years between 1992 and 2004. Here we only look at sales data for the 12 therapeutic classes listed in footnote 18. For both reasons the sample size for these tests is smaller than for previous tests.

According to testable prediction 3, the expected market share when the new drug reaches the market should be positively correlated with the probability of in-house development. That is, it is not the market share at the time of the development decision which should matter, but rather the expectation at the time of the development decision concerning the market share that the firm

will have once the new product is introduced. In Table 8 we examine how market share along with our patent existence measures affects the in-house versus outsource decision. The top panel of the table reports results for the logit specification in equation (5).

$$(5) \qquad \text{Prob}(Y_{ijkt}=1) = \Lambda(\zeta_0+\zeta_1 MSP_{ijkt}+\zeta_2 EOP_{ijkt}+\zeta_3 X_{jkt}+\zeta_4 Z_{jkt}+\zeta_5 W_{jt}+C_k+T_t)$$

For each drug under development belonging to an originating firm j and therapeutic class k, MSP_{ijkt} is the market share based on the current year-t sales of patented drugs for the same firm and class. The other regressors are defined the same way as in equation (1). In the bottom panel, we estimate a similar equation except the current MSP measure is replaced with projected future MSP. The specific equation estimated is given in equation (6).

$$(6) \qquad \text{Prob}(Y_{ijkt}=1) = \Lambda(\eta_0+\eta_1 MSP_{ijkt+\tau}+\eta_2 EOP_{ijkt}+\eta_3 X_{jkt}+\eta_4 Z_{jkt}+\eta_5 W_{jt}+C_k+T_t)$$

As indicated, the value for MSP in this specification is the expected value at the current date of what MSP will be at the date the new product is introduced given successful development. Thus, the number of years in the future the expectation concerns depends on the development phase of the observation. We base this value on DiMasi et al. (2003) which estimates the average time a drug spends in each development phase.[22]

The top panel shows that the current year MSP is positively correlated with originators choosing in-house development decisions, but the effect is not statistically significant. In contrast, in the bottom panel we find that expected future MSP is positive and statistically significant in all specifications. We also find that in both the top and bottom panels the coefficient on the patent existence variable is always positive, but it is only statistically significant in column 2 where the variable is EOP1 and in both panels it is statistically significant in that column at least at the five

[22] Based on findings in DiMasi et al. (2003), for observations in the pre-clinical phase the expectation is ten years after the year of the observation, for observations in phase I it is also ten years, and for phase II it is eight years.

percent level. These findings, especially those in column 2, are consistent with testable prediction 3.

In Table 9 we rerun the tests in Table 8 but replace the patent existence variables with the patent length variables. The results are similar. In the top panel the coefficient on current MSP is positive but never statistically significant, while in the bottom panel the coefficient on expected future MSP is always positive and statistically significant at the five percent level. Also, in both the top panel and bottom panel the coefficients on the patent length variables are always positive, but there is only strong statistical significance in column 2 in which LOP1 is the patent length variable. In particular, in that column this coefficient is statistically significant at the five percent level in the top panel, and at the one percent level in the bottom panel.

Note further that the results concerning expected future MSP in Tables 8 and 9 suggest that the effect of expected future MSP on in-house development is economically as well as statistically significant. That is, the various coefficients on expected future MSP in the bottom panels of Tables 8 and 9 indicate that a one percentage point increase in the expected future MSP variable is associated with an increase in the probability of in-house development between 9.6 and 12.1 percentage points.

D) Robustness Checks

In this subsection we consider the robustness of our results in two respects. We first consider whether results are robust to how we categorize whether an originator chooses in-house or outsourced development. We then consider whether results are robust to how we define therapeutic categories.

In the analysis above a drug is defined as developed in-house if it satisfies one of two conditions: (a) there was never a licensing agreement concerning the development process between the originator and another firm; or (b) the earliest licensing agreement occurs after the beginning of phase III trials. One might argue, however, that the design and nature of a drug is mostly fixed as early as the completion of phase I testing. With this in mind, in Tables 10 and 11 we rerun tests reported in Tables 4 and 5 with the single change that the development process is categorized as being in-house given no license agreement prior to the beginning of phase II trials instead of no license agreement prior to the beginning of phase III trials.

Table 10 reports results using our patent existence variables and our preferred specification that includes controls for both experience and scope. The results are similar to the results in Table 4. The coefficient on the patent existence measure is positive in each regression and also statistically significant at least at the five percent level in each regression. Comparing the two tables, we see that the results are, in fact, stronger when in-house development is defined as no license agreement prior to the beginning of phase II trials. In particular, the coefficient on the patent existence variable is always larger in Table 10 than in the corresponding regression in Table 4.

Table 11 reports results for our patent length variables. These results are similar to the results in Table 10. The coefficient on the patent length variable is positive in every regression, and is statistically significant at the five percent level or better in every regression. Comparing Tables 5 and 11 we see that the results are stronger when in-house development is defined as no outsource agreement prior to the beginning of phase II trials. That is, the size of the coefficient on the patent length variable is always larger in Table 11 than in Table 5.

As indicated above, our second set of robustness tests concerns the way we define therapeutic categories. In particular, one might be concerned that our therapeutic categories are too coarse to accurately capture the cannibalization effect that our theory focuses on. That is, if a firm is currently developing one drug that we classify as being in the same therapeutic class as another patented drug the firm owns, but in reality the two drugs are in different markets, then choosing to develop the new drug in-house will not be be due to the firm's incentive to limit cannibalization and protect the value of the patent on the other drug.

To address this concern, we first redefine our main explanatory variables, i.e., patent existence and patent length of other drugs owned by originators in the same therapeutic class, by using a set of narrower therapeutic classifications. For example, whereas before anti-arrhythmic drugs and cardiostimulant drugs were classified as being in the same therapeutic class, now they are in separate classes. Second, we rerun our main patent existence and patent length tests including experience and scope variables. Results are reported in Table 12. The coefficients on the patent existence and patent length variables are all positive, and in most regressions the coefficient is similar in size to the corresponding coefficient when the broader classification scheme was employed. But statistical significance, on average, across the six regressions is weaker than for the analogous six regressions which employed the broader classification scheme. This is likely due to smaller sample sizes per cell when the narrower classification scheme is employed.

VII. ALTERNATIVE EXPLANATIONS

There are two major alternative explanations for why firms choose to conduct research and development for some products in-house, while outsourcing these activities for other products. One explanation is that the decision depends on a trade-off between providing incentives for

research effort and minimizing finance costs. The basic argument, put forth initially in Aghion and Tirole (1994), is that an integrated structure is chosen when providing incentives for research effort is the more important concern. Note that this theoretical approach is quite different from ours. The focus in that analysis is the probability of successful development, while our argument concerns the new product's location in product space and how that might affect the value of existing products through cannibalization.

While we do not doubt that the perspective developed by Aghion and Tirole (1994) is important in many real world integration decisions concerning research and development, we feel that this perspective is unlikely to be the correct explanation for the results we find. According to that theory, firms with existing successful patents should be less financially constrained. Therefore, consistent with our findings, in that argument a firm with an existing patent should be more likely to choose in-house development because financing costs will be less of a concern. However, this alternative argument does not explain why current market share of existing patented drugs in the same product class owned by the originators is less successful in predicting in-house development than is future expected market share as shown in Tables 10 and 11. So our argument concerning reducing costs of cannibalization does a better job of explaining our empirical findings than does the Aghion and Tirole argument.

Another potential explanation for the in-house versus outsource decision is that firms may choose to develop some products in-house in order to capitalize on past specific investments. Even though in our empirical analysis we control for experience and scope, one might still suspect that the correlation we find between our patent existence and patent length variables and in-house development to some extent reflects past specific investments. This interpretation, however, fails to explain why the positive correlations between in-house development and the various patent

existence and patent length variables are weaker when the market is crowded with competing patented drugs ownd by other firms, as found in a number of our tests. Furthermore, the specific investment argument also does not explain our empirical findings concerning expected market share of other patented drugs owned by the originator. In summary, neither of the major alternative explanations can account for our empirical findings as fully as the theory we developed in Section III.

VIII. CONCLUSION

Our paper focuses on the idea that limiting cannibalization of existing patented products owned by originators is important for understanding a firm's decision concerning whether to develop new products in-house or outsource the development process. We begin by constructing and analyzing a theoretical model in which ownership of existing patented products in the same product class as a new product increases the incentive for an originator to develop the new product in-house. The logic is that a licensee has less of an incentive than the originator to avoid cannibalizing the value of current patented products owned by the originators, so in-house development is preferred when avoiding such cannibalization is important. The model generates testable predictions concerning in-house development, patent existence, patent length, and market share for existing patented products.

We employ data from the pharmaceutical industry to investigate the model's predictions. Our empirical findings are consistent with the theoretical predictions. First, controlling for firm characteristics and unobserved therapeutic class heterogeneity, we find that an originator with existing patented products as the product under development is more likely to develop the new product in-house. Second, the probability of in-house development also increases with the patent length of patented drugs owned by the originator that are in the same therapeutic class as the drug

under development. Third, the relationship between in-house development and our patent existence and patent length variables is weaker when there are more same class patented drugs owned by other firms, holding fixed the total number of same class patented drugs on the market. Fourth, using market share data based on drug sales, we find that the probability a new drug is developed in-house increases with the originating firm's expected future market share of its existing same class patented drugs at the date the new drug is expected to reach the market.

These findings suggest that avoiding cannibalization of existing products is an important factor in determining whether a new product is developed in-house or outsourced. In this paper, the focus has been on the incentive for in-house development when the originator owns existing patented products in the same product class, and the originator wants to control the design of the new product. A complementary perspective is that in-house development is also important when the originator owns existing patented products in the same product class that are about to expire, and as a result it is important for the originator to control the timing of the new product introduction. We feel this is an interesting topic for future research.

APPENDIX

Proof of Lemma 1: The proof of Lemma 1 implicitly assumes that $l^M=½$. For a formal proof of this statement, see the proof of Lemma 2 i) below. Consider a game between the originator and the licensee in which, at the very beginning of each period $2<t\leq T$, both players can make suggestions about who is assigned production, sales and pricing rights, and the (potentially negative) transfer from the originator to the licensee. These proposals may condition on the quantity of the new product sold each period but not on the exact price or location in product space due to inherent non-verifiability. The sequence of proposals in any given period is immaterial. A subgame-perfect

Nash equilibrium featuring a renegotiation proof contract demands that there is an equilibrium in behavioral strategies that coincides with the equilibrium strategies chosen at t=0. Moreover, it requires that there is no equilibrium in behavioral strategies that institutes an alternative contract at any t>0.

The first result that can be established, is that production of the new product in any Nash equilibrium of the game is always assigned to the developer in all periods since $c_2^+ > c_2$. Suppose this is not true in period t. Then, any firm could suggest at the beginning of period t to change the producer—as production is immaterial to incentives—and split the additional profits in any interior way while adhering otherwise to the original contract. This proves the production part of both i) and ii).

Moreover, we can establish that, for every period t>2, sales and pricing rights are allocated in a renegotiation proof subgame-perfect Nash equilibrium in such a way as to maximize total surplus. For if not, the argument provided above regarding the allocation of production applies. Note that, after the patented product has expired, both firms can only make positive profits from the new product.

That means that the originator drops any concerns about the patented product for all periods $t>t_E$ and ex ante solely focuses on maximizing profits from the new product in these periods. This aligns the incentives at $t>t_E$ between the originator and the licensee fully for any contract that is individually rational for the licensee, i.e., under which the licensee expects to break even (and only those are naturally equilibrium candidates in the first place). As such, if allocated at at t, $t>t_E$, to maximize total surplus, it is immaterial who is assigned pricing and sales rights in these periods.

The same cannot be said for periods t such that $2<t\leq t_E$. Since total surplus of the contractees in period t is maximized by optimally setting prices of both products while internalizing variable

cost, sales and pricing rights for any such period t are necessarily allocated to the originator in any renegotiation proof contract that constitutes a subgame-perfect Nash equilibrium. Moreover, to fully internalize variable cost, the originator transfers each of these periods' quantity of the new product multiplied by c_1 to the licensee. If both rights were allocated to the licensee instead, that licensee would simply maximize profits from the new product, in fact as a competitor of the originator. In such a scenario the originator could propose to take over sales and pricing while being able to split the additional total profits in any interior way. Alternatively, it also not feasible to allocate pricing to the originator whereas the licensee is responsible for sales. As the position and price are non-contractible, the licensee could only be directed to transfer a fixed amount or an amount that depends on sales numbers to the originator. Such a transfer, however, incentivizes the originator to maximize profits of the patented product only, violating maximization of surplus. This proves the sales and pricing rights allocation part of i).

Since transfers are price and thus revenue independent, the licensee—if determined to develop the product—will in general not choose k as to maximize the total surplus to be shared among itself and the originator. If the licensee's profits are independent of total surplus, the licensee clearly chooses an investment level k=0. If the licensee's profits only depend on the performance of the new product, the licensee chooses a k_L with $0<k_L<k^*$, where k^* is the investment level that maximizes total profits for both products. This is because on a Salop circle, the profits of the new product increase in the expected distance from the patented product in product space. This can be seen from acknowledging that a smaller expected distance increases the average willingness to pay of any customer group buying the product. Since $p'(0)=\infty$, it is then immediate that $k_L>0$. If, on the other hand, k is chosen to maximize both the total profits from both products for t, $2<t\leq t_E$, and for the new product for t, $t>t_E$, the benefits from a decreased mean

distance are higher than if k is chosen only to maximize profits form the new product for t, $t>t_E$. This establishes that $k_L<k^*$ and, likewise, that total surplus increases in k with $0\leq k_L<k^*$. As a result, in every subgame-perfect Nash equilibrium, the licensee wants to optimize profits emanating from the new product. Thus, pricing and sales rights for $t>\max\{2,t_E\}$are allocated to the licensee, establishing this part of ii).

Since multiple licensees present the originator with take it or leave it contracts, the ex ante expected profit of a licensee necessarily equals 0. For if not, another licensee would undercut the contract offer on the table. In i) and ii) we established a unique division of production, sales and pricing rights across all periods t. This pins down $k=k_L$ uniquely as well as the quantity of the new product sold as a function of the realization of ε. To fully internalize revenues and costs in periods t, $2<t\leq t_E$, as discussed above, the originator is required to transfer the production of these goods to the licensee. Furthermore, the originator needs to transfer k_L to the licensee while the licensee transfers the ex ante expected profits made in periods t_E to T to the originator, for if not, the ex ante zero profit condition of the licensee is violated and individual rationality dictates for the licensee to reject the contract. This establishes iii).

Proof of Lemma 2: It is immediate to see that on a a Salop circle both total surplus from selling both goods for positive profits in periods t, $2<t\leq t_E$, as well as only the new product in periods t, $\max\{2,t_E\}$, increases in the distance of the location of the new product from the patented one in product space. This, again is due to the average customer willingness to pay increasing in the distance of the products in product space. As choosing the mean is not costly, any developer that is maximizing profits in any period chooses $l^M=½$, establishing i).

Condition ii) follows directly since the originator, if developing the product in house, maximizes the profits of the new product only when $t_E \leq 2$ and condition ii) of Lemma 1.

$K(I,t_E)>K(O,t_E)$ for t, $t_E>2$, follows from the discussion in the proof of condition ii) in Lemma 1. The licensee, if developing the new product, chooses k as to maximize the profits of the new product in periods t, $t>t_E$. while the originator as developer maximizes total surplus from both products until $t=t_E>2$ and from the new product thereafter. As a consequence, the originator invests more in k as a developer than the licensee does under these conditions. On the other hand, if K(O,T)=0, the licensee never controls pricing and sales and thus does not benefit from the profit of either good. As a result, the licensee has no incentive to invest. This completes the proof of iii).

The later the patent expires, the longer the originator as developer invests in maximizing expected distance between the two products for the sake of both products' profits. By the discussion above, it follows that this investment increases in the number of periods with a valid patent. The last step follows from ii) above. This argument proves condition iv).

Finally, the reverse is true if the licensee is developing the product. The licensee's outcome only depends on profits in periods $t>t_E$. Therefore, the licensee invests more in maximizing expected distance between the products, the more periods it benefits from profits. This establishes v).

Proof of Proposition 1: We will prove three conditions, which when combined establish the claim. First, we will show that, for every vector of admissible parameters, there is an equilibrium of the subgame that is initiated when the originator chooses to outsource development of the new product. Second, the originator's expected equilibrium profit in this subgame is unique. And, if an equilibrium of the one-player subgame in which the originator develops the new product internally

exists, it is unique as well. Finally third, there is a unique subgame perfect equilibrium of the entire game as it pertains to expected outcomes for both parties and this equilibrium is a cutoff equilibrium of the form described in the claim of the proposition.

First, by assumption we focus on parameter values under which all consumers buy either of the two products in every period and it is profitable to have a licensee develop the new product even if the licensee were to choose an investment level k=0. We know that whenever development is outsourced, in this subgame the licensee controls production while sales and pricing rights are in the hands of the originator for t, $2<t\leq t_E$, and under the control of the licensee for periods t, $\max\{2,t_E\}$. To ensure the existence of an equilibrium in this subgame, we have to establish that there is no vector of parameters such that allocating sales and pricing in any period $t\leq t_E$ to the licensee results in higher total surplus by motivating the licensee to choose a more efficient k. While Lemma 1 i) shows that such a contract cannot constitute an equilibrium, we have not ruled out that, for some parameters, it may constitute a profitable deviation from the contract, causing a lack of equilibria in this subgame. Assume that sales and pricing are allocated to the licensee in some period $t\leq t_E$, and that the potential gain from a larger k outweighs the loss from price competition between the originator with the patented product and the licensee with the new product in t. This logic, however, is flawed. Once ε has been realized, there is always a follow-up contract that would make the licensee better off giving up sales and pricing rights in t with $2<t\leq t_E$. As such, the licensee would not choose a socially better k than the above described k_L in the first place. It follows that this subgame always has an equilibrium.

Second, points i) and ii) of Lemma 1 together with the first part of point iii) of Lemma 1 establish the uniqueness of this subgame equilibrium in terms of profits, since all rights are unambiguously assigned every single period and the expected profit of the licensee equals 0. While

the timing of fixed payments is innocuous as there must be always one party objecting to a contract change reducing its profits, this pins down the expected subgame equilibrium profits of the originator uniquely. Now consider the one-player subgame initiated by the originator choosing to develop the new product internally. In this scenario, the originator retains all rights for all periods and thus chooses the socially optimal k=k*. As a consequence, this subgame clearly has a unique equilibrium if the resulting expected profit for the originator is positive.

Third, let the expected profit of the originator from outsourcing equal $\Pi(O)$ while the expected profit from internally developing the new product is denoted as $\Pi(I,F_O)$. It follows from the discussion above that for any given set of parameters, if the subgame initiated by the originator choosing to develop the new product internally has an equilibrium, $\Pi(I,F_O)$ equals a positive constant minus F_O. The assumption about the feasibility of positive profits when outsourcing coupled with the fact that the originator chooses the socially efficient k=k* implies that $\Pi(I,F_L)> \Pi(O)>0$. Thus, by the continuity of $\Pi(I,F_O)$ in F_O, there necessarily is an $F^*> F_L$ such that $\Pi(I,F^*)= \Pi(O)>0$. It follows that the subgame perfect equilibrium of the overall game—that is unique up to timing of transfers if development is outsourced as argued above—has the originator choose to develop the new product internally if $F_O \leq F^*$ and to outsource if $F_O > F^*$. Defining $\Delta^*=F^*-F_L$ establishes the first part of the claim. Moreover, uniqueness follows trivially.

Finally, it remains to be shown that a.) $\Delta^*=0$ for t_E, $t_E \leq 2$, and b.) Δ^* is strictly increasing for t_E, $t_E \geq 2$. a.) follows from Lemma 2 ii) and the fact that the expected profit of the licensee equals 0. Now consider Lemma 2, points ii), iv) and v). Together these statements imply that the k chosen for t_E, $t_E \geq 2$, by the licensee when developing the product always falls short of k*, the optimal k as chosen by the originator when developing the new product in-house. What is more, they collectively imply that $K(I,t_E)-K(O,t_E)$ strictly increases in t_E, $t_E \geq 2$. As a consequence, the nominal

welfare loss from delegating the development to the licensee strictly increases in t_E for fixed T. This establishes b.).

REFERENCES

Aghion, P. and J. Tirole (1994), "The Management of Innovation," *Quarterly Journal of Economics*, 109, pp. 1185-1209.

Akerlof, G.A. (1970), "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism," *Quarterly Journal of Economics*, 84, pp. 488-500.

Azoulay, P. (2004), "Capturing Knowledge Within and Across Firm Boundaries: Evidence from Clinical Development," *American Economic Review*, 94, pp. 1591-1612.

Balakrishnan, S. and B. Wernerfelt (1986), "Technical Change, Competition and Vertical Integration," *Strategic Management Journal*, 7, pp. 347-359.

Barrera, C. and M. Waldman (2019), "Vertical Integration, Secrecy, and High Tech Industries," mimeo, Johnson Graduate School of Management, Cornell University.

Choi, J.P. (1994), "Network Externality, Compatibility Choice, and Planned Obsolescence," *Journal of Industrial Economics*, 42, pp. 167-182.

Danzon, P.M., S. Nicholson, and N.S. Pereira (2005), "Productivity in Pharmaceutical Biotechnology R&D: The Role of Experience and Alliances," *Journal of Health Economics*, 24, pp. 317-339.

Demougin, D.M. (1989), "A Renegotiation-Proof Mechanism for a Principal-Agent Model with Moral Hazard and Adverse Selection," *Rand Journal of Economics*, 20, pp. 256-267.

Dewatripont, M. (1988), "Commitment through Renegotiation-Proof Contracts with Third Parties," *Review of Economic Studies*, 55, pp. 377-390.

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, pp. 151-185.

Galdon-Sanchez, J.E., R. Gil, and A. Bao-Moriones (2015), "Outsourcing of Peripheral Services: Evidence from Spanish Manufacturing Plant-Level Data," *European Economic Review*, 78, pp. 328-344.

Gibbons, R. (2005), "Four Formal(izable) Theories of the Firm?," *Journal of Economic Behavior and Organization*, 58, pp. 200-245.

Gil, R. and C.A. Ruzzier (2018), "The Impact of Competition on 'Make-or-Buy' Decisions: Evidence from the Spanish Local TV Industry," *Management Science*, 64, pp. 1121-1135.

Grossman, S.J. and O. Hart (1986), "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration," *Journal of Political Economy*, 94, pp. 691-719.

Hart, O. (2017), "Incomplete Contracts and Control," *American Economic Review*, 107, pp. 1731-1752.

Hart, O. and J. Moore (1988), "Incomplete Contracts and Renegotiation," *Econometrica*, 56, pp. 755-785.

Hart, O. and J. Moore (1990), "Property Rights and the Nature of the Firm," *Journal of Political Economy*, 98, pp. 1119-1158.

Higgins, M.J. and D. Rodriguez (2006), "The Outsourcing of R&D Through Acquisitions in the Pharmaceutical Industry," *Journal of Financial Economics*, 80, pp. 351-383.

Klein, B., R.G. Crawford, and A.A. Alchian (1978), "Vertical Integration, Appropriable Rents, and the Competitive Contracting Process," *Journal of Law and Economics*, 21, pp. 297-326.

Lafontaine, F. and M. Slade (2007), "Vertical Integration and Firm Boundaries: The Evidence," *Journal of Economic Literature*, 45, pp. 629-685.

Lam, M.D. (2004), "Why Alliances Fail," *Pharmaceutical Executive*, June 1.

Lerner, J. and R.P. Merges (1998), "The Control of Technology Alliances: An Empirical Analysis of the Biotechnology Industry," *Journal of Industrial Economics*, 46, pp. 125-156.

Lerner, J. and U. Malmendier (2010), "Contractibility and the Design of Research Agreements," *American Economic Review*, 100, pp. 214-246.

Levy, D.J. (1985), "The Transactions Cost Approach to Vertical Integration: An Empirical Examination," *Review of Economics and Statistics*, 67, pp. 438-445.

Mills, L. (2004), "Great Science Not All that Matters Developing Treatments," *Financial Times*, November 10, p. 5.

Moorthy, S. and I.P.L. Png (1992), "Market Segmentation, Cannibalization, and the Timing of Product Introductions," *Management Science*, 38, pp. 345-359.

Nahm, J. (2004), "Durable-goods Monopoly with Endogenous Innovation," *Journal of Economics & Management Strategy*, 13, pp. 303-319.

Nicholson, S., P. M. Danzon, and J. McCullogh (2005), "Biotech-pharmaceutical Alliances as a Signal of Asset and Firm Quality," *Journal of Business,* 78, pp. 1433-1464.

Novak, S. and S. Stern (2009), "Complementarity Among Vertical Integration Decisions," *Management Science,* 55, pp. 311-332.

Pan, Y. (2016), *Essays in Organizations and Health Economics*, Unpublished PhD Dissertation, Cornell University.

Patel, D. (2017), "The Pros and Cons of Outsourcing," *Forbes.com,* 17 July.

Raineri, S. (2019), "Top Outsourcing Disadvantages," *The Balance SmallBusiness*, 28 February.

Salop, S. C. (1979), "Monopolistic Competition with Outside Goods," *Bell Journal of Economics,* 10, pp. 141-156.

Schulze, U. M. Baedeker, Y.T. Chen, and D. Greber (2014), "R&D Productivity: On the Comeback Trail," *Nature Reviews Drug Discovery*, 13, pp. 331-332.

Siebert, R. (2015), "Entering New Markets in the Presence of Competition: Price Discrimination versus Cannibalization," *Journal of Economics & Management Strategy*, 24, pp. 369-389.

Tucker, I.B. and R.P. Wilder (1977), "Trends in Vertical Integration in the U.S. Manufacturing Sector," *Journal of Industrial Economics,* 26, pp. 81-94.

Waldman, M. (1993), "A New Perspective on Planned Obsolescence," *Quarterly Journal of Economics,* 108, pp. 273-283.

Waldman, M. (1996), "Planned Obsolescence and the R&D Decision," *Rand Journal of Economics,* 27, pp. 583-595.

Williamson, O.E. (1975), *Markets and Hierarchies: Analysis and Antitrust Implications,* Free Press: New York.

Williamson, O.E. (1979), "Transaction-Cost Economics: The Governance of Contractual Relations," *Journal of Law and Economics,* 22, pp. 233-261.

Williamson, O.E. (2010), "Transaction Cost Economics: The Natural Progression," *American Economic Review*, 100, pp. 673-690.

TABLES

Table 1: Summary of Development Phases

Drug Development Stage	Description (taken from FDA website)
Pre-clinical trial	Submission of investigational new drug application for the FDA to review. Companies need to show results of pre-clinical testing on laboratory animals and propose plans for human testing
Phase I trial	Usually conducted in healthy volunteers to determine the most frequent side effects, as well as how the drug is metabolized and excreted. Number of subjects range from 20 to 80. Emphasis is on safety
Phase II trial	Obtain preliminary data on whether the drug treats a certain disease or condition. Number of subjects range from a few dozen to about 300. Continues to evaluate safety and short term side effects
Phase III trial	The FDA and the sponsors meet to determine how large-scale studies in Phase III should be done. Gather more information on safety and effectiveness. Studies different populations, dosages and combined usage of other drugs. Number of subjects ranges from several hundred to about 3,000 people

Source: http://www.fda.gov/drugs/resourcesforyou/consumers/ucm143534.htm

Table 2: Definition of Constructed Variables

Variable	Description
In-house	Indicator equals 1 if compound is never licensed out by the originating firm or if its earliest licensing deal was made after the start of Phase III trials
Existence of Patents	
EOP1	Indicator equals 1 if at least one other compound in the same therapeutic class and same firm is patented
EOP2	Number of other patented compounds in the same therapeutic class and same firm
EOP3	Expected number of other patented compounds in the same therapeutic class and same firm that would reach the market
Length of Patents	
LOP1	Length of the longest patent among other compounds in the same therapeutic class and same firm
LOP2	Sum of the patent length among other compounds in the same therapeutic class and same firm
LOP3	Weighted sum of patent length among other compounds in the same therapeutic class and same firm according to their probability of becoming an approved drug
Experience	Cumulative count of compound-year observations within a firm for a therapeutic class corresponding to the compound of interest
Scope	Sum of the squares of the percentage of compounds being developed for each therapeutic class within a firm in a given year
PDM	Number of patented drugs on the market in the same therapeutic class but not the same firm as the compound of interest
TDM	Total number of drugs on the market in the same therapeutic class as the compound of interest
MSP	Market share based on sales for existing patented drugs in the same class and same firm as the compound of interest

Table 3: Descriptive Statistics on In-house Development and Patent Profile

Number of compounds	11739
Number of firms	610
Years covered	1989-2004

Variable	Overall Mean	Mean Outsourced Compounds	Mean In-House Compounds	Min.	Max.
	Level of Observation: compound-year (67,215 obs.)				
In-house	0.759			0	1
	(0.428)				
Existence of Patents					
EOP1	0.725	0.638	0.753	0	1
	(0.446)	(0.481)	(0.431)		
EOP2	10.504	4.567	12.393	0	101
	(17.213)	(9.521)	(18.626)		
EOP3	1.833	1.049	2.083	0	15.12
	(2.746)	(1.970)	(2.906)		
Length of Patents					
LOP1	11.605	9.587	12.248	0	20
	(7.854)	(7.952)	(7.712)		
LOP2 (x10)	10.950	4.759	12.920	0	106.2
	(17.811)	(9.717)	(19.288)		
LOP3 (x10)	1.736	1.005	1.969	0	12.38
	(2.498)	(1.796)	(2.640)		

Note: EOP1 is an indicator for at least one other patented compound in the same therapeutic class and same firm as the compound of interest. EOP2 is the number of other patented compounds in the same therapeutic class and same firm. EOP3 is the expected number of other patented compounds in the same therapeutic class and same firm that would reach the market. LOP1 is the length of the longest patent among other compounds in the same therapeutic class and same firm as the compound of interest. LOP2 is the sum of patent lengths among other compounds in the same therapeutic class and same firm. LOP3 is the weighted sum of patent lengths among other compounds in the same therapeutic class and same firm according to their probability of becoming an approved drug.

Table 4: Logit Models of In-house Development: Existence of Patents

	(1)	(2)	(3)	(4)	(5)	(6)
Existence of Patents						
EOP1	0.757***	0.357***				
	(0.107)	(0.117)				
EOP2			0.072***	0.050***		
			(0.013)	(0.016)		
EOP3					0.291***	0.115*
					(0.052)	(0.070)
Experience		0.033***		0.013		0.027***
		(0.007)		(0.011)		(0.010)
Scope		-0.702***		-0.671***		-0.704***
		(0.199)		(0.196)		(0.198)
Phase I	-1.036***	-1.011***	-0.983***	-0.987***	-1.009***	-1.005***
	(0.143)	(0.142)	(0.142)	(0.142)	(0.142)	(0.142)
Phase II	-1.170***	-1.168***	-1.123***	-1.136***	-1.161***	-1.164***
	(0.128)	(0.128)	(0.128)	(0.129)	(0.128)	(0.128)
Constant	5.280***	5.770***	5.446***	5.755***	5.439***	5.845***
	(0.712)	(0.728)	(0.715)	(0.729)	(0.713)	(0.730)
Observations	47831	47831	47831	47831	47831	47831

Note: Dependent variable is one if a compound is developed in-house, and zero otherwise. All specifications include a full set of therapeutic category and year indicators. EOP1 is an indicator for at least one other patented compound in the same therapeutic class and same firm as the compound of interest. EOP2 is the number of other patented compounds in the same therapeutic class and same firm. EOP3 is the expected number of other patented compounds in the same therapeutic class and same firm that would reach the market.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard error (in parentheses) are heteroskedasticity-robust and clustered on the compound level.

Table 5: Logit Models of In-house Development: Length of Patents

	(1)	(2)	(3)	(4)	(5)	(6)
Length of Patents						
LOP1	0.044***	0.019***				
	(0.007)	(0.007)				
LOP2			0.066***	0.043***		
			(0.012)	(0.014)		
LOP3					0.306***	0.140**
					(0.052)	(0.062)
Experience		0.034***		0.018*		0.027***
		(0.008)		(0.009)		(0.009)
Scope		-0.718***		-0.687***		-0.708***
		(0.198)		(0.195)		(0.197)
Phase I	-1.018***	-1.007***	-0.973***	-0.988***	-0.989***	-1.003***
	(0.141)	(0.142)	(0.141)	(0.142)	(0.141)	(0.142)
Phase II	-1.138***	-1.160***	-1.102***	-1.134***	-1.128***	-1.160***
	(0.128)	(0.128)	(0.128)	(0.129)	(0.128)	(0.128)
Constant	4.920***	5.778***	5.042***	5.757***	4.952***	5.813***
	(0.567)	(0.729)	(0.569)	(0.728)	(0.567)	(0.729
Observations	47831	47831	47831	47831	47831	47831

Note: Dependent variable is one if a compound is developed in-house, and zero otherwise. All specifications include a full set of therapeutic category and year indicators. EOP1 is an indicator for at least one other patented compound in the same therapeutic class and same firm as the compound of interest. EOP2 is the number of other patented compounds in the same therapeutic class and same firm. EOP3 is the expected number of other patented compounds in the same therapeutic class and same firm that would reach the market.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard error (in parentheses) are heteroskedasticity-robust and clustered on the compound level.

Table 6: Logit Models of In-house Development: Existence of Patents with Interaction Effect

	(1)	(2)	(3)
Existence of Patents			
EOP1	0.874***		
	(0.188)		
EOP1 × PDM	-0.020***		
	(0.006)		
EOP2		0.081***	
		(0.026)	
EOP2 × PDM		-0.002**	
		(0.001)	
EOP3			0.324***
			(0.098)
EOP3 × PDM			-0.007***
			(0.002)
PDM	0.017	0.008	0.004
	(0.028)	(0.026)	(0.027)
TDM	-0.003	-0.003	0.001
	(0.013)	(0.012)	(0.013)
Experience	0.035***	0.023**	0.030***
	(0.007)	(0.011)	(0.011)
Scope	-0.596***	-0.622***	-0.652***
	(0.202)	(0.198)	(0.198)
Phase I	-1.001***	-0.984***	-1.007***
	(0.143)	(0.142)	(0.143)
Phase II	-1.157***	-1.126***	-1.157***
	(0.128)	(0.129)	(0.128)
Constant	5.489***	5.626***	5.698***
	(0.733)	(0.736)	(0.734)
Observations	47831	47831	47831

Note: Dependent variable is one if a compound is developed in-house, and zero otherwise. All specifications include a full set of therapeutic category and year indicators.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard error (in parentheses) are heteroskedasticity-robust and clustered on the compound level.

Table 7: Logit Models of In-house Development: Length of Patents with Interaction Effect

	(1)	(2)	(3)
Length of Patents			
LOP1	0.049***		
	(0.011)		
LOP1 × PDM	-0.001***		
	(0.000)		
LOP2		0.062***	
		(0.023)	
LOP2 × PDM		-0.001*	
		(0.001)	
LOP3			0.324***
			(0.098)
LOP3 × PDM			-0.007***
			(0.002)
PDM	0.017	0.008	0.009
	(0.028)	(0.027)	(0.027)
TDM	-0.004	-0.004	-0.002
	(0.013)	(0.013)	(0.012)
Experience	0.035***	0.022**	0.030***
	(0.008)	(0.009)	(0.010)
Scope	-0.608***	-0.642***	-0.651***
	(0.203)	(0.199)	(0.197)
Phase I	-0.997***	-0.984***	-1.006***
	(0.143)	(0.142)	(0.142)
Phase II	-1.149***	-1.125***	-1.152***
	(0.128)	(0.129)	(0.128)
Constant	5.471***	5.638***	5.659***
	(0.734)	(0.738)	(0.734)
Observations	47831	47831	47831

Note: Dependent variable is one if a compound is developed in-house, and zero otherwise. All specifications include a full set of therapeutic category and year indicators.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard error (in parentheses) are heteroskedasticity-robust and clustered on the compound level.

Table 8: Logit Models of In-house Development: Market Share and Existence of Patents

	(1)	(2)	(3)	(4)
Panel A				
Current MSP	5.950	4.498	6.278	1.855
	(10.135)	(9.145)	(9.993)	(6.966)
EOP1		0.592**		
		(0.234)		
EOP2			0.043	
			(0.028)	
EOP3				0.308*
				(0.171)
Observations	26688	26688	26688	26688
Panel B				
Future MSP	24.447**	20.347**	24.221**	24.064**
	(11.866)	(10.097)	(11.428)	(10.945)
EOP1		1.385***		
		(0.522)		
EOP2			0.004	
			(0.026)	
EOP3				0.024
				(0.245)
Observations	7317	7317	7317	7317

Note: Dependent variable is one if a compound is developed in-house, and zero otherwise. All specifications include experience, scope, development phase indicators, as well as a full set of therapeutic category and year indicators.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard error (in parentheses) are heteroskedasticity-robust and clustered on the compound level.

Table 9: Logit Models of In-house Development: Market Share and Length of Patents

	(1)	(2)	(3)
Panel A			
Current MSP	4.649	7.111	3.838
	(9.251)	(10.349)	(8.294)
LOP1	0.039**		
	(0.015)		
LOP2		0.038	
		(0.023)	
LOP3			0.280*
			(0.158)
Observations	26688	26688	26688
Panel B			
Future MSP	18.843**	24.072**	23.185**
	(9.391)	(11.272)	(10.565)
LOP1	0.085***		
	(0.028)		
LOP2		0.004	
		(0.020)	
LOP3			0.056
			(0.191)
Observations	7317	7317	7317

Note: Dependent variable is one if a compound is developed in-house, and zero otherwise. All specifications include experience, scope, development phase indicators, as well as a full set of therapeutic category and year indicators.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard error (in parentheses) are heteroskedasticity-robust and clustered on the compound level.

Table 10: Logit Models of a Different Measure In-house Development: Existence of Patents

	(1)	(2)	(3)
Existence of Patents			
EOP1	0.406***		
	(0.133)		
EOP2		0.065***	
		(0.021)	
EOP3			0.140**
			(0.083)
Experience	0.031***	0.007	0.024**
	(0.008)	(0.012)	(0.012)
Scope	-0.672***	-0.615***	-0.659***
	(0.222)	(0.217)	(0.220)
Phase I	-1.013***	-0.986***	-1.005***
	(0.144)	(0.143)	(0.143)
Constant	6.942***	6.904***	7.033***
	(1.223)	(1.225)	(1.225)
Observations	42962	42962	42962

Note: Dependent variable is one if a compound is developed in-house, and zero otherwise. All specifications include a full set of therapeutic category and year indicators.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard error (in parentheses) are heteroskedasticity-robust and clustered on the compound level.

Table 11: Logit Models of a Different Measure In-house Development: Length of Patents

	(1)	(2)	(3)
Length of Patents			
LOP1	0.25***		
	(0.008)		
LOP2		0.058***	
		(0.018)	
LOP3			0.189**
			(0.076)
Experience	0.031***	0.012	0.022**
	(0.008)	(0.011)	(0.010)
Scope	-0.687***	-0.637***	-0.663***
	(0.221)	(0.215)	(0.218)
Phase I	-1.009***	-0.986***	-1.003***
	(0.144)	(0.143)	(0.143)
Constant	6.933***	6.898***	6.985***
	(1.223)	(1.225)	(1.225)
Observations	42962	42962	42962

Note: Dependent variable is one if a compound is developed in-house, and zero otherwise. All specifications include a full set of therapeutic category and year indicators.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard error (in parentheses) are heteroskedasticity-robust and clustered on the compound level.

Table 12: Patent Profile Variables Defined on Finer Therapeutic Classifications

	(1) EOP1	(2) EOP2	(3) EOP3	(4) LOP1	(5) LOP2	(6) LOP3
Existence of Patents (EOP)	0.284***	0.072*	0.044			
	(0.120)	(0.038)	(0.133)			
Length of Patents (LOP)				0.018**	0.058*	0.078
				(0.008)	(0.030)	(0.123)
Experience	0.036***	0.034***	0.041***	0.037***	0.035***	0.040***
	(0.008)	(.008)	(0.008)	(0.008)	(0.008)	(0.008)
Scope	-0.733***	-0.738***	-0.733***	-0.747***	-0.747***	-0.739***
	(0.198)	(0.197)	(0.199)	(0.197)	(0.197)	(0.199)
Phase I	-1.007***	-0.998***	-1.005***	-1.005***	-0.997***	-1.005***
	(0.142)	(0.142)	(0.142)	(0.142)	(0.142)	(0.142)
Phase II	-1.169***	-1.158***	-1.162***	-1.162***	-1.155***	-1.167***
	(0.128)	(0.128)	(0.128)	(0.128)	(0.128)	(0.128)
Constant	5.838***	5.819***	5.835***	5.835***	5.835***	5.910***
	(0.730)	(0.733)	(0.731)	(0.731)	(0.732)	(0.742)
Observations	47831	47831	47831	47831	47831	47831

Note: Dependent variable is one if a compound is developed in-house, and zero otherwise. All specifications include a full set of therapeutic category and year indicators.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard error (in parentheses) are heteroskedasticity-robust and clustered on the compound level.