

# Capturing Knowledge within and across Firm Boundaries: Evidence from Clinical Development

By PIERRE AZOULAY\*

*How do firm boundaries influence employees' acquisition of information? Using detailed project-level data and qualitative evidence, I document that pharmaceutical firms are more likely to outsource the coordination of data-intensive clinical trials, while they are more likely to assign knowledge-intensive trials to internal teams. Managers do not choose between market and hierarchy, but between the hierarchy of the firm—in which subjective performance evaluations are combined with flat explicit incentives—and the hierarchy of its subcontractor—whose virtue stems precisely from the ability to provide high-powered incentives on a narrow set of monitorable tasks. (JEL J41, L22, L65, O32)*

Managers often emphasize the difficulties involved in capturing knowledge created within the confines of their suppliers' hierarchy, particularly if this knowledge is complex or idiosyncratic. Echoing this intuition, early research on firm boundaries viewed firms primarily as communities specialized in the creation and transfer of knowledge (Kenneth J. Arrow, 1974). But why should managers “know more” about the activities of their own employees than they do about those of their subcontractors? In this paper, I study how employees' incentives change with the integration decision, and how these incentives in turn influence the acquisition of different types of information.

I focus on the outsourcing of a particular information-intensive activity: the coordination

of clinical trials sponsored by pharmaceutical companies. Since the mid-1980s, pharmaceutical firms have partly contracted out the operational aspects of clinical trials to specialized contractors called contract research organizations (CROs). I argue that variation in project characteristics leads to variation in the relative importance of employees' effort on two tasks that compete for their attention: *data production*—the routine manipulation, storage, and transfer of symbolic information within established categories; and *knowledge production*—the establishment of novel conceptual categories, hypotheses, and causal associations (Lars Osberg et al., 1989). Firms respond to this variation by choosing to outsource or insource the project. Each option entails a different allocation for a key decision right that defines the ways in which firms affect the employee's choice of effort and its allocation across different tasks. Since it is harder to measure performance in knowledge production than to measure performance in data production, the firm chooses to assign its own employees to projects for which the former are relatively more important than the latter. In so doing, it ensures that incentives for both tasks are kept in balance. Conversely, data-intensive projects are more likely to be outsourced, as distorting the allocation of effort is less costly for these projects. This argument implies that heterogeneity in project characteristics will be systematically related to the integration decision.

\* Graduate School of Business, Columbia University, 3022 Broadway, 704 Uris Hall, New York, NY 10027 and National Bureau of Economic Research (e-mail: pa2009@columbia.edu). This paper is a revised version of the second chapter of my dissertation. I thank my committee—Rebecca Henderson, Robert Gibbons, and Scott Stern—for their guidance, as well as Nick Argyres, David Autor, Ernie Berndt, Matthew Bidwell, Rajeev Dehejia, Ray Fisman, Jeff Furman, Luis Garicano, Tom Hubbard, Casey Ichniowski, Josh Lerner, Chiaki Moriguchi, Brian Silverman, Manuel Trajtenberg, Birger Wernerfelt, Sid Winter, and two anonymous referees for valuable comments and suggestions. Financial support from the Program of the Pharmaceutical Industry, the Industrial Performance Center, and the Center for Innovation and Product Development at MIT is gratefully acknowledged. The usual disclaimer applies.

Using detailed data on 6,826 clinical trials sponsored by 53 firms between 1995 and 1999, I document that even after controlling for a number of alternative explanations, *knowledge-intensive* projects are more likely to be assigned to internal teams, while *data-intensive* projects are more likely to be outsourced. The credibility of this finding hinges on the ability to distinguish empirically between these two types of projects. Variation in the location of research activity provides a rich source of heterogeneity at the project level, as clinical trials typically involve a variety of different research sites: academic medical centers, private practices, community hospitals, or even dedicated, for-profit research centers. Because of the peculiar career incentives in academic medicine, I argue that the proportion of academic investigators in a clinical study captures the relative importance of knowledge-production activities relative to data-production activities. Although this variable provides the most robust findings, several other project characteristics yield similar conclusions. These results fit with Bengt Holmström's (1999) subeconomy view of the firm, according to which balanced yet "sluggish" incentives constitute firms' mixed blessing.

The statistical exercise is complemented by in-depth interviews with personnel of six pharmaceutical and biotechnology firms, as well as employees of seven CROs. The qualitative evidence enables me to reconcile the subeconomy view with practitioners' intuition that low-powered *explicit* incentives need not be inconsistent with creativity and initiative inside firms. These characteristics figure prominently in accounts of life inside organizations (Kirk Monteverde, 1995), and firms' superior ability to create and transfer knowledge are often ascribed to them. I find that firms use relational contracts to ensure that employees' incentives are both balanced and relatively high-powered. The choice is not between market and hierarchy, but between the hierarchy of the firm—in which subjective performance evaluations are combined with flat explicit incentives—and the hierarchy of its subcontractor—whose virtue stems precisely from the ability to provide high-powered incentives on a narrow set of monitorable tasks.

The rest of the paper proceeds as follows. In Section I, I describe the institutional setting and

the informational constraints on contracting in clinical development. Drawing on the qualitative evidence, Section II develops specific hypotheses relating project heterogeneity to the integration decision. Section III presents the data, descriptive statistics, and main econometric results. Concluding remarks are offered in Section IV.

## I. Tapered Integration in Clinical Development

In order to gain regulatory approval for market introduction, the Food and Drug Administration (FDA) requires that a pharmaceutical company provide substantial evidence of a drug's effectiveness, through adequate and well-controlled clinical investigations. Although the precise details of these requirements have evolved over the years, proof of effectiveness must be provided by the results of randomized controlled trials.

In contrast to early-stage discovery research, pharmaceutical firms contract out experimental human studies to independent physicians called *clinical investigators*, and rely on *clinical monitors* to recruit, coordinate, and supervise the large number of investigators required to conduct experimental human studies. Clinical monitoring is an important element of "Good Clinical Practices" (Department of Health and Human Services [DHHS], 1997), a set of rules mandated and enforced by regulatory authorities around the world. Since this paper studies the outsourcing of this activity, it is worth understanding precisely its characteristics as a distinct occupation.

### A. Clinical Development as an Information-Intensive Activity

Clinical monitors are prototypical information workers. Their output can be condensed into the package insert that comes bundled with the prescription drug sold to final consumers. The information produced in clinical development is of two types: data—information that can be transmitted reliably once the rules required for decoding it are known; and knowledge—the establishment of novel conceptual categories, hypotheses, and causal associations.

According to their narrow job description, clinical monitors ensure that the information

that originates from the doctors participating in a clinical study is accurate (in the sense that the case report forms submitted to the regulatory authorities reflect exactly the source documents, i.e., the original patient records and charts) and is structured in accordance with standards laid out by the regulatory agency and the particular firm sponsoring the trial. These unglamorous tasks are critical to the extent that skimping on data quality can jeopardize the scientific credibility of a new drug application and damage the reputation of the firm with regulators. As such, clinical monitoring involves the routine manipulation, storage, and transfer of symbolic information within established categories (Osberg et al., 1989).

Clinical monitors also produce knowledge—search rules or heuristics that identify problems and the elements consisting of their solution (Bruce Kogut and Udo Zander, 1992). Much of it often remains tacit, in that the formal expression of the solution is unlikely to capture fully the procedures that led to it (Michael Polanyi, 1966). The process of knowledge production consists of interactive tasks—such as mediating disputes or gaining the trust of third parties and subordinates—that are very difficult to reduce to a sequence of “if-then-do” steps. According to this broader job definition, clinical monitoring cannot be reduced to the mass-production of clinical data geared toward regulatory approval.

The significance of knowledge production in clinical development is probably best understood through example. The following anecdote was reported by the vice president for clinical affairs at a biotechnology company. To perform a clinical trial for a drug used to treat pulmonary hypertension (a life-threatening condition), this company recruited investigators from academic medical centers and private clinics. Puzzled by a very slow patient enrollment in the study, the clinical team leader asked the monitors to investigate and identify the root cause of the problem. The monitors noticed a pattern in the patient enrollment statistics: the for-profit centers were lagging far behind the academic centers. Upon further investigation, the monitors determined that the delays were caused by the requirement that patients undergo a lung biopsy to determine their eligibility. In academic centers, this test could be performed and the results provided to patients on the same day, while in

the private clinics, confirmation of eligibility was postponed for several days because of the need to ship biopsy samples to an outside laboratory for analysis. With this information, the study team amended the protocol to substitute a type of specialized X-ray for the biopsy requirement. The ability to perform the screening of patients in a single day resulted in a much-improved enrollment process.

This anecdote illustrates the importance of knowledge production in clinical development. A particularly important challenge is to acquire, synthesize, and assimilate information generated locally by clinical investigators, either through direct interaction with patients or by purposeful analysis. Since they are in contact with the complete set of investigators participating in a clinical study, monitors can bring information and problem-solving capabilities together at a single locus (Eric von Hippel, 1998), sometimes resulting in the production of idiosyncratic knowledge about experimental compounds and the diseases they are intended to treat.<sup>1</sup>

The distinction between data and knowledge production is essential, as monitors need to allocate effort across both tasks, and these activities differ in the extent to which inputs can be monitored and performance measured and contracted upon. The problem of balancing incentives across tasks is brought under sharper focus by the outsourcing practices of pharmaceutical firms.

### *B. Stylized Facts of Clinical Development Outsourcing*

Since the mid-1980s, monitoring and data management (along with other ancillary operational tasks) have been increasingly outsourced to CROs. The contract research industry is currently growing at the brisk pace of 20 percent per year, and it is estimated that up to one-fifth of overall clinical development budgets are captured by these independent firms, representing

<sup>1</sup> In the extreme, clinical monitors may become conduits for serendipitous medical findings. Annetine C. Gelijns et al. (1998) provide numerous instances of unexpected benefits from medical research, of which the most recent example is probably the case of Viagra and heart disease (James Kling, 1998).

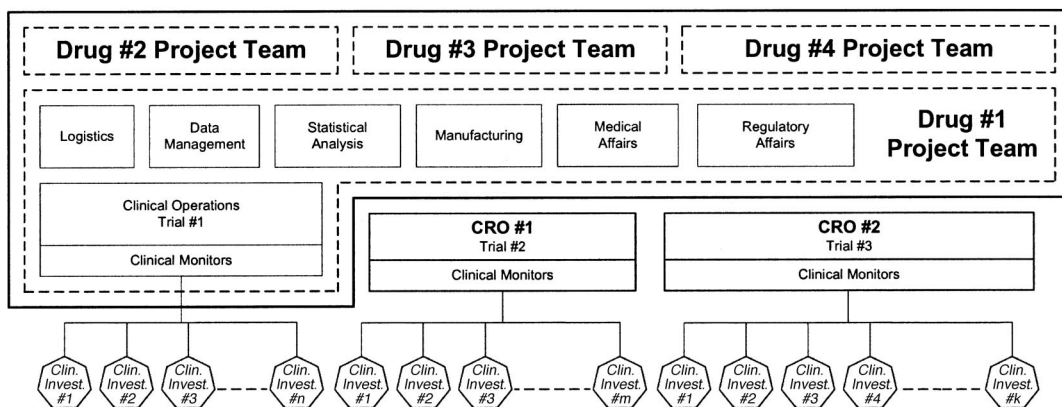


FIGURE 1. ORGANIZATION OF THE CLINICAL TRIALS PROCESS

slightly less than \$7 billion in 2001 (Parexel, 2003/2004). Clinical development outsourcing differs from other arm's-length contractual settings in that the boundaries of the firm can shift on a project-by-project basis.<sup>2</sup> CRO teams are generally in charge of key operational aspects of the study, with little or no input into knowledge production. Moreover, for a given study, sponsors may choose to retain some functions in house while contracting out others to a single or multiple CROs. Other characteristics of vertical relationships in clinical development deserve mention.<sup>3</sup>

Most important, clinical development provides an instance of *tapered integration*, whereby a firm produces some of its requirements using its own labor, while contracting for the rest on the market (Dennis W. Carlton, 1979). Using data that will be described in more

<sup>2</sup> Throughout the paper, the words *study*, *trial*, and *project* are synonymous. Similarly, the words *drug*, *compound*, and *molecule* are used interchangeably.

<sup>3</sup> Figure 1 provides a summary schema of the clinical trials process. It depicts a pharmaceutical firm (whose boundary is drawn using the thick solid line) concurrently developing three drugs, using three distinct cross-functional product teams (whose boundaries are drawn using the thin dotted line). The box corresponding to team #1 has been enlarged, revealing the different functions necessary to the successful development of drug #1. For each drug, there is a portfolio of trials (three in the case of drug #1, two of which—trial #2 and trial #3—are outsourced to two different CROs). The clinical monitors (inside both the pharmaceutical firm and the CROs) oversee the gathering of clinical data that takes place in a variety of clinical sites (hospitals, private clinics, academic medical centers) under the stewardship of clinical investigators.

detail below, I compute the fraction of activity that is outsourced in a given year as the proportion of clinical sites in the firm's portfolio corresponding to projects coordinated by CROs. The distribution of the resulting variable, *%CRO*, is displayed in Figure 2. It is relatively uniform, but with a mass point at 0: a sizable number of firms outsource none of their trials *in any given year*. Figure 3, which graphs the distribution of the mean outsourcing propensity for the period 1995–1999, makes clear that this mass point is not an artifact of firm effects. Only five firms refrain from using CROs during the whole period, while only two contract out all of their trials. Most firms seem to use CROs moderately: the mean outsourcing propensity is 29 percent, the median 20 percent.

Three issues that are often the focus in studies of firm boundaries—liability protection, asset ownership, and appropriability—are minor considerations in pharmaceutical firms' outsourcing decisions. First, CROs do not shield pharmaceutical firms from responsibility for the integrity of the studies they sponsor (DHHS, 1997). In this regard, these suppliers are very different from accounting firms, which can be held liable for the accuracy of their audits.<sup>4</sup>

<sup>4</sup> In a phone conversation with Dr. Murray Lumpkin, deputy director of the FDA's Center for Drug Evaluation Research, it was apparent that regulators do not perceive situations where sponsors monitor their own studies as raising conflict-of-interest issues. Although the FDA collects information about study monitoring, FDA medical examiners are generally unaware whether a study they are reviewing has been outsourced. Moreover, pharmaceuti-

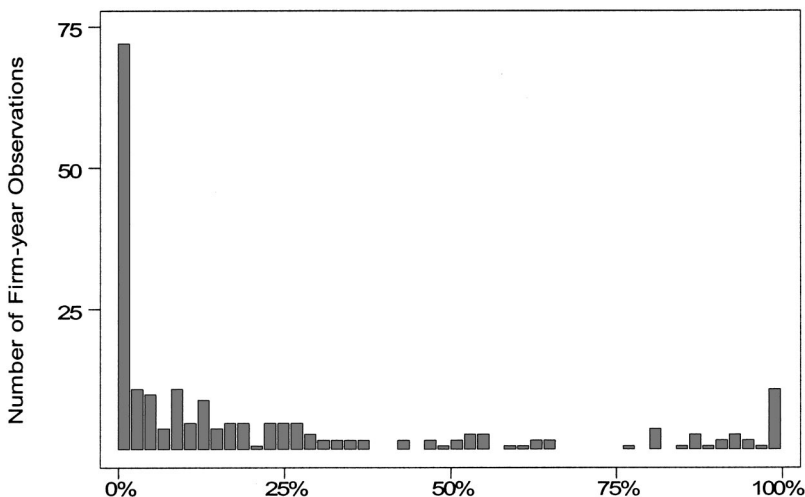


FIGURE 2. PROPORTION OF CLINICAL SITES IN OUTSOURCED TRIALS, 1995-1999

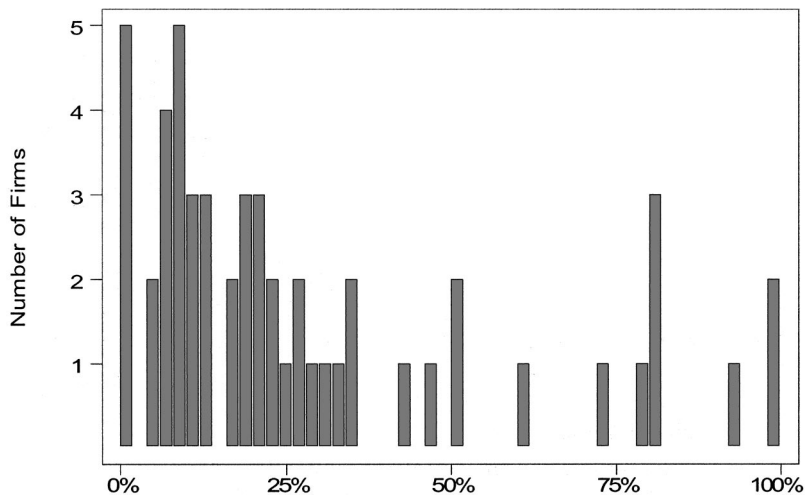


FIGURE 3. MEAN FIRM-LEVEL OUTSOURCING PROPENSITY, 1995-1999

Second, the physical assets used in production (such as computers) are nonspecific, and asset ownership does not appear relevant for incentives. This is important insofar as the recent literature on firm boundaries identifies integration with residual rights of control over physical assets (Oliver Hart, 1995). Finally, appropri-

ability concerns do not loom large in the outsourcing decision. Clinical development outsourcing contracts, unlike licensing deals between established pharmaceutical firms and biotechnology startups, involve no sharing of intellectual property rights. Moreover, CROs are not viable competitors on the product market.

Outsourcing in clinical development is not defined by asset ownership, certification, or the right to determine output or prices. This asser-

cal firms themselves do not appeal to a credibility rationale to justify their decision to outsource.

tion makes it all the more imperative to state precisely what distinguishes a subcontractor from an internal division. This is discussed in the next section.

## II. Outsourcing, Incentives, and Project Heterogeneity

Field research revealed that employee assignment was a significant bone of contention between buyers and suppliers. Outsourcing clinical development involves relinquishing a key decision right to suppliers: the ability to match particular employees to specific projects. In fact, it is useful to take the transfer of this decision right from the buyer to the supplier as the definition of outsourcing in clinical development. Interviews indicated that the identity, skill level, and job design of CRO employees could not be easily contracted upon. In the words of an outsourcing manager, "CROs keep giving us bad people to choose from, and there is nothing I can do about it." Others highlighted instances where a team of seasoned veterans had been promised by a CRO, but "rookies" had been substituted for the "A-team" at the last minute. This bait-and-switch tactic appeared to be relatively common. Formal contracts did not solve the issue because buyers—or a court of law—could not distinguish voluntary turnover from opportunistic reassignment to another client.

Focusing on employee assignment is also useful because it brings into focus how employee incentives change with the integration decision. Data production activities can be partitioned into a set of simple, monitorable tasks whose output can be contracted upon (albeit in a distortionary fashion). Knowledge production, on the other hand, is heavily influenced by employees' ability to learn from the previous stages of a project and to apply this knowledge creatively during subsequent stages. Moreover, project-specific learning is deeply embedded in the particular routines, structures, and processes used by the firm to translate knowledge into decisions. This means that simply measuring performance is arduous; inferring effort devoted to knowledge production from observable signals of performance is even more so. In this respect, clinical development can be analyzed through

the lens of a multitask agency problem (Holmström and Paul Milgrom, 1991): when two substitute tasks compete for the attention of an agent, and these tasks vary in the extent to which effort is observable, it may be optimal to provide low-powered incentives for both tasks, since high-powered rewards may skew the agent's allocation of effort away from the task that is costly to monitor. In the limit, with complete unobservability of effort on one task, it may be justified to provide no contingent rewards for any of the tasks. While performance on the monitorable task suffers, no effort distortion is induced by the incentive scheme.

Holmström and Milgrom (1994) elaborate on this framework and argue that principal ownership of assets, flat performance incentives, and exclusion from alternative tasks are complementary instruments that jointly characterize the employment relationship. They show that such a system of incentives will be used when balancing incentives is valuable, i.e., when it is hard or costly to measure performance on important tasks. Slight modifications are required to adapt these insights to the setting of drug development. I argue that formal contracts can specify which party gets to make skill-project matches, but not how these matches are made as the project unfolds. As pointed out by Eric Maskin and Jean Tirole (1999), a decision right with this characteristic can play a role akin to that of asset ownership in Holmström and Milgrom (1994) and Hart (1995). Similarly, limiting the agent's attention to a single project is equivalent to excluding alternative tasks.

Could pharmaceutical firms deal with multi-task incentive problems within their boundaries, for example by providing high-powered incentives to employees in their "data" division, and lower-powered—but balanced—incentives to employees in their "knowledge" division? Firms would probably find it difficult to allocate irrevocably decision rights regarding employee assignment to their "data" division. Practitioners were skeptical of dual internal labor markets for two additional reasons. First, they worried about the legal consequences of such arrangements, since courts may not grant firms the right to treat workers sharing the same narrow job description differently, even if actual

job requirements differ substantially between the two categories of employees. Further, they worried that equity norms would lead to confusion and discontent among employees of both divisions. The fact that CRO employees were never mixed with firm employees within a project is consistent with the idea that equity concerns lead pharmaceutical firms to segregate their data workers in a separate firm with which they enjoy an arm's-length relationship.

#### A. *Qualitative Evidence at the Firm Level*

Over a two-month period, I visited six pharmaceutical and biotechnology companies and interviewed representatives of seven suppliers over the phone or in industry conferences. The interview data provide rich (sometimes vivid) *firm-level* evidence regarding the relationship between job design and incentives in clinical development. It is less convincing as a source of *project-level* evidence, but is helpful in suggesting hypotheses that can be fruitfully tested using project-level data.

The managers making staff deployment decisions inside the pharmaceutical firm and their counterparts inside the CRO did not share the same objectives. Inside the firm, managers strived to produce all the information required to reach the correct "go/no-go" decision for a particular *drug*, subject to a budget constraint which includes "time costs" (i.e., foregone sales opportunities because the drug, whose patent clock is ticking, lingers in development). Inside the CRO, managers minimized time-to-project completion, subject to two constraints: data quality had to exceed some minimum standard, and billable hours for the firm as a whole had to reach some target. This second constraint arose from the fact that CROs aggregate their clients' peak demand for clinical development services. This could be profitably achieved only to the extent that monitors were drawn from a common pool and assigned on a "first-come-first-served" basis to projects.<sup>5</sup> In

<sup>5</sup> This discussion assumes that in-house managers do not worry about keeping staff busy. To a large extent, this reflects the fact that CROs act as buffers against pipeline shocks. While a more complete exploration of this idea lies outside the scope of this paper, it is studied in depth elsewhere (Azoulay, 2003a).

contrast, in-house monitors specialized therapeutically, and assignment decisions took idiosyncratic expertise into account. For example, if a monitor had worked on previous studies for a drug, he or she was very likely to be assigned to the follow-on study as well.

Through its effect on the ability to match individual skills with specific projects, outsourcing skewed clinical monitors' attention and effort away from knowledge production toward data-processing tasks. A manager overseeing suppliers at a large pharmaceutical firm described the unbalanced incentives of CRO employees in these terms:

"... We have a line-by-line definition of the CRO's responsibilities. That means that the CRO is less likely to notice stuff that might be going on at the sites. There are no incentives for the individuals at CROs for capturing soft data, unlike here, where you get rewarded at every level. At a CRO, you might work for two or three sponsors at the same time. So it's all about hard deliverables. Anything beyond the contract you do not get."

CRO employees confirmed the importance of individual, explicit incentives. The data-production process generates a large number of observable performance measures. For example, it is possible to monitor the number of case report forms checked by a clinical monitor in a given amount of time; data inconsistencies give rise to queries that reflect directly on the thoroughness of source-document verification; and query resolution times measure the efficiency of the troubleshooting process when data inconsistencies arise. While pay was not directly contingent on these metrics, I found that bonuses and advancement in the CROs' internal labor markets were heavily dependent on them. Indeed, in interviews with former CRO employees, suppliers were characterized as "data sweatshops," and their employees as "data mules" or "CRF collection machines."

Buyers often included penalties into contracts for on-time delivery of the data, along with stipulations regarding acceptable data quality. These incentives trickled down to the supplier's individual employees. Rewarding speed in clinical development could be counterproductive: a slowly performed clinical trial justifying FDA

approval contributes much more to firm value than a rapidly performed trial leading to rejection of the new drug application by the regulator. Pharmaceutical sponsors appeared aware of the fact that “paying for speed” may have unintended consequences, and therefore used rules of thumb to determine which studies were good outsourcing candidates—studies for which effort distortions would not be too costly.

In contrast to CROs, pharmaceutical firms relied on subjective performance evaluations as a source of employee incentives. Since knowledge production was difficult for employees and pharmaceutical firms to contract upon, it tended to be rewarded through the use of implicit incentives such as subjective bonuses and promotions. The ability to provide such implicit rewards appeared inherently tied to the ability to match individual skills with specific projects.

Moreover, job design differed across buyers and suppliers along one critical dimension. While employees of pharmaceutical firms dedicated their time fully to a single project (or at least to a single drug), CROs assigned employees to multiple projects at a time and shifted them laterally from project to project as the need arose. By increasing the informativeness of commonly used performance measures, multi-project assignment kept down the cost of monitoring data-production activities. What is known in the industry as the *data-management process* provides a good illustration of this phenomenon. After the data have been collected, but before the study database is “locked,” a number of cross-checks are applied to verify the data’s integrity. Whenever an inconsistency arises, a query is generated, and the source document (the original patient chart) is checked for accuracy. This is a very time-consuming and costly process. Queries can be traced back to particular monitors and reflect on the thoroughness of their work, but are a noisy measure of individual performance. The number of queries is also influenced by the quality of the coordinating staff at the clinical sites, whether patients had unusual complications, etc. CRO managers can, however, potentially meter queries originating from multiple sites monitored by the same employee over the same time period. For data-production activities, extracting signal from noise is therefore easier inside the CRO than inside the pharmaceutical firm, where

monitors are assigned to a single project at a time.

In summary, the qualitative evidence is consistent with the view that flat performance incentives, subjective performance evaluations, and single-project assignments characterize the employment relationship inside pharmaceutical firms. In contrast, CRO hierarchies are characterized by steep individual incentives for speed and data quality and simultaneous assignment to multiple projects for multiple clients. This analytic framework parallels that used by George P. Baker and Thomas N. Hubbard (2003). In the context of the trucking industry, these authors argue that outsourcing provides strong incentives with respect to utilizing trucks intensively, and this interferes with job designs that give drivers non-driving responsibilities. Here, I argue that outsourcing provides CROs strong incentives that interfere with job designs that give clinical monitors knowledge-production responsibilities.

One might wonder why relational contracts between pharmaceutical firms and their suppliers cannot govern the assignment of workers to projects. In practice, there is overwhelming evidence that buyers use a large number of different suppliers, often replace current suppliers with new ones, and more generally seize the opportunities provided by intense competition upstream (Azoulay, 2003b). Clients may find it to their advantage to segregate their project portfolio between insiders and CRO employees precisely because outsourcing distorts the allocation of effort across tasks. CROs may be “data sweatshops,” but sweatshops have their uses: the provision of high-powered incentives on a narrow set of monitorable tasks.

### B. *Project-Level Hypotheses*

In order to analyze outsourcing patterns at the transaction level of analysis, I argue that variation in project characteristics leads to variation in the relative importance of employee effort on data-production and knowledge-production activities. It also leads to variation in the extent to which effort on each of these activities is observable. Since it is harder to measure performance in knowledge production than in data production, the firm chooses to assign its own employees to projects for which the marginal

return to the former is higher than the marginal return to the latter. In so doing, it ensures that incentives for both tasks are kept in balance for these knowledge-intensive projects. Conversely, data-intensive projects—those studies for which the marginal return to data-production effort is particularly high—are more likely to be outsourced, as distorting the allocation of effort is less costly and may even be beneficial because of the high-powered incentives faced by CRO employees. Similarly, project-level variation in the observability of effort influences the integration decision. Holding the marginal returns to effort constant, less costly monitoring of data production (knowledge production), should lead to a higher (lower) propensity to outsource a project. This leads to the following hypotheses:

**Hypothesis 1:** The higher the marginal return on data-production effort relative to knowledge-production effort, the more likely is the project to be outsourced, *ceteris paribus*.

**Hypothesis 2a:** The propensity to outsource is increasing in the project-level costs of monitoring data production, *ceteris paribus*.

**Hypothesis 2b:** The propensity to outsource is decreasing in the project-level costs of monitoring knowledge production, *ceteris paribus*.

In the next section these hypotheses are tested using detailed project-level data for a large sample of firms.

### III. Empirical Model and Results

The foregoing discussion suggests that the integration decision will be a function of project and firm characteristics. I address these predictions by estimating a binary choice model. Because the estimation strategy is affected by data availability, this section begins with a description of the data. Results from the econometric specification are then presented.

#### A. Data and Sample Characteristics

The data used in this essay come from a dataset of clinical investigator contracts made available by FastTrack Systems, Inc. Since

1991, clinical trial information has been gathered from subscribing pharmaceutical companies in order to help them better plan and negotiate investigator grants. Moreover, since 1995, FastTrack has recorded whether a CRO was involved in each trial. The breadth of the data's longitudinal and cross-sectional coverage is impressive, with 137,000 clinical investigator contracts corresponding to 17,000 clinical trials for the period 1991–1999. While the sample of firms is unbalanced (coverage improved as the service gained popularity), it comprises 90 subscribing units. Although no company can be identified by name under our confidentiality agreement, the sample includes nearly all large pharmaceutical companies (U.S.- and foreign-based), as well as most large biotechnology firms. Given the propensity of biotechnology start-ups to rely on established firms' clinical capabilities, the sample is highly representative of the industry as a whole.<sup>6</sup>

Since information about outsourcing is available only for the last five years of data, aggregating observations to the clinical-trial level results in 6,826 clinical studies, sponsored by 53 firms, for the development of 1,039 experimental compounds. For each project, the dataset records a number of study characteristics, the name of the molecule, an identifier for the firm sponsoring the study, and a dummy variable that denotes the involvement of a CRO. I complement these data with information from company annual reports.

The strengths of the data lie in their broad industry coverage and fine-grained level of detail, but they are not without limitations. The information on outsourcing is summarized by a

<sup>6</sup> I thank Mark Hovde, Anca Serban, and Mariann Best for explaining the subtleties of the data. Differences between “subscribers” and “firms” stem from the fact that certain firms have several subscriptions (e.g., ABC Pharma-UK and ABC Pharma-US). Firms resulting from a merger are assigned a distinct subscription identifier, a convention the usefulness of which cannot be stressed enough in the context of the pharmaceutical industry. The sample comprises data from 9 companies out of the top 10 firms, 26 out of the top 30 firms, and 33 out of the top 50 firms, where the rankings reflect R&D spending listed in annual reports to shareholders in the year 2000. Companies in the sample spent a total of \$41.4 billion in R&D that year. This corresponds to 82 percent of the aggregate amount reported by the top 45 biggest spenders.

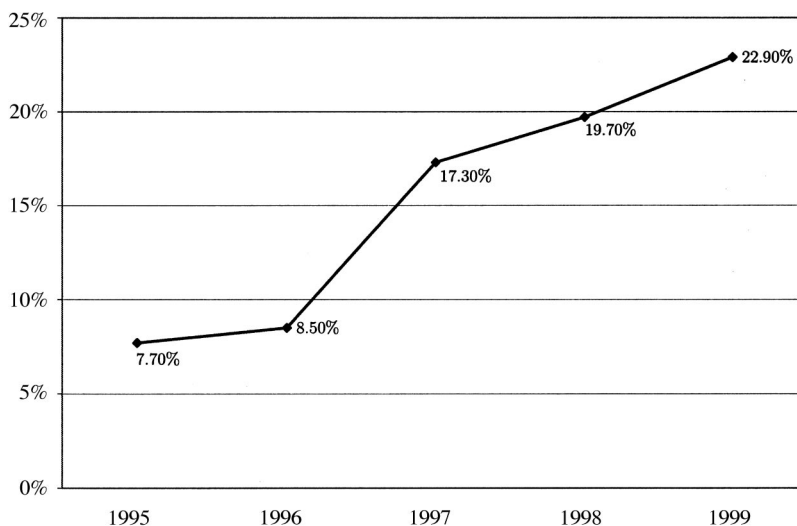


FIGURE 4. PROPORTION OF CLINICAL STUDIES OUTSOURCED, 1995–1999

dummy variable, which does not capture all the richness and heterogeneity of outsourcing practices at the project level (e.g., no information is available to describe the substance of the duties performed by the CRO during the project). Moreover, the clinical trials included in the dataset do not constitute the universe of studies sponsored by these firms. Excluded from the data are any trials for which investigator contracts are negotiated directly by the CRO. In other words, the statistical exercise can reveal information only on a specific contracting margin: the decision to outsource, *conditional on having retained the selection of clinical investigators in-house.*<sup>7</sup>

### B. Descriptive Statistics

For the industry as a whole, Figure 4 shows that the proportion of clinical studies outsourced tripled during the sample period, from 7.7 percent in 1995 to 22.9 percent in 1999.

<sup>7</sup> According to practitioners, fully outsourced studies are more likely to be performed post approval and/or to examine the consequences of long-term maintenance treatment. One would expect these trials to be more data-intensive on average than the trials composing the sample analyzed here. If this is the case, excluding these studies makes it harder to detect an effect of knowledge considerations on the propensity to outsource.

This dramatic increase provides a useful reminder that outsourcing practices in drug development are relatively new and still evolving.

Tables 1 and 2 present descriptive statistics. Table 1 contains such information on firm-level variables as outsourcing intensity, pipeline shocks, and overall clinical research experience. Table 2 displays information pertaining to project-level variables, such as whether or not the trial was outsourced, the phase of the trial, the proportion of investigators involved who are academics, and other observable project characteristics.

Since traditional measures of firm size (such as sales or assets) do not seem appropriate, I use the number of clinical sites as the basic unit of activity in clinical development. Trials differ extensively in size (between 1 and 275 sites in this sample), and practitioners often use rules of thumb that assign a given number of sites to clinical monitors. Therefore, I measure firm size as the log of the total number of clinical sites in the firm's portfolio, averaged over the last three years; pipeline shocks are computed as the deviation between the total number of sites in the firm's portfolio from the past-three-year moving average; and firm-level experience in drug development is measured by the cumulative stock of clinical sites for the firm.

A cursory analysis of Table 2 reveals a number of empirical regularities. Outsourced studies

TABLE 1—DESCRIPTIVE STATISTICS  
Firm-Level Variables (53 firms, 211 firm-year observations)  
153 clinical studies per firm on average, min. = 1, max. = 929

		Obs.	Mean	Std. dev.	Min.	Max.	Firm mean ( <i>n</i> = 53)
<i>%CRO</i>	Proportion of clinical sites in outsourced trials for firm <i>j</i> in year <i>t</i>	211	0.242	0.318	0	1	0.290
<i>BIOTECH</i>	Firm not present on the product market as of 1995	211	0.052	0.223	0	1	0.113
<i>US PRESENCE</i>	U.S. subsidiary as of 1995	211	0.843	0.364	0	1	0.849
<i>SCOPE</i>	Therapeutic scope of firm <i>j</i> in year <i>t</i>	211	3.134	1.571	1	7.571	2.911
<i>SITES</i>	Number of clinical sites for firm <i>j</i> in year <i>t</i> (×100)	211	3.354	3.739	0.010	23.420	3.191
<i>SIZE</i>	Size of firm <i>j</i> in year <i>t</i> , three-year site moving average (×100)	211	3.301	3.224	0.057	19.863	3.061
<i>SHOCK</i>	Deviation of <i>SITES</i> from past-three-year moving average (×100)	211	-0.110	2.702	-9.900	12.211	0.012
<i>EXPERIENCE</i>	Cumulative number of clinical sites for firm <i>j</i> up to year <i>t</i> (×100)	211	18.290	16.390	0	61.74	16.110

TABLE 2—DESCRIPTIVE STATISTICS  
Project-Level Variables (6,826 obs.)

		Overall mean	Mean insourced projects ( <i>n</i> = 5,642)	Mean outsourced projects ( <i>n</i> = 1,184)	Min.	Max.
<i>CRO</i>	Was the study outsourced?	0.173 (.379)			0	1
<i>%AMC</i>	Proportion of academic clinical sites in the study	0.619 (.423)	0.653 (.411)	0.454 (.439)	0	1
<i>PHASE I</i>	Phase I study	0.244 (.429)	0.224 (.417)	0.340 (.474)	0	1
<i>PHASE II</i>	Phase II study	0.234 (.423)	0.243 (.429)	0.187 (.390)	0	1
<i>PHASE III</i>	Phase III study	0.437 (.496)	0.451 (.498)	0.374 (.484)	0	1
<i>PHASE IV</i>	Phase IV study	0.086 (.280)	0.083 (.275)	0.101 (.300)	0	1
<i>WELL-TREATED</i>	Top 10 best-treated diseases	0.115 (.319)	0.106 (.308)	0.157 (.364)	0	1
<i>INPATIENT</i>	Inpatient study	0.346 (.476)	0.331 (.471)	0.418 (.493)	0	1
<i>#PROCEDURES</i>	Number of distinct medical procedures	15.869 (9.320)	15.615 (9.128)	17.079 (10.106)	0	57
<i>#SITES</i>	Number of clinical sites	7.861 (15.270)	7.386 (14.102)	10.121 (19.768)	1	275
<i>HERF</i>	Herfindahl index of patient concentration across sites	0.514 (.400)	0.505 (.393)	0.556 (.429)	0.004	1
<i>#COUNTRIES</i>	Number of countries in which testing takes place	1.190 (1.002)	1.164 (.918)	1.312 (1.323)	1	23
<i>DURATION</i>	Study duration in weeks (as per clinical protocol)	30.054 (48.121)	31.885 (49.940)	21.329 (37.082)	0.143	520

tend to be larger or shorter in duration than studies coordinated in-house staff. They involve a lower proportion of academic sites and are more procedure-intensive.

Table 3 displays the distribution of outsourcing intensity by trial phase. Drug development is a sequential process beginning with Phase I safety trials, continuing with Phase II “proof of princi-

TABLE 3—DISTRIBUTION OF OUTSOURCING ACROSS TRIAL PHASE

	Mean outsourcing level	Std. dev.	Freq.
<i>PHASE I—Overall</i>	0.242	0.428	1,663
<i>PHASE I—Oncology</i>	0.094	0.292	213
<i>PHASE I—Other therapeutic areas</i>	0.263	0.441	1,450
<i>PHASE II</i>	0.139	0.346	1,594
<i>PHASE III</i>	0.148	0.356	2,985
<i>PHASE IV</i>	0.204	0.404	584

ple” trials, and ending with larger-scale, efficacy Phase III trials designed to validate Phase II results in an environment as similar as possible to that of regular medical practice. Phase IV studies are performed post-approval, often in an effort to ensure acceptance of the new drug by prescribing physicians. Uncertainty regarding the compound’s toxicity, side effects, and other idiosyncrasies is resolved upon completion of each stage, so that one would expect knowledge-production activities to assume decreasing prominence (relative to data-production activities) as development unfolds.

The raw data broadly support this intuition, but with some important caveats. While outsourcing is most frequent in Phase IV studies, there does not seem to be a difference in the level of outsourcing between Phase II and Phase III trials. One explanation for this lack of difference is that Phase III trials are more likely than Phase II trials to be outsourced on a “turnkey” basis, so that the difference observed in the sample understates the true difference.<sup>8</sup> Phase I trials are outsourced more often than either Phase II or Phase III trials, but this is expected. Phase I trials correspond to projects whose degree of complexity varies widely, from the most sophisticated (such as “first-in-man” pharmacokinetic and pharmacodynamic studies) to the most routine and codified (such as bioavailability and bioequivalence studies which can take place at any time along the path to regulatory submission). Unfortunately, the data at hand make it difficult to disentangle the “rou-

<sup>8</sup> Another explanation pertains to a relatively recent trend in clinical development that has lead some firms to blur the definitions of Phase II and Phase III trials and to conduct so called “Phase II/III” studies.

TABLE 4—DISTRIBUTION OF OUTSOURCING ACROSS BROAD THERAPEUTIC AREAS

	Mean outsourcing level	Std. dev.	Freq.
<i>Cardiovascular</i>	0.225	0.418	725
<i>Gastrointestinal</i>	0.195	0.397	297
<i>Central nervous system</i>	0.169	0.375	1,143
<i>Anti-infective</i>	0.174	0.379	586
<i>Oncology</i>	0.103	0.304	1,004
<i>Immuno-modulation</i>	0.156	0.363	563
<i>Dermatology</i>	0.163	0.371	208
<i>Endocrinology</i>	0.137	0.344	641
<i>Hematology</i>	0.185	0.390	146
<i>Ophthalmology</i>	0.099	0.300	121
<i>Genito-urinary system</i>	0.217	0.412	466
<i>Respiratory system</i>	0.205	0.404	611
<i>Pain/anesthesia/musculoskeletal</i>	0.329	0.461	213

tine” from the “complex” Phase I studies. Phase I oncology studies constitute an exception. Because of their harmful side effects, nearly all cancer drugs are first tested on patients—as opposed to healthy volunteers—so that one can be fairly sure that these studies correspond to “first-in-man” experiments. Only 9.5 percent of these studies are outsourced. Table 4 reports the distribution of outsourcing across 14 broad therapeutic areas. While this is a crude classification (for instance, the anti-infective class covers a wide spectrum of medical indications, from HIV to run-of-the-mill bacterial infections), outsourcing is most frequent in the pain and cardiovascular areas, which are characterized by large average trial sizes and relatively well-understood technologies. Conversely, it is very infrequent in oncology, a therapeutic class for which clinical protocols are scientifically involved and the data collection requirements extremely complex. Similar patterns emerge at the *therapeutic indication* level of analysis. The projects analyzed here pertain to 361 indications, and only for a small number of those are there enough observations in the sample for a meaningful descriptive look. As an illustration, 34 percent of 59 ulcer studies are outsourced, but only 11 percent of 229 transplantation studies.

Although the tabulations above do not make it possible to draw firm conclusions, they point to sources of variation that need to be controlled for in multivariate analyses.

### C. Variation in the Composition of Investigator Teams

The clinical trials process is managed through a three-tiered hierarchy: the product team inside the pharmaceutical firm (who must integrate the result of each study into the overall development plan for the experimental drug); the clinical monitors (who oversee the quality of the data and attend to the smooth unfolding of the trial from an operational standpoint); and the set of clinical investigators (who enroll patients in the study and treat them according to an agreed-upon experimental protocol). Below, I infer the relative importance of knowledge-production and data-production activities for clinical monitors (the middle rung of the hierarchy) from project-level variation in the composition of the investigator team (the lower rung of the hierarchy). The empirical validity of this approach hinges on a detailed understanding of the investigator-selection process by pharmaceutical firms.

Variation in the location of research activity provides a rich source of heterogeneity at the project level, as clinical trials typically involve a variety of different research sites: academic medical centers, private practices, community hospitals, or even dedicated, for-profit research centers. The distinction between academic and nonacademic testing sites is important. Over the past 15 years, academic medical centers have gradually ceased to be the preferred locus of industry-sponsored drug development activities.<sup>9</sup> The academic and nonacademic sectors differ in the relative emphasis put on knowledge production (as opposed to data production) by clinical investigators. In addition to industry-supported clinical trials, academic investigators also carry out "basic" clinical investigations, which are rewarded by publications, NIH grants, academic prestige, and promotion. Career concerns may induce academic physicians

to shift their attention away from careful record-keeping and other activities directed at ensuring the quality and integrity of the data. In contrast, in commercial sites, investigators' allocation of effort is not skewed away from data production by competing incentives.

This diversity provides pharmaceutical firms with the opportunity to match carefully the composition of the investigator team with the type of problems most likely to arise during the clinical study. For example, when the focus of the study is on confirmation of a scientific hypothesis, the objectives of investigators in the commercial sector will be more aligned with sponsors' interests. In contrast, when hypothesis generation is more valuable or when the product team "is ignorant about what it is ignorant about," then encouraging investigators to follow their scientific intuition might become comparatively more valuable.

I argue that the same unobserved project characteristics lead pharmaceutical firms to select a team of investigators dominated by academics (respectively by nonacademics) *and* to value relatively more knowledge-production effort (respectively data-production effort) by clinical monitors. As a result, a higher proportion of academic investigators should decrease the probability that a given trial will be outsourced.

Investigator selection is, like outsourcing, a decision variable of the firm. One might worry that particular types of investigators (e.g., for-profit) are matched with particular types of clinical monitors (e.g., CRO employees) because of some other unobserved factor. Such endogenous matching (Daniel A. Akerberg and Maristella Botticini, 2002) threatens the validity of the interpretation above.

Although the data do not provide a credible source of exogenous variation for the choice of investigators, a number of alternative explanations can be ruled out using econometric methods, as explained below. As a complement, I have also pursued the strategy of examining qualitatively the investigator selection process at each of the six companies visited for this study. Interviews with personnel in charge of investigator recruitment provide strong evidence that the outsourcing decision and investigator choice are *not* jointly determined. While these firms carefully chose the mix of academic

<sup>9</sup> The data used in this paper show that the proportion of academic clinical sites at the industry level decreased steadily from 66 percent in 1993 to 57 percent in 1999. FDA records show that the number of U.S. investigative sites tripled between 1990 and 1995, from 6,159 sites in 1990 to 18,557 sites in 1995, with nearly all of this growth taking place in the for-profit sector (Mark Hovde and Robert Seskin, 1997).

and nonacademic investigators participating in the studies they sponsored, they did so in response to opportunities to learn about the disease and the experimental molecule, not as a function of the integration decision. Another consideration appeared to play a role in the investigator selection process: “credentializing” the drug in the eyes of regulatory authorities and potential prescribing physicians. Whereas the first rationale has implications for the outsourcing of clinical monitoring, the second rationale does not. This observation forms the basis for an additional check on the validity of the paper’s empirical strategy: if investigator mix is a valid proxy for knowledge, the outsourcing-investigator relationship should hold less strongly when one has independent reasons to believe that “credentializing” the drug looms larger as a concern than knowledge acquisition.<sup>10</sup>

#### D. Econometric Considerations

Estimating the effect of project and firm characteristics on the integration decision requires a procedure that accommodates its discrete nature. A standard approach to this type of problem is to model a latent variable,  $CRO^*$  (the relative inefficiency of outsourcing compared with integration) as

$$CRO_{ijt}^* = \beta' \mathbf{X}_{ijt} + \gamma' \mathbf{Z}_{jt} + \varepsilon_{ijt}$$

where  $i$  indexes projects,  $j$  indexes firms,  $t$  denotes calendar time, and  $\mathbf{X}$  (respectively  $\mathbf{Z}$ ) is a vector of observable project-level (respectively firm-level) characteristics. In reality, rather than observing  $CRO^*$ , I observe the integration decision itself:

$$CRO_{ijt} = \begin{cases} 1 & \text{if } \varepsilon_{ijt} > -\beta' \mathbf{X}_{ijt} - \gamma' \mathbf{Z}_{jt} \\ 0 & \text{otherwise} \end{cases}$$

<sup>10</sup> Other factors mentioned by my respondents as determinants of the investigator mix include the type of clinical venues favored by the target patient population and the social connections of senior staff. In these data, the pharmaceutical firm is handling the selection of investigators *whether or not the project is outsourced*. Therefore, social networks between pharmaceutical team members and academic physicians are unlikely to drive the association between outsourcing and reliance on academic medicine.

I assume that the  $\varepsilon$ ’s are drawn from a logistic distribution; this choice is motivated by the need to accommodate more sophisticated error structures, as explained below. The next few paragraphs detail the construction of the variables included in  $\mathbf{X}$  and  $\mathbf{Z}$ .

The clinical development process is plagued by uncertainty. The compound’s potency, associated toxicities, adverse interactions, and side effects are revealed gradually as development unfolds. Ex ante, firms can form priors as to the difficulty of “letting the compound show its true value.” This prior will be based on characteristics of the study, some of which are observable, such as project phase (exploratory vs. confirmatory), compound novelty (“first-in-class” vs. “me-too”), complexity of the target indication, etc. The applied researcher’s challenge is to find high-quality, project-level proxies that act as shifters of (i) the marginal returns to knowledge- and data-production activities; and (ii) the difficulty of observing knowledge- and data-production effort.

*Marginal Return to Knowledge- and Data-Production Effort.*—I construct the variable %AMC, the proportion of investigators in the trial affiliated with academic medical centers. A higher proportion of academic investigators should make it less likely that a given trial will be outsourced, since CROs and their employees are not rewarded, explicitly or implicitly, for the acquisition of knowledge from the clinical sites.

I employ the set of trial phase dummies discussed earlier (the Phase IV category is omitted). Phase I oncology trials should be less likely to be outsourced than Phase II trials, which in turn should be less likely to be outsourced than Phase III trials. All trials on the critical path to regulatory approval should be less likely to be outsourced than Phase IV trials (which take place after a drug has been approved).

Clinical studies taking place in hospital settings are likely to require more careful monitoring of serious adverse events—a knowledge-intensive task—than outpatient studies. As a result, inpatient studies should be less likely to be outsourced, *ceteris paribus*. I include the *INPATIENT* dummy to capture this effect.

The more balanced incentives faced by internal monitors should also influence the team’s

ability to adapt to changes in circumstances surrounding clinical projects as they unfold. I use the protocol length of a clinical study (*DU-RATION*) as a proxy for the probability of change orders. I expect shorter studies to lend themselves more easily to outsourcing.

Finally, I include two variables expected to shift *downward* the marginal return to knowledge-production effort relative to data-production effort. The number of distinct medical procedures in the study (*#PROCEDURES*) proxy for the data-processing burden associated with the clinical trial. A similar argument can be made regarding project size (*#SITES*). Knowledge-production activities tend to be underemphasized in larger trials where logistical issues take center stage. High values for both variables should increase the likelihood of outsourcing.

*Observability of Data- and Knowledge-Production Effort.*—Clinical studies for which data-production activities are less costly to monitor are likely to be good candidates for outsourcing. I argue that data-production monitoring costs are a decreasing function of the concentration of patients across clinical sites. Regardless of whether a study is outsourced, pharmaceutical firms audit the quality of their clinical data by selecting a sample of sites in the study (generally those with the largest patient enrollment) and replicating source-document verification. A higher Herfindahl index of patient concentration, *HERF*, should translate into more accurate estimates of data quality and lower auditing costs.

Unfortunately, the data do not provide good proxies for the costs of monitoring knowledge production. One would expect that keeping track and assimilating idiosyncratic knowledge generated at clinical sites would become harder in trials involving a large number of clinical investigators. On this ground, larger trials should be less likely to be outsourced. Since study size also shifts the relative marginal return to data- and knowledge-production effort, its predicted effect on the likelihood of integration is ultimately ambiguous.<sup>11</sup>

While project-level variation generates nuanced hypotheses regarding the effect of multitasking incentives on the integration decision, Figures 1 and 2 suggest that there are differences in firms' average propensity to outsource. These firm-level patterns can be explained by invoking other rationales for outsourcing.

*Adjustment Costs.*—Practitioners often justify outsourcing to CROs by referring to the “peaks and valleys problem”—the ability to start large trials quickly and, more important, the flexibility afforded by the ability to cancel a service contract in the eventuality of a pipeline “dry spell.” Drugs enter and exit the pipeline frequently, creating wide variation in the number of employees needed to staff clinical-study teams. Confronted with large sunk costs of hiring and firing employees, partial vertical disintegration enables pharmaceutical firms to cope with sudden and lumpy shifts in the level of demand for clinical trial services. This is the case because technology shocks tend to be uncorrelated across firms, creating a role for a free-standing upstream supplier to pool labor demands more efficiently (Dennis W. Carlton, 1979; Katharine G. Abraham and Susan K. Taylor, 1996). Similarly, larger firms or firms that are developing drugs in multiple therapeutic areas may be better able to redeploy employees from terminated projects to new ones. The specification includes firm-level variables meant to capture the influence of pipeline fluctuations on the project-level outsourcing decision: *SHOCK*, *SIZE*, and *SCOPE*. These variables may have less explanatory power when analyzing the project-level outsourcing decision than when investigating the extent of outsourcing at the firm level. Nonetheless, it seems important to account for their effect at this finer-grained level of analysis.

*Heterogeneous Firm Capabilities.*—Firms that have developed extensive internal capabilities in clinical development may face higher relative costs of outsourcing than less experi-

<sup>11</sup> To a large extent, *HERF* and *#SITES* reflect exogenous factors related to disease prevalence and geography. Moreover, regulatory authorities provide guidelines in sta-

tistical design that severely constrain firms' latitude in deciding upon the level of these variables. In what follows, they will be treated as exogenous.

enced firms. To capture the importance of overall development expertise, I compute *EXPERIENCE*, the cumulative count of clinical sites involved in the firm's portfolio of projects up to year  $t$ . I also include *BIOTECH* (whether the firm was present on the product market as of 1995) and *U.S. PRESENCE* (whether the firm was U.S.-based or, in the case of foreign firms, whether it owned a U.S. subsidiary with clinical development facilities as of 1995).

In recent years, there has been a marked trend in favor of multinational clinical trials. This evolution results from a greater regulatory harmonization that enables pharmaceutical companies to seek simultaneous approval in a larger number of countries. Many large CROs claim to provide "global reach" in their marketing brochures, and specialized suppliers may be in a better position than their upstream clients to amortize the sunk infrastructure costs required to manage multinational studies. The likelihood of outsourcing is expected to increase in the project's geographic span, *#COUNTRIES*.

### E. Results

Table 5 reports estimates from the specification above. I begin by examining the hypotheses pertaining to incentives, adjustment costs, and heterogeneous firm capabilities separately (columns [1], [2], and [3]), before proceeding to specifications that investigate these arguments concurrently (columns [4] and [5]). All these models include year, therapeutic class, and region fixed effects. Fixed effects at the regional and therapeutic class level control for the array of cross-class and cross-region differences in history, technological opportunities, and state of knowledge that have been constant over the sample period, while year effects control for the secular increase in the prevalence of outsourcing at the industry level.

Model (1) yields coefficient estimates generally in line with Hypotheses 1 and 2b. *%AMC* is strongly negatively correlated with outsourcing, as expected. Projects in pre-approval stages of development are less likely to be outsourced than Phase IV, post-approval projects (the omitted phase dummy in the specification). Shorter and more procedure-incentive studies are also more likely to be outsourced, as expected. The coefficients before *INPATIENT* are not statisti-

cally significant. I find strong support for Hypothesis 2b: patient concentration across study sites is a strong predictor of outsourcing. As noted above, the interpretation of the effect of *#SITES* is ambiguous. The finding that the likelihood of nonintegration is increasing in study size is consistent with Hypothesis 1, but inconsistent with Hypothesis 2a.

Model (2) investigates the effect of adjustment costs. Surprises in the level of activity (*SHOCK*) increase the likelihood of outsourcing. Moreover, this effect is less pronounced for large firms, as evidenced by the effect of the interaction term between *SHOCK* and  $\ln(\text{SIZE})$ . Finally, the propensity to outsource is increasing in the firm's therapeutic scope (but not in the firm's size).

Model (3) focuses on the role of heterogeneous firm capabilities. The likelihood of outsourcing any given project decreases with the firm's overall experience in clinical development. Biotechnology firms and firms without a U.S. presence are more likely to outsource. Geographic span does not appear to have any effect.

Of course, if multiple rationales for outsourcing exist, statistical tests that account for only one class of explanations will be misspecified. In column (4), I check whether the results above are robust to the simultaneous inclusion of all predictor variables. Quite remarkably, I find that this is the case. Some results are even strengthened compared to columns (1) through (3). For example, the coefficient estimates for the *PHASE* variables are ordered monotonically. While this result has intuitive appeal and is very robust across specifications, an inequality test cannot reject the null hypothesis that they are in fact equal. On the other hand, the magnitudes and statistical significance of the coefficient estimates for the variables meant to capture the influence of adjustment costs are weakened.

The interpretation of the statistical estimates in column (4) is subject to caution, since it does not account for the effect of unobserved firm practices related to both observable study characteristics and the integration decision. For example, pharmaceutical firms have been shown to exhibit heterogeneity in their "taste for science" in the setting of drug discovery (Iain M. Cockburn et al., 2000). If pharmaceutical firms

TABLE 5—LOGIT MODELS OF THE OUTSOURCING DECISION

	(1)	(2)	(3)	(4)	(5)
<i>%AMC</i>	-0.613** [0.118]			-0.689** [0.137]	-0.772** [0.145]
<i>PHASE I</i> × <i>OTHERS</i>	0.079 [0.270]			-0.217 [0.269]	-0.259 [0.300]
<i>PHASE I</i> × <i>ONCOL.</i>	-0.514 <sup>†</sup> [0.294]			-0.692* [0.304]	-0.759** [0.289]
<i>PHASE II</i>	-0.366* [0.164]			-0.481** [0.180]	-0.495** [0.185]
<i>PHASE III</i>	-0.340* [0.159]			-0.378* [0.160]	-0.504** [0.171]
<i>INPATIENT</i>	-0.154 [0.111]			-0.111 [0.122]	-0.181 [0.124]
<i>#PROCEDURES</i>	0.018** [0.006]			0.017** [0.006]	0.019** [0.006]
<i>Ln(DURATION)</i>	-0.104* [0.041]			-0.068 [0.043]	-0.071 [0.050]
<i>Ln(#SITES)</i>	0.589** [0.090]			0.287 <sup>†</sup> [0.164]	0.779** [0.155]
<i>HERF</i>	1.553** [0.285]			0.799* [0.392]	1.778** [0.389]
<i>Ln(SIZE)</i>		-0.160 [0.113]		-0.020 [0.139]	0.175 [0.197]
<i>SCOPE</i>		-0.225** [0.061]		-0.166** [0.063]	-0.028 [0.064]
<i>SHOCK</i>		0.295** [0.071]		0.207** [0.074]	-0.032 [0.068]
<i>Ln(SIZE)</i> × <i>SHOCK</i>		-0.108** [0.032]		-0.078* [0.030]	0.031 [0.030]
<i>EXPERTISE</i>			-0.015** [0.003]	-0.004 [0.006]	0.014** [0.005]
<i>BIOTECH</i>			0.957* [0.398]	0.883 <sup>†</sup> [0.491]	
<i>US PRESENCE</i>			-1.089** [0.235]	-0.672* [0.290]	
<i>#COUNTRY</i>			0.011 [0.046]	-0.001 [0.047]	0.013 [0.045]
<i>Constant</i>	-3.164** [0.452]	-1.027** [0.365]	-1.124** [0.379]	-1.769** [0.601]	-4.286** [0.727]
Firm fixed effects	No	No	No	No	Yes
Log likelihood	-2,832.47	-2,821.48	-2,825.47	-2,692.23	-2,481.63
Nb. of observations	6,689	6,689	6,689	6,689	6,689

Notes: All Models include year, therapeutic class, and region fixed effects. Phase IV is the omitted Phase dummy.

Standard errors (in parentheses) are heteroskedasticity-robust and clustered by firm-year cells.

<sup>†</sup> Significant at the 10% level.

\* Significant at the 5% level.

\*\* Significant at the 1% level.

with a strong scientific culture are both more closely tied to academia and take a more negative view of CROs because of their lack of scientific credentials, then *%AMC* and outsourcing will covary in the data, but this need not reflect the influence of monitors' incentives.

Since I observe a number of projects that differ widely along observable characteristics (such as *%AMC*) but are sponsored by the same

firm, a number of alternative interpretations can be ruled out. In column (5), I add a full set of firm fixed effects to the specification.<sup>12</sup> Ac-

<sup>12</sup> Fixed-effect estimators applied to nonlinear models are biased because of the incidental parameter problem. An important issue is to determine whether the number of observations per firm in the dataset is large enough for this estimation strategy to be applied with confidence. Using

TABLE 6—LOGIT MODELS OF THE OUTSOURCING DECISION  
(Robustness checks)

	(6)	(7)	(8)	(9)
%AMC	-0.823** [0.149]	-0.735** [0.111]	-0.813** [0.113]	-0.740** [0.126]
PHASE I × OTHERS	-0.368 [0.307]	-0.219 [0.186]	-0.022 [0.234]	-0.126 [0.222]
PHASE I × ONCOL.	-0.887** [0.295]	-1.027** [0.344]	-0.545 [0.346]	-1.204** [0.447]
PHASE II	-0.557** [0.186]	-0.522** [0.165]	-0.366* [0.169]	-0.502* [0.200]
PHASE III	-0.544** [0.169]	-0.404** [0.144]	-0.413** [0.150]	-0.395* [0.172]
INPATIENT	-0.111 [0.124]	-0.219† [0.113]	-0.257* [0.125]	-0.290* [0.137]
#PROCEDURES	0.019** [0.007]	0.019** [0.005]	0.020** [0.005]	0.023** [0.006]
Ln(DURATION)	-0.099* [0.049]	-0.122** [0.040]	-0.040 [0.041]	-0.059 [0.048]
Ln(#SITES)	0.848** [0.173]	0.848** [0.147]	0.775** [0.143]	0.721** [0.152]
HERF	1.894** [0.426]	1.923** [0.362]	1.733** [0.354]	1.566** [0.384]
Firm fixed effects	Yes	Yes	Yes	No
Other controls	Firm-specific time trends	Firm-class interaction fixed effects (CML)	Indication fixed effects (CML)	Molecule fixed effects (CML)
Log likelihood	-2,377.88	-1,910.56	-2,091.92	-1,487.55
Nb. of observations	6,689	5,861	6,192	4,734

Notes: All models include year, therapeutic class, and region fixed effects;  $Ln(SIZE)$ ,  $SCOPE$ ,  $Ln(SIZE) \times SHOCK$  (the “adjustment costs” variables);  $EXPERTISE$  and  $\#COUNTRY$  (the “firm capabilities” variables). Phase IV is the omitted Phase dummy. Standard errors (in parentheses) are heteroskedasticity-robust and clustered by firm-year cells. In fixed-effect non-linear models, units for which there is no within-variation in the dependent variable drop out of the estimation sample. To make sure that these particular subsamples are not driving the results, models (1) through (6) were reestimated using the subset of the data used to estimate model (9). The results were qualitatively unchanged.

† Significant at the 10% level.

\* Significant at the 5% level.

\*\* Significant at the 1% level.

counting for unobserved firm heterogeneity has a dramatic effect on the results: the effect of variables pertaining to adjustment costs and capabilities disappears, while the hypotheses pertaining to the multitask agency problem still find strong support in the data. The differences between columns (4) and (5) illustrate the perils involved in interpreting the results of purely cross-sectional tests.

I choose model (5) as a benchmark specification and compute the magnitudes for a num-

ber of variables. At the mean of the data, increasing %AMC from 61 percent (the mean value) to 100 percent (a mass point corresponding to 45 percent of the observations) lowers the probability of outsourcing from 17.3 percent to 14.00 percent—a nearly 20-percent drop; increasing  $HERF$  by a standard deviation increases the probability of outsourcing from 17.3 percent to 20.7 percent; while a Phase IV study would be 5.62 percent more likely to be outsourced than a Phase III study.<sup>13</sup>

Table 6 displays models that probe the robustness of specification (5) to more elaborate

Monte Carlo simulations for the fixed-effect probit model, James J. Heckman (1981) shows that above a critical panel length of eight periods, the bias is negligible, although this critical length depends on the frequency of the event being studied. Fortunately, the number of observations available to identify the firm effects is well above this number here.

<sup>13</sup> In unreported regressions, I explored whether nonlinearities in the effects of %AMC, #SITES, and  $HERF$  could materially affect the results of Table 5. This was not the case.

forms of unobserved heterogeneity. Model (6) adds firm-specific time trends to the specification, thus addressing the concern that firm-level outsourcing fads might contaminate the results. Established firms are often organized into “therapeutic silos” enjoying a large degree of autonomy. If some therapeutic divisions within a firm exhibit more of a “taste for science” than others, then fixed effects at the firm level will not properly account for such heterogeneity. Model (7) reestimates the model with a full set of firm-class interaction fixed effects. The results obtained in column (5) of Table 5 still hold under these more restrictive decompositions of the error term.

Another class of alternative interpretations relies on the importance of unobserved disease or drug characteristics. For some diseases, the supply of clinical investigators is limited to academic medical centers. In such cases, a small number of academic physicians act as essential gatekeepers to the patient community, and pharmaceutical firms might prefer to monitor such studies with their own employees as a way to nurture these relationships. To rule this out, model (8) replaces broad therapeutic class controls by a set of narrow therapeutic indication fixed effects. For instance, whereas cystic fibrosis and diabetes trials were part of the same endocrine disease group in model (5), model (8) assigns to both sets of studies its own fixed effect. The structure of the data can even be exploited to account for unobserved heterogeneity at the molecule level. There is an average of 7 studies per molecule, with a minimum of 3 and a maximum of 132 studies. Accordingly, model (9) adds molecule fixed effects to the regression, sweeping out any idiosyncratic variation in outsourcing propensity related to experimental compounds. In these specifications, the multitasking interpretation of the results is subjected to a very stringent test, as only within-indication or within-drug information is used to identify the effect of covariates on the likelihood of outsourcing. Estimation is performed by conditional maximum likelihood, using Gary Chamberlain’s (1984) fixed-effect logit model. Although some coefficients are estimated less precisely, the results confirm the earlier estimates.

As a final check on the plausibility of interpreting the effect of investigator mix in terms of the multitasking framework, I interact  $\%AMC$

with two dummy variables corresponding to studies for which the selection of academics is more likely to reflect a desire to build up the drug’s credentials than knowledge acquisition concerns. The first subsample comprises 547 Phase IV trials. I single out Phase IV trials since they are performed post approval, often in an effort to ensure acceptance of the drug by prescribing physicians. Using medical opinion leaders in academia is a way to shape prescribers’ perception of the new treatment as legitimate (“seeding the market” in the industry lingo). The second subsample comprises 786 trials corresponding to drugs intended for the 10 most well-treated diseases.<sup>14</sup> For such diseases, more is presumably known about best practices in protocol design, potential enrollment bottlenecks, acceptable surrogate markers, etc. A firm might still want to include prominent academics in these studies, so as to ensure a more expedited review by regulatory authorities.

If the knowledge-vs.-data interpretation of the  $CRO/\%AMC$  relationship is legitimate, this association should be less likely to hold for Phase IV studies and for trials corresponding to well-treated diseases. This is indeed the case, as can be seen in Table 7, which reports the results of regressions that take model (5) of Table 5 and model (9) of Table 6 as baselines, and add the interaction terms  $\%AMC \times PHASE\ IV$  (models [10] and [11]) and  $\%AMC \times WELL-TREATED$  (models [12] and [13]) to these specifications. While the main effect of  $\%AMC$  remains significant and similar in magnitude to the results reported in Table 6, the interaction terms are not significant. Although this test does not rule out all possible endogenous matching mechanisms, it greatly narrows the scope of plausible alternative interpretations.

## F. Discussion

Extant theoretical perspectives on outsourcing take contract incompleteness and opportu-

<sup>14</sup> These are: otitis media, insomnia, pneumonia, bronchitis, asthma, rheumatoid arthritis, pain, urinary tract infections, skin and soft tissue infections, and hypertension. To select these diseases, I drew from a list of icd9 codes and associated drugs provided to me by Frank Lichtenberg. These 10 diseases have the largest number of already approved treatments.

TABLE 7—LOGIT MODELS OF THE OUTSOURCING DECISION  
(Interaction effects)

	(10)	(11)	(12)	(13)
%AMC	-0.803** [0.152]	-0.731** [0.132]	-0.725** [0.144]	-0.702** [0.134]
WELL-TREATED	—	—	0.252 [0.191]	0.505* [0.254]
%AMC × PHASE IV	0.294 [0.345]	0.029 [0.356]	—	—
%AMC × WELL-TREATED	—	—	-0.249 [0.253]	-0.097 [0.254]
Other controls	Firm fixed effects	Molecule fixed effects (CML)	Firm fixed effects	Molecule fixed effects (CML)
Log likelihood	-2,480.48	-1,483.74	-2,479.82	-1,481.20
Nb. of observations	6,689	4,734	6,689	4,734

Notes: Columns (10) and (12) take the specification of column (5), Table 5, as baseline. Columns (11) and (13) take the specification of column (9), Table 6, as baseline. All models include year, therapeutic class, and region fixed effects; standard errors (in parentheses) are heteroskedasticity-robust and clustered by firm-year cells.

† Significant at the 10% level.

\* Significant at the 5% level.

\*\* Significant at the 1% level.

nistic behavior as their starting point. Yet they tend to fall into two distinct categories. The first approach views the problem of drawing firm boundaries as one of facilitating ex post adaptation to changing circumstances (Oliver E. Williamson, 1985; Steven Tadelis, 2002). The second approach focuses on the provision of appropriate ex ante incentives for specific investments (Hart, 1995) or the allocation of effort across tasks (Holmström and Milgrom, 1994; Andrea Shepard, 1993). In the first approach, the dissipation of quasi-rents through haggling is purely wasteful. In the second, investment/effort choices are set below their first-best level but parties bargain their way to efficiency conditional on these initial choices.<sup>15</sup>

This paper analyzes the problem of outsourcing in clinical development through the lens of the latter perspective. This is a simplification, as issues of ex post adaptation do arise between CROs and their clients. The contract negotiations are often protracted, with suppliers attempting to limit

buyers' ability to introduce modifications during the execution stage. Of course, the unpredictability of ongoing development projects makes renegotiation a likely occurrence, and conflicts over "change orders" often ensue.

Unfortunately, the data used here are not well suited to test arguments relating bargaining difficulties over post-contractual amendments to the choice of organizational form. One exception may be provided by the influence of *DURATION*, the length of time patients are enrolled in the study. If the probability of change orders increases with protocol length, we would expect longer studies to be monitored by employees if bargaining costs are lower inside the firm (Birger Wernerfelt, 1997; Duncan Simester and Mark Knez, 2002). Although the relationship between the likelihood of outsourcing and *DURATION* goes in the expected direction, the statistical significance of this effect is not robust across specifications.

Finally, the multitask agency approach provides too simplified and static a view of insiders' incentives. Taken literally, this perspective implies that they would face no incentives to provide either type of effort if one of the tasks—knowledge production—was infinitely costly to monitor. Yet the qualitative evidence reveals that pharmaceutical firms bear little in common with government agencies or other bureaucracies plagued by lack of employee initiative.

<sup>15</sup> The clinical monitors studied here perform ongoing services for the pharmaceutical firms, and it is hard to identify large discrete investments of the sort that might take place at the beginning of a relationship (Jonathan Levin, 2003). As a result, I have found it easier to couch the paper's argument in agency terms, rather than through the lens of the hold-up problem and specific investments, but this distinction may be mostly semantic.

Internal monitors are highly motivated, but this stems primarily from such career-based rewards as subjective bonuses and promotions. By combining the absence of piece rates with subjective evaluations, pharmaceutical firms ensure that employees' incentives are both balanced and relatively high powered.

#### IV. Conclusion

This paper performs an empirical analysis of outsourcing in clinical development. I argue that variation in project characteristics leads to variation in the importance of employee effort on two tasks: knowledge production and data production. Since it is harder to measure performance in knowledge production than to measure performance in data production, pharmaceutical firms choose to assign their own employees to projects for which the former are relatively more important than the latter. In so doing, they ensure that incentives for both tasks are kept in balance. Conversely, data-intensive projects are more likely to be outsourced, as distorting the allocation of effort is less costly for these projects.

Consistent with this argument, I document that even after controlling for a number of alternative explanations, knowledge-intensive projects are more likely to be assigned to internal teams, while data-intensive projects are more likely to be outsourced. These results fit with Holmström's (1999) subeconomy view of the firm, whereby organizations design incentive systems in a manner that internalizes some of the contractual externalities present in markets due to asymmetric information. Focusing on the costs of communicating knowledge—as opposed to the incentives to acquire it—would lead to an alternative perspective on the optimal size and scope of hierarchies; Luis Garicano (2000) provides an interesting model along these lines. How communication costs and incentive problems interact to shape firm boundaries is an important topic for future research.

#### REFERENCES

- Abraham, Katherine G. and Taylor, Susan K.** "Firms' Use of Outside Contractors: Theory and Evidence." *Journal of Labour Economics*, 1996, 14(3), pp. 394–424.
- Akerberg, Daniel A. and Botticini, Maristella.** "Endogenous Matching and the Empirical Determinants of Contract Form." *Journal of Political Economy*, 2002, 110(3), pp. 564–91.
- Arrow, Kenneth J.** *The limits of organization*. New York: W. Norton & Co., 1974.
- Azoulay, Pierre.** "A Portfolio Approach to Transaction Costs Economics: Evidence from Clinical Development." Unpublished Paper, 2003a.
- Azoulay, Pierre.** "Agents of Embeddedness." National Bureau of Economic Research, Inc., NBER Working Paper: 10142, 2003b.
- Baker, George P. and Hubbard, Thomas N.** "Make versus Buy in Trucking: Asset Ownership, Job Design, and Information." *American Economic Review*, 2003, 93(3), pp. 551–72.
- Carlton, Dennis W.** "Vertical Integration in Competitive Markets under Uncertainty." *Journal of Industrial Economics*, 1979, 27(3), pp. 189–209.
- Chamberlain, Gary.** "Panel Data," in Z. Griliches, ed., *Handbook of econometrics volume II*. Amsterdam: North-Holland, 1984, pp. 1248–318.
- Cockburn, Iain M.; Henderson, Rebecca M. and Stern, Scott.** "Untangling the Origins of Competitive Advantage." *Strategic Management Journal*, 2000, 21(10–11), pp. 1123–45.
- Department of Health and Human Services.** "Good Clinical Practice: Consolidated Guideline." *Federal Register*, 1997, 62(90), pp. 25692–709.
- Garicano, Luis.** "Hierarchies and the Organization of Knowledge in Production." *Journal of Political Economy*, 2000, 108(5), pp. 874–904.
- Gelijns, Annetine C.; Rosenberg, Nathan and Moskowitz, Alan J.** "Capturing the Unexpected Benefits of Medical Research." *New England Journal of Medicine*, 1998, 339(10), pp. 693–98.
- Hart, Oliver.** *Firms, contracts, and financial structure*. Clarendon Lectures in Economics. Oxford: Oxford University Press, 1995.
- Heckman, James J.** "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process," in C. F. Manski and D. McFadden, eds., *Structural analysis*

- of discrete data with econometric applications. Cambridge: MIT Press, 1981, pp. 179–95.
- Holmström, Bengt.** “The Firm as a Sub-economy.” *Journal of Law, Economics, and Organization*, 1999, 15(1), pp. 74–102.
- Holmström, Bengt and Milgrom, Paul.** “Multi-task Principal–Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design.” *Journal of Law, Economics, and Organization*, Special Issue 1991, 7(0), pp. 24–52.
- Holmström, Bengt and Milgrom, Paul.** “The Firm as an Incentive System.” *American Economic Review*, 1994, 84(4), pp. 972–91.
- Hovde, Mark and Seskin, Robert.** “Selecting U.S. Clinical Investigators.” *Applied Clinical Trials*, 1997, 6(2), pp. 34–42.
- Kling, James.** “From Hypertension to Angina to Viagra.” *Modern Drug Discovery*, 1998, 1(2), pp. 31–38.
- Kogut, Bruce and Udo, Zander.** “Knowledge of the Firm, Combinative Capabilities, and the Replication of Technology.” *Organization Science*, 1992, 3(3), pp. 383–97.
- Levin, Jonathan.** “Relational Incentive Contracts.” *American Economic Review*, 2003, 93(3), pp. 835–57.
- Maskin, Eric and Tirole, Jean.** “Two Remarks on the Property-Rights Literature.” *Review of Economic Studies*, 1999, 66(1), pp. 139–49.
- Monteverde, Kirk.** “Technical Dialog as an Incentive for Vertical Integration in the Semiconductor Industry.” *Management Science*, 1995, 41(10), pp. 1624–38.
- Osberg, Lars; Wolf, Edward N. and Baumol, William J.** *The information economy: The implications of unbalanced growth*. Halifax, Nova Scotia: The Institute for Research on Public Policy, 1989.
- Parexel.** *Pharmaceutical R&D statistical sourcebook*. Waltham: Parexel International Corporation, 2003/2004.
- Polanyi, Michael.** *The tacit dimension*. Garden City: Doubleday, 1966.
- Shepard, Andrea.** “Contractual Form, Retail Price, and Asset Characteristics in Gasoline Retailing.” *RAND Journal of Economics*, 1993, 24(1), pp. 58–77.
- Simester, Duncan and Knez, Marc.** “Direct and Indirect Bargaining Costs and the Scope of the Firm.” *Journal of Business*, 2002, 75(2), pp. 283–304.
- Tadelis, Steven.** “Complexity, Flexibility, and the Make-or-Buy Decision.” *American Economic Review*, 2002, (*Papers and Proceedings*) 92(2), pp. 433–37.
- von Hippel, Eric.** “Economics of Product Development by Users: The Impact of ‘Sticky’ Local Information.” *Management Science*, 1998, 44(5), pp. 629–44.
- Wernerfelt, Birger.** “On the Nature and Scope of the Firm: An Adjustment–Cost Theory.” *Journal of Business*, 1997, 70(4), pp. 489–514.
- Williamson, Oliver E.** *The economic institutions of capitalism*. New York: Free Press, 1985.