Quality Risk and Contract Manufacturing:
Theory and Empirical Evidence

JOHN V. GRAY
Fisher College of Business
The Ohio State University
2100 Neil Avenue
Columbus, OH 43210
Phone: 614-247-8021, Fax: 614-292-1272
e-mail: gray.402@osu.edu

ALEDA V. ROTH
College of Business and Behavioral Sciences
Department of Management
Clemson University
132E Sirrine Hall
Clemson, SC 29634-1305
Phone: 864-973-4144, Fax: 864-656-3091
e-mail: aroth@clemson.edu

BRIAN TOMLIN
Tuck School of Business
Dartmouth College
100 Tuck Hall
Hanover, NH 03755
Phone: 603-646-2136, Fax: 603-646-1308
e-mail: brian.tomlin@tuck.dartmouth.edu

March 6, 2012
Quality Risk and Contract Manufacturing:
Theory and Empirical Evidence

Abstract

This paper investigates quality risk and contract manufacturing. Contract manufacturers (CMs) produce products to another firm’s specifications, whereas internal plants (IPs) produce products to their own firm’s specifications. While CMs are being increasingly used in many industrial sectors, there is a dearth of empirical information regarding their potential for producing nonconforming products relative to IPs. Drawing heavily from agency theory, we argue that CMs will have a tendency to operate with a higher quality risk than IPs, controlling for measurable plant characteristics (i.e., size, product type, etc.) as long as the finished product cannot be perfectly tested for all quality defects. Based primarily on this theoretical lens, we also articulate plant-level characteristics which may moderate this effect. We test our hypotheses using a plant-level measure of quality risk based on Food and Drug Administration (FDA) inspection data on a sample of 152 plants considered as drug manufacturers by the FDA. Our results provide some empirical evidence that, in this setting, CMs operate at a higher level of quality risk than internally-owned facilities, on average and controlling for plant characteristics. We also find that the relationship between contract manufacturing and quality risk is significantly moderated by several factors in ways that are consistent with extant theory. These moderating factors imply that the quality risk differences between CMs and IPs are reduced or eliminated when the CMs primary product line is subject to more intense regulation (and therefore face more scrutiny from regulators), are larger (and therefore are more frequently inspected by customers), and are older (and therefore have developed capabilities through experience with multiple customers).

Key Words: quality risk, contract manufacturing, supply chain management

“If we can buy it cheaper than we can make it, then of course that’s what we’re going to do.”
GlaxoSmithKline CEO Dr. J.P. Garnier, quoted in Barnes (2007)

1. Introduction

In this paper, we develop a theoretical foundation for the examination of the question of whether and when contract manufacturers (CMs) and internal plants (IPs) might tend to operate with different levels of quality risk. Further, we examine moderating factors that shed light on the contingencies of this difference. Managers at manufacturing firms often organize their production networks to effectively
deliver products at the lowest possible total cost. Since at least the early 1990s, more and more firms have chosen to employ CMs to produce some or all of their complete finished products (Hayes et al., 2005; Tully 1994) in industries ranging from electronics (Sturgeon 2002) to automobiles (Edmondson 2003). The use of CMs can provide many benefits to brand-owning firms, such as faster speed-to-market, capital avoidance, and per-unit cost savings due to economies of scale or scope. However, using CMs also removes control of the manufacturing operation from the brand-owning firm. As such, using CMs may also pose certain risks that may or may not be widely understood by decision-makers in industry (Amaral et al., 2006). For example, Ben Venue’s Bedford, Ohio contract manufacturing facility was recently shut down by the FDA due to quality lapses, which had resulted in metal particles in a brand-owning customer’s product and concerns with sterility concerns in others’ products (Loftus 2011). Scholars in economics (Alchian and Demsetz 1972; Williamson 1991) and in business strategy (Grant 1996; Leiblein 2003) have theoretically articulated factors that tend to affect the relative total costs of utilizing another firm to perform an activity relative to performing the activity in-house. However, only a handful of studies directly test the performance implications of different organizational forms (Macher and Richman 2008). Further, none focus on the performance dimension of quality risk. Examining the effect of outsourced manufacturing on technical performance, Leiblein et al. (2002, p. 830) stated that there is an opportunity to “consider how firms’ boundary decisions influence other performance dimensions, such as overall firm profitability, excess cash flow, or risk.”

For semantic clarity, we first define the primary independent variable in this research, the type of manufacturing plant: CM or IP. A *contract manufacturer's plant* (CM) is an establishment that manufactures finished or nearly finished products to other companies’ specifications. An *internal plant* (IP) is an establishment that manufactures finished or nearly finished products under the brand name and specifications of its own company. Our dependent variable is *quality risk*. In this paper, we apply Gray et al.’s (2011, p.738) definition of quality risk which is “the propensity of a manufacturing establishment to fail to comply with good manufacturing practices. Quality risk is, therefore, a proxy for the likelihood that
a product shipped from a given establishment will fail to perform as intended due to manufacturing-related issues.”

Our setting is the pharmaceutical industry, which includes regulated cosmetics and both over-the-counter and prescription pharmaceuticals. This is a particularly interesting sector in which to assess the relative quality risks of CMs and IPs for several reasons. First, conformance quality is of paramount importance to this industry. For example, Johnson & Johnson (J&J) estimated $900 million in lost sales in 2010 due to recalls (Hobson 2011). J&J has also paid substantial fines and litigation expenses, and has lost untold brand equity due to these quality issues (Voreacos et al., 2011). While the majority of recalls and press coverage focused on J&J’s internal plants, many recalls also originated at CMs. It should be pointed out that CMs are rarely mentioned by name in press reports, as both the Food and Drug Administration (FDA) and J&J considered their identities to be “trade secrets” (Taylor 2010a). Yet, the FDA has recently issued guidance clarifying that brand-owning firms will be held accountable for violations at their CMs (Taylor 2010b). Second, the industry is experiencing a rapid growth in contract manufacturing—a trend driven largely by a slowdown in growth and decreasing margins caused by expiring patents and increasing pricing pressure (PRWeb 2011; Wilhelmsson 2004). Indeed, the global pharmaceutical CM market was estimated to be over $40 billion in 2009, and expected to reach over $70 billion by 2014 (Cosel 2011). Given the critical importance of outgoing product quality and the increased reliance on CMs, it is imperative to get insights into whether and when CMs may tend to operate at a higher level of quality risk than IPs.

The consensus of the analytic supply chain papers seems to be this: Due to double marginalization, differences in consequences of an external product quality failure, and unobservable quality effort at the CM, CMs will tend to invest less in product quality than a vertically-integrated manufacturer unless contracts and inspection policies are carefully designed (Baiman et al., 2000; Balachandran and Radhakrishnan 2005; Chao et al., 2009). On the other hand, a business strategy “core competence” lens may lead to the opposite view; i.e., CMs focus on manufacturing, and therefore, should be likely to operate with a lower level quality risk than an otherwise comparable IP. These two polar perspectives
pose a managerial dilemma as there are unresolved tensions between the classical operations management literature, which is largely based on agency theory, and the “core competence” perspective from business strategy literature, which is based on organizational learning and attention.

This paper differs from and contributes to the existing academic literature in several ways. First, we draw upon and integrate literature from OM/QM and economics/strategy to make a theoretical point: Subtle differences in incentives to invest in the resources required to operate with low quality risk will cause CMs to tend to operate at a higher level of quality risk than comparable internal plants, on average, \textit{ceteris paribus}, as long as the finished-product cannot be perfectly tested. Second, based on this theory, we develop a model, whereby we articulate how measurable plant characteristics may serve as proxies for factors that moderate this relationship; and thereby, establish contingencies for our theory development. Third, we apply a secondary-source, plant-level measure of quality risk as a dependent variable to test the hypotheses resulting from this theory development. Our measure of quality risk draws upon secondary data collected by the Food and Drug Administration (FDA), which is from a different source than our independent variables, and thus, our results are not subject to common methods variance (Podsakoff et al., 2003; Siemsen et al., 2010). We find moderate empirical support for our first-order hypothesis that CMs operate at a higher quality risk than IPs, on average. Further, this difference is moderated by plant size, plant age and primary product line produced in the plant in ways that can be considered consistent with theory. We propose that our findings are apt to apply in other manufacturing industries where finished product testing is not sufficient to ensure outgoing product quality.

The rest of this paper is organized as follows. In Section 2, we review the relevant literature; we develop our hypotheses in Section 3. In Section 4, we discuss our empirical setting and measures. We present and discuss the results of our analysis in Section 5, and we conclude in Section 6.

2. Related Literature

To motivate our model, we first review the relevant quality management (QM) literature, followed by related analytic and empirical research.
2.1 Theoretical Underpinnings of Quality Management and Risk

Before proceeding, we note that “quality risk” differs from “quality management” (QM) in the following sense: quality risk is an outcome variable related to the level of adherence to procedural requirements designed to ensure that any product released to the trade was produced in conformance to process specifications. QM includes several organizational-level constructs, such as customer focus (Nair 2006). Some of these QM constructs may be considered antecedents of low quality risk, whereas some more directly relate to other performance dimensions.

Operating with low quality risk is especially important in cases where the finished product cannot be perfectly tested. If the finished product can be perfectly tested for all possible defects, then the brand-owning firm can simply return defective products to the CM, or never take shipment of them if testing is performed at the plant. Thus, we restrict our theoretical development to the common settings where manufacturing defects related to process non-compliance, which is difficult to monitor, may possibly go undetected. We further develop this concept of “testability” of the finished product and its implications in to our research question in Section 4.

While there exists a plethora of literature on QM within factory walls (Nair 2006; Sousa and Voss 2002), few studies have specifically investigated the impact of different supply chain configurations on manufacturing quality (Flynn and Flynn 2005; Foster 2008). There is a construct in the QM literature pertaining to best practices in managing quality at suppliers, under the rubric of “supplier management” or similar (Forker 1997, Trent and Monczka 1999). This construct is one component of a Total Quality Management (TQM) program. The importance of supplier management to an effective QM program implies that if the buyer does not invest in good supplier management practices, suppliers may not operate at a level of quality risk that is optimal for the buyer. However, this implicit assumption is not examined. It is yet to be determined empirically whether and under what conditions systematic differences in quality risks are present when we contrast “pure” IPs and “pure” CMs. This empirical assessment is a key contribution of this paper to the OM/QM, supply chain, and business strategy literatures.
It is important to note that a buyer can conduct quality audits or surveys to assess “observable” practices (e.g., the presence of a Statistical Quality Control (SQC) program, written procedures for operations and change management, and/or documented training). However, a key insight from the OM and QM literatures is this: the mere existence of these observable practices is not sufficient to guarantee an effective quality management program (Dow et al., 1999; Giffti et al., 1990; Powell 1995). Rather, to mitigate risks, quality practices must be rigorously followed by employees in their daily work without shortcutting the established processes. These behaviors occur when leadership creates a culture for quality in which knowledge sharing, troubleshooting to determine root causes, and ongoing learning occur regularly among empowered employees on the production floor (Roth et al., 1994). Such a ‘realized’ quality culture, by nature, may be difficult for brand-owning firms to monitor, and in turn, influence when contracting out production to another business entity.

2.2 Related Analytic Literature

Our theoretical arguments draw primarily upon agency theory,¹ which considers settings where there exists both information asymmetry and goal misalignment between a principal and an agent (Eisenhardt 1989). When both information asymmetry and goal misalignment are present, theory would suggest that the agent will tend to optimize its private interests, which in turn, may be suboptimal for the principal (Laffont and Martimort 2002). In our study, we consider the owners and, by proxy, the corporate management of the brand-owning firm to be the principal. An important literature uses agency theory to consider approaches to align the incentives of top management with the shareholders/owners (Shapiro 2005). Here, we presume the interests of corporate management are aligned with the best interests of the firm, and consider the agency problem with regard to how this management can incentivize its plants to operate with these best interests in mind. The plant, either a CM or an IP, can be considered the agent. The key difference, by definition, between IPs and CMs is that in the case of CMs a firm boundary exists between the principal and the agent for production, whereas in the case of IPs there is no such boundary.

¹ Other perspectives are relevant to our research question. While agency theory is our focus and the basis for H1-H3, we do draw from other perspectives to develop H4.
The relationship between firm boundaries and quality has been modeled and analyzed in both the operations and accounting game-theoretic literatures. Some have focused on the effect of vertical disintegration on quality defined as the consumer’s valuation of the product (Economides 1999; Kaya and Özer 2009). Others have focused on conformance quality, which is more germane to our research. Typically a supplier-buyer dyad is used whereby the buyer procures finished product (or a component) from an external supplier; and conformance quality is operationalized as the percentage of parts that are defective. Recognizing that decentralization can lead to misaligned objectives due to differences in profit margins and consequences of external product failures, Reyniers and Tapiero (1995a and 1995b) and Wan and Xu (2008) examine how operational measures (i.e., incoming inspections) and contract structure (e.g., damage-sharing contracts in which costs of external failures are shared with the supplier) can be used to eliminate or mitigate the quality loss due to misaligned objectives. Baiman et al. (2000), Balachandran and Radhakrishnan (2005) and Hwang et al. (2006) focus on the moral hazard challenge in the dyad, whereby process compliance efforts on the part of the supplier and the quality monitoring efforts on the part of the buyer may be unobservable to the other party. These analytic papers also examine how contract structure and/or quality appraisal methods influence the conformance quality outcome. Adverse selection, due to information asymmetry regarding a supplier’s capabilities, is examined in Chao et al. (2009) and Lim (2001). The latter paper also incorporates moral hazard concerns, and examines if and how contracts can be structured to mitigate the quality loss due to decentralization. We draw upon these notions, and the more general principal-agent literature, to develop our hypotheses. Our contribution to the analytic literature is both a test of the implications of the assumptions, in part, by uncovering whether CMs, on average, have realized higher quality risks than IPs. Further, by including moderating variables we develop insights into managerially relevant contingencies.

2.3 Related Empirical Work

While the consensus of the analytic literature is that CMs will be prone to shirking with regard to quality, empirical scrutiny of actual quality differences between IPs and CMs has only been found in service management, not manufacturing. The service management literature provides tentative evidence that
vertically integrated service supply chains tend to have higher “quality” than disintegrated ones. For example, Harris and Winston (1983) reported that service quality was higher in railroads after vertical mergers. Michael (2000) found hotel and restaurant chains with more franchised outlets had lower quality performance, which they attributed at least partly to difficulty in contracting for quality. Hsieh et al. (2010) found that vertically integrated international courier services had less variability in delivery times than disintegrated ones. Taken together, these papers demonstrate that in service settings, where quality risk may be less tangible, firms find it difficult to overcome possible misaligned incentives such that the existence of an organizational boundary between the brand-owning firm and the operation relates to decreased quality performance. More closely related to our work, in a practitioner-oriented report, Macher and Nickerson (2006) used data from 26 pharmaceutical plants from 14 companies to find that plants that engage in any contract manufacturing (versus those that do not) tend to have poorer internal metrics, such as cycle time. However, they do not theorize about the drivers of this result nor the propensity for outgoing quality risk.

3. Hypotheses

In this section, we first draw from the above literature and the logic of agency theory to posit that, on average, a CM will tend to operate at a higher quality risk than a comparable IP. This hypothesis is a direct comparison of two types of plants (“pure” CMs versus “pure” IPs). We consider both plants agents to the principal, the corporate management of the brand-owning firm. In either case, there will be some level of information asymmetry between the principal and the agent, whether the agent is a CM or an IP. Appendix A lists several behaviors directly related to quality risk that would be difficult for the principal to observe directly. It is rather straightforward to argue that there will be more information asymmetry between the principal and the CM than there will be between the principal and the IP with regard to these issues. With IPs, the principal maintains decision rights over personnel, procedures, and investments, can monitor the operations as often as it desires, without pre-announcement or restrictions on access. In contrast, CMs maintain the decision rights regarding their hiring and internal reward systems. Further, with CMs, the brand-owning firm’s personnel often provide advance notice of their facility inspections.
Moreover, the monitoring of CM behaviors vs. IP behaviors may be less complete due to issues with confidentiality and trade secrets. Given these considerations, information asymmetry will be greater between the principal and the CM relative to that between the principle and the IP.

We make four separate, but related, arguments supporting the notion that goal misalignment between the CM and the principal is greater than that of the IP and the principal. Recall that for both plants the principal is the management at the brand-owning firm acting in the best interest of the brand-owning firm. First, typically the responsibility for any quality failure is shared between the CM and the brand-owning firm in their outsourcing relationship, especially in the presence of shared decision rights over production factors that may affect quality risk (e.g., joint process development). As stated by Mayer et al. (2004, p. 1065), “contracting for supplier liability when the buyer’s actions influence the size of the liability or the size of the liability is difficult to measure is not readily feasible because of verification difficulties. Such verification problems give rise to a moral hazard in which a supplier underprovides effort and hence quality.” Second, the brand-owning firm will tend to bear greater external failure costs (e.g., lost customer goodwill, brand reputation, etc.) than will be borne by the CM’s firm (see, for example, Rockoff and Lublin 2012). In a study at a brand-owning Fortune 500 company in the electronics industry, Nagar and Rajan (2001) found that one dollar of measurable external failure cost equated to a total profit loss of ten dollars, much of which would be difficult to assign to a CM even if the CM were responsible for the external failure. As press coverage of quality issues typically focuses on the brand-owning firm, the CM’s firm generally does not suffer the same public reputational effect. Third, we note that compliance with good manufacturing practices (GMPs) (which relates directly to our construct of quality risk) has been shown to be costly in the short term–as high as 25 percent of finished drug manufacturers’ operating costs (Bruttin and Dean 2004; Cosel 2011). Since CMs (versus IPs) may be more cost sensitive or may envision a shorter production horizon due to contracting periods, on average, they may be less incented to make such an investment for a specific product or process. The fourth argument is a bit more subtle. As gleaned from Appendix A, many of the behaviors necessary to operate with low quality risk are difficult

---

2 To our knowledge, no one has determined the long-term cost differences of different levels of compliance.
to observe. As articulated by Terlaak and King (2006, p. 583): a “supplier’s quality is often a function of intangible organizational characteristics that cannot be observed.” Consciously or unconsciously aware that quality behaviors are less observable and difficult to monitor, but that cost and delivery will be constantly be observed, CMs face a multi-dimensional incentive problem that may lead to shirking in the less visible performance dimension (Alchian and Demsetz 1972; Barzel 1982; Holmstrom and Milgrom 1991). Based on the above arguments, we assert that there is more goal misalignment between the principal and the agent (plant) when the agent is a CM than when the agent is an IP.

Due to the higher level of information asymmetry and goal misalignment regarding quality risk between the principals and CMs relative to that between the principals and IPs, we hypothesize:

H1: Contract manufacturer’s plants will operate with a higher level of quality risk than internal plants, on average, ceteris paribus.

The logic leading up to H1 indicated that due to agency issues with regard to the performance dimension of quality risk, CMs will tend to operate at a higher level of quality risk than an otherwise comparable IP, ceteris paribus. Based on this overarching theory, some plant characteristics may moderate the relationship between CM and quality risk. We examine three salient moderating factors: 1) primary product type that is a proxy for the intensity of government regulations faced by both types of plants; 2) plant size that we conceptualize as a proxy for the intensity of customer inspections faced by a CM. Both of these factors directly relate to agency theory. Finally, 3) plant age, relates to the knowledge-based view (Grant 1996) and organizational learning theories (Argote et al., 2003), which we discuss further in the lead-up to H4.

We first discuss the primary product type that the plant produces. We assert that primary product type relates to the intensity and consistency of external regulation. If brand-owning firms’ corporate offices are less able to align incentives with regard to quality risk with their CMs relative to their IPs, external regulation may serve as an incentive alignment mechanism. Based on our theory, the prospect of facing more intense regulatory scrutiny should have less of an effect on IPs (than CMs), as IPs were naturally more incented to operate with low quality risk. Arguably, CMs, who may be less naturally incentivized to
consistent with low quality risk based on the logic behind H1, will respond more to the presence of external regulation than a comparable IP. More formally:

**H2:** *There will be a negative interaction effect between the amount of regulatory scrutiny faced by a plant and CM on quality risk.*

Monitoring is one of the key mechanisms to align incentives in principal–agent relationships. We expect that all customers of CMs will, to some degree, monitor their CMs with finished product inspections and/or facility audits. And, CM size will tend to relate to the number of customers that a CM serves, on average. Therefore, larger CMs will tend to undergo more frequent external (customer) audits. All else equal, these audits will serve to focus the attention of the CM being audited to quality, at least during and shortly after each inspection (Anand et al., 2011; Ocasio 1997). Both large IPs and small IPs only have one primary customer—their own firm—and therefore, should have less variation in inspections relative to size. Therefore:

**H3:** *There will be a negative interaction effect between plant size and CM on quality risk.*

The age of the plant may have a different effect for CMs (versus IPs), in a way that does not directly relate to agency theory, but rather to learning and the knowledge-based view. Younger plants are faced with an initial state where the know-how is limited. IPs do not have an organizational boundary between R&D and manufacturing, whereas CMs do. Thus, the knowledge to produce the plant’s product(s) with low quality risk would have a tendency to be more easily transferred to an IP than to a CM (Kogut and Zander 1992). IPs will be able to more effectively engage in activities such as “Design for Manufacturing” (DfM) (Ettlie and Stoll 1990) and “Quality by Design” (Juran 1992) with the relevant R&D organization. IPs would be more likely to employ equipment and procedures designed specifically for the product produced at the facility. Thus, IPs may have a tendency to start-up with a high level of operational capability, and low quality risk, relative to a CM.

By the same logic, CMs may face steeper challenges initially, as each new product requires exchange of product and process knowledge across organizational boundaries. Further, as CMs generally wish to be capable of serving multiple customers with different requirements, they are more
likely to employ general-purpose equipment. Both of these factors could lead to a lower initial state of quality risk relative to an otherwise comparable IP. However, CMs will tend to face more product launches, with diverse organizations, over a given period of time relative to an IP, on average. This breadth of experience gleaned from serving multiple customers may lead to rapid improvement of operating knowledge, assuming that CMs continue to learn from each customer and each new product startup. The CM’s dynamic environment may lead to more learning opportunities than in an otherwise similar IP. We posit that the initial lower starting point and faster rate of learning will manifest themselves in the following way:

H4: There will be a negative interaction effect between plant age and CM on quality risk.

The arguments supporting H3 and H4 imply that size and age relate to quality risk differently for CMs relative to IPs. This is due to CMs starting at a higher quality risk because of incentives and knowledge issues, as well as the incentive alignment effects of frequent customer inspections and the learning benefits of developing processes for multiple customers’ products. Indeed, it may be that these drivers synergistically support each other, implying a three-way interaction. Formally:

H5: There will be a negative interaction effect between plant age, plant size, and CM on quality risk.

4. Empirical Setting and Measures

We test our hypotheses using manufacturing plants in the mainland United States classified as drug manufacturers by the FDA. We earlier mentioned “testability,” and noted that if products were perfectly testable incentives can be aligned by contract so our agency-related arguments do not apply. We now further develop our “testability” concept. We first divide testability into two dimensions: coverage (i.e., the percentage of products actually tested for quality defects) and thoroughness (i.e., the percentage of possible defects actually tested for in a given test). We use these dimensions to create Figure 1. This figure demonstrates how the quality risk in contract manufacturing—that is, the increased probability of a defective product leaving the plant when a CM is doing the work versus an IP—depends critically upon these testability dimensions. The figure also shows how mitigation approaches may differ depending on
the level of each testability dimension. The pharmaceutical industry can likely be categorized in the lower left corner of Figure 1. This is because drug production tends to be high volume and testing tends to be destructive, therefore only a small portion of all products are tested at all. Further, many drug products may be sensitive to process changes in ways that may not be apparent until the products age. And, it is impossible to test for all possible contaminants. Indeed, in Cosel’s (2011) Pew Institute report, Section 1.4.3 (p. 39-40) was entitled “Challenges involved in testing” and stated that “designing a test to capture any unexpected substance in a drug is essentially impossible” (p.39). Taken together, the low “testability” of drug production makes it a setting for which we expect agency issues to be prevalent. Therefore, based on our theory, there should exist a tendency for a higher quality risk in CMs relative to IPs. In contrast, a setting approaching complete testability may not face the agency issues discussed above. An example in practice is the manufacture of some electronics, where the final step in the production process is a functionality test. In such settings, 100% of products are tested for a large percentage of possible defects.

4.1 FDA Inspection Database

This study is part of a research program designed to assess the relationship between manufacturing plant characteristics and quality risk. This study on contract manufacturing involves a comparison of two types of domestic plants (CMs versus IPs), whereas prior research on offshoring studied matched onshore-offshore pairs of plants from the same company. Thus, we employ different theory and a distinctively different sample to address different research questions than Gray et al. (2011)’s study of offshore production. The two studies do draw from the same FDA inspection database and share the same heuristic for transforming a plant’s raw inspection history into a plant-level quality risk measure. In Appendix B, we describe both the FDA inspection data and the process used to develop the heuristic to measure plant-level quality risk for both this paper and Gray et al. (2011).

4.2 Plant Classification: Contract Manufacturer or Internal Plants

A challenging aspect of this study was the need to manually assign each plant as either a “pure” IP or “pure” CM. Our raw database was very large, and contained no information regarding whether the plant
was a CM or IP. As described in Appendix C, we employed a systemic process to classify plants. Upon completion of the classification procedure, our sample contained 64 “pure” CMs and 88 “pure” IPs.

4.3 Control and Moderating Variables

We controlled for all reliably-available plant-level characteristics. Secondary data at the plant level are difficult to obtain for any firm, and many of the firms in our study are private, further compounding the problem (Roth et al., 2008a). The National Establishment Time Series (NETS) database was our primary source for data on plant size, plant age, and primary product (SIC code). Plant size is measured as the number of employees ($Emp$). Plant age ($Age$) is the age of the plant as of 2007, in years. As a plausible proxy for the degree of external regulatory scrutiny, we code plant as one of two product types, based on primary SIC code. SIC code 283 ($Pharm=1$) reflects situations where the primary product is “drugs,” capturing over-the-counter and prescription pharmaceutical products. Products classified as “toilet goods/cosmetics” (SIC code 284, $Pharm=0$) are often regulated as drugs if they are designed to treat, cure, or prevent some illnesses (e.g., dandruff shampoo), and thus appear in the FDA’s database. While all plants in our data set are regulated by the FDA, pharmaceutical plants ($Pharm=1$) tend to face more consistent regulatory attention, and it is likely that all plant employees are acutely aware that they are working in a regulated environment. Conversely, plants producing primarily cosmetics ($Pharm=0$) often have a single regulated product (e.g., dandruff shampoo), and thus, may face less intense and frequent scrutiny. And, some such plants may not have been regulated for their entire history. We also included the company size, measured as total sales in billions of U.S. dollars ($TSales$) to control for company-level infrastructural support received at the plant-level. $TSales$ was obtained from the online D&B database or alternative sources, as it was not included in the plant-level NETS database. An examination of histograms of the continuous independent variables (i.e., $TSales$, $Emp$, and $Age$) indicated a right skew (Hair et al., 1998). Further, in the case of these variables, the effect of a one unit increase in the value on quality risk logically decreases as the value of the variable becomes larger. Thus, we log transformed these variables prior to all analyses; and named the transformed variables $\ln(Emp)$, $\ln(Age)$, and $\ln(TSales)$. Also, we centered all of the continuous variables at their means, prior to analysis.
4.4 Descriptive Statistics

The descriptive statistics and correlations for all variables are given in Table 1.  

As shown in Table 2, t-tests and z-tests reveal that CMs and IPs differ significantly on all control variables \( p < .01 \), as well as on quality risk. These statistical results underscore the importance of controlling for plant characteristics.

5. Model Estimation, Results, and Discussion

Our full model (i.e., Model 6 in Table 3) is represented by the equation:

\[
\ln(QRisk) = \beta_0 + \beta_1CM + \beta_2 \ln(Age) + \beta_3 \ln(TSales) + \beta_4 \ln(Emp) + \beta_5 Pharm + \\
\beta_6 (Pharm \times CM) + \beta_7 (\ln Emp \times CM) + \beta_8 (\ln Age \times CM) + \\
\beta_9 (\ln Age \times \ln Emp \times CM) + \varepsilon
\]  

The details are given in 4.1 followed by the presentation of results and discussion.

5.1 Econometric Specification and Results

Due to the restriction of our sample to plants that have been inspected three or more times by the FDA, the distribution of the dependent variable was reasonably continuous and unimodal, in spite of the fact that individual inspection outcomes are discrete. This distribution allowed us to employ ordinary least squares (OLS) regression. As discussed later in this section, we perform regression diagnostics to confirm that the model meets the assumptions of OLS.

To avoid losing 13 observations due to missing data in \( TSales \) and \( Age \), we employ multiple imputation (Allison 2002). We utilize 5 imputations, which is deemed to be more than adequate given the small amount of missing data present (Rubin 1987). Further, since the 152 plants came from 116 companies, we address concerns with non-independent standard errors among the plants by employing the Huber-White “sandwich” estimate of variance using Stata’s “cluster” option (Rogers, 1993; Williams 2000). In Table 3, we present results of the OLS analyses.
Model 1 results indicate moderate support for H1. Controlling for measurable plant characteristics, CMs have a higher quality risk than internal plants ($\beta=.115, p<.10$), on average. Models 2-6 show that this first-order result is substantially moderated by plant-level factors, consistent with our hypotheses. The significant interaction effects in Models 2-4 show that all three plant-level factors—primary product ($Pharm$), plant size ($\ln(Emp)$), and age ($\ln(Age)$)—relate to plant-level quality risk differently for CMs than they do for IPs. Model 5 indicates that there is also a significant three-level interaction between CM, age, and size. Model 6 includes all variables in a single model, and is the best-fitting model which we therefore now discuss. From Model 6, we see that primary product ($H2, Pharm*CM, \beta_6=-.268, p<.05$), plant size ($H3, \ln(Emp)*CM, \beta_7=-.188, p<.01$), and age ($H4, \ln(Age)*CM, \beta_8=-.153, p<.10$) all negatively moderate the relationship between plant type (CM) and $\ln(QRisk)$. Further, there is a significant three-way interaction between age, size, and CM ($H5, \ln(Age)*\ln(Emp)*CM, \beta_9=-.098, p<.01$). Taken together, we have at least moderate support for each of our hypotheses.

Standard regression diagnostics were run on each of the 5 imputations of Model 6. As would be expected, given the small amount of missing data present, each of the 5 regression models were very similar. In all cases, the residual-vs-fitted plot did not indicate any nonlinearity or heteroskedasticity. In no imputation were more than five standardized residuals greater than 2 (about 7 would be expected by chance if the assumptions are met). Further, both the histogram and q-q plot of the residuals indicated approximate normality. We formally checked for evidence of excessive multicollinearity in Model 6 using commonly accepted rules of thumb; i.e., variance inflation factors (vif) and condition indices (Hair et al., 1998). We found no evidence of problematic multicollinearity as no vif of any variable in any imputation was greater than four. Furthermore, the null hypothesis of correct model specification per the Ramsey RESET test was not rejected ($p>.20$) for all 5 imputations (Wooldridge 2009).

To refine our understanding of the interaction effects present in Model 6, we performed a post hoc analysis with individual models for the CMs and IPs. The regression models are given in Equations 2a and 2b, and the results are depicted in Tables 4 and 5 below.
Table 4 and 5 provide more in-depth information about how measurable plant-level characteristics relate to quality risk, for CMs and IPs, separately. Note that the average R² for the CMs in substantially higher than that for the IPs (29.4% vs. 7.8%), indicating that these characteristics explain much more of the variance in CMs relative to IPs. Table 4 demonstrates that CM size relates significantly to quality risk ($\beta_{3CM} = -.149$, $p<.01$), and that age and size interact ($\beta_{4CM} = -.095$, $p<.05$). For IPs (Table 5), no plant-level characteristics relate to quality risk; however, company size has a moderate effect ($\beta_{2IP} = -.020$, $p<.10$).

<<Insert Tables 4 and 5 About Here>>

5.2 Discussion of Results

Our empirical results provide moderate empirical evidence that contract manufacturers (CMs) operate at a higher level of quality risk than internal plants (IPS), on average and controlling for measurable characteristics. However, this difference is moderated by several plant-level factors. Drawing upon OM-based theory, the principals of a brand-owning firm may find it difficult to monitor the CM’s plant operations, and hence, to determine the true quality risk of a CM relative to an IP. This monitoring difficulty combines with misaligned incentives between the CM and the brand-owning firm to create a quality risk difference between IPs and CMs. This quality risk difference should exist in other industries where the finished product is not completely testable; we leave the empirical replication in other industries to future research.

We found evidence that the relationship between being a CM and quality risk was moderated by the level of external regulatory scrutiny that the products face (see Figure 2). Specifically, in pharmaceutical plants, as expected, there was less of a difference in quality risk between IPs and CMs than existed in cosmetics plants. While all plants were regulated to some degree, the cosmetics plants often produce mainly non-regulated products, with only a few regulated by the FDA. Indeed, in this sample, cosmetics
plants have been less consistently inspected by the FDA, on average. The pharmaceutical plants in our sample were inspected an average of ten times, the cosmetics plants an average of four times during the time period. Our theory would indicate that internal plants would be more likely to invest in systems and procedures to operate at low quality risk, on average, regardless of the nature of external regulation. While we are cautious about generalizing beyond our data, it is reasonable to suspect that in otherwise similar industries with minimal or no external regulation that there may be a greater difference between IPs and CMs with regard to quality risk. Again, this is an area prime for future research.

Plant size also affects IPs differently than CMs. The analysis indicates that larger CMs have lower quality risk than smaller CMs. As illustrated in Figure 3, large CMs in fact achieve a predicted level of quality risk comparable to large IPs. Our agency-theory related explanation in H3 is that larger CMs are likely to have many customers, and therefore, more frequent external audits than a smaller CM. Frequent inspections act as a mechanism to focus attention on quality and align incentives.

The three-way interaction effect among plant age, plant size, and CM in Model 6 of Table 3 (and the significant negative interaction in Model 1 of Table 5) both indicate that the negative effect of size on quality risk for CMs is enhanced by plant age. Figure 4 depicts this for CMs. The contour plot shows a relationship from high predicted quality risk (lower portion of the graph) to lower predicted quality risk (upper right corner) with both CM age and size. A similar graph for IPs, not presented, has no discernible pattern.

Overall, our analyses indicate this: While the quality risk of IPs is not very sensitive to the plant-level characteristics that served as controls, the quality risk of CMs is sensitive to them. Two of the effects—size and the level of regulatory scrutiny of the primary products—can both be thought of as factors that increase the level of external monitoring. This monitoring acts to reduce, and eventually overcome, the agency-related concerns with contract manufacturing expressed in H1. The third moderating variable,
age, relates to learning. Our results are consistent with the knowledge-based view of the firm (Grant 1996). These empirical results are of great importance in establishing contingencies under which insights from the analytic model assumptions hold or do not hold. Further, they help to resolve some of the tension between quality and strategic management theories.

5.3 Robustness Checks

In this section, we examine the robustness of our results to a different specification of the dependent variable, and also to our treatment of missing data. While the Delphi panel reached a consensus on the specific dependent variable used in our analysis, the inclusion and measurement of the trend adjustment was a debated topic. Thus, we also check our results where the dependent variable is just a simple average quality risk score, with no trend adjustment. With the exception of the \( \ln(\text{Age}) \times \text{CM} \) interaction (H4), all hypotheses remain supported in this robustness check (at least \( p < .10 \)). Further, we note that when listwise deletion is employed to deal with missing data instead of multiple imputation, all hypotheses except H4 are at least moderately supported. Taken together, these robustness checks indicate that the agency theory-related hypotheses (H1, H2 and H3) are the most consistently supported.

6. Conclusion

In this section, we first discuss the limitations of our study and consider alternative explanations. Then, we present some opportunities for future research. Finally, we review the implications for research and practice (including policy), and conclude.

6.1 Limitations and Future Research

This study used carefully collected secondary data from multiple sources to examine the quality risk of CMs, which is of mounting importance to operations, supply chain management, and business strategy. We argued through the lens of agency theory that CMs would tend to operate with a higher quality risk than internal plants, but that various factors would moderate this difference. We generally found support for our hypotheses. The data sources and approach both have limitations, some of which can be addressed in future work.
We first discuss the measurement of our variables. Regarding our dependent variable, while FDA inspections are thorough, standard, and designed explicitly to assess quality risk, the combination of two inspection results (i.e., Form 483 yes/no and District Decision) used in this study is a coarse measure. In addition, there is variability in FDA inspectors (Macher et al., 2011) as well as inspection purpose. Although imperfect, we assert that the FDA inspection history is the best available measure of quality risk available for many plants producing non-identical products. The fact that a panel of experts was employed to develop this measure for this research stream, as described in Section 4.1 and Appendix B, provides substantial face validity to the measure. We also note that the inspection procedure is not different for CMs and internal plants, as all drug investigations are guided by the *Investigator Operations Manual*, Chapter 5 (FDA 2012). This manual explicitly states that it applies “to all individuals who perform field investigations.” Further, inspectors can be considered to be assigned randomly to plants (Macher et al., 2011). Regarding control and moderating variables, their reliability cannot be readily assessed (Roth et al., 2008a) and for this plant-level analysis these variables were only consistently available at the end of the time period of the inspections use to construct the quality risk measure. However, we assert that our measures are the best available representations of a plant’s size, age, primary product, and total sales available through secondary sources, especially considering that our measures are at the plant level.

A strict and technically correct interpretation of our H1 result is this: conditional upon the variables included in the model, CMs tend to operate with moderately higher levels of quality risk than IPs. This finding is consistent with the findings of Macher and Nickerson (2006)’s practitioner-oriented report, which found that plants engaging in contract manufacturing had lower internal metrics than “pure” IPs. Further, H2-H5 indicate that this first-order result is substantially moderated by plant size, plant age, primary product line, and the interaction of plant age and plant size. The inclusion of control variables and interactions captures many possible alternative explanations for our results. However, only a controlled experiment can unambiguously determine the effect of an X on a Y; causality cannot be confirmed with certainty in this observational study.
Our research has identified areas that point toward interesting follow-up studies. Studies similar to ours in other industries, and/or with additional moderators, would enhance the generalizability of these findings and also test the validity of our theoretical explanations of the key drivers. However, in industries not regulated by the FDA, new establishment-level measures of quality risk would be necessary. Another ripe opportunity for research would involve collecting primary data from buyers and CMs to obtain insights into the management of quality risk across organizational boundaries. This research could uncover how transaction-level characteristics, as well as the nature of the internal CM systems and buyer behaviors, influence the average quality risk of CMs. Gray and Handley (2011) and Handley and Gray (2012) have begun conducting such research.

6.2 Implications for Research and Practice

This paper offers several contributions to the research community. First, we articulate a theoretical explanation as to the key drivers of differences in the quality risk performance of CMs and IPs. That is, when finished products cannot be perfectly tested, a brand-owning firm may not be able to align the CM’s incentives regarding quality risk with its own. Second, we provided moderate empirical evidence that in the pharmaceutical industry, CMs tend to operate with a higher quality risk than internal plants, on average, controlling for some measurable plant characteristics. This empirical finding is consistent with the majority of the analytic literature, but is novel in the empirical literature. Third, we found that this relationship is significantly moderated by the plant’s primary product type (an indicator of external regulatory scrutiny), plant size and plant age. Plant size (\(\ln(\text{Emp})\)) was the strongest and most consistently supported moderator. Regarding primary product and plant size, we proposed that the moderation is driven by an increased threat of external inspections (more FDA inspections in the case of pharmaceutical CMs vs. cosmetic CMs, and more facility audits by buyers/customers in the case of large CMs vs. small CMs). Our theory indicates that this inspection threat has a greater impact on CMs due to their tendency to invest less in quality risk-reducing activity relative to IPs who will more naturally operate with lower quality risk. Importantly, these moderating effects are consistent with the agency theory lens. Regarding plant age, we speculate both that CMs start at a higher level of quality risk due to the presence of
organizational boundary between themselves and the product experts, but that they learn more rapidly than IPs due to producing more products and serving multiple companies (i.e., they may more rapidly develop absorptive capacity (Cohen and Levinthal, 1990)). These moderating factors provide a rich picture of the relationship between quality risk and the presence of a firm boundary between brand-owning firm and its manufacturing operation.

Our study also offers managerial insights. First, the underlying drivers of a quality risk difference between IPs and CMs—including the difficulty of monitoring quality-risk lowering behaviors—are difficult to eliminate. As a result, firms availing of contract manufacturing should recognize the potential quality risk associated with this strategy. While difficult to completely eliminate, buying firms should commit substantial resources to monitor CMs (e.g., put resources in CM’s plants, increase audits) and also provide contractual clauses to align incentives (Handley and Gray 2012). Second, brand-owning firms using CMs should be particularly wary of small and young CMs. In contrast, large and experienced CMs can achieve quality risk levels comparable to similar IPs (at least in this regulated industry). If true, this seems to indicate that CMs—serving many customers vs. just one—actually improve with regard to quality risk more quickly, albeit from a lower starting point, than IPs. In industries such as electronics with large and established CMs, this may indicate that there is no increased quality risk in employing CMs, on average. Third, the moderating influence of external regulatory scrutiny would suggest that the incentive differences driving our main hypothesis may be significantly more pronounced in industries that are not subject to external regulation. Fourth, Figure 1 and the concept of “testability” can be utilized by managers to help determine whether and how there may be a tendency for CMs to operate a higher quality risk than a comparable IP, ceteris paribus. This factor, testability, is not explicitly covered in most treatments of make-buy decisions, and is not directly related to commonly-employed constructs in the make-buy literature, such as asset specificity (Williamson 1985). Testability is, however, related to the “measurement” perspective in the theory of the firm (Alchian and Demsetz 1972; Barzel 1982).

Finally, our research provides some policy guidance to the FDA. The FDA’s holding of buying firms accountable for the actions of their CMs (Taylor 2010b) may have the unintended consequence of
exacerbating the agency theory issues discussed earlier. However, the benefits in terms of encouraging buying firms to more closely monitor their CMs likely outweigh this drawback. We also note that our research timeframe pre-dates the following from Cosel (2011, p.34): “A surge in FDA warning letters to both foreign and domestic contract manufacturers illustrates the agency’s concerns….An FDA official noted that these increases specific to contractor problems probably were not attributable to changes in FDA activity, as the agency had not stepped up oversight of contract manufacturers as a category.” Our research indicates that the FDA should perhaps increase its focus on CMs, particularly those that are small and young.

In conclusion, this research provides a theory-based explanation and some empirical evidence of a tendency for CMs to operate with a higher quality risk than that of similar IPs, on average. We find interesting moderation effects that are generally consistent with our agency theory lens, as well as organizational learning perspectives. We believe this work adds an important strategic operations perspective to the theory of the firm, with the plant as the unit of analysis. We hope it will serve as a point of departure for future research on the quality implications of different supply chain configurations and also provides managers with an approach to weigh quality risk when exploring a contract manufacturing strategy.
## APPENDICES

### Appendix A: Examples of Difficult to Monitor Quality-Risk Reducing Practices

<table>
<thead>
<tr>
<th>Category</th>
<th>Practices</th>
</tr>
</thead>
</table>
| **Prior to Startup**      | ✓ Training: Assuring all production and laboratory personnel know proper operating procedures, process conditions, and what to do in unusual circumstances.  
                            | ✓ Process Validation: testing and ensuring the production and laboratory processes will perform as intended, including under unusual conditions  
                            | ✓ Raw Material Qualification: Ensuring vendor knows acceptable range of raw material specifications, and has capability and systems to continuously meet those specifications  
                            | ✓ Computer System Qualification: Ensuring all automatic systems are validated across the range of scenarios they may possibly face  |
| **During Normal Production** | ✓ Following Procedures: Ensuring written procedures are continuously followed and any deviation is justified  
                              | ✓ Calculation of Yield: Ensuring two individuals independently account for all materials  
                              | ✓ Addition of Materials: Ensuring two individuals independently confirm the proper amount and type of ingredient is added  
                              | ✓ Testing of Product: Ensuring proper items tested at proper intervals; ensuring two individuals test that product meets specifications  
                              | ✓ Hygiene: Ensuring personnel and plant are clean and sanitary; including free of pests  
                              | ✓ Release Authority: Ensuring only trained quality control personnel are authorized to release product to the trade  |
| **During Changeover/Shutdown** | ✓ Ensuring that proper cleaning and sanitation procedures are in place and followed for transition from Product A to Product B or extended shutdown  
                                | ✓ Ensuring maintenance procedures are consistently followed and documented  |
| **Change Management**     | ✓ Ensuring all changes are reviewed by qualified QA personnel and validations and training redone as needed  |
| **Out of Specification Product** | ✓ Ensuring proper procedure followed after the discovery of any out of specification results  
                                   | ✓ Ensuring QA has the power to prevent the release of product until cause of out-of-specification is determined  |

*Source: Code of Federal Regulations, 21 CFR-Parts 210 and 211, also known as the “Good Manufacturing Practices” (GMPs)*
Appendix B: Measure of Plant-Level Quality Risk from FDA Inspection Data

Some of the following is paraphrased from Gray et al. (2011), where the interested reader can find a more detailed description. The FDA is responsible for inspecting the plants that produce drug products for consumption in the U.S. to ensure compliance with GMPs. The inspectors are trained to consistently follow the *Investigations Operations Manual* (FDA 2012) to conduct these inspections. For this research stream, we submitted a Freedom of Information Act request to the FDA to obtain data on the outcomes of all plant inspections from January 1994-January 2007. The time period represents a time of relative stability of FDA inspection guidelines, between the 1993 increase in FDA authority (Farley 1993) and the full implementation of risk-based inspections and increased focus on enforcement under the Obama administration (Sharfstein 2009).

For each inspection conducted at a particular facility, there are two separate assessments of its quality risk. The first is an indication as to whether the inspector issued a “Form 483” with the establishment inspection report (EIR). A Form 483 is issued “when in the investigator's ‘judgment,’ conditions or practices observed, indicate that any food, drug, device, or cosmetic have been adulterated or are being prepared, packed, or held under conditions whereby they may become adulterated or rendered injurious to health” (FDA 2012, Section 5.2.3). The EIR, with or without a Form 483, is then sent to the FDA district office. The district examines the EIR, and also checks other sources of information regarding the plant’s compliance (e.g., consumer complaints and/or tests of samples purchased from the trade), and makes one of three determinations. An “official action” is the most severe determination. In this case, the FDA has concluded that regulatory or administrative sanctions will be recommended if issues are not resolved in a timely manner. Next is the “voluntary action” category. This means that while one or more objectionable conditions were found at the facility, the FDA district office is not going to recommend any further administrative action beyond routine re-inspection. The third is a recommendation of “no action,” which indicates that no follow-up is required. For this research stream, a panel of four industry experts with extensive experience at both the FDA and FDA-regulated plants went through a Delphi Process (Linston and Turoff 1975) to develop the quality risk score heuristic. The panel biographies are given in Gray et al.
they included a former FDA district director, two individuals with substantial operational experience in FDA-regulated facilities, and one individual with significant quality consulting experience. After a series of independent interviews, the panel agreed on a “score” for each possible set of inspection outcomes. These scores are given in Table A1, along with percentages of each combination of outcomes for our sample in this paper.

Having determined appropriate single-inspection scores, the next challenge was to determine a method to handle the fact that all plants used in this study have been inspected multiple times to create a plant-level score. The panel settled on this simple transformation in our sample for the quality risk variable:

\[
QRisk = \frac{\sum_{j=1}^{n} QR_j}{n} + \frac{\sum_{j=1}^{n-1} QR_{j+1} - QR_j}{n+1}
\]

where \(QR_j\) is the Single Inspection Quality Risk score from Table 1 for inspection \(j\); \(n\) = the number of inspections, and \(j\) = the indicator for a specific inspection (ordered chronologically). \([\text{from Gray et al., 2011}]\)

The first term is the simple average of the single-inspection quality risk scores, and the second term is a simple trend adjustment.

**Appendix C: Classification of Plants as IPs or CMs**

The raw database was large (29,955 inspections of 14,029 plants). We carefully eliminated several types of plants before manually coding so that the time-consuming activities could be performed on a smaller subset of plants more likely to be relevant for the final analysis. First, we deleted “unusable” audits—those without a district decision, which lowered the number of inspections to 29,371 and the number of plants to 13,662. Then, we eliminated all plants with two or less usable audits, which resulted in the elimination of 10,628 plants and 12,659 inspections. This was primarily done to make the manual search of plants tractable, but had some additional benefits. For example, a large number of plants unlikely to be applicable (e.g., those that in fact are not engaging in manufacturing—which would be discovered after one audit) were eliminated from further consideration. And, we reasoned our plant-level measure would be more reliable if the FDA had multiple inspections during the time period under consideration. With
this smaller sample size, we performed a preliminary manual check of the plants to further determine status. For example, any companies that were obviously primarily wholesalers, doctor’s offices, pharmacies, retailers, pure R&D, home medical devices, etc. were removed from the data set, with a resulting data set of 10,241 inspections of 1556 plants. Next, we eliminated all foreign plants, including those from Puerto Rico and the U.S. Virgin Islands, leaving 8116 inspections of 1164 plants. From these remaining 1164 plants, we manually matched names and addresses to merge those that could be found in a subset of the National Establishment Time Series (NETS) database that included all manufacturers with which had a primary or non-primary SIC code of 2834 or 2844 anytime during the years 1990-2009. After merging, our database for consideration was comprised of 546 plants. Of these, 415 had primary SIC code 283; 105 had SIC 284, the remaining 28—no more than 5 in any three-digit SIC code—had another 3-digit primary SIC code. Examination of secondary SIC codes, combined with internet searches, led to the assignment of all of these plants to either SIC code 283 or 284. Available secondary information was searched following written guidelines to determine whether the plant engaged in contract manufacturing and/or production of branded products. To be included in our sample for this study, plants had to be classified “pure” CMs or “pure” IPs. For the CMs, there could be no evidence on the plant’s website of any company-owned brands produced at the plant. And, finished- or nearly-finished products had to be in the plant’s offering of capabilities, i.e., the plant could not be primarily engaged in active pharmaceutical ingredients (APIs) or bench-scale R&D. CM classifications were reasonably straightforward, as “pure” CMs tend to advertise their capabilities. To be classified as an IP, in addition to finding evidence that the plant produced products under their own consumer brands, there had to be no evidence of contract manufacturing activity. As branded-firms don’t always advertise their CM activities, we only considered a plant a “pure” IP if we found no evidence of CM activity on the company’s website or through a specific internet search protocol that included checking several CM directories. Finally, we discarded plants with less than 5 employees. Upon completion of the classification procedure, our sample contained 64 “pure” CMs and 88 “pure” IPs.
REFERENCES


### Table 1: Descriptive Statistics and Pearson Bivariate Correlations

<table>
<thead>
<tr>
<th>Name</th>
<th>N</th>
<th>Mean</th>
<th>StDev</th>
<th>QRisk</th>
<th>CM</th>
<th>TSales</th>
<th>Emp</th>
<th>Age</th>
<th>Pharm</th>
</tr>
</thead>
<tbody>
<tr>
<td>QRisk</td>
<td>152</td>
<td>1.11</td>
<td>.81</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CM</td>
<td>152</td>
<td>.421</td>
<td>.495</td>
<td>.25***</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TSales</td>
<td>148</td>
<td>14.8</td>
<td>25.0</td>
<td>-.22***</td>
<td>-.30***</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emp</td>
<td>152</td>
<td>306.2</td>
<td>581.7</td>
<td>-.16**</td>
<td>-.24***</td>
<td>.25***</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>143</td>
<td>41.3</td>
<td>34.1</td>
<td>-.12</td>
<td>-.34***</td>
<td>.14</td>
<td>.21**</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Pharm</td>
<td>152</td>
<td>.645</td>
<td>.480</td>
<td>-.07</td>
<td>-.23***</td>
<td>.17**</td>
<td>.05</td>
<td>.14*</td>
<td>1.00</td>
</tr>
</tbody>
</table>

* p < .10; ** p < .05; *** p < .01. Listwise deletion used for correlation table. QRisk is defined by Equation (1); CM=1 if plant is a CM, 0 for IP; other variable descriptions can be found in Section 3.3.

### Table 2: Mean Differences and Differences in Proportions

<table>
<thead>
<tr>
<th>Name</th>
<th>Mean (IP)</th>
<th>Mean (CM)</th>
<th>Test statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>QRisk</td>
<td>1.35</td>
<td>.94</td>
<td>-3.18***</td>
</tr>
<tr>
<td>TSales</td>
<td>5.99</td>
<td>21.3</td>
<td>b 3.84***</td>
</tr>
<tr>
<td>Emp</td>
<td>140</td>
<td>427</td>
<td>3.09***</td>
</tr>
<tr>
<td>Age</td>
<td>28.3</td>
<td>51.8</td>
<td>b 4.34***</td>
</tr>
<tr>
<td>Pharm</td>
<td>.516</td>
<td>.739</td>
<td>2.84***</td>
</tr>
</tbody>
</table>

*p<.10, ** p< .05, *** p <.01 We used t-statistic for continuous variables (difference in means) and z-statistic for 0-1 variables (difference in proportions). Missing data on 4 plants (3 IPs, 1 CM) for TSales and 9 plants (all IPs) for Age. Observations with missing values were excluded from the t-tests.

### Table 3: Plant-Level OLS Regression on Quality Risk (DV=ln(QRisk))(n = 152, 5 imputations)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
<th>Model 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>ln(Age)</td>
<td>.601**</td>
<td>.509**</td>
<td>.590**</td>
<td>.590**</td>
<td>.585**</td>
<td>.488**</td>
</tr>
<tr>
<td>ln(TSales)</td>
<td>-.033</td>
<td>-.038</td>
<td>-.011</td>
<td>-.011</td>
<td>-.035</td>
<td>.037</td>
</tr>
<tr>
<td>ln(Emp)</td>
<td>-.015</td>
<td>-.012</td>
<td>-.019</td>
<td>-.018</td>
<td>-.018</td>
<td>-.022</td>
</tr>
<tr>
<td>Pharm</td>
<td>.036</td>
<td>.158**</td>
<td>.034</td>
<td>.031</td>
<td>.029</td>
<td>.157</td>
</tr>
<tr>
<td>CM(H1)</td>
<td>.115†</td>
<td>.278*</td>
<td>.081</td>
<td>.089</td>
<td>.072</td>
<td>.245</td>
</tr>
<tr>
<td>Pharm*CM (H2)</td>
<td>-.254* (.142)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ln(Emp)*CM (H3)</td>
<td>-.167** (.045)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ln(Age)*CM (H4)</td>
<td>-.181*(.102)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ln(Age)*ln(Emp)*CM (H5)</td>
<td>-.095* (.040)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Average R²: 8.4% a 10.7% b 16.4% c 10.7% d 19.1% e 21.6% f
Average Adj R²: 5.3% a 7.0% b 13.0% c 7.0% d 14.6% e 16.7% f

Notes: *p<.10, * p< .05, ** p <.01; two-tailed test for control variables, one-tailed test when sign of coefficient hypothesized.

a 5 imputations; range 8.0-9.0%; b 10.3-11.3%; c 15.9-16.8%; d 10.3-11.3%; e 18.6-19.8%; f 21.1-22.2%
Table 4: Plant-Level OLS Regression on Quality Risk for CM SubSample  
(DV=ln(QRisk)) (n = 64, 5 imputations)  
(Nonstandardized Parameter Estimates; all continuous variables mean-centered)

<table>
<thead>
<tr>
<th>Model</th>
<th>Variable</th>
<th>β (SE)</th>
<th>Average R²</th>
<th>Average Adj R²</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(constant)</td>
<td>.745**(.082)</td>
<td>29.4%</td>
<td>23.2%</td>
</tr>
<tr>
<td></td>
<td>ln(Age)</td>
<td>-.117 (.080)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ln(TSales)</td>
<td>-.015 (.015)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ln(Emp)</td>
<td>-1.49**(.042)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pharm</td>
<td>-.129 (.098)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ln(Age)*ln(Emp)</td>
<td>-.095* (.040)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: †p<.10, * p<.05, ** p <.01; two-tailed test for all variables  
*5 imputations; range 29.4-29.4%

Table 5: Plant-Level OLS Regression on Quality Risk for IP Subsample  
(DV=ln(QRisk)) (n = 88, 5 imputations)  
(Nonstandardized Parameter Estimates; all continuous variables mean-centered)

<table>
<thead>
<tr>
<th>Model 1</th>
<th>Variable</th>
<th>β (SE)</th>
<th>Average R²</th>
<th>Average Adj R²</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(constant)</td>
<td>.492**(.093)</td>
<td>7.8%</td>
<td>2.2%</td>
</tr>
<tr>
<td></td>
<td>ln(Age)</td>
<td>.031 (.061)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ln(TSales)</td>
<td>-.020† (.012)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ln(Emp)</td>
<td>.055 (.035)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pharm</td>
<td>.152 (.104)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ln(Age)*ln(Emp)</td>
<td>-.014 (.041)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: †p<.10, * p<.05, ** p <.01; two-tailed test for all variables  
*5 imputations; range 7.1-8.7%

Table A1: Distribution of Possible Plant Inspection Outcomes – Single Inspection

<table>
<thead>
<tr>
<th>Form 483</th>
<th>District Decision</th>
<th>Percent of Inspections for Plants Used in Study, 1994-2007 (n=1253)</th>
<th>Delphi Panel Consensus Quality Risk (QR)Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>No Action</td>
<td>39.3%</td>
<td>0</td>
</tr>
<tr>
<td>No</td>
<td>Voluntary Action</td>
<td>5.6%</td>
<td>0.5</td>
</tr>
<tr>
<td>No</td>
<td>Official Action</td>
<td>1.4%</td>
<td>3</td>
</tr>
<tr>
<td>Yes</td>
<td>No Action</td>
<td>3.0%</td>
<td>1</td>
</tr>
<tr>
<td>Yes</td>
<td>Voluntary Action</td>
<td>35.7%</td>
<td>1.5</td>
</tr>
<tr>
<td>Yes</td>
<td>Official Action</td>
<td>15.0%</td>
<td>3.5</td>
</tr>
</tbody>
</table>
**FIGURES**

**Figure 1:** How the Two Dimensions of Testability Affect the Ability to Manage Quality Risk in Contract Manufacturing

<table>
<thead>
<tr>
<th>Coverage</th>
<th>Thoroughness</th>
<th>Incomplete Inspections</th>
<th>Complete Inspections</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>Low</td>
<td>Of Most Units</td>
<td>Of Most Units</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Systemic issues may persist on non-testable dimensions</em></td>
<td><em>Can effectively contract for quality</em></td>
</tr>
<tr>
<td>Low</td>
<td>High</td>
<td>Of Few Units</td>
<td>Of Few Units</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Difficult to effectively contract for quality</em></td>
<td><em>Can use statistical/random sampling to understand risk</em></td>
</tr>
</tbody>
</table>

**Figure 2:** Average Predicted ln(QRisk), by plant type (CM vs. IP) and primary product (Pharmaceutical vs. Cosmetic)
Figure 3: Linear Fit of Predicted $\ln(QRisk)$, by plant type (CM vs. IP)

Figure 4: Contour graph of Predicted Quality Risk, for CMs