NBER WORKING PAPER SERIES

THE ORGANIZATION OF INNOVATION: INCOMPLETE CONTRACTS AND THE OUTSOURCING DECISION

Thomas Jungbauer
Sean Nicholson
June Pan
Michael Waldman
Lucy Xiaolu Wang

Working Paper 28379
http://www.nber.org/papers/w28379

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2021, Revised July 2025

We would like to thank Catherine Barrera, Benno Buehler, Christopher Cairns, Yi Chen, Iain Cockburn, Florian Ederer, Chris Forman, Soheil Ghili, George Jakubson, Shulamit Khan, Francine Lafontaine, Jorge Lemus, Scott Masten, Marcel Preuss, Mark Satterthwaite, Christian Schmid, Jagadeesh Sivadasan, Scott Stern, Chad Syversen, Rainer Widmann, Heidi Williams, unnamed seminar participants, in particular at Cornell University and the University of Michigan, as well as those of the first virtual 2020 Annual Conference of the Society for Institutional and Organizational Economics (SIOE 2020), the 2021 ZEW Annual Conference of the Mannheim Centre for Competition and Innovation (MaCCI 2021), the 2021 Annual Conference of the Royal Economic Society (RES 2021), the 2024 International Industrial Organization Conference (IIOC 2024), and the 2025 Productivity Lunch Seminar at NBER for helpful comments, Kiran Butte, Yuhan Chen, Yujie Feng, Qiwei He, Rebecca Herzberg, and Douglas Weinstein for valuable research assistance, and Wyatt Gotbetter for valuable industry insights. We are particularly grateful to three anonymous referees and the editor, Daniel Yi Xu. All remaining errors are our own. Earlier drafts of this paper were circulated under the titles "The Organization of Innovation: Theory and Evidence from the Pharmaceutical Industry," and "The Organization of Innovation: Property Rights and the Outsourcing Decision." The authors have contributed equally and are listed alphabetically. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

The Organization of Innovation: Incomplete Contracts and the Outsourcing Decision
Thomas Jungbauer, Sean Nicholson, June Pan, Michael Waldman, and Lucy Xiaolu Wang
NBER Working Paper No. 28379
January 2021, Revised July 2025
JEL No. D23, L24, L65, O32

ABSTRACT

Why do firms outsource research and development (R&D) for some products while conducting R&D in-house for similar ones? An innovating rm risks cannibalizing its existing products. The more profitable these products, the more the firm wants to limit cannibalization. We apply this logic to the organization of R&D by introducing a novel theoretical model in which developing in-house provides the firm more control over the new product's location in product space. An empirical analysis of our testable predictions using pharmaceutical data concerning patents, patent expiration, and outsourcing at various stages of the R&D process supports our theoretical approach.

Thomas Jungbauer
Cornell University
jungbauer@cornell.edu

Michael Waldman
Cornell University
mw46@cornell.edu

Sean Nicholson
Cornell University
Department of Policy Analysis and Management
and NBER
sn243@cornell.edu

Lucy Xiaolu Wang
University of Massachusetts, Amherst
xiaoluwang@umass.edu

June Pan
VISA
junepan89@gmail.com

A online appendix is available at http://www.nber.org/data-appendix/w28379

1 Introduction

A firm selling products for which its market power is substantial will want to deter the introduction of close substitutes that would cannibalize sales of these already profitable products.[1] One possibility is that this threat comes from a product introduced by a competitor.[2] Another possibility, however, is that the threat comes from new products the firm itself introduces, in which case the firm will want to control the development process in order to limit cannibalization. We show how this basic argument helps explain the organization of the innovation process. In particular, we analyze a model of new product development by a firm with an existing patented product in the same product class, and show both theoretically and empirically that a firm will sometimes choose in-house development over outsourcing to limit cannibalization of its existing product portfolio. As discussed in detail below, our argument can be thought of as a variant of the well-known property rights theory of the firm (Grossman and Hart, 1986; Hart and Moore, 1990) in which vertical integration better aligns incentives for investment decisions.

We start by constructing and analyzing a theoretical model that formalizes our basic argument. In our model, a firm, called the originator, develops a new product and chooses whether to conduct the remaining 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 realized location from the mean product 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 owns an existing patented product in the same product class as the new one. Also, in our main analysis the originator is a monopolist, while in an extension we introduce a rival.

In the model, when the originator outsources there is a contract between the originator and a firm, called the licensee, that conducts the R&D. Our focus is renegotiation proof contracts, where the contract specifies for each period who produces the product, who sells the product, and who sets the new product price. In our model, the cost of cannibalization is a function of the existing product's patent expiration date. For each possible period of patent expiration, the analysis compares the optimal investment in location precision when the R&D is conducted in-house with the investment when it is outsourced.

Our analysis generates four testable predictions. First, compared to firms that do not own existing patented products in the same product class as the one under development, a firm that

[1] See Arrow (1962) for an early related discussion, while a more recent related analysis appears in Igami (2017). We discuss the latter paper in the literature review section.

[2] In a recent paper, Cunningham et al. (2021) investigate competition that can lead to 'killer acquisitions,' i.e., firms with market power acquiring products in development and then discontinuing the R&D efforts.

does is less likely to outsource the R&D. Second, the probability of outsourcing is negatively related to the remaining patent duration of the existing patented product. Third, the probability of outsourcing is also negatively related to the market share of the existing patented product or products. For each of these predictions, the basic logic behind the prediction follows the same path. The incentive to outsource is lower the more profitable is the cannibalized product in the absence of a new product introduction.

The fourth prediction follows from a basic premise of the property rights theory of the firm, which is that ownership provides a level of control that cannot be achieved through contracting. That is, when the originator owns an existing patented product in the same product class as the product being developed, relative to outsourced R&D, in-house R&D will result in fewer product category changes concerning the nature of the product being developed. The logic here is that with in-house R&D the originator has more control over the development process, and as a result, the product developed has a higher probability of being close in product space to the optimal location from the originator's perspective.

We test these predictions using data from the pharmaceutical industry. In terms of the pharmaceutical industry, our argument is based on the idea that a compound can frequently be tested for its efficacy concerning multiple related conditions, or specific types of patients within the same condition (e.g., early-stage versus late-stage prostate cancer), and thus—consistent with Lerner and Merges (1998)—decisions taken early in the development process can affect a drug's location in product space. In turn, a pharmaceutical firm with relevant existing patented drugs has an incentive to conduct R&D in-house to better control the development process limiting the possible cannibalization of its existing patented drugs.

In our main empirical analysis we use therapeutic classifications in the Pharmaprojects dataset to construct measures concerning whether the originator of a new product owns existing patented products in the same class. This analysis strongly supports all four testable predictions. We find that the probability of outsourcing is lower when the originating firm has one or more existing patented compounds in the same therapeutic class as the compound under development. In addition, given the presence of existing patented compounds in the same therapeutic class, the probability of outsourcing is negatively correlated with the remaining patent lengths of these compounds. We also supplement the main data source with sales data, and show that the originator is less likely to outsource development the higher its market share for its same-class patented drugs. Finally, given the originator owns an existing patented product in the same therapeutic class as the product being developed, the number of development "pivots" during the development process is lower with in-house development. This is consistent with the fourth prediction if the initial posi-

tion in product space of the product being developed is the optimal location. Note that we discuss alternative explanations for these findings in Section 6.

We also conduct a number of robustness checks including the following two. The first involves the sample of development decisions that we assume can affect the new product's location in product space, while the second relates to how we define whether the originator of a new product owns an existing patented product in the same product class. In each case results again support the testable predictions.

One important message of this paper is that the vertical integration decision can depend on factors not covered by the mainstream theories describing boundaries of the firm. Specifically, models of vertical integration typically focus either on the characteristics of an input and/or the final product(s).[3] Our argument, however, is fundamentally different. In our model, whether or not to outsource the development process depends on the characteristics of products in the firm's product portfolio for which the R&D is not an input. 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 resulting in a cost of outsourcing development decisions. Our empirical analysis suggests that this consideration is an important driver of the in-house versus outsourcing decision regarding the development process.

The outline for the paper is as follows. The remainder of this section discusses our contribution in light of the relevant literature. Section 2 then presents the model and provides a preliminary analysis. In Section 3, we conduct a full equilibrium analysis and discuss our testable predictions. Section 4 then describes the pharmaceutical data, before Section 5 presents the empirical analysis. Thereafter, in Section 6 we discuss the extent to which our findings can be explained by alternative theories or empirical factors. Finally, Section 7 concludes. Proofs of formal statements in the text are found in Appendix A. Appendix B generalizes our theoretical results introducing competition and multiple R&D activities. Appendix C provides additional tables.

Related Literature. This paper contributes to the literature on incomplete contracting, and, in particular, is closely related to the property rights theory of the firm which starts with the seminal analysis of Grossman and Hart (1986) (see also Hart and Moore, 1990).[4] The property rights theory of the firm postulates that contracts are incomplete, and asset ownership grants residual control rights to the owner of the asset. In Grossman and Hart (1986), two parties make non-contractible investments ex ante while utility is non-transferable. They show that one party

[3] 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. They provide two possible explanations for the result, neither of which, however, is related to our argument.

[4] See Gibbons (2005) and Lafontaine and Slade (2007) for surveys focused on the vertical integration decision including the property rights theory of the firm.

purchases the other's assets whenever the former's investment is more important than the latter's.

Our paper relates to 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. We show that the originator chooses to vertically integrate and retain the ability to influence the new product's location instead, 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 important 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), 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 chooses a higher investment level 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 higher investment exceeds any reduced fixed costs associated with outsourcing.

A related study by Aghion and Tirole (1994) employs the incomplete contracting framework to examine the organization of innovation. In their analysis, the firm's key choices are R&D efforts and financing. Their results suggest that when R&D efforts are more important, 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 consistent with Aghion and Tirole's theory. We further discuss the relationship of Aghion and Tirole (1994) with our theoretical and empirical analysis in Section 6.

This paper also contributes to the literature on the role of patents in market economies (see Hall et al. (2014) for a survey), focusing on a variety of topics including the choice of patenting versus secrecy, optimal design of patent systems, and the role of patent pools. We show that patenting also contributes to understanding the organization of firms—in particular, whether R&D is conducted in-house or outsourced. Because owning a patent is valuable, protecting the value of that ownership can be important. Therefore, the possibility of cannibalization of products currently owned may lead a firm with valuable patents to conduct R&D in-house due to the additional control gained.

We also contribute to the mostly empirical literature that relates vertical integration with product market competition. 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. Hortaçsu and Syverson (2007) employ data from cement and ready-mixed concrete plants to study a related issue which is whether vertical integration is used for foreclosure and to increase market power. They find that instead vertical integration primarily lowers costs and leads to lower prices. More recently, Galdon-Sanchez et al. (2015) concerning services and Gil and Ruzzier (2018) focusing on the Spanish television industry find a negative relationship between vertical integration and product market competition. We develop a theory that predicts a positive relationship between competition and the frequency of outsourcing. This positive relationship is due to reduced incentives for outsourcing when the originator owns an existing patented product in the same product class as the product under development. We further provide empirical testing using pharmaceutical data that supports the predicted relationship.

Another literature this paper contributes to is the large literature that employs Salop's circle model, introduced in Salop (1979), to investigate a variety of issues. It is well known in this literature that a monopolist selling multiple products maximizes profits by locating those products equidistant from each other along the circle. We employ this result in the two-product case to investigate the outsourcing decision concerning development investments when there is uncertainty concerning the new product's location in product space.[5] This result is similar to our finding (in an extension of our main model) that the incentive to avoid cannibalization increases with the firm's current market share. Note that, relative to this literature, what is novel about our paper is our focus on how the cannibalization issue influences the choice of how development is conducted.

There are also a number of papers that contribute to our understanding of the cannibalization issue. For example, Petrin (2002) shows that consumer welfare is helped by competition across firms concerning the introduction of new products where such entry serves to cannibalize the profits of other firms. A paper closer to ours is Igami (2017). That paper estimates a dynamic oligopoly model using data from the hard disk drive industry and shows that, relative to entrants, incumbents innovate less because of the costs of cannibalizing their own products.

Similar to the focus of our paper, several papers analyze the ability of firms to limit cannibalization when introducing a new product. Moorthy and Png (1992) show that, when consumers

[5] A similar logic concerning the need to optimally locate new products can arise when brand proliferation by a monopolist is used to deter entry — see Schmalensee (1978) for a related analysis and discussion. We abstract away from the entry deterrence issue since incorporating it would complicate the analysis without changing the main testable predictions we focus on.

are impatient, 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 price of a higher quality product. In Siebert (2015) a firm's optimal strategy in entering a new market in a duopoly setting with vertical differentiation is to introduce a single product. This result arises when avoiding cannibalization is more important than price discrimination. The literature on planned obsolescence in which renting is used to avoid time inconsistency concerning new product introductions such as Waldman (1993, 1996), Choi (1994) and Nahm (2004) is also closely related. We contribute to this literature by showing how the desire to limit cannibalization can affect the internal organization of the firm.

Finally, the paper contributes to an existing literature concerning how R&D is organized and the role of such organizational decisions in the pharmaceutical industry. Nicholson et al. (2005) show that biotech companies send positive signals to investors by forming alliances with larger pharmaceutical firms, while Danzon et al. (2005) find that success rates of complex phase II and phase III trials are higher for products developed in an alliance. Azoulay (2004) finds that pharmaceutical firms are more likely to outsource data-intensive clinical trials, while knowledge-intensive trials are typically conducted in-house. Also, as discussed above, Lerner and Merges (1998) study control rights in biotechnology alliances and find that control rights are more likely to be assigned to the R&D firm when the firm has superior financial resources. We are the first to offer a patent protection perspective on the choice of pharmaceutical alliance decisions at the project level.

2 Model and Preliminary Analysis

In this section, we present a multi-period model of a firm's decision to conduct R&D either in-house or through outsourcing, and provide preliminary results regarding both scenarios.

2.1 The Model

There is a single risk neutral firm—the originator—that owns an existing patented product and has decided to develop a new product in the same product class. The originator exhibits a constant marginal cost c_1 for producing a unit of its existing patented product. 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 generic producers, there is pool of $N \geq 2$ identical risk-neutral licensees. Licensees command 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

which is a random draw from the probability density function $f(\cdot)$ with support $[F_{min}, \infty)$, while a licensee incurs 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., $\Delta = F - F_L$. We further assume economies of scope between developing and producing the new 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 $c_2^+ > c_2$.[6]

We assume that there are T total periods, $T \geq 3$, with the new product being developed in period two, where sales begin in period three, and the patent for the new product lasting 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 equilibrium behavior changes as a function of t_E.[7]

If the originator chooses in-house development, it develops the new product, chooses the new product price each period, and produces and sells the new product each period. If the originator chooses to outsource development, on the other hand, then the originator and the licensee enter into a contract. This 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 (potentially negative) payment each period from the originator to the licensee which may depend on that period's new product quantity.[8]

We also assume the contract is renegotiation-proof, i.e., after the contract is signed, there is no alternative contract in later periods both parties agree to that induces a Pareto-improvement (makes both parties better off and at least one strictly).[9] Furthermore, actions and outcomes associated with the development process itself are assumed to be non-verifiable and thus non-contractible. This means that, in case the originator chooses in-house development, then it makes all 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.[10]

[6] We ignore the possibility that a licensee could use its knowledge obtained in the development process to become a rival in the output market in a future period. See Novak and Stern (2009) for a related empirical analysis.

[7] In order to simplify exposition, we disregard discounting.

[8] Implicitly, we are assuming that the new product price is non-verifiable and thus not contractible. We further assume that the payment in period t does not depend on the product quantity in different periods. Neither assumption is essential for our main qualitative results, but they serve to simplify the description of equilibrium behavior.

[9] Focusing on renegotiation-proof contracts is a standard approach taken by many contracting papers. For early discussions focusing on how the possibility of renegotiation affects equilibrium contracting, see, for example, Dewatripont (1988), Hart and Moore (1988), and Demougin (1989). Assuming contracts are renegotiation proof does not affect equilibrium behavior concerning final outcomes such as prices charged to consumers, investment levels in location precision (see below), and in-house versus outsourcing decisions. Without this assumption, however, there are multiple equilibria due to the possibility of equilibrium renegotiations that occur after an initial outsourcing contract is signed.

[10] See Casas-Arce et al. (2019) for a related analysis in which details of the development process are left incomplete in a setting in which renegotiation is possible. These authors point out that this approach is consistent with

Following Salop (1979), the product space is characterized by a unit circle, in which the location of the new product on the unit circle depends on non-contractible development decisions. That is, the firm developing the new product (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 mean 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

$$l = l_M + \varepsilon, \tag{1}$$

where ε is a random draw from one of the following two uniform distributions:

$$U[-\alpha, \alpha] \text{ and } U[-\beta, \beta], \quad \alpha < \beta < \frac{1}{2} - (c_2 - c_1),^{11}$$

where α and β are model parameters and the investment level, denoted k, determines the probability that the random draw is from the uniform distribution with the smaller range. To be precise, $p(k)$ denotes the probability that ε is drawn from $U[-\alpha, \alpha]$ given the investment level equals k. We assume that $p(\cdot)$ is continuously differentiable, $p(0) = 0$, $p'(0) = \infty$, $p'(k) > 0$ and $p''(k) < 0$ for all $k \geq 0$, and $p(\infty) < 1$.[12]

Due to the importance of development decisions being non-contractible in our model, it is worthwhile providing some additional discussion concerning this assumption. Implicitly, our assumption implies it is easier for a firm to ensure that the employees whose decisions affect the new product location provide efficient effort when development is conducted in-house rather than outsourced. When development is in-house, the originator has direct control over compensation and other incentives (such as promotion decisions) and can use those to elicit optimal effort levels. On the contrary, when development is outsourced, the originator does not have this type of direct control over employee incentives, and it is difficult to write a contract at a level of detail such that the outside firm elicits efficient effort from the relevant employees. Note that, this basic argu-

commonly observed contracting practices in R&D intensive industries. Also, see Lerner and Malmendier (2010) for an investigation of contract design in the biotechnology industry when R&D is non-contractible.

[11] The assumption $\beta < \frac{1}{2} - (c_2 - c_1)$ is imposed to ensure that some consumers on both 'sides' of the new product's location purchase the new product in every period t, $2 < t \leq T$.

[12] Uncertainty concerning a new product's location in product space is one of a number of important aspects of uncertainty concerning real-world R&D investments. We abstract away from most other types of uncertainty in our modeling. We do this in order to make the relationship between uncertainty involving the new product's location in product space, cannibalization of existing patented products, and the outsourcing decision more transparent.

ment is consistent with the property rights approach and also various descriptions of outsourcing in real-world markets (Lam, 2004; Patel, 2017).

On the demand side, there is a unit mass of consumers 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. To be precise, a consumer's marginal valuation of the qth unit of a product is given by

$$V(q) = V^+ - vq, \tag{2}$$

where V^+ is a constant which represents the marginal valuation of the initial unit. Note that this valuation function is characterized by decreasing marginal utility of consumption. A consumer also faces a travel cost for consuming a product not at the consumer's exact location in product space. The travel cost a consumer incurs from consuming a unit located a distance s from the consumer's location in 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 buy 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 optimal quantity choices.

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 is realized and publicly revealed.[13] In the first period, the originator also decides whether to develop the new product in-house or outsource development to a licensee. If the originator chooses to outsource, 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 development contract to the originator and the originator chooses a licensee.[14]

At the beginning of the second period, the originator chooses a second period price for the existing patented product and consumers again make purchase decisions. If the patent has not expired, then the originator sets the monopoly price, while if it has expired then competition with

[13]The assumption that the realization of F is publicly revealed is not essential for our results.

[14]The simultaneous take-it or leave-it development contract offers are the equivalent of giving all the bargaining power to the originator, i.e., when outsourcing occurs the licensee earns zero expected profits and the expected surplus all goes to the originator.

generic producers means the price equals marginal cost c_1. The developer (originator or licensee) also chooses a mean value for the new product's location in product space, l_M, and an investment level in location precision, k. These choices are private information of the developer. After these choices have been made, the noise term is drawn from the respective distribution. Thus, by the end of the second period the new product's location in product space is determined. We assume this location to be publicly observable but not verifiable by the courts.

In the third period, the new product is brought to the market. In case the patent on the existing product expires before the third period, then the price for the existing product is at marginal cost due to Bertrand competition among generic producers, and the firm with control rights for the pricing of the new product takes this price as given when choosing a price for the new product.[15] 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. On the other hand, if the patent on the existing patented product has not expired and the licensee has control rights, then the two prices are determined by Bertrand competition between the two firms, given product differentiation due to the different locations in product space. In the following periods, if any, prices are determined using the same rules as in the third period. Also, our focus throughout is Subgame Perfect Nash equilibrium.

Finally, there are three aspects of our model worth noting. First, in this paper we assume that R&D investments associated with developing the new product are successful with certainty. This is clearly a simplification, since in real-world settings R&D investments frequently fail. Introducing a probability of failure, however, would complicate the model without changing the qualitative nature of equilibrium. Second, an implicit assumption, which is important for our analysis, is that the originator cannot sell the patent for the original product to a licensee. One reason such a strategy might not be feasible is the existence of private information on the part of the originator concerning the value of related products, and the resulting adverse selection problem, as in Akerlof (1970), which prohibits the possibility of trade.[16] Third, in our main model just described, the originator is a monopolist in the therapeutic class. In an extension of the model found in Appendix B and described in subsection 3.2, we explore how results change when we move away from the monopoly case. We employ this extension to derive one of our testable predictions.

[15] When assuming a different mode of competition, such as, for example, Cournot competition, the existing product's price would be above marginal cost. This, however, would not materially affect our qualitative findings concerning the originator's outsourcing decision.

[16] Allowing the originator to sell the patent for the new product would not change the analysis since there is never a benefit in doing so, given we assume the patent for the existing patented product cannot be sold.

2.2 Preliminary Results

We start with results concerning the nature of the equilibrium contract when the originator chooses to outsource. As captured in Lemma 1 below, 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.[17]

Lemma 1 *Consider an equilibrium to the game in which the originator chose to outsource development of the new product. Then, the contract between the originator and the licensee satisfies i) through iii).*

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_2.

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 licensee sum to the fixed amount that results in zero expected profits for the licensee.

All formal proofs can be found in Appendices A and B. The logic for part i) is as follows. First, the licensee is assigned production of the new product, because of economies of scope between development and production. Second, in any period prior to patent expiration of the existing patented product, the joint profits of the originator and licensee are maximized by giving sales and control rights over pricing to the originator, so that it can choose prices that maximize the joint profits of the two products. Given the contract is renegotiation-proof, this means that sales and control rights for the pricing decision must be assigned to the originator. Also, having the payment from the originator to the licensee be a fixed amount plus the number of new units sold multiplied by c_2 means higher joint profits, because in choosing prices the originator will internalize all the returns associated with the pricing decisions.

[17] In the pharmaceutical industry, for example, development control rights such as the design of drug trials are often allocated to the licensee. In addition, contracts commonly assign marketing and commercialization rights to licensees with robust financial resources (Lerner and Merges, 1998). A 2024 report by investment bank Harris Williams argues that this form of outsourcing is generally more frequent after the expiration of patents (https://www.harriswilliams.com/our-insights/hcls-outsourced-pharma-services-new-data-shows-resurgence-is-underway). Our theory about control and self-cannibalization provides one potential explanation for these observations.

Now consider part ii) of the lemma. 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. 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 from competition among licensees.

The next step is to consider decisions concerning new product location. Let $L(j, t_E)$ denote the mean distance between the new product and the existing patented product 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 parameters fixed, if the in-house versus outsource decsion is taken as given, then i) through v) describe* $L(j, t_E)$ *and* $K(j, t_E)$.

i) $L(I, t_E) = L(O, t_E) = \frac{1}{2}$ *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. 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 $L(I, t_E) = \frac{1}{2}$ 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 after patent expiration the originator does not benefit from the existing product. So the investment level increases the later is patent expiration of the existing patented product, i.e.,

$$K(I, T) > K(I, T-1) > ... > K(I, 2) = K(I, 1). \tag{3}$$

Suppose the originator opts for outsourcing and $t_E = T$. In this case, in each period $t > 2$, joint surplus is maximized if the originator receives the returns associated with new product sales and has control rights over the pricing decision. As a consequence, that is what is specified in a renegotiation-proof contract. In turn, this means that the licensee has no incentive to invest in location precision, so $L(I, T) = L(O, T) = \frac{1}{2}$ and $K(I, T) > K(O, T) = 0$. Note that the mean location specified in the contract is still $\frac{1}{2}$ as it maximizes 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.,

$$K(O,1) = K(O,2) > K(O,3) > ... > K(O,T) = 0. \tag{4}$$

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. Lastly, similar to the other cases, $L(O,t_E) = \frac{1}{2}$ given $2 < t_E < T$.

3 Equilibrium Analysis and Testable Predictions

In this section we provide a characterization of the in-house versus outsourcing decision and thereafter present three testable predictions.

3.1 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 the reduced fixed cost of development by the amount Δ.[18] The cost of outsourcing, on the other hand, is that—as shown in the previous section—the expected investment in location precision is suboptimal if $t_E > 2$. 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.[19]

[18] As a reference point, consulting firm Weaver, for instance, reports that outsourcing of R&D in the pharmaceutical industry can lead to 25-30% reduction in quality-related investigation costs, 30-35% reduction in compliance-related capital expenditure, and additionally significant decreases in quality control staffing and training expenses (https://weaver.com/resources/outsourcing-in-pharmaceutical-manufacturing-enhancing-quality-while-managing-risk/).

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

Proposition 1 *Holding all other parameters fixed, all equilibria are characterized by a value* Δ^* *such that the originator chooses in-house development when* $\Delta \leq \Delta^*$ *and chooses outsourcing otherwise, 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, equilibria exist and all satisfy results in Lemmas 1 and 2, where the equilibrium contract given* $\Delta > \Delta^*$ *is unique up to the timing of the payments described in iii) of 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 section, when this is the case there is no disadvantage in terms of investments in location precision from choosing outsourcing. This means the investment in location precision is independent of whether the originator chooses in-house development or outsources. As a consequence, 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 $t_E > 2$. From the analysis in the previous section, we know that for these parameterizations, a licensee's incentive to invest in location precision falls short of the originator's given in-house development. Given this advantage associated with in-house development, the originator only chooses outsourcing if there is a sufficiently large reduction in the fixed cost of development associated with outsourcing, i.e., $\Delta^* > 0$ if $t_E > 2$.

Finally, consider two values for t_E, t' and $t' + 1$, where $t' \geq 2$. From the analysis in the previous section, we know that the investment in location precision increases with t_E given in-house development, while it decreases with t_E given outsourcing. In other words, the expected loss due to a lower investment in location precision when the originator chooses outsourcing is higher when $t_E = t' + 1$. It follows that 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 in t_E.

3.2 Testable predictions

We now discuss testable predictions of our theory. The first three testable predictions follow immediately from results stated in Proposition 1.

Testable Prediction 1 *A firm developing a new product has a lower probability of choosing outsourcing 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 follows from Proposition 1, i.e., $\Delta^* = 0$ given $t_E = 1$ or 2 and $\Delta^* > 0$ for all $t_E > 2$. Recall that Δ^* determines the probability that outsourcing is chosen, with a higher value for Δ^* translating into a lower probability that the choice is to outsource. 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. If $t_E > 2$, however, 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 meaningful cost advantage.

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 lower is the probability the firm chooses to outsource.*

Testable Prediction 2 is the Proposition 1 result that Δ^* increases with an increase in t_E. As before, Δ^* determines the probability of outsourcing with a higher value for Δ^* reflecting a lower probability. The proposition states that an increase in t_E increases Δ^*, with the underlying logic being that an increase in t_E raises investment incentives given in-house development, but does not given outsourcing. Thus, 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 increases, Δ^* rises which translates into a lower probability of outsourcing.

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 expected distance in product space between the new product's realized location and the optimal location will be smaller if development is conducted in-house rather than outsourced.*

Testable Prediction 3 follows from results in Lemma 2. This lemma states that, given the originator owns an existing patented product and the new product is expected to reach the market before this patent expires, the investment in location precision is higher when development is conducted in-house. In turn, this immediately tells us that the new product's expected location in product space is closer to the optimal location than is the case with outsourcing.

The fourth and final prediction concerns market share. In our base model, the originator is a monopolist in the product class. Suppose that, however, 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 avoiding cannibalization of its own patented products would be high. In Appendix B we formalize this logic by extending our formal analysis to a setting in which the product space is an equilateral triangle and there is a rival that sells a competing product in the same product class as the originator's products. We formally show that, because of investment incentives concerning product location, the probability the originator chooses outsourcing decreases with the originator's market share at the date of the new product introduction (see Proposition 2 in Appendix B).

Below we translate this result into a testable prediction.

Testable Prediction 4 *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 outsourcing will be lower 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. Thus, when that market share is higher, we expect a lower probability of outsourced development, since outsourcing decreases investments in location precision.

There are two additional points concerning testable predictions worth noting. First, in the model, there is a single in-house versus outsource development decision for each product. In the data, in contrast, products are associated with multiple development decisions, each of which can be conducted either in-house or outsourced. With this in mind, in Appendix B we extend the model to allow for multiple R&D investments for a single product and show that the testable predictions hold for each development decision in this extension of our analysis.

Second, in the model we assumed that licensees have the same capabilities concerning producing the product and determining the optimal price as the originator. We made this modeling choice to highlight that our basic argument does not depend on the two firms having different capabilities. In

real world market of the sort we are modeling, however, licensees may be specialized in development in which case contracts would never assign production or pricing decisions to the licensees. In terms of testable predictions, we thus focus on predictions of our approach that are valid whether or not licensees are specialized in development.[20]

4 Data from the Pharmaceutical Industry

The pharmaceutical industry is an excellent setting in which to test 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. Second, the industry heavily relies on patents. This allows us to create measures of the cost of cannibalization based on the number and expiration dates of existing patents owned by the originator that are in the same therapeutic class as the drug under development. Third, uncertainty in the development process implies that the pharmaceutical industry is a prime example of an environment in which firms cannot perfectly control the exact location of new products in the product space.[21] 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 to test our theory.[22]

4.1 Main Data Source and Descriptive Statistics

Our principal data source is the Pharmaprojects dataset. This dataset was assembled by the company Informa and contains information concerning the development of new pharmaceutical projects throughout the world (PharmaProjects, 1989-2004). The dataset covers information for the time period 1989 through 2004, for drug compounds that were originated by companies that were publicly traded at some point between 1989 and 2004.[23] For each chemical compound under development, the dataset contains the name of its originator, the therapeutic class, active ingredi-

[20] We can show that our four testable predictions all hold even if licensees are specialized in development and thus never produce or choose prices.

[21] There are differences between the nature of uncertainty in our theoretical model and the nature of uncertainty concerning location in product space in the pharmaceutical industry. For example, in our model product space is continuous, while in the pharmaceutical industry product space can be thought of as discrete where a location concerns the condition or conditions a drug is approved to treat, and the types of patients within a condition (e.g., treatment-naive patients or patients who have already failed another treatment). We chose to develop our theoretical model and predictions employing a continuous product space because this better matches prior literature focused on analyses related to location in product space. But most of our theoretical predictions could be derived in a model in which product space is discrete.

[22] See Lakdawalla (2018) for a survey of the literature on the pharmaceutical industry focused primarily on innovation, pricing, and marketing decisions.

[23] Publicly traded firms, which typically have significant experience conducting clinical trials, account for a majority of compounds under development, making the decision to outsource relevant (i.e., they have the capability to run a phase II or III trial if they choose).

ents, patent number and its filing date, if any, whether the development was outsourced and, if so, the names of these outside firms, and the beginning and end dates of development contracts and stages.

A key issue for our empirical analysis is to define whether development decisions are in-house versus outsourced. In particular, whether a development decision is in-house or outsourced depends on whether the originator ever signs a development contract and, if it did, the stage of the development process at which the earliest development contract was signed.[24] If there was never a development contract, then all decisions are categorized as in-house. On the other hand, if there was a development contract at some point sufficiently early in the development process, we categorize decisions made prior to the contract as in-house, and those made after the first development contract as outsourced. This is consistent with the finding in Lerner and Merges (1998) that in the pharmaceutical industry, control rights concerning development decisions are assigned to the development firm when a contract is signed.[25]

Table C.1 in Appendix C 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 typically conducted on a relatively small scale, with the former recruiting around 20 to 80 subjects and the latter between a few dozen and about 300 subjects. Phase III continues testing on safety and effectiveness by employing a much larger sample, usually ranging from several hundred to $3,000$, and following the patients for a longer time period than in phase II. Conceivably, a developer could affect a new drug's location in product space during phase III by recruiting specific patient population groups and testing for specific efficacy measures and side effects. However, in most cases, once phase II is completed and the FDA meets with the developer to discuss plans for phase III, it is 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, drug-compound years that are before the first development contract are classified as in-house observations. If the first development contract is prior to the end of phase II, then drug-compound years are classified as outsourced starting with the year of the first development contract. If the first development contract is after the end of phase II, then all observations are classified as in-house.[26]

We construct two measures of whether the originator of a new product owns existing patented

[24] In addition to the information contained in the Pharmaprojects dataset, we compiled a list of mergers and acquisitions and assigned compounds to the correct originating firms to avoid problems related to misclassification.

[25] They find this is the case as long as the development firm has sufficient financial resources.

[26] As reported in Subsection 5.5, we also provide a robustness test that shows that results are qualitatively unchanged if the end of phase I rather than the end of phase II is used to define in-house versus outsourced observations.

products in the same product class.[27] The first measure, $EOP1$, 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. The second measure, $EOP2$, equals the number of other patented products owned by the originator that are in the same product class.

We also construct patent length variables. The first patent length variable, $LOP1$, is the remaining length of the patent with the largest remaining length of all the patents owned by the same firm in the same therapeutic 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 therapeutic class owned by the same firm as the drug under development.

Some of our tests also include two control variables suggested by the analysis in Danzon et al. (2005). First, we define experience, denoted Z, as a firm's cumulative experience in developing drugs within a therapeutic area, measured as the firm's cumulative compound-year observations developing drugs in a therapeutic class by a given year (including both in-house and outsource years) during our sample period. Including this variable in our tests allows us to control for the possibility that learning-by-doing reduces the cost of developing a new drug in a therapeutic class in which the originator has prior development experience. Second, we construct a Herfindahl-Hirschman Index (HHI), denoted W, for each originator'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 HHI, the more concentrated is the firm's development portfolio in terms of therapeutic class. Including this variable controls for economies of scope which is an additional cost-side reason that an originator might choose in-house development.

[27] We crosswalked the raw therapeutic classes from the Pharmaprojects dataset to the widely-used Anatomical Therapeutic Chemical (ATC) system at the first and second levels. This standardization enhances the characterization of drug therapeutic categories and facilitates easier data merging with IMS data in subsequent analyses. Table C.2 reports the main (first level) ATC distribution in our sample, both at the panel and compound levels.

Table 1: Definition of Constructed Variables

Variable	Description
In-house	Indicator equals 1 if compound is never contracted out by the originating firm or if its earliest Development Contract 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
Length of Patents	
LOP1	Length of the longest patent among compounds in the same therapeutic class and same firm
LOP2	Sum of the patent lengths among compounds in the same therapeutic class and same firm
Other	
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

We also construct two variables related to market-level variability in a firm's incentive to avoid cannibalization. We employ these variables primarily in our tests of prediction 4, which is based on our model extension that introduces rivalry into the analysis. 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 are owned by firms other than the originator of the drug under development. The second variable, 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 including these two variables is 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. There are two main reasons this could be the case. First, if cannibalization is mostly in terms of other firms' products, then being close to the optimal location in product space is less important (this is related to testable prediction 4). Second, increased competition might lower prices in the therapeutic class again decreasing the incentive for the originator to locate the product near the optimal location.[28] Table 1 provides a full list of our constructed variables along

[28]There also may be other reasons that the level of competition affects the choice of in-house versus outsourced

with their definitions.

Table 2: Descriptive Statistics on In-house Development and Patent Profile

Number of compounds					11,493
Number of firms					532
Years covered					1989-2004
	Outsourced Compounds Mean (SD)	In-house Compounds Mean (SD)	Overall Mean (SD)	Min	Max
	Level of Observation: compound-year (109,115)				
In-house	0	1	0.785	0	1
	(0)	(0)	(0.411)		
Existence of Patents					
EOP1	0.586	0.766	0.727	0	1
	(0.493)	(0.424)	(0.445)		
EOP2	4.370	11.030	9.597	0	64
	(8.550)	(14.007)	(13.312)		
Length of Patents					
LOP1	10.490	12.967	12.434	0	20
	(7.881)	(7.223)	(7.440)		
LOP2	56.827	135.479	118.553	0	884
	(103.235)	(169.847)	(161.177)		

Table 2 reports descriptive statistics for the development decision and patent data. The sample used in the main analysis contains 11,493 compounds originated between 1989 and 2004.[29] At the compound-year level, 78.5 percent of the observations are classified as in-house development. The data for the first patent existence measure, $EOP1$, show that in 76.6 percent of the compound-year observations characterized by in-house development and 58.6 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 ($EOP2$). The same patterns hold for measures of the length of patents ($LOP1$ and $LOP2$).

4.2 Secondary Data and Descriptive Statistics

To supplement the analysis, we also use the IMS dataset to create a market share measure based on drug sales (IMS, 1992-2004). The IMS dataset includes a list of drugs and their annual sales in the US between 1992 and 2004. We merge the sales data into the principal dataset from

development that are independent of cannibalization concerns.

[29] Table C.3 in Appendix C further reports in-house and outsource status by firm type. About 95% of originators in our sample are biopharmaceutical companies, and 78.8% of their projects are in-house (similar to the share in the full sample).

Pharmaprojects based on the therapeutic class, firm name, and whether the drug is branded. We calculate for each compound under development in the principal dataset the market shares for the originator's other patented compounds in the same class in each year. This variable, referred to as MSP (unit: %), allows us to test how market share based on branded drug sales data affects a firm's incentive to choose in-house development.[30] Table 3 provides descriptive statistics for the control variables at the compound-year panel level.

Table 3: Descriptive Statistics on the Control Variables (1989-2004)

Variable	**Obs.**	**Mean**	**Std. Dev.**	**Min.**	**Max.**
Pre-clinical	109,115	0.792	0.406	0	1
Phase I	109,115	0.061	0.239	0	1
Phase II	109,115	0.082	0.275	0	1
Phase III	109,115	0.044	0.204	0	1
Launched	109,115	0.021	0.145	0	1
Experience	109,115	191.436	304.518	1	1,974
PDM	109,115	26.637	12.313	0	42
Scope	109,115	0.097	0.119	0.011	1
TDM	109,115	51.376	24.818	2	81
MSP (with imputations)	109,115	3.952	8.311	0	85.326
MSP (without imputations)	51,439	8.383	10.457	0	85.326

5 Empirical Tests

In this section, we empirically test the theoretical predictions derived in Section 3. We start with the two predictions concerning the role of other patents owned by the originator, then consider the predictions involving pivots and market share. Robustness tests are presented thereafter.

5.1 Patent Existence

To investigate whether the dataset is consistent with the first testable prediction which concerns patent existence, we estimate the following Logit specification:

$$Prob(Y_{ijkt} = 1) = \Lambda\left(\alpha_0 + \alpha_1 EOP_{ijkt} + \alpha_2 X_{ijt} + \alpha_3 Z_{jkt} + \alpha_4 W_{jt} + C_k + T_t\right). \quad (5)$$

$\Lambda(\cdot)$ 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

[30] The IMS data classifies drugs using the Anatomical Therapeutic Chemical (ATC) Classification System. We employ two versions of MSP in the analysis: the first includes the full sample, with missing MSP imputed as zero, as firms with no sales in a therapeutic class are naturally recorded as missing; the second uses only the merged sample with positive sales and without imputation, acknowledging that IMS data do not contain the universe of drug sales.

a patent existence variable, where in some specifications it is the indicator measure $EOP1$, while in other tests it is the indicator measure $EOP2$. X_{ijt} is a vector of development phase indicator variables. For example, the indicator variable for phase I takes on a value of 1(0) when compound i is (is not) in phase I tests in year t. In the analysis, years during the pre-clinical testing phase are the omitted base group, and controls for years during phase I, phase II, phase III trials, and launched are included. Z_{jkt} measures the originator's cumulative experience (compound-years) in the therapeutic class corresponding to the compound of interest, and W_{jt} is the originator's therapeutic scope, both as defined in Table 1 (experience and scope variables).

Equation (5) also includes therapeutic class fixed effects, C_k, to control for unobserved class characteristics that affect both patent existence and development integration decisions. Year fixed effects, T_t, control for cross-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., there is a predicted positive correlation between owning an existing patented product in the same product class as the product under development and choosing in-house development.[31]

Table 4 reports the regression results. Each patent existence measure is estimated under two different specifications. In the first specification therapeutic experience and therapeutic scope are omitted, whereas they are included in the second. All regressions employ robust standard errors clustered at the compound level to account for heteroskedasticity and potential correlation concerning the in-house versus outsource decision for a particular compound across observations.

The first two columns report results for $EOP1$. The coefficient on $EOP1$ is positive and statistically significant at the one-percent level in each regression. The coefficient on the experience variable is also 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 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 increasing costs which makes in-house development less likely (recall that our scope variable is such that a higher value means a less diversified portfolio of projects).

Columns 3 and 4 consider the same set of tests focusing on our patent count variable. The results are similar to those in columns 1 and 2. The only difference worth noting is that in the column 4 specification the coefficient on the experience variable is negative and not statistically significant.

The baseline probability that the development of a drug will be observed to be outsourced by

[31] We do not include firm fixed effects in our main specification because many firms have very few products under development in a therapeutic class, and including firm fixed effects would eliminate some projects and firms that are disproportionately small and young. Nevertheless, we include firm fixed effects in a robustness check (Table C.10).

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

	(1) In-house	(2) In-house	(3) In-house	(4) In-house
EOP1	0.899***	0.505***		
	(0.0550)	(0.0600)		
EOP2			0.0632***	0.0595***
			(0.00430)	(0.00531)
Phase I	-0.649***	-0.674***	-0.622***	-0.640***
	(0.0799)	(0.0801)	(0.0796)	(0.0799)
Phase II	-1.031***	-1.111***	-1.017***	-1.034***
	(0.0692)	(0.0704)	(0.0711)	(0.0712)
Phase III	-1.425***	-1.542***	-1.436***	-1.462***
	(0.0906)	(0.0931)	(0.0943)	(0.0936)
Launched	-2.018***	-2.174***	-1.996***	-2.021***
	(0.126)	(0.129)	(0.127)	(0.126)
Experience		0.00151***		-0.000167
		(0.000142)		(0.000132)
Scope		-1.467***		-1.360***
		(0.150)		(0.148)
Constant	0.567***	1.353***	0.929***	1.400***
	(0.215)	(0.222)	(0.206)	(0.216)
Observations	109,115	109,115	109,115	109,115

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 indicates 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. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Standard errors (in parentheses) are heteroskedasticity-robust and clustered at the compound level.

the originator in a year prior to phase III is 21.5 percent. Employing the coefficient on $EOP1$ in column 2, the probability of contracting prior to phase III is 6.5 percentage points lower 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. Also, calculations employing analogous coefficients from other columns yield similar results.

5.2 Patent Length

To investigate whether the length of patents owned by originators in the same product class as the drug under development is positively correlated with the originator's decision to develop the product in-house, we estimate the following Logit specification:

$$Prob(Y_{ijkt} = 1) = \Lambda\left(B_0 + B_1 LOP_{ijkt} + B_2 X_{ijt} + B_3 Z_{jkt} + B_4 W_{jt} + C_k + T_t\right). \tag{6}$$

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

	(1) In-house	(2) In-house	(3) In-house	(4) In-house
LOP1	0.0518***	0.0271***		
	(0.00333)	(0.00356)		
LOP2			0.00525***	0.00407***
			(0.000353)	(0.000334)
Phase I	-0.657***	-0.680***	-0.631***	-0.652***
	(0.0792)	(0.0797)	(0.0798)	(0.0800)
Phase II	-1.043***	-1.120***	-1.022***	-1.058***
	(0.0693)	(0.0706)	(0.0711)	(0.0713)
Phase III	-1.466***	-1.569***	-1.452***	-1.501***
	(0.0902)	(0.0930)	(0.0942)	(0.0942)
Launched	-2.009***	-2.174***	-1.995***	-2.065***
	(0.123)	(0.128)	(0.125)	(0.126)
Experience		0.00157***		0.000439***
		(0.000144)		(8.58e-05)
Scope		-1.602***		-1.333***
		(0.149)		(0.149)
Constant	0.429**	1.341***	0.821***	1.344***
	(0.215)	(0.223)	(0.207)	(0.217)
Observations	109,115	109,115	109,115	109,115

Note: Dependent variable is one if a compound is developed in-house, and zero otherwise. All specifications include a full set of therapeutic categories and year indicators. LOP1 is the length of the longest patent among compounds in the same therapeutic class and same firm. LOP2 is the sum of the patent lengths among compounds in the same therapeutic class and same firm. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Standard errors (in parentheses) are heteroskedasticity-robust and clustered at the compound level.

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 specifications it is the sum of the patent lengths of the drugs the originator owns in the same class as the drug under development, $LOP2$. The control variables for development phase X_{ijt}, therapeutic experience Z_{jkt}, scope W_{jt} and fixed effects for therapeutic class C_k and year T_t are the same as in Equation (5). From Testable Prediction 2, we expect B_1 to be positive, i.e., an originator with longer aggregate patent life for drugs it owns in the same therapeutic class as the drug under development should be more likely to choose in-house development.

Table 5 reproduces the regressions 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, and statistically significant

at the one-percent level in each regression. Second, 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 $LOP2$ exhibit similar patterns, both in the sign of coefficients and the statistical significance levels.

Consider the coefficient in column 2 on $LOP1$. This coefficient tells us that a one standard deviation increase in the length of the longest patent held by the originator in the same product class as the product under development is associated with a 3.6 percentage point increase in the probability a compound will be developed in-house in a year prior to phase III relative to the base level. We can also conduct similar exercises using results from other columns. For example, employing the coefficient on $LOP2$ in column 4 yields that an increase in the sum of patent lengths of patented drugs in the same product class owned by the originators of one standard deviation is associated with an increase in the probability of in-house development by 13.1 percentage points relative to the base level.

5.3 Development Pivots

This subsection considers tests related to Testable Prediction 3 which remember is the following. Suppose the originator owns an existing patented product in the same product class as the product being developed and the new product is expected to reach the market before the patent expires. In such a case the new product's realized location in product space will, on average, be closer to the optimal location if development is conducted in-house.

In terms of the pharmaceutical industry, we interpret this prediction as applying to the therapeutic classes a compound is tested on during the development process that are in addition to the initial or primary class. We refer to this number as the number of development pivots. In other words, our interpretation of the prediction is that the number of development pivots, given the relevant conditions are satisfied, should be smaller when development is in-house. The idea here is that the optimal location is the initial or primary class, and development pivots represent movements away from the optimal location.

Our exact methodology is to define the number of development pivots as the total number of fine therapeutic classes a compound is tested on minus one, i.e., minus the initial or primary class. We also construct for each compound an indicator for having existing patented drugs in the same product class as the initial or primary class, and for each compound-year observation, an interaction term of this patent indicator with in-house status.[32] We ran OLS regressions where

[32]Throughout the paper patent existence and patent length are defined in terms of first-level ATC. Pivots are

the dependent variable is the number of development pivots, while the independent variables are the patent variable, the interaction term, and observable characteristics. The testable prediction is that the coefficient on the interaction term should be negative.

Table 6 reports results. In column 1 the independent variables are the in-house variable, the patent variable, and the interaction term. In column 2 we add therapeutic class fixed effects (defined in terms of the initial or primary therapeutic class), while in column 3 we further add firm type fixed effects. The main finding is that, consistent with the prediction, the coefficient on the interaction term is negative and statistically significant at the five percent level in each regression.

We also find that the coefficient on the in-house variable is negative and statistically significant at the one percent level in all three regressions, while the coefficient on the patent variable is negative and statistically significant at the five percent level in all three regressions.[33] The result concerning the in-house variable is consistent with our theoretical approach, although not predicted by the specific model we investigate. That is, because we assume consumers are uniformly distributed around the circle, if the originator does not own an existing patented product in the same product class as that of the product being developed, then all locations are equally desirable and so in-house development and outsourced development result in the same investment in location precision, i.e., zero. This theoretical result is inconsistent with the negative and statistically significant coefficients on the in-house variable in Table 6.

However, if in the theoretical model we moved away from the uniform distribution of consumers, then the model would be consistent with the negative and statistically significant coefficients on the in-house variable. The reason is that the originator would have an incentive to locate the new product where consumer density is higher, and this means the expected distance between the realized location and the optimal location would be smaller with in-house development.

Why there is a statistically significant positive coefficient on the patent variable in each regression seems less clear cut. That is, the positive and statistically positive coefficients on the patent variable is puzzling to us and it suggests that future research focused on the determinants of the number of development pivots is warranted.

5.4 Market Share

This subsection considers tests related to Testable Prediction 4, which is that outsourcing should be less common when the originator's market share in the therapeutic class is larger. We conduct

defined in terms of second-level ATC. However, results are similar when pivots are defined in terms of first-level ATC.

[33] We have also conducted tests similar to those found in Table 6 but without the interaction term and the patent variable. In these regressions, we also find that the coefficient on the in-house variable is negative and statistically significant at the one percent level in all three regressions.

Table 6: Number of Pivots by In-house Development Status

	(1) # pivot	(2) # pivot	(3) # pivot
In-house	-0.249***	-0.216***	-0.217***
	(0.0397)	(0.0381)	(0.0381)
In-house x Patent	-0.108**	-0.117**	-0.113**
	(0.0520)	(0.0503)	(0.0503)
Patent	0.111**	0.116**	0.117**
	(0.0481)	(0.0463)	(0.0464)
Constant	0.726***	0.183***	-0.136
	(0.0362)	(0.0359)	(0.108)
Therapeutic class FE		Yes	Yes
Firm type FE			Yes
Observations	11,493	11,493	11,493

Note: Each unit is a drug project. Dependent variable is the number of pivots, calculated as the total number of other additional therapeutic classes (second level ATC) a drug is tested or intended to be tested for, in addition to the main therapeutic class the project is seeking approval for or being approved. Column 1 reports the regression with in-house, patent existence, and the interaction term. Column 2 further includes a set of primary therapeutic class fixed effects. Column 3 further adds firm type fixed effects to the specification in Column 2. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Standard errors (in parentheses) are heteroskedasticity-robust and clustered at the compound level.

two sets 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 or outsourcing. In the second, we use IMS sales data to construct market share measures for an originator, and then directly test how expected market share affects the choice of in-house versus outsourced development.

As indicated, we start with tests concerning the number of drugs in the therapeutic 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:

$$\begin{aligned} Prob(Y_{ijkt} = 1) = \Lambda(\gamma_0 + \gamma_1 EOP_{ijkt} + \gamma_2 EOP_{ijkt} PDM_{jkt} + \gamma_3 PDM_{jkt} \\ + \gamma_4 TDM_{kt} + \gamma_5 X_{ijt} + \gamma_6 Z_{jkt} + \gamma_7 W_{jt} + C_k + T_t). \end{aligned} \tag{7}$$

TDM is the total number of drugs on the market that are in the same therapeutic class as the drug under development, while PDM is the number of patented drugs on the market in the same therapeutic class as the drug under development but owned by different firms than the originator of

the compound of interest. Given that we control for TDM, the theory predicts that the correlation between in-house development and the existence of a patent owned by the originator that is in the same therapeutic class as the product under development should be smaller when there is a higher number of competing patented products 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 (5) and (6), and our focus is the specification that includes controls for experience and scope.

Table C.4 in Appendix C reports results for estimating Equation (7). 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. Column 2 shows results for the same test, except that in column 2 $EOP2$ is the patent existence variable. The results in this test are similar to these in the column 1 test, both in the sign of estimates and the associated statistical significance.

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 (7), except now the explanatory variable of interest is a measure of patent length rather than patent existence. We again focus on the specification which includes controls for experience and scope.

Results are reported in Columns 3-4 in Table C.4. Column 3 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. Column 4 considers the same test as in column 3, except that in column 4 LOP2 is the patent length variable. The pattern of results in column 4 is the same as in column 3. Overall, we find results consistent with the fourth testable prediction.

We now consider a second approach for measuring how competition from other firms' patented products in the same therapeutic 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 the market share of the originator's existing patented products in the same therapeutic class as the product under development. Note that construction of our market share measures requires IMS data, which only covers the years between 1992 and 2004.

According to Testable Prediction 4, 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 a development decision which should matter, but rather the expectation at the time of a development decision concerning the market share that the firm

will have once the new product is introduced. Of course, we do not have data that allows us to directly measure the expected market share when the new drug is expected to reach the market. Our approach is to use two different proxies for this expected market share. Our first approach is that for an observation in any year t the expected market share is proxied by year t's market share, i.e., the current market share. Our second approach is to assume that the firm has perfect foresight concerning future market share. The details and results concerning these two approaches are described below.

In Table C.5 in Appendix C we examine the correlation between current market share and in-house development. The top panel of the table reports results for the Logit specification in Equation (8):

$$\begin{aligned} Prob(Y_{ijkt} = 1) = \Lambda(\zeta_0 + \zeta_1 MSP_{ijkt} + \zeta_2 EOP_{ijkt} \\ + \zeta_3 X_{ijt} + \zeta_4 Z_{jkt} + \zeta_5 W_{jt} + C_k + T_t). \end{aligned} \tag{8}$$

For each drug in 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 therapeutic class. The other regressors are defined the same way as in Equation (5).

In the top panel (Panel A), we impute missing sales in the dataset as zero sales. Here we find that the coefficient on the market share variable is positive and statistically significant at the one percent level in all five regressions. In the bottom panel (Panel B), we take a more conservative approach. In particular, in constructing the market share measure we only include firms for which sales are observed. Here we find that the coefficient on the market share variable is positive in all five regressions, statistically significant at the five percent level in three of the regressions, and significant at the ten percent level in the remaining two regressions. In addition, in columns 1 and 2, we find that in both the top and bottom panels, the coefficient on the patent existence variable is always positive, and it is statistically significant at least at the five-percent level in all four regressions. In columns 4 and 5, we rerun the tests in columns 2 and 3, but replace the patent existence variables with the patent length variables, yielding similar results. These findings are consistent with Testable Prediction 4.

In Table C.6 in Appendix C we redo the tests in Table C.5 employing future market share measures. Results are similar, although statistical precision is, on average, somewhat less for the market share coefficients. That is, all ten market share coefficients are positive as was the case in Table C.5, and in the top panel all the market share coefficients are significant at the one percent level which is also the same as in Table C.5. In the bottom panel, however, only one of the market

share coefficients is statistically significant at the five percent level, while the other four are all statistically significant at the ten percent level.

5.5 Robustness Checks

In this subsection, we consider the robustness of our results in a few respects. In the analysis above we define in-house compound years as years prior to the first outsourcing contract if that contract occurs prior to the beginning of phase III trials, or all years if there is not contract prior to the beginning of phase III trials. One might argue, however, that the design and nature of a drug are mostly fixed as early as the completion of phase I testing. With this in mind, in Table C.7, we rerun tests reported in Tables 4 and 5 with the single change that in defining in-house compound years we substitute the beginning of phase II trials for the beginning of phase III trials. These results are similar to those in Tables 4 and 5. Table C.7 columns 1 and 2 report results using our patent existence variables and our preferred specification. The coefficient on the patent existence measure is positive in each regression and statistically significant at the one-percent level in each regression. Columns 3 and 4 report results for our patent length variables. Here we similarly find that the coefficient on the patent length variable is positive and statistically significant at the one-percent level in each regression.

Our second set of robustness tests concerns the way we define therapeutic class. In particular, one might be concerned that our therapeutic classes are too coarse to accurately capture the cannibalization effect that our theory focuses on. To address this concern, we redefine our main explanatory variables, i.e., patent existence and patent length of other drugs owned by originators in the same top Anatomical Therapeutic Chemical (ATC) class, by using a set of narrower therapeutic classifications based on the second-level ATC codes. For example, instead of the therapeutic class of "cardiovascular system (C)", drugs are classified into finer classes, including antihypertensive (C02), diuretics (C03), peripheral vasodilators (C04), vasoprotectives (C05), agents acting on the renin-angiotensin system (C09), and lipid modifying agents (C10). Results for this modified specification are reported in Table C.8 in Appendix C. The coefficients on the patent existence and patent length variables are all positive and statistically significant at the one-percent level.

We next explore how results vary with firm type, since it is possible that the behavior of the originator concerning the in-house versus outsource decision may depend on firm type. Table C.9 shows the results accounting for firm type employing fixed effects and focused subsamples. Panel A reports results when firm-type fixed effects are included in our specification, where the firm-type distribution is reported in Table C.3. Panel B reports results in a subsample that excludes any Contract Research Organizations/Contract Development and Manufacturing Organizations

(CROs/CDMOs) that participated in a project to address the concern that CROs/CDMOs may focus more on execution and may act with less autonomy as licensees/collaborators, thus operating differently than standard pharmaceutical companies.[34] Panel C studies the subsample of biopharmaceutical firms (which comprise the largest share of firms in our sample). Across panels, our results remain very similar.

In Table C.10, we report results incorporating a rich set of firm-level measures to address the concern that firm-specific characteristics may be driving the results. Panel A reports results using a specification that includes extra firm portfolio measures, including time-varying measures on the number of products in a firm's portfolio in a given therapeutic class in a year and the number of competing products in the same therapeutic class per year. Panel B reports results including firm fixed effects, which would eliminate firms that only have a single or very few products across different therapeutic areas. Here we find that the coefficient of interest is positive and statistically significant for EOP2 and LOP2, but not for EOP1 and LOP1. We suspect that including firm fixed effects eliminates much of the relevant variation for the regressions concerning EOP1 and LOP1.

Finally, we show that our results are robust in various subsample analyses, as reported in Table C.11. Panel A focuses on the panel eliminating observations that occur after the end of phase II to show that the results are robust to ignoring observations from later development stages. Panel B shows results when we exclude compound years with ownership changes (e.g., due to the originator being either acquired or merged). Results stay very similar. Panel C drops observations that are potentially right-censored if a drug began a phase within the 95 percent completion threshold but has not yet completed it. All subsample results support our predictions.

6 Alternative Explanations

In this paper, we provide a novel theory for why firms choose to outsource research and development for some new products while they choose to conduct R&D for other similar products in-house. Two potential alternative explanations for this phenomenon are the following. First, in many instances the decision between in-house development and outsourcing 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, and vice versa. Note that this theoretical approach differs substantially from ours. Their focus is the probability of successful

[34] CROs and CDMOs provide outsourced services to biopharma companies, with CROs focusing on research, clinical trials, and regulatory support, and CDMOs handle drug development, scale-up, and manufacturing. CROs/CDMOs allow biopharmaceutical companies to streamline operations without sufficient in-house capabilities.

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 an important factor in many real-world integration decisions concerning research and development, we feel that their argument is an unlikely explanation for our findings. According to that theory, firms with existing successful patents should be less financially constrained. Therefore, consistent with our findings, a firm with an existing patent should be more likely to choose in-house development as financing costs are less of a concern. However, it does not account for why patents in the same product category as the product under development should be particularly important for the in-house versus outsource decision. It also does not explain our results concerning the frequency of pivots during the development process.

Another potential explanation for the in-house versus outsource decision concerning R&D is a learning curve argument. Firms may choose to develop some products in-house because of lower costs associated with learning-by-doing. Even though in our empirical analysis we include variables designed to control for lower costs due to experience and scope, one might still suspect that the correlations at least to an extent reflect lower costs due to earlier R&D investments. For example, a learning curve argument is potentially consistent with our finding that an existing patent in the same product class as that of the product being developed increases the likelihood of in-house development. However, this alternative explanation does not easily explain our empirical findings concerning development pivots. Remember, we do not just find that in-house development is associated with a smaller number of development pivots. We also find that the negative effect concerning the number of development pivots is larger when the originator owns an existing patented product in the same therapeutic class as the product under development. This is exactly what our theory predicts, and it is unclear why, if cannibalization is not an issue, this would be the case.[35]

In addition, there are a number of other alternative theories for the in-house versus outsource decision, none of which appears to be a good match for our empirical findings however. One argument, for example, is that learning-by-doing on the part of developers can be important. In particular, an originator may choose to outsource because a potential outside firm has significant experience with developing products in the relevant product category, and due to learning-by-doing it is the low-cost developer. This theory neither explains our main findings concerning patent existence and patent length, nor our findings concerning pivots during the development process.

[35]The fact, however, that firms do frequently choose to develop new products in the same therapeutic class as a patented product the firm already owns does suggest that learning-by-doing is a factor. Otherwise, firms would avoid developing new products in the same therapeutic class as a patented product the firm already owns in order to fully avoid cannibalization. We would like to thank one of the referees for suggesting this argument to us.

Another argument, put forth initially in Azoulay (2004), is that data intensive R&D activities are more likely outsourced, while knowledge intensive R&D activities are typically conducted in-house. This argument also does not explain our results concerning patent existence and patent length, or our development pivot findings. In summary, the literature does not provide an alternative theory of the in-house versus outsource decision that accounts for our empirical findings as well as our model of limiting cannibalization.

7 Conclusion

This paper focuses on the idea that limiting cannibalization of existing patented products is important for understanding a firm's decision concerning whether to develop a new product in-house or outsource. In the first part of the paper we construct and analyze a theoretical model in which ownership of existing patented products in the same product class as a new product decreases the incentive for an originator to outsource development. The logic is that a licensee has a smaller incentive to avoid cannibalizing the value of the originator's current patented products, so outsourcing is suboptimal when avoiding cannibalization is important. We employ this model to derive four testable predictions which concern the choice of in-house versus outsourced development and the frequency of development pivots.

In the latter part of the paper we employ data from the pharmaceutical industry to investigate the testable predictions. Our findings are consistent with the predictions. For example, controlling for firm characteristics and therapeutic class, we find that an originator with existing patented products in the same class as the product under development is less likely to outsource development. We also find evidence consistent with our prediction concerning development pivots. Finally, we also show that our results are robust to alternative specifications of outsourcing and therapeutic class.

We focus on the incentive for in-house development rather than outsourcing when the originator owns existing patented products in the same product class, and wants to control new product design. A complementary perspective is that in-house development is also important when the originator owns existing patented products in the same product class 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.[36] Additional interesting directions for future research concern other factors that are endogenous in real-world markets that are treated as exogenous in

[36] Williams (2013) analyzes how intellectual property protection affects the rate of subsequent innovation. In addition, we have explored whether in-house development speeds up the development process and found that this does seem to be the case (see Table C.10 Panel C in the Appendix). Further research along these lines focused on optimal timing rather than solely speed of the development process seems to us of particular interest.

our analysis. For example, we take as exogenous the choice of the product class of the product being developed. Similarly, another factor treated as exogenous in our analysis is the quality of the innovation. We feel that focusing on these factors as endogenous outcomes is an interesting topic for future research. Another direction for future research concerns heterogeneity regarding different types of licensing and outsourcing decisions. Some types of outsourcing may be more important than others in terms of risks associated with cannibalization, and thus our predictions concerning in-house development versus outsourcing may be more important for these types of development decisions. Finally, focusing on how rivalry in the innovation process itself might affect our basic conclusions also seems worthwhile.[37]

References

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

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

Arrow, K. (1962). Economic Welfare and the Allocation of Resources for Invention. In Universities-National Bureau Committee for Economic Research, Committee on Economic Growth of the Social Science Research Council (Ed.), *The Rate and Direction of Inventive Activity: Economic and Social Factors*, pp. 609–626. NJ: Princeton University Press.

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

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

Benkard, C. L. (2004). A Dynamic Analysis of the Market for Wide-Bodied Commercial Aircraft. *The Review of Economic Studies 71*, 581–611.

Budish, E., B. N. Roin, and H. Williams (2015). Do Firms Underinvest in Long-Term Research? Evidence from Cancer Clinical Trials. *American Economic Review 105*, 2044–2085.

[37] There is a large literature on rivalry in the innovation process. See, for example, Gans and Stern (2000), Benkard (2004), Goettler and Gordon (2011), Igami (2017), Igami and Uetake (2020), and Yang (2020). Also, see Budish et al. (2015) and Krieger et al. (2017) for analyses focused on the nature of investments.

Casas-Arce, P., T. Kittsteiner, and F. A. Martinez-Jerez (2019). Contracting with Opportunistic Partners: Theory and Application to Technology Development and Innovation. *Management Science 65*, 842–858.

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

Cunningham, C., F. Ederer, and S. Ma (2021). Killer Acquisitions. *Journal of Political Economy 129*, 649–702.

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*, 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*, 256–266.

Dewatripont, M. (1988). Commitment Through Renegotiation-Proof Contracts with Third Parties. *Review of Economic Studies 55*, 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*, 151–185.

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

Gans, J. S. and S. Stern (2000). Incumbency and R&D Incentives: Licensing the Gale of Creative Destruction. *Journal of Economics & Management Strategy 9*, 485–511.

Gibbons, R. (2005). Four Formal(izable) Theories of the Firm? *Journal of Economic Behavior & Organization 58*, 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*, 1121–1135.

Goettler, R. L. and B. R. Gordon (2011). Does AMD spur Intel to Innovate More? *Journal of Political Economy 119*, 1141–1200.

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

Hall, B., C. Helmers, M. Rogers, and V. Sena (2014). The Choice between Formal and Informal Intellectual Property: A Review. *Journal of Economic Literature 52*, 375–324.

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

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

Hortaçsu, A. and C. Syverson (2007). Cementing Relationships: Vertical Integration, Foreclosure, Productivity, and Prices. *Journal of Political Economy 115*, 250–301.

Igami, M. (2017). Estimating the Innovator's Dilemma: Structural Analysis of Creative Destruction in the Hard Disk Drive Industry, 1981-1998. *Journal of Political Economy 125*, 798–847.

Igami, M. and K. Uetake (2020). Mergers, Innovation, and Entry-Exit Dynamics: Consolidation of the Hard Disk Drive Industry, 1996–2016. *The Review of Economic Studies 87*, 2672–2702.

IMS (1992–2004). IMS Health Market Research Data.

Krieger, J., D. Li, and D. Papanikolaou (2017). Developing New Drugs. *mimeo*.

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

Lakdawalla, D. N. (2018). Economics of the Pharmaceutical Industry. *Journal of Economic Literature 56*, 397–449.

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

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

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

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

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

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

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

Novak, S. and S. Stern (2009). Complementarity Among Vertical Integration Decisions: Evidence from Automobile Product Development. *Management Science 55*, 311–332.

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

Petrin, A. (2002). Quantifying the Benefits of New Products: The Case of the Minivan. *Journal of Political Economy 110*, 705–729.

PharmaProjects (1989–2004). PharmaProjects Research Database.

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

Schmalensee, R. (1978). Entry Deterrence in the Ready-to-Eat Breakfast Cereal Industry. *The Bell Journal of Economics 9*, 305–327.

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

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

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

Waldman, M. (1996). Planned Obsolescence and The R&D Decision. *RAND Journal of Economics 27*, 583–595.

Williams, H. L. (2013). Intellectual Property Rights and Innovation: Evidence from the Human Genome. *Journal of Political Economy 121*, 1–27.

Yang, C. (2020). Vertical Structure and Innovation: A Study of the SoC and Smartphone Industries. *The RAND Journal of Economics 51*, 739–785.

Appendices

Appendix A Proofs

Proof of Lemma 1. In any period t in which only the existing product is available, i.e., $t < 3$, each consumer maximizes

$$U(Q) = \int_0^Q (V^+ - vq)dq - (P + ds)Q \tag{A.1}$$

with respect to the quantity Q.[38] If more than one product is available, the consumer maximizes Equation (A.1) once he or she has determined which of the two products she buys. This optimal quantity is

$$\boldsymbol{Q}(P, s) = \frac{V^+ - P - ds}{v}, \tag{A.2}$$

leading to an equilibrium utility of

$$U(\boldsymbol{Q}) = \frac{(V^+ - P - ds)^2}{2v}, \tag{A.3}$$

where boldface characters indicate equilibrium values. Observe from Equation (A.3) that a consumer chooses product i over product j whenever

$$P_i + ds_i < P_j + ds_j, \tag{A.4}$$

where s_i refers to the distance of the consumer from product i in product space.

Note that the originator will always choose the monopoly price for the existing product in periods t, $t < 3$, and that the originator can only earn profits from the new product in periods t after which the existing product's patent has expired, $t > t_E$.

As such, the most interesting problem concerns periods t, $2 < t \leq t_E$, in which both products will be sold at a profit. For a given location l for the new product and prices P_O (for the originator's existing product) and P_I (for the new product or innovation), profits can be defined by means of the indifferent consumers on both sides of either product. Let

$$\theta_{OI}(P_O, P_I) = \frac{P_I - P_O}{2d} + \frac{l}{2} \tag{A.5}$$

denote the indifferent consumer located between 0 and l and

$$\theta_{IO}(P_O, P_I) = \frac{P_O - P_I}{2d} + \frac{1}{2} + \frac{l}{2} \tag{A.6}$$

the indifferent consumer located between l and 1. Then, profits accrued by product O are given by

$$\Pi_O(P_O, P_I) = \left[\int_0^{\theta_{OI}(P_O,P_I)} \boldsymbol{Q}(P_O, y)\, dy + \int_0^{1-\theta_{IO}(P_O,P_I)} \boldsymbol{Q}(P_O, y)\, dy \right] (P_O - c)^{39} \tag{A.7}$$

[38] Recall that we assume V^+ to be large enough such that it is always optimal to serve the entire market.

[39] Throughout this section we set $c \equiv c_1 = c_2$ significantly simplifying algebra without affecting any of the qualitative results presented in this paper.

while product I's profits are

$$\Pi_I(P_O, P_I) = \left[\int_0^{l-\theta_{OI}(P_O,P_I)} \boldsymbol{Q}(P_I, y)\, dy + \int_0^{\theta_{IO}(P_O,P_I)-l} \boldsymbol{Q}(P_I, y)\, dy \right] (P_I - c). \tag{A.8}$$

At the very beginning of each period t, $2 < t \leq T$, 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 constitutes 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

$$c_2^+ > c_2. \tag{A.9}$$

Suppose this is not true in period t. Then, either firm could suggest at the beginning of period t to reassign production—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, $t > 2$, sales and pricing rights are allocated in a renegotiation-proof subgame-perfect Nash equilibrium in such a way as to maximize total surplus in a given period. For if not, a Pareto-improving renegotiation is possible.

This immediately implies that sales and pricing rights for the new product can only ever be allocated to the licensee if the licensee internalizes variable cost and revenue of all products held by the originator in a given period. It follows that sales and pricing rights at t, $2 < t \leq t_E$, necessarily reside with the originator. This establishes the sales and pricing part of i). That is to say, the originator solves

$$\max_{P_O, P_I} \Pi_{OI}(p_O, p_I) = \Pi_O(p_O, p_I) + \Pi_I(p_O, p_I) \tag{A.10}$$

in periods 3 to t_E.

Clearly, the overall profit in any given period depends on the location of the new product l. Π_{OI} can be shown to have three local maxima in the (P_O, P_I) space. The global maximum, however, is the same for each $l \in [0, 1)$. Equilibrium prices in period t, $2 < t \leq t_E$, are given by

$$\boldsymbol{P_O} = \boldsymbol{P_I} = \frac{4(V^+ - c) - d + 2dl(1-l)}{8} \tag{A.11}$$

for any realized location of the new product l. As such,

$$\Pi_{OI}(\boldsymbol{P_O}, \boldsymbol{P_I}) = \frac{\left(4(V^+ - c) - d + 2dl(1-l)\right)^2}{64v}, \tag{A.12}$$

and

$$\frac{\partial \Pi_{OI}(\boldsymbol{P_O}, \boldsymbol{P_I})}{\partial l} = 0 \tag{A.13}$$

at $l = \frac{1}{2}$ only with

$$\frac{\partial^2 \Pi_{OI}(\boldsymbol{P_O}, \boldsymbol{P_I})}{\partial^2 l} < 0 \tag{A.14}$$

if V^+ is large enough. It follows that the originator strictly prefers $l = \frac{1}{2}$.

If the licensee, however, never internalizes variable costs and revenue, it does not care about the location of the new product. Therefore it chooses $k = 0$. Can the originator do better? Indeed, it can.

By allocating sales and pricing rights of the new product in periods t, $t > t_E$, the originator forces the licensee to internalize variable cost and revenue and thus maximize the profits of the new product. This induces the licensee to care about the location of the new product. Once the patent of the originator's existing product has expired, product O is sold at price c. For large enough V^+ the licensee chooses to price the new product at

$$\frac{8V^+ + 10c + 2d + \sqrt{64{V^+}^2 - 128cV^+ + 64c^2 + 40cd - 40dV^+ + 22d^2 - 36d^2 l(1-l)}}{18} \tag{A.15}$$

and its profits are uniquely maximized at $l = \frac{1}{2}$. As a consequence, the licensee chooses $k > 0$ in period 1 as $p'(0) = \infty$.

Finally, there is a competitive pool of licensees. As such, the licensee necessarily expects to make zero economic profit ex ante. This means that for each period t, $2 < t \leq t_E$, the originator guarantees to pay the licensee c times the quantity of the new product sold. Moreover, the licensee agrees to pay to the originator the expected profit from the new product in periods $t_E + 1$ to T minus the incurred fixed cost F_L. This establishes the remainder of i) and ii) as well as iii). ■

Proof of Lemma 2. We have established in the proof of Lemma 1 above that the originator always prefers to locate the new product at $l = \frac{1}{2}$ in periods t, $2 < t \leq t_E$. The same is true for the originator, and equivalently, for the licensee in periods $t > t_E$. i) follows.

ii) follows directly since the originator and a licensee face the same optimization problem in periods $t > t_E$. If $t_E < 3$, the existing and new product are never sold under a patent in the same period and only periods t, $t > t_E$, affect the incentives of the developer to invest in location precision.

The originator and licensee have the same incentive to invest in location precision for all periods t, $t > t_E$. However, from the proof of Lemma 1 it is clear that the originator faces an additional incentive to invest in location precision to increase its profits through period t_E. Moreover, if $K(O, T) = 0$, the licensee never controls pricing and sales and thus never internalizes variable costs or revenue. As a result, the licensee has no incentive to invest. This completes the proof of iii).

The later the patent of the existing product 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 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 firm 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, $t > \max\{2, t_E\}$. 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. 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 in the first place. It follows that this subgame always has an equilibrium.

Second, i) and ii) of Lemma 1 together with the first part of 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 always be 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.

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 $\Pi(I, F)$. 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, $\Pi(I, F)$ equals a positive constant minus F. 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

$$\Pi(I, F_L) > \Pi(O) > 0. \tag{A.16}$$

Thus, by the continuity of $\Pi(I, F)$ in F, there necessarily is an $F^* > F_L$ such that

$$\Pi(I, F^*) = \Pi(O) > 0. \tag{A.17}$$

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 $F \leq F^*$ and to outsource if $F > F^*$. 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

$$\frac{\partial \left(K(I, t_E) - K(O, t_E)\right)}{\partial t_E} > 0, \tag{A.18}$$

i.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). ■

Appendix B Extensions

In this section we provide two extensions of our basic model introduced in Section 2. The first extension addresses the presence of competition and the market share of existing patents as a predictor of the outsourcing decision. The second extension concerns a single firm that faces multiple outsourcing decisions.

B.1 Competition

We formally present a 2-period model without discounting that simplifies along several dimensions, and argue why the underlying logic applies under more general circumstances. Each consumer is interested in buying a single product in each period only. Moreover, there is a now a competitor we call the rival who offers a competing product in the same product category starting in both periods.

More specifically, consider the product space to be an equilateral triangle with circumference 1, which, naturally, is homeomorphic to a circle but lends itself better for describing the outsourcing decision of an innovating firm in the presence of a competitor. Label its corners clockwise starting at the top by A, B, and C, i.e., in terms of location $A = 0$, $B = \frac{1}{3}$, and $C = \frac{2}{3}$. Consumers with a total mass of 1 are uniformly distributed along the perimeter of the triangle.

Assume the originator O to have a patented product located at A and a rival R having a patented product located at B. In period 1, only products O and R vie for the consumers located around the triangle by simultaneously setting prices.

At the end of period 1, the originator innovates and introduces a second product. We denote this product by the subscript I. We further assume that the developer can only position its product in product space between A and C due to technological feasibility or the patent that the rival holds. The product's final location is determined by the product developer's location choice as well as a random process just as in Section 2 that depends on the developer's investment in location precision k. In fact, just as before, the new product's location is $l = l_M + \varepsilon$, where ε is drawn from either of two uniform distributions $U[-\alpha, \alpha]$ and $U[-\beta, \beta]$ with $\alpha < \beta \leq \frac{1}{12}$. If the developer chooses to locate the product nearby A or B, ε is drawn from the resulting conditional distribution truncated at the respective end of the line connecting A and B. In addition, we simplify by assuming that all three products can be produced at 0 marginal cost. If, however, a firm that did not develop product I produces it, it exhibits a positive marginal cost.

Just as in Section 2 the originator chooses whether to outsource the development of product I to a firm from a competitive pool of licensees or to develop and produce in-house. This choice depends on the realization of the originator's stochastic fixed cost of development F in relation to the known licensee's fixed cost of development F_L. If the originator decides to outsource, a renegotiation-proof contract is signed in period 1 before the development of the product.

We assume that the new product is introduced after period 1, and offered alongside products

O and R. The following intuition carries over from Sections 2. Naturally, if the originator decides to develop product I there is no outsourcing, and the originator obtains production, sales, and pricing rights of product I. If, however, the originator decides to outsource, the licensee is assigned production of product I, but never pricing and sales rights since product O and I are both offered in period 2 and the licensee would not internalize the effects of its pricing choice on product O.

A consumer's utility from buying product I is determined by the product's utility V_i^+ (where we assume $V_I^+ = V_O^+$, i.e., technological innovation takes place in product space, not affecting the product's base utility), its price P_i and the consumer's distance from the product's location in product space. That is, for example a consumer located at θ on the line between A and B receives a utility when buying from the originator of

$$U_O(\theta) = V_O^+ - P_O - \delta(\theta) \text{ and } U_R(\theta) = V_R^+ - P_R - \delta(\frac{1}{3} - \theta) \tag{B.1}$$

when buying from its rival, where $\delta(\cdot)$ denotes the distance cost function. Throughout this subsection we assume $\delta(\cdot)$ to be the identity function and V_O^+ and V_R^+ to be sufficiently large such that every consumer buys some product in equilibrium.

Below we show that in this environment, the originator is less likely to outsource development and production of the new product the larger its initial market share in the first period, i.e., its predicted market share in period two in the absence of innovation.

Proposition 2 *In the unique equilibrium outcome of the 2-period innovation game with competition, the originator is more likely to choose in-house development the higher its market share before the introduction of the new product.*

Proof of Proposition 2. It is straightforward to see that competition in period 1 in this model is equivalent to competition of two competitors with fixed positions on a Salop circle. In equilibrium, the originator chooses a price of

$$P_O^1 = \frac{V_O^+ - V_R^+}{3} + \frac{1}{2} \tag{B.2}$$

to obtain market share

$$MS_O^1 = \frac{V_O^+ - V_R^+}{3} + \frac{1}{2}, \tag{B.3}$$

such that the originator's equilibrium market share strictly increases in V_O^+ and strictly decreases in V_R^+.[40]

Let us now assume that both the originator and a licensee would locate product I at $l_M = C$, an assumption that we will justify below. Let θ_{OR} denote the consumer who is located between A and B and indifferent between buying from O and R, θ_{RI} the consumer located between B and $C + \epsilon$, where ϵ refers to the realization of ε, and indifferent between buying from R and I, and θ_{IO} the consumer located between $C + \epsilon$ and A and indifferent between buying products I or O.

In period 2, the originator maximizes

$$\begin{aligned}\Pi_{OI}(P_O^2, P_I^2, P_R^2) = &\left[\theta_{OR}(P_O^2, P_R^2) + \left(1 - \theta_{IO}(P_I^2, P_O^2)\right)\right] P_O^2 \\ &+ \left(\theta_{IO}(P_I^2, P_O^2) - \theta_{RI}(P_R^2, P_I^2)\right) P_I^2\end{aligned} \tag{B.4}$$

by choosing P_O^2 and P_I^2, with the superscript indicating the period.

[40] We assume throughout this subsection that $|V_O^+ - V_R^+|$ is sufficiently small such that all two (three) products exhibit positive demand in period 2 (3). We refer to such an equilibrium as an interior solution.

In equilibrium, the originator chooses to price its products at

$$\boldsymbol{P_O^2} = \frac{12V_O^+ - 12V_R^+ + 20 - 9\epsilon}{36}, \tag{B.5}$$

and

$$\boldsymbol{P_I^2} = \frac{12V_O^+ - 12V_R^+ + 20 - 3\epsilon}{36}, \tag{B.6}$$

respectively, while its rival prices at

$$\boldsymbol{P_R^2} = \frac{6V_R^+ - 6V_O^+ + 8 + 3\epsilon}{18}, \tag{B.7}$$

with bold variables indicating equilibrium values.

Substituting these equilibrium prices into Equation (B.4) and derivation with respect to location precision results in

$$\frac{\partial \Pi_{OI}(\boldsymbol{P_O^2}, \boldsymbol{P_I^2}, \boldsymbol{P_R^2})}{\partial \epsilon} = \frac{144V_R^+ - 144V_O^+ + 126\epsilon - 240}{1296}, \tag{B.8}$$

an expression that is always negative for an interior solution. This establishes that the originator prefers product I to be located at $l = C$, and, in fact, always prefers the product to be closer to C.

In addition, we can now see that

$$\frac{\partial^2 \Pi_{OI}(\boldsymbol{P_O^2}, \boldsymbol{P_I^2}, \boldsymbol{P_R^2})}{\partial \epsilon \partial (V_O^+ - V_R^+)} = -\frac{1}{9}. \tag{B.9}$$

This implies that location precision is more important for the originator's profits the larger its period 1 market share.

Since a licensee as developer does not internalize revenue or variable cost since it is paid a fixed amount, it does not invest at all in location precision, i.e., $k = 0$. It chooses, however, $l_M = C = \frac{2}{3}$ in equilibrium, as, otherwise, the originator contracts with another licensee. The originator as developer on the other hand chooses the efficient investment in location precision k^*. Note that $k^* > 0$ due to $p'(0) = \infty$ and Equation (B.8). Moreover, by the argument above based on Equation (B.9), it increases in the originator's period 1 market share.

Just as in Section 2, the payments from the originator to a licensee are such that the licensee earns zero economic profits, i.e. they amount to F_L in this case. Denote the originator's expected equilibrium profits in period 2 as a function of the investment level $\mathbb{E}\boldsymbol{\Pi_{OI}^2}(k)$, and note that

$$\frac{\partial \mathbb{E}\boldsymbol{\Pi_{OI}^2}(k)}{\partial k} > 0 \tag{B.10}$$

for all $k \in [0, k^*)$.

It follows that the originator chooses in-house development if

$$\mathbb{E}\boldsymbol{\Pi_{OI}^2}(k^*) - F > \mathbb{E}\boldsymbol{\Pi_{OI}^2}(0) - F_L. \tag{B.11}$$

This gives rise to a

$$\Delta^* = \mathbb{E}\boldsymbol{\Pi_{OI}^2}(k^*) - \mathbb{E}\boldsymbol{\Pi_{OI}^2}(0), \tag{B.12}$$

such that the originator chooses to outsource development if and only if

$$\Delta = F - F_L \tag{B.13}$$

exceeds Δ^*. It follows directly from Equation (B.9) that $\mathbb{E}\boldsymbol{\Pi}^{\mathbf{2}}_{\boldsymbol{OI}}(k^*)$ increases more in

$$V_O^+ - V_R^+, \tag{B.14}$$

and therefore in MS_O^1, than does $\mathbb{E}\boldsymbol{\Pi}^{\mathbf{2}}_{\boldsymbol{OI}}(k^*)$, and thus Δ^* increases in MS_O^1. This proves the claim. ■

The two-period model introduced in this subsection establishes that if both the existing as well as a new product hold patents at the same time, the originator is more likely to choose in-house development the higher the market share of its existing product in the first place, thus rationalizing Testable Prediction 3. Clearly this insight generalizes to the multiple periods setting in Section 2 as periods in which only one of the products holds a patent do not affect the rationale underlying this result. It can also be generalized in terms of the distance cost function, and for consumers who buy more than one unit of a product such as in Section 2. Moreover, the model can be extended to one in which consumers populate the interior of the triangle.

B.2 Multiple R&D investments

Consider the model presented in Section 2 and assume that the originator faces two product development tasks, tasks 1 and 2, when developing a new product, each of which may be outsourced to companies from a pool of competitive licensee, or undertaken in-house. The originator now faces two random draws from probability density functions $f^1(\cdot)$ and $f^2(\cdot)$, the realizations of which determine its fixed cost of taking on the respective task in-house. licensee on the other hand incur known fixed costs of F_L^1, and F_L^2 respectively. Let Δ^i refer to the originator's fixed cost disadvantage related to task i. Likewise we assume that a firm that has undertaken one of the development tasks has a competitive advantage in production, i.e., its marginal production cost c satisfies $c_2 < c_2^+$, with c_2^+ being the cost of a company that has not fulfilled either of the development tasks.

The location of the new product in product space is determined by the location choice and investment in location precision of both firms undertaking a development task. Specifically assume that both developing companies face the familiar investment decision in location precision. The more they invest, the likelier it is that the random error regarding the location of the product is drawn from a favorable distribution, i.e., $U[-\alpha, \alpha]$ instead of $U[-\beta, \beta]$. We continue to assume $\alpha < \beta \leq \frac{1}{2} - (c_2 - c_1)$. The final location of the new product in product space is then determined by the convex combination

$$l = \phi_1 \left(l_M^1 + \varepsilon^1 \right) + \phi_2 \left(l_M^2 + \varepsilon^2 \right), \tag{B.15}$$

where $\phi_i > 0$, $\phi_1 + \phi_2 = 1$, denotes the contribution of task i to the final location of the product. Furthermore, assume that both firms make their location and investment decisions at the same time that the originator decides who controls production and pricing.

If, in equilibrium, the originator ends up outsourcing either none or one of the tasks, the equilibrium solution and contracts are—bar minor details—given by the solutions presented in Sections 2 and 3. As a consequence, we focus in this subsection on the equilibrium outcome in which the originator outsources both development tasks to licensees and argue under which circumstances this outcome arises. Note that the basic intuition from Section 2 about contracts carries over. As such, it has to be true that either of the licensees (we refer to them as licensee 1 and licensee 2 according to their task), but not the originator, undertakes production of the new

product. Furthermore, the originator will not delegate pricing rights to licensees before the patent of its existing product expires. Which licensee will be allocated the right to price the new product and collect revenue?

Proposition 3 *In the unique equilibrium of the innovation game in which the originator chooses to outsource multiple tasks, the originator assigns production and pricing rights to each licensee with positive probability. Moreover, the originator is less likely to outsource a development task i) if it has an existing patented product in the same product class, and ii) the longer this patent is expected to be valid after the new product reaches the market.*

Proof of Proposition 3. First consider the case in which the originator outsources both tasks but assigns pricing w.l.o.g. with certainty to licensee 1. In this scenario, licensee 2 will not invest at all in location precision. In this scenario, the originator understands that as

$$p'(0) = \infty,\ p'(k) > 0\ \forall k > 0,\ p(\infty) < 1 \text{ and } \phi_i > 0 \text{ for } i \in \{1, 2\}, \tag{B.16}$$

assigning production and pricing of the new product with an infinitesimal probability to licensee 2 pushes l in expectation towards $\frac{1}{2}$, the originator's strictly preferred location of the new product. This follows since licensee 2 prefers the location $\frac{1}{2}$ if there is any chance that it collects the revenue of the new product for at least some periods while paying a fixed amount to the originator. Thus, and due to $p'(0) = \infty$, licensee 2 invests a positive amount in location precision. As $p(\cdot)$ is continuously differentiable and $p'(0) = \infty$, there is $\zeta > 0$ small enough such that

$$\phi_2 * p'(\zeta) > \phi_1 p'(K_1) \tag{B.17}$$

for any positive investment level K_1 of licensee 1 and any $\phi_2 > 0$. This establishes that the originator assigns production and pricing to both licensees with positive likelihood.

The originator never assigns task i to a licensee if

$$F^i < F^i_L \Leftrightarrow \Delta^i < 0. \tag{B.18}$$

However, even if $\Delta^i > 0$, the originator does not necessarily outsource task i. This follows from the argument in Section 2 laying out that even if licensee i chooses the optimal location of $\frac{1}{2}$ it does in general not choose the investment level preferred by the originator as it does not internalize the cannibalization of the existing product.

Naturally, if the patents of the existing and the new product do not overlap, the originator's and the licensee's objectives align perfectly. This establishes i). Moreover, just as in the base model described in Section 2, the originator is more likely to assign development task i to a licensee when $\Delta^i > 0$ and there is smaller number of periods in which the licensee does not internalize cannibalization, i.e., if the patent length of the existing product is shorter, establishing ii). ■

Appendix C Tables

Table C.1: Summary of Drug Development Phases

Development Stage	Description (according to the FDA)
Pre-clinical trial	Submission of investigational new drug application for the FDA to review. Companies need to show the 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 ranges 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 ranges 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.

Note: More details are available at the sources: http://www.fda.gov/drugs/resourcesforyou/consumers/ucm143534.htm.

Table C.2: Therapeutic Classification: Panel Data vs. Compound-Level Data

Description	Panel Data		Compound-Level	
	Freq.	Percent	Freq.	Percent
A Alimentary tract and metabolism	6,229	5.71	700	6.09
B Blood and blood forming organs	4,167	3.82	410	3.57
C Cardiovascular system	10,924	10.01	975	8.48
D Dermatologicals	2,400	2.20	230	2.00
F Formulations	1,756	1.61	171	1.49
G Genito urinary system and sex hormones	3,460	3.17	358	3.11
H Systemic hormonal preparations (excl. sex hormones and insulins)	2,280	2.09	200	1.74
J Antiinfectives for systemic use	17,476	16.02	1,820	15.84
L Antineoplastic and immunomodulating agents	27,167	24.90	3,084	26.83
M Musculo-skeletal system	5,916	5.42	646	5.62
N Nervous system	19,482	17.85	2,147	18.68
P Antiparasitic products, insecticides and repellents	451	0.41	43	0.37
R Respiratory system	4,662	4.27	460	4.00
S Sensory organs	988	0.91	92	0.80
V Various	1,757	1.61	157	1.37
Total	109,115	100.00	11,493	100.00

Note: This table reports the main therapeutic class distribution in our sample. The therapeutic categorization used is the main level of the standard Anatomical Therapeutic Chemical (ATC) Classification System developed by the World Health Organization. For a very small share of observations, we cannot map Pharmaproject therapeutic class to ATC, and we used the F group for this Pharmaproject-only group. More information on the ATC system is available at https://atcddd.fhi.no/atc_ddd_index/.

Table C.3: Firm Types: Overall, In-house, and Outsource

Originator Firm Type	Overall		In-house		Outsource	
	Freq.	Percent	Freq.	Percent	Freq.	Percent
Biopharmaceuticals	102,986	94.38	81,162	94.78	21,824	92.94
Chemicals	3,406	3.12	2,537	2.96	869	3.70
Health (broad)	2,507	2.30	1,825	2.13	682	2.90
Academia/research/NPOs	136	0.12	78	0.09	58	0.25
CRO/CDMO	80	0.07	31	0.04	49	0.21
Total	109,115	100.00	85,633	100.00	23,482	100.00
Licensee Firm Type	Overall			Outsource		
1{} indicators	Obs	Mean	SD	Obs	Mean	SD
Biopharmaceuticals	109,115	0.216	0.412	23,482	0.784	0.412
Chemicals	109,115	0.015	0.123	23,482	0.053	0.223
Health (broad)	109,115	0.012	0.109	23,482	0.038	0.191
Academia/research/NPOs	109,115	0.005	0.069	23,482	0.021	0.143
CRO/CDMO	109,115	0.002	0.046	23,482	0.008	0.089

Note: This table reports firm types across originators and licensees, listed by descending order. Panel A presents the originator firm type distribution (overall, in-house, and outsourced). Each outsourced project can have multiple licensees, summarized by firm type indicator variables in Panel B. The "Health (broad)" group includes firms that contribute to broad health sectors through medical devices, diagnostics, and health-related services rather than focusing primarily on pharmaceuticals, chemicals, research, or contract work. Acronyms: CRO: Contract Research Organization, CDMO: Contract Development and Manufacturing Organization, NPO: Nonprofit Organization.

Table C.4: Logit Models of In-house Development: Patents with Interaction Terms

	(1) In-house	(2) In-house	(3) In-house	(4) In-house
EOP1	1.164***			
	(0.140)			
EOP1 × PDM	-0.0251***			
	(0.00447)			
EOP2		0.129***		
		(0.0170)		
EOP2 × PDM		-0.00239***		
		(0.000484)		
LOP1			0.0588***	
			(0.00809)	
LOP1 × PDM			-0.00120***	
			(0.000255)	
LOP2				0.00959***
				(0.00125)
LOP2 × PDM				-0.000186***
				(3.60e-05)
PDM	-0.000178	0.0279**	-0.00346	0.0356***
	(0.0118)	(0.0126)	(0.0119)	(0.0125)
TDM	-0.0122	-0.0214**	-0.0187*	-0.0292***
	(0.0101)	(0.0104)	(0.0103)	(0.0104)
Experience	0.00157***	0.000162	0.00161***	0.000669***
	(0.000144)	(0.000133)	(0.000146)	(9.32e-05)
Scope	-1.434***	-1.394***	-1.566***	-1.351***
	(0.149)	(0.149)	(0.149)	(0.149)
Phase I	-0.669***	-0.646***	-0.676***	-0.655***
	(0.0800)	(0.0793)	(0.0796)	(0.0795)
Phase II	-1.110***	-1.047***	-1.123***	-1.069***
	(0.0706)	(0.0713)	(0.0706)	(0.0715)
Phase III	-1.546***	-1.463***	-1.575***	-1.501***
	(0.0940)	(0.0935)	(0.0936)	(0.0941)
Launched	-2.208***	-2.049***	-2.199***	-2.078***
	(0.130)	(0.127)	(0.128)	(0.126)
Observations	109,115	109,115	109,115	109,115

Note: Dependent variable is one if a compound is developed in-house, and zero otherwise. All specifications include a full set of therapeutic categories and year indicators. EOP1 indicates 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. LOP1 is the length of the longest patent among compounds in the same therapeutic class and same firm. LOP2 is the sum of the patent lengths among compounds in the same therapeutic class and same firm. $*p < 0.1$, $**p < 0.05$, $***p < 0.01$. Standard errors (in parentheses) are heteroskedasticity-robust and clustered at the compound level.

Table C.5: Logit Models of In-house Development: Market Share and Patents

	(1) In-house	(2) In-house	(3) In-house	(4) In-house	(5) In-house
Panel A					
Current MSP	0.0377***	0.0330***	0.0248***	0.0333***	0.0234***
	(0.00578)	(0.00565)	(0.00559)	(0.00568)	(0.00543)
EOP1		0.408***			
		(0.0610)			
EOP2			0.0492***		
			(0.00515)		
LOP1				0.0207***	
				(0.00371)	
LOP2					0.00341***
					(0.000319)
Observations	101,586	101,586	101,586	101,586	101,586
Panel B					
Current MSP	0.0126**	0.0107**	0.0107**	0.00949*	0.00971*
	(0.00526)	(0.00522)	(0.00531)	(0.00523)	(0.00528)
EOP1		0.627***			
		(0.195)			
EOP2			0.0135**		
			(0.00555)		
LOP1				0.0435***	
				(0.0118)	
LOP2					0.00129***
					(0.000375)
Observations	51,439	51,439	51,439	51,439	51,439

Note: Dependent variable is one if a compound is developed in-house, and zero otherwise. All specifications include experience, scope, development phase indicators, and a full set of therapeutic category and year indicators. We control for current market share based on sales for existing patented drugs in the same class and same firm as the compound of interest. Panels A and B use merged samples of our pipeline and sales data. Panel A includes a larger sample due to imputing missing MSP values as zero for firms without positive sales in a given therapeutic area, assuming their absence indicates a lack of sales activity. However, it is slightly smaller than the full pipeline dataset, as our sales data starts three years later. Panel B focuses on firms with positive sales in the therapeutic area of interest and does not impute MSP for those without sales data, recognizing that IMS data may not capture the universe of drug sales. EOP1 indicates 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. LOP1 is the length of the longest patent among compounds in the same therapeutic class and same firm. LOP2 is the sum of the patent lengths among compounds in the same therapeutic class and same firm. Standard errors (in parentheses) are heteroskedasticity-robust and clustered at the compound level. $*p < 0.1$, $**p < 0.05$, $***p < 0.01$.

Table C.6: Logit Models of In-house Development: Future Market Share and Patents

	(1) In-house	(2) In-house	(3) In-house	(4) In-house	(5) In-house
Panel A					
Future MSP	0.0431***	0.0351***	0.0322***	0.0364***	0.0298***
	(0.00849)	(0.00806)	(0.00852)	(0.00818)	(0.00833)
EOP1		0.650***			
		(0.0916)			
EOP2			0.0731***		
			(0.0116)		
LOP1				0.0276***	
				(0.00495)	
LOP2					0.00455***
					(0.000670)
Observations	14,468	14,468	14,468	14,468	14,468
Panel B					
Future MSP	0.0165**	0.0130*	0.0145*	0.0126*	0.0135*
	(0.00758)	(0.00748)	(0.00779)	(0.00750)	(0.00778)
EOP1		0.976***			
		(0.244)			
EOP2			0.0295**		
			(0.0117)		
LOP1				0.0589***	
				(0.0138)	
LOP2					0.00193***
					(0.000731)
Observations	8,165	8,165	8,165	8,165	8,165

Note: Dependent variable is one if a compound is developed in-house, and zero otherwise. All specifications include experience, scope, development phase indicators, and a full set of therapeutic category and year indicators. We control for future market share based on future drug sales in the same therapeutic category by the same firm, assuming perfect fullsight and average development length as in DiMasi et al. (2003). Panels A and B use merged samples of our pipeline and sales data. Panel A includes a larger sample due to imputing missing MSP values as zero for firms without positive sales in a given therapeutic area, assuming their absence indicates a lack of sales activity. Panel B focuses on firms with positive sales in the therapeutic area of interest and does not impute MSP for those without sales data, recognizing that IMS data may not capture the universe of drug sales. EOP1 indicates 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. LOP1 is the length of the longest patent among compounds in the same therapeutic class and same firm. LOP2 is the sum of the patent lengths among compounds in the same therapeutic class and same firm. Standard errors (in parentheses) are heteroskedasticity-robust and clustered at the compound level. $*p < 0.1$, $**p < 0.05$, $***p < 0.01$.

Table C.7: Logit Models of Alternative In-house Development: Existence and Length of Patents

	(1) In-house	(2) In-house	(3) In-house	(4) In-house
EOP1	0.500***			
	(0.0605)			
EOP2		0.0585***		
		(0.00539)		
LOP1			0.0266***	
			(0.00358)	
LOP2				0.00403***
				(0.000341)
Experience	0.00150***	-0.000140	0.00156***	0.000447***
	(0.000146)	(0.000134)	(0.000148)	(8.80e-05)
Scope	-1.402***	-1.302***	-1.537***	-1.274***
	(0.150)	(0.149)	(0.149)	(0.149)
Phase I	-0.648***	-0.614***	-0.654***	-0.625***
	(0.0810)	(0.0807)	(0.0806)	(0.0808)
Phase II	-0.834***	-0.756***	-0.844***	-0.780***
	(0.0723)	(0.0728)	(0.0724)	(0.0728)
Phase III	-1.453***	-1.374***	-1.481***	-1.412***
	(0.0930)	(0.0935)	(0.0929)	(0.0940)
Launched	-2.092***	-1.943***	-2.093***	-1.986***
	(0.128)	(0.124)	(0.126)	(0.124)
Observations	109,115	109,115	109,115	109,115

Note: Dependent variable is one if a compound is developed in-house by the end of phase I, and zero otherwise. In contrast to our main dependent variable, which indicates whether a compound is developed in-house by the end of phase II, this alternative in-house measure aims to address the concern that the design and nature of a drug may be fixed as early as the completion of phase I testing. All specifications include experience, scope, development phase indicators, as well as a full set of therapeutic category and year indicators. Standard errors (in parentheses) are heteroskedasticity-robust and clustered at the compound level. $*p < 0.1$, $**p < 0.05$, $***p < 0.01$.

Table C.8: Patent Profile Variables Defined on Finer Therapeutic Classifications

	(1) In-house	(2) In-house	(3) In-house	(4) In-house
EOP1	0.644***			
	(0.0607)			
EOP2		0.0965***		
		(0.0120)		
LOP1			0.0300***	
			(0.00370)	
LOP2				0.00622***
				(0.000766)
Experience	0.00175***	-1.04e-05	0.00220***	0.000967***
	(0.000309)	(0.000300)	(0.000331)	(0.000245)
Scope	-1.606***	-1.732***	-1.804***	-1.748***
	(0.166)	(0.165)	(0.167)	(0.166)
Phase I	-0.698***	-0.660***	-0.702***	-0.675***
	(0.0828)	(0.0823)	(0.0822)	(0.0824)
Phase II	-1.088***	-1.030***	-1.107***	-1.057***
	(0.0717)	(0.0717)	(0.0716)	(0.0717)
Phase III	-1.522***	-1.469***	-1.564***	-1.506***
	(0.0944)	(0.0942)	(0.0941)	(0.0940)
Launched	-2.196***	-2.097***	-2.208***	-2.138***
	(0.129)	(0.127)	(0.127)	(0.127)
Observations	107,098	107,098	107,098	107,098

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. This table reports the results of constructing drug profile variables based on the second Anatomical Therapeutic Chemical (ATC) level, capturing finer therapeutic classifications. The sample is slightly smaller than in the main analysis, as some observations are dropped due to the more demanding fixed effects of the finer categories. Standard errors (in parentheses) are heteroskedasticity-robust and clustered at the compound level. $*p < 0.1$, $**p < 0.05$, $***p < 0.01$.

Table C.9: Robustness Checks: Firm Types and Focused Samples

	(1) In-house	(2) In-house	(3) In-house	(4) In-house
Panel A: with firm type fixed effects				
EOP1	0.503***			
	(0.0603)			
EOP2		0.0599***		
		(0.00537)		
LOP1			0.0271***	
			(0.00358)	
LOP2				0.00411***
				(0.000338)
Observations	109,115	109,115	109,115	109,115
Panel B: exclude CROs/CDMOs				
EOP1	0.504***			
	(0.0605)			
EOP2		0.0602***		
		(0.00538)		
LOP1			0.0269***	
			(0.00360)	
LOP2				0.00410***
				(0.000340)
Observations	108,846	108,846	108,846	108,846
Panel C: sample of biopharmaceutical firms				
EOP1	0.523***			
	(0.0624)			
EOP2		0.0593***		
		(0.00537)		
LOP1			0.0287***	
			(0.00368)	
LOP2				0.00411***
				(0.000339)
Observations	102,986	102,986	102,986	102,986

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. Panel A reports the results after adding firm type fixed effects to account for potential differences in originator's business models. Panel B excludes compounds developed by or with assistance from CROs/CDMOs (Contract Research Organizations/Contract Development and Manufacturing Organizations). Panel C restrict the sample to projects developed by biopharmaceutical firms. Standard errors (in parentheses) are heteroskedasticity-robust and clustered at the compound level. $*p < 0.1$, $**p < 0.05$, $***p < 0.01$.

Table C.10: Robustness Checks: Firm Portfolios, Firm Fixed Effects, and Speed Test

	(1) In-house	(2) In-house	(3) In-house	(4) In-house
Panel A: with extra firm portfolio measures				
EOP1	0.406***			
	(0.0626)			
EOP2		0.0773***		
		(0.0100)		
LOP1			0.0198***	
			(0.00378)	
LOP2				0.00514***
				(0.000568)
Observations	109,115	109,115	109,115	109,115
Panel B: with firm fixed effects				
EOP1	0.0396			
	(0.0859)			
EOP2		0.0206***		
		(0.00598)		
LOP1			-0.00275	
			(0.00496)	
LOP2				0.00157***
				(0.000371)
Observations	105,224	105,224	105,224	105,224
Panel C: development speed test (drug-level)				
	pre-III	pre-II	pre-I	II-III
In-house	-0.848***	-0.783***	-0.655***	-0.380**
	(0.199)	(0.128)	(0.114)	(0.161)
Observations	593	1,202	1,318	436

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. Panel A reports results when controlling for additional portfolio measures, including the total number of products in a firm's pipeline in a given therapeutic category each year and the total number of competing products each year in the same therapeutic category of the compound of interest by other firms. Panel B reports results when including firm fixed effects, and with reduced observations due to projects owned by firms with single/small projects. Panel C reports regression in drug-level data using years between phases as the dependent variable and in-house development status as the main covariate of interest. Columns 1-4 have outcome variables as years between pre-clinical to phase III, pre-clinical to phase II, pre-clinical to phase I, and phase II-III, respectively. Standard errors (in parentheses) are heteroskedasticity-robust and clustered at the compound level. $*p < 0.1$, $**p < 0.05$, $***p < 0.01$.

Table C.11: Robustness Check: Other Subsample Analyses

	(1) In-house	(2) In-house	(3) In-house	(4) In-house
Panel A: focused sample by the end of phase II				
EOP1	0.526***			
	(0.0607)			
EOP2		0.0612***		
		(0.00559)		
LOP1			0.0277***	
			(0.00359)	
LOP2				0.00425***
				(0.000355)
Observations	102,037	102,037	102,037	102,037
Panel B: subsample with no ownership changes				
EOP1	0.496***			
	(0.0631)			
EOP2		0.0692***		
		(0.00635)		
LOP1			0.0278***	
			(0.00373)	
LOP2				0.00492***
				(0.000408)
Observations	101,352	101,352	101,352	101,352
Panel C: adjust for potential censoring per phase				
EOP1	0.572***			
	(0.0658)			
EOP2		0.0584***		
		(0.00571)		
LOP1			0.0296***	
			(0.00390)	
LOP2				0.00393***
				(0.000348)
Observations	98,842	98,842	98,842	98,842

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. Panel A reports the results when focusing on the subsample containing all observations by the end of phase II. Panel B examines the subsample with no ownership changes, such as merge and acquisitions. Panel C report a subsample where we drop the observations that are potentially right censored given the lengthy drug development process. Standard errors (in parentheses) are heteroskedasticity-robust and clustered at the compound level. $*p < 0.1$, $**p < 0.05$, $***p < 0.01$.