Authors: *Gray, Roth, and Tomlin*
Revision 1 submitted to *Management Science; no. MS-00017-2008*

Quality Risk in Contract Manufacturing: Evidence from the U.S. Drug Industry

John V. Gray
Fisher College of Business, The Ohio State University, Columbus, Ohio 43210, gray.402@osu.edu

Aleda V. Roth
College of Business and Behavioral Science, Clemson University, Clemson, South Carolina 29634, aroth@clemson.edu

Brian Tomlin
Kenan-Flagler Business School, University of North Carolina, Chapel Hill, North Carolina 27599, brian_tomlin@unc.edu

This paper investigates quality risk in contract manufacturing, focusing on prescription and over-the-counter drug manufacturing plants in the United States. Applying the Delphi method with a panel of industry experts, we derive a new plant-level measure of quality risk based on Food and Drug Administration (FDA) inspection data. We posit, based on theory, that contract manufacturers' plants pose a higher quality risk than internal plants, on average. Our model is tested empirically on a sample of 154 plants classified as drug manufacturers by the FDA. The results indicate that the outsourcing of a product to contract manufacturers does present a significant quality risk to buying firms, but that a plant's age is an important contingency factor. This finding is contrary to conventional wisdom based on the "core competence" argument, which promotes production outsourcing to allow brand-owning drug firms to focus on product development and marketing capabilities. In contrast, an operations perspective reveals that crucial tacit aspects of an effective quality management program cannot be easily codified in contracts and may not be visible to the buying firm. This lack of visibility, combined with misaligned incentives, is a key driver of the observed quality risk associated with contract manufacturing.

Key Words: quality management, outsourcing, contract manufacturing, drug industry, supply chain management

Introduction

Outsourcing the production of parts, components, and systems has long been a tradition in operations and supply chain management (Hayes et al. 2005). Since the early 1990s, the outsourcing of the production of finished products to contract manufacturers has intensified in many industries (Tully 1994), including electronics (Sturgeon 2002), consumer products (Jeffries 2003), and even automobiles (Edmondson 2003). While the outsourcing of production to contract manufacturers may provide many potential benefits to the buying firm, it also poses possible risks (Quelin and Duhamel 2003, Hayes et al. 2005). Scholars in economics (e.g., Williamson 1991, Grossman and Hart 1986) and business strategy (e.g., Barney 1999, Leiblein 2003) have theoretically and empirically examined the conditions under which outsourcing is generally the preferred choice for a given activity. However, much of the extant literature has focused on direct economic risks, such as opportunistic contracting and costly renegotiations. Only a handful of studies in these fields directly test the performance implications of different organizational forms (Macher and Richman 2008). As stated by Leiblein et al. (2002, p. 830), there remains a need to "consider how firms' boundary decisions influence other performance dimensions, such as overall firm profitability, excess cash flow, or risk."

The goal of this paper is to empirically investigate the connection between quality risk and manufacturing outsourcing. The analysis contrasts quality risk in contract manufacturers and internal plants. We restrict attention to domestic manufacturers in order to avoid confounding the effects of outsourcing with those of offshoring. Specifically, we address the following question: Do contract manufacturers' plants pose a higher quality risk than internal plants, on average? For semantic clarity, we operationally define the three main variables associated with this question. A *contract manufacturer's plant* (CM) is an establishment that manufactures finished or nearly finished products to another company's 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. *Quality risk* is a construct

defined here as the propensity for product shipped from a given establishment to fail to perform as intended due to manufacturing-related issues.

The U.S. over-the-counter and prescription drug industry (hereafter drug industry) is a particularly interesting sector in which to study the quality risk associated with outsourcing to contract manufacturers. Clearly, conformance quality is of paramount importance to the industry. Since 2001, off-quality products and related compliance issues have cost the industry more than $700 million in fines, plus billions more in lost revenues (D'Souza et al. 2007). In addition, the industry is rapidly moving in-house production to contract manufacturers (Wilhelmsson 2004) – a trend driven largely by a slowdown in growth and decreasing margins caused by the lack of successful innovation (Hirschler 2007). Given the importance of product quality and the increased reliance on contract manufacturing, it is essential to rigorously evaluate whether contract manufacturers pose a greater quality risk.

While there exists a mature literature base on quality management within factory walls (Sousa and Voss 2002; Nair 2006), few studies have investigated the impact of different supply chain configurations on manufacturing quality (Robinson and Malhotra 2005). Empirical studies in services, where there are more intangible components, provide some evidence that vertically integrated service supply chains tend to have higher quality than disintegrated supply chains (see, for example, Harris and Winston [1983] on railroads; Michael [2000] on the hotel and restaurant sector; and Hsieh et al. [2004] on international courier services). Analytic researchers have shown that double marginalization and the assumed difficulty of enforcing performance quality levels through contracts may lead vertically-integrated firms to produce finished products and components with higher levels of performance quality than two separate firms in a vertical chain (Kaya and Özer 2005, Economides 1999). Many scholars have studied the relative effectiveness of approaches to managing quality at existing suppliers or CMs analytically (e.g., Sheopuri and Zemel 2005, Hwang et al. 2006, Reyniers and Tapiero 1995, Baiman et al. 2000, Mayer et al. 2004) and empirically (e.g., Forker 1997, Trent and Monczka 1999). None of these previous studies have explicitly considered the quality risk impact of the outsourcing decision. Finally, Macher and Nickerson

(2006) surveyed 26 finished product plants from 14 companies and found that plants that engage in any contract manufacturing (versus those that do not) tend to have poorer internal metrics, such as cycle time and change in yield; however, these authors do not provide an explanation as to why this may be the case, nor do they assess other non self-reported metrics, such as the quality risk of product shipped to the trade.

This paper contributes to the academic literature and practice in several ways. First, we create an innovative measure of plant-level quality risk using the Delphi method (Linston and Turoff 1975). A panel of experts with significant industry and FDA experience guided the development of a heuristic that, when applied to historical records of publicly-available FDA inspection data, provides an objective measure of quality risk. Second, we draw upon and integrate literature and theory from multiple disciplines to argue that, in spite of the potential benefits realized from both the CM and IP focusing on their respective "core competencies," outsourcing to contract manufacturers poses a quality risk. Third, using our quality risk metric as a dependent variable, and controlling for several potentially relevant variables (e.g., plant size, plant age, product, and firm size), we empirically test our model using data from FDA-regulated plants in the U.S. drug industry. Overall, our results provide baseline empirical evidence that production by CMs does pose a significantly greater quality risk than internal plant manufacturing; however, unexpectedly, this effect is mitigated with plant age.

The rest of this paper is organized as follows. In Section 2, we develop our hypotheses. In Section 3, we discuss our data source and the measurement of variables in the study. We present and discuss the results of our analysis in Section 4, and we conclude in Section 5.

2. Hypothesis Development

In this section, we develop our two main hypotheses. In Section 2.1, we examine whether there might be a quality risk in outsourcing from three perspectives: capabilities, cooperation, and coordination. The capabilities perspective focuses on the relative abilities of the two parties—the firm that owns the brand and the contract manufacturer—in determining whether outsourcing poses a quality risk. This perspective, and the related core competence paradigm, may lead to the conclusion that outsourcing to

CMs should improve quality, on average. Drawing from the quality management literature and agency theory, we then look at challenges in attaining cooperation when outsourcing. Finally, pulling from organizational theory and the emerging literature on new product development across multiple organizations, we examine challenges in achieving coordination across firm boundaries. Summarizing these views, we posit that there is a quality risk in outsourcing. In Section 2.2, we examine whether or not ISO 9000 certification has a direct influence on quality risk.

2.1 Quality Risk in Outsourcing

2.1.1 Capabilities: Core Competence Perspective

Firms outsource production for many reasons. One of the primary justifications for outsourcing any function or activity is to allow firms to focus on their "core competencies" (Prahalad and Hamel 1990, Quinn and Hilmer 1994). It is often assumed that for brand-owning firms that sell directly to consumers, marketing and/or innovation are the core competencies, whereas manufacturing competence is not. Chesbrough and Teece (1996) discuss the trade-offs that outsourcing presents to a firm's innovation capabilities. They note that firms can safely outsource to CMs when specifications can be codified, as in these cases little tacit knowledge can be gained from production itself (see also Monteverde 1995). In contrast to brand-owning firms, the core competence of a CM should clearly include manufacturing prowess and a high quality capability. The basic conclusion from the capabilities perspective is this: firms which prioritize innovation and/or marketing would favor outsourcing production, all else being equal. Whether and how quality risk is affected by firms focusing on their core competencies is not discussed in this literature. Presumably this gap in the literature is due to the implicit assumption that quality will not suffer, and may be even be enhanced because managers in the two separate firms will each concentrate on developing their unique core competencies.

2.1.2 Cooperation: The Agency Perspective

In reviews of the quality management (QM) literature, Sousa and Voss (2002) and Nair (2006) observed that the rote implementation of quality practices alone is not sufficient to drive quality performance.

Instead, the "infrastructural" components of QM are also necessary to drive quality performance (Anderson et al. 1994, Dow et al. 1999, Flynn et al. 1995, Powell 1995, Giffi et al. 1990, Samson and Terziovski 1999). QM infrastructural practices include many 'soft' and intangible cultural elements, such as executive commitment, open organization, employee empowerment, and a zero-defects mentality. These elements are difficult to observe across organizational boundaries (Terlaak and King 2006).

It is important to note that a buyer can observe through audits or surveys whether a supplier (1) utilizes process management tools, such as Statistical Quality Control (SQC); (2) has written procedures for daily processes and change management; (3) documents the training of its employees; and/or (4) has received external process certification. However, the key insight from the QM literature discussed in the previous paragraph would suggest that the mere existence of these observable practices is not sufficient to guarantee a reduction in quality risk. Rather, to mitigate risks, quality practices must be rigorously followed by employees daily. These behaviors occur when leaders create a culture for quality in which knowledge sharing, learning, and troubleshooting to root causes occur regularly among empowered employees on the production floor (Roth et al. 1994). Unfortunately for the buying firm, such critical aspects of QM programs, and therefore low quality risk, are difficult to observe directly.

Relating these insights from QM to the principal-agent literature, there are some CM (agent) behaviors that are critical to reducing quality risk and require effort by the CM (agent) that cannot be observed by the buyer (principal). We posit that this lack of observability is a critical driver of the quality risk in outsourcing. As noted by Laffont and Martimort (2002, p. 145): "By the mere fact of delegation [to an agent], the principal often loses any ability to control those actions that are no longer observable." Assuming that the output produced by the CM cannot be perfectly tested, the presence of important unobservable actions presents a "moral hazard." Eisenhardt (1989) specifically proposes that when outcome measurability is low, behavior-based contracts (which can only be enforced through hierarchical governance—i.e., vertical integration) are preferred, ceteris paribus.

Beyond unobservability, for moral hazard to be a concern, the two organizations need to have different objectives. There are plausible reasons to believe that a CM will place less emphasis on quality risk than a branded firm. First, there is the valuation of the brand name. In theory, employees at the brand-owning company will tend to have more concern for protection of the brand than those at a contractor. Second, there will usually be shared responsibility for any quality failure when outsourcing. Buyers will often specify raw material suppliers and the production processes in addition to the product specifications. To some degree in all cases, especially in those previously mentioned, responsibility for off-quality issues will not always be easy to assign. A contractor knows that the likelihood that she will bear the entire cost of a quality failure is small; an internal plant knows that his company will bear this entire cost. Third, in an internal plant, employees would generally expect that when they set up a production process for their own company's product, they are doing so for the life of the product, which is less likely to be the case at the CM. This difference will likely engender more robust system designs, training, and validation on the front-end at an internal plant than would occur at a CM.

Another agency-related argument, which does not directly depend on the "first-order" differences in valuation of low quality risk, is based on the work of Holmstrom and Milgrom (1991). When multidimensional incentives are present, difficulty in the measurement of one dimension can lead actors to allocate their attention to the more easily measured dimensions. In an outsourcing relationship, cost and delivery time are significantly more easily (and constantly) measured than the level of quality risk at the CM, as they will manifest themselves quickly, often directly resulting in contractual penalties. The presence of these more measurable performance dimensions in outsourcing can lead to unintentional shirking in quality. Several conceptual articles in the economics literature have specifically articulated the effect that measurement difficulties have on organizational form. Alchian and Demsetz (1972) state that when output cannot be easily attributed to an individual, as in team production, a vertically integrated firm can reduce shirking (relative to a disintegrated firm) by at least partially rewarding team members based on input behaviors. Building on this logic, Barzel (1982, p. 42) noted that "firms will form and

trade with each other at junctures where output can be readily measured, but where output is difficult to measure the different steps will be performed within the firm." If we accept from the QM literature that quality risk can be difficult to measure, these arguments would encourage maintaining production in-house (all else being equal) so that input behaviors could be directly observed.

Taken together, the above arguments indicate that from an agency perspective it may be challenging to achieve cooperation such that a CM will invest in the creation of sufficiently low quality risk in their manufacturing plant. The subsequent moral hazard is driven by the unobservability of the behaviors of the CM, and misaligned incentives with regard to quality between the CM and the buyer in the short-run.

2.1.3 Coordination: The Organizational Theory Perspective

While cooperation refers to the alignment of incentives, coordination is the "alignment of actions" (Gulati et al. 2005, p. 417).[1] Gulati et al. (2005, p.419) note that coordination challenges "can persist even when interests are aligned; i.e., when cooperation is achieved." These challenges are well-documented in the product development literature (e.g., Parker and Anderson 2002, Sosa et al. 2004, Anderson et al. 2008). They can arise from many factors, including lack of rich communications (Daft and Lengel 1986), lack of shared organizational routines (Nelson and Winter 1982), and lack of shared knowledge (Grant 1996). The importance of coordination is highest in situations with high task interdependence and uncertainty (Gulati et al., 2005), such as product development. As noted by Anderson et al. (2008, p. 263): "product development involves a greater degree of process, marketing, creative, and technical uncertainty than typically found in other settings such as production management."

While typically less interdependent and uncertain than new product development, a production outsourcing relationship still will involve several instances in which tacit knowledge and understanding must be exchanged to ensure low quality risk. Indeed, the initial startup of the production process in a CM's facility essentially *is* a development project. Equipment specifications must be agreed upon;

[1] We note that the term does not have the same meaning typically utilized by the analytical supply chain literature in which contracts are devised to "coordinate" the supply chain (as reviewed by Cachon [2003]). The contracts in that literature are usually devised to align financial incentives, with the implicit assumption that actions will then also be aligned.

process instructions must be designed and validated; and procedures and tests for startup, shutdown, cleaning, etc. must be devised, validated, and implemented. The expectations of the buying firm regarding the content and thoroughness of these procedures, the design of the equipment, and the on-going compliance with the guidelines must be shared across organizations. Further, in the course of a production outsourcing relationship, interdependent interactions may frequently arise due to unforeseen complications in supply or production. If communicated to the buyer at all, the appropriate remedies must be agreed upon across organizational boundaries. Furthermore, any product innovations devised by the buying firm after the relationship begins will need to be implemented at the CM, resulting in a new process development project across two organizations. Thus, while production may seem separable, it is clear that much coordination is needed between the buyer and CM both at the start of a contract and during the many interdependent interactions that often occur throughout the duration of the relationship.

2.1.4 Resolving the Puzzle

The above insights do not completely resolve the puzzle of whether firms incur a quality risk in outsourcing to contract manufacturers, as conflicting arguments are presented: the capabilities perspective would indicate that outsourcing may reduce quality risk; whereas issues posed by achieving cooperation and coordination would indicate that outsourcing can increase quality risk. We now attempt to resolve the conflicting arguments primarily by viewing this puzzle through an operations strategy lens—and in so doing weakening the core competence argument. Both the practitioner literature and the operations and supply chain management literature consistently show that cost is the main driver of outsourcing decisions in manufacturing (e.g., Doig et al. 2001, Gray et al. 2009b).[2] Contract manufacturers are acutely aware of this fact. Further, once production begins, CMs will see immediate savings (increased expenses) from any decrease (increase) in production costs; and they can incur immediate penalties for late deliveries. However, product quality must simply be "acceptable."

[2] This combined with the unobservability of quality risk, can lead to "adverse selection," which refers to situations where the agent misrepresents her capabilities in the presence of information asymmetry (Eisenhardt, 1989). Examples include coaching of employees and significant cleaning of equipment when a perspective buyer visits, or overstating quality capabilities on a buyer survey. While moral hazard, discussed earlier, motivation of the buyers, occurs after the relationship is underway, adverse selection occurs during the CM selection phase.

Quality can *appear* to be acceptable despite a less-than-full implementation of quality best practices, as many of the essential infrastructural aspects of strong quality management systems, such as culture, are difficult to observe. Furthermore, in most industries, product quality failures are low-probability events, and managers tend to overvalue certain short-term gains relative to uncertain long-term gains (Laverty 1996). Thus, while we agree that the 'intended' core competence of a CM is manufacturing, we posit that its actual (or 'realized') core competence is slightly more subtle: the development of manufacturing capability to win and keep orders from potential customers. In CM, cost is an order winner while conformance quality would often be what Hill (1994) would call an "order qualifier." Thus, in the pursuit of contracts, the added organizational costs may deter the CM from fully developing the systems and culture that are required to lower the quality risk to a level that would be optimal for the buying firm.

Weighing all of the above arguments, we posit that the issues with achieving cooperation and coordination will overwhelm the "core competence" argument with regard to the quality risk associated with contract manufacturing. More formally:

H1: *Contract manufacturer's plants are more likely to be operating with a higher level of quality risk than internal plants, ceteris paribus.*

2.2 ISO 9000 Certification

As of December 2006, nearly 900,000 ISO 9000 certificates had been issued in 170 countries worldwide. In the United States alone, about 45,000 certificates had been issued (Helberling 2007). Despite ISO 9000's broad reach, the evidence regarding quality performance improvement due to ISO 9000 certification remains inconclusive. The few studies that have found a link between ISO 9000 certification and quality performance conclude that simply implementing ISO 9000 does not guarantee improved quality performance; rather, the firm must go beyond the standard to see improvement (Martinez-Costa et al. 2009, Naveh and Marcus 2004, Voss and Blackmon 1998, Anderson et al. 1999). Interestingly, some authors have shown financial improvements correlated with ISO 9000 certification (Corbett et al. 2005, Sharma 2005), but these studies do not examine the specific effect of ISO 9000 certification on product quality. While ISO 9000 has not been shown to translate directly into improved quality performance,

researchers have found that certification often serves as a signal of high quality to potential buyers (Terlaak and King 2006); and this has a marketing benefit (Anderson et al. 1999). We believe that these inconsistent conclusions regarding the effect of ISO 9000 on quality performance are driven by the fact that true quality risk is unobservable, as discussed in Section 2.1.2. Thus, lowering quality risk requires a significant and ongoing investment in quality systems and practices that go beyond the documentation required to obtain ISO 9000 certification. Therefore, we hypothesize:

H2: *ISO 9000 certification will not affect the likelihood of lowering plant-level quality risk, ceteris paribus.*

3. Empirical Setting and Measures

In this section, we document the empirical methodology used to operationalize the variables used in the study. The development of a valid metric for quality risk is an important contribution of this work. Boyer et al. (2005, p. 446) observed that in operations management (OM) research, "identifying alternative, innovative sources of data is becoming increasingly important." However, no objective measure capable of comparing quality risk across a broad range of plants was found in an extensive review of OM metrics pertaining to quality (Roth et al. 2008a). We next discuss the FDA's Field Accomplishments and Compliance Tracking System (FACTS) database, which provided the raw data for our quality risk measure. We then describe the process we used in this study to develop a heuristic to transform the FDA inspection data into a numerical measure that is a reasonable proxy for plant-level quality risk. Finally, we document the sources and processes used to determine source of production (CM vs. IP), ISO 9000 certification, and the other independent control variables.

3.1 FDA Inspection Database

U.S. regulated industries are subject to inspection by certain governmental organizations. In the case of FDA-regulated industries (e.g., drugs, food, and medical devices), inspections in U.S. plants are thorough and broadly consistent between plants. FDA inspectors provide in-depth, expert judgments of the potential quality risk posed by a particular manufacturing facility. The average inspection lasts almost 100 hours, with most inspections taking between two days and three weeks to complete. Inspections

usually involve spot checks of records, conversations with random employees, and tours of the manufacturing facilities. Importantly, FDA inspectors have the legal authority to ask for any information related to the production of any regulated product. We found no better assessment of manufacturing-related quality risk available for a large number of plants with diverse product lines.

The FDA inspection data are available in its FACTS database, which can be obtained under the Freedom of Information Act (FOIA). Although the data is publicly accessible, obtaining the complete data set required for this study was time-consuming, as FOIA requests must pass through multiple levels of bureaucracy and must be answered in the order received. Data on every establishment inspection in the drug industry in the United States from January 1994 through April 2006 was obtained. We chose 1994 as the starting date because the FDA's regulatory authority was increased, as was the overall quality of the inspection data, after a major legal decision in 1993 involving Barr Laboratories (Farley 1993).

To guide our discussion of the heuristic for assessing quality risk, we provide a brief description of the process by which the FDA creates and codes the inspection data into the FACTS database. The database contains the company name and establishment location, the date of the inspection(s) at each establishment, and for each inspection, two distinct indicators of quality risk. First is a "Yes/No" variable that specifies whether or not a Form 483 was issued by the inspector; and second is a district decision code. A Form 483 is issued to an establishment if, at the conclusion of the inspection, the inspector believes there are deviations from "Good Manufacturing Practices" (GMPs, outlined in the Code of Federal Regulations [21 CFR, mainly Parts 210 & 211]) that are significant enough to warrant formal documentation. A Form 483 was issued in 56 percent of all plant inspections in the total drug database from 1994–2006. Next, a complete Establishment Inspection Report (EIR), including a Form 483 if issued, is sent to the district office. The district office reviews the EIR, as well as other information related to the plant, such as customer complaints, and makes a decision on the inspection. Based on the total information provided, the FDA district office indicates one of five possible actions: (1) No Action, (2) Voluntary Action, (3) Official Action, (4) Pending, and (5) Referred to State officials. Of these, the

first three types of district decisions are useful indicators of relative quality risk. "No Action" means that there were no significant problems uncovered related to the plant. "Voluntary Action" implies that objectionable conditions or problems are found, but that the district would neither recommend regulatory action nor require a response from the establishment. "Official Action" indicates that either regulatory or administrative sanctions will be recommended.[3] Together, the Form 483 decision and district decision provide tangible evidence of quality risk.

In addition, we obtained ancillary information about inspection-related seizures, injunctions, and recalls related to plants included in our study from the Enforcement section of the FDA website. The actual cost to companies for receiving any of these notices of violations are very incident dependent. Some violations may simply require employee retraining or the repair of a leaky roof. Others may be much more costly. Significant fines for non-compliance are fairly rare, but can be as high as $5 million per year as in the case of Texas CM Pharmafab (Pharma Marketletter, 2007). And, of course, if violations impact new product launch and/or brand equity, revenue losses can be very significant.

3.2 The Delphi Process: Development of a Quality Risk Metric

The use of raw FDA data presents some measurement difficulties. It is obviously desirable to utilize both the FDA inspector's decision (483 Yes/No) and the district decision for each inspection, as well as any inspection-related seizures, injunctions, or recalls. But it is not obvious how an overall plant-level quality risk metric should be developed. Also, it is important to know how to assess a plant with multiple inspections over the time period. In order to develop a heuristic that results in a reasonable proxy variable for quality risk, we used the Delphi method with a panel of four expert judges (Linston and Turoff 1975). We drew upon their extensive experience with both manufacturing in regulated industries and FDA inspections. One panelist had 30 years of experience with the FDA, including serving as District Director. Each judge participated voluntarily.

[3] Note that it is possible, although rare, for the district to issue official action even if the inspector did not issue a Form 483 (see Table 1). This could occur if highly objectionable data were obtained from consumer complaints or finished-product samples outside of the inspection process.

In three Delphi rounds, the experts provided independent judgments that led to the creation of the quality risk score heuristic. For each round, we created a script and interviewed each expert individually, following the script. We then compiled a round summary and returned the summary to each expert. Calls were recorded and reviewed after each round prior to sending out summaries. After three rounds, we conducted a fourth "wrap-up" conference call with the experts. The expert panel concluded that their consensus heuristic provided a reasonable assessment of a plant's overall quality risk. We now proceed to discuss the development of the quality risk metric.

3.3 Operationalization of *Quality Risk* Using FDA Inspection Data

Quality risk was defined in Section 1 as the *propensity* for product shipped from a given establishment to fail to perform as intended due to manufacturing-related issues. To develop a metric that taps into this construct, we assess the plant level FDA Form 483 decisions (Yes/No) and three meaningful district decisions (No Action, Voluntary Action, Official Action), which together result in six possible outcomes of an FDA establishment-level inspection in Table 1. Additionally, and only in cases where Official Action was indicated, the FDA can take enforcement action (Seizure/Injunction/Recall) during or as an immediate result of the inspection, which is a seventh possible outcome depicted in Table 1.

Table 1: Distribution of Possible Plant Inspection Outcomes – Single Instance

Form 483	District Decision	Percent of Inspections for 154 Plants Used in Study, 1994-2006 (N=980)	Delphi Panel Consensus Quality Risk (QR)Score*
No	No Action	39.3%	0
No	Voluntary Action	6.6%	0.5
No	Official Action	1.1%	3
Yes	No Action	1.8%	1
Yes	Voluntary Action	37.3%	1.5
Yes	Official Action	12.8%	3.5
Enforcement	Official Action	1.0%	10

*Estimation of the quality risk of each inspection outcome is based on a final consensus score obtained from the Delphi approach, using our 4 expert judges.

We asked the panel of experts to rate each of the seven inspection outcome categories in terms of relative quality risk from a single audit. Delphi single-inspection quality risk consensus scores were determined by the expert judges in 2 rounds. As depicted in Table 1, the judges believed that a district

decision of Official Action indicated a significantly higher quality risk than a district decision of Voluntary Action. Not surprisingly, the judges placed a substantially higher weight on seizures/injunctions/recalls relative to the next highest actions.

The judges next discussed how best to handle the multiple inspections of a single plant, given that plants will vary in number and timing of inspections. The baseline option was simply to average the quality risk scores for a single plant. This starting point led to two key questions regarding multiple inspections: (1) To what extent do multiple FDA inspections indicate an increased quality risk? (2) To what extent does the trend in individual inspection scores affect the quality risk? Regarding question (1), through the Delphi summaries the judge with the most FDA experience informed the panel that the majority of inspections are part of either annual work plans or the pre-approval process for new drugs. Furthermore, "for cause" inspections, which are performed because of a perceived quality risk, will usually be the result of, or result in, Official Action. The expert panel subsequently determined that raising a plant's quality risk score due to the number of inspections would inappropriately raise the score of large plants with many new drugs and process changes. In answering question (2), the panel concurred that plants that have shown the ability to improve their quality systems over time presented a lower quality risk than those whose individual inspection scores appeared to be deteriorating over time. The panel, therefore, converged on adjusting the adjust quality risk score for trend, but not frequency. The panel arrived at the following heuristic that offers a plant-level quality risk score:

$$\text{QRisk}_p = \frac{\sum_{j=1}^{n} QR_j}{n} + \frac{\sum_{j=1}^{n-1} QR_{j+1} - QR_j}{n+1} \qquad (1)$$

where QR_j is the Single Inspection Quality Risk score from Table 1for inspection j; n= the number of inspections, and j = the indicator for a specific inspection (ordered chronologically).

The first term is simply the average of the quality risk scores from the individual inspections. The second term is a simple trend adjustment. Note that the second term is divided by $n + 1$, when there are only $n - 1$ terms in the numerator. This adjustment ensures that quality risk scores for plants with few inspections are not modified excessively for their improvement or decline in inspection scores. As a

validity check, the panelists reviewed the resulting quality risk scores for a selection of plants and collectively determined that the $QRisk_p$ metric matched their "subjective" quality risk assessment, based on a detailed examination of inspection results per plant. We note that our results are insensitive to the specific form of quality risk measurement, as discussed in Section 4.3.

There is precedent in the academic literature for the use of government inspection data to measure an organization's propensity to comply. In the areas of safety and environmental compliance, data from the Occupational Safety and Health Administration (OSHA) and the Environmental Protection Agency (EPA) have been utilized. The challenges faced in these areas are similar to those presented by the use of FDA data; however, we found no studies that incorporated the use of industry experts and/or the Delphi method to transform the richness of multiple raw measures into a single plant-level score. In a recent example from the safety literature, Filer and Golbe (2003, p. 366) state that the "vast majority" of the safety literature assumes that "detected violations are a reasonable proxy for actual violations." They utilize inspection-level counts of violations as their dependent variable to study how a firm's financial condition affects its safety performance. Gray and Shadbegian (2005) investigate the effect of various plant- and firm-level factors on OSHA compliance. Available EPA data, as well as the challenges involved in their use, are discussed in Gerde and Logsdon (2001). Examples of research using EPA data include Grant and Jones' (2004) investigation of the effect of right-to-know programs on environmental performance, incorporating the average weighted chemical releases as the dependent variable; and King and Lenox's (2001) study of the effect of lean practices on environmental performance using emissions relative to size in a given year as the dependent variable. We know of only one study that uses FDA data as a proxy for quality compliance (Macher et al. 2006). These authors use FDA inspection data both to demonstrate the heterogeneity of regulators (based on experience and level of training) and to propose methods to reduce this heterogeneity in inspections.[4]

[4] We note that there is no indication that this heterogeneity would introduce bias in our results. That is, there does not appear to be any evidence that the FDA assigns inspectors with differing levels of training and experience to CMs vs. internal plants. This was confirmed in an e-mail and phone conversation with a high-level FDA employee.

3.4 Plant Classification: Contract Manufacturer or Internal Plants

For this study, we selected plants that were solely producing products for their own companies ("pure" IP) and those which were solely producing products for other companies ("pure" CM). Focusing on these two extremes reduces confounding effects and simplifies the interpretation of our results. To do this, another key task of this research was to conduct a comprehensive search of the FDA plant database to identify a sample of these "pure" CMs and IPs. First, we systematically sampled one of every four plants in the 5,637-plant FDA database and carried out a methodical web and literature search on each of the resulting sample of 1,409 plants in order to assess the nature of the manufacturing performed using eight distinct classifications.[5] Of these categories, the two which could be candidates for this study were "pure" CM or a "pure" IP. Separately, published lists of contract manufacturers and brands in the OTC, regulated cosmetics, and pharmaceutical industries were used to identify possible plants for our sample. Finally, local OTC store shelves were manually searched as an additional source for internal plants. Not surprisingly, there was a considerable lack of transparency in companies' reporting of the source of their production; and many internal plants also performed some level of contract manufacturing. For these reasons, the majority of the plants in the sample database could not be classified as either pure contract manufacturers or pure internal plants.

The first round identified 345 potential plants as either a pure CM or a pure IP. From this initial classification, we independently rechecked each of these 345 plants by employing additional news searches in Lexis-Nexis and making phone calls to plants in order to obtain more detailed evidence of the nature of plant operations. This validation process yielded a final judgmental sample of 154 plants for which we could be reasonably certain of the reliability of our classification of plants into one these two "pure" groups. The final sample of 77 CMs and 77 IPs represent 116 different companies; it is by chance that the final sample was evenly balanced.

[5] In addition to CM and IP, these classifications include: Generics manufacturer, "upstream"/chemical manufacturer, medical gas manufacturer, pharmacy, mixed, and "unable to classify."

3.5 ISO 9000 Certification

The Quality Digest[6] database was our source of information on the ISO 9000 certification status of each of the 154 sample plants. The Quality Digest data derives from the ISO 9000 certification registrars. Because our sample spans several years, we included only those plants for which the registration could have affected their inspection scores. For example, we did not want to include a plant that was only inspected prior to 2003 but certified in 2005. We coded plants as certified if at least one-half of their inspections occurred after they began the process to become ISO 9000 certified.[7]

3.6 Other Control Variables

To address our research questions, we controlled for a number of factors that might also affect quality risk. The following control variables were deemed to be important: company size, plant size, plant age, type of product manufactured, and three dummy variables that indicate whether the plant produces primarily regulated products, whether the parent company is under financial strain, and whether the parent company is publicly or privately held. 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 (Ojala 2004). After an extensive and unsuccessful search for free public data sources, we purchased access to a Dun and Bradstreet (D&B) database.[8] The D&B data was our primary source for data on product, plant age, plant size, total company sales, and ownership status (i.e., *Public* =1 if public and 0, if private parent firm). In addition, ReferenceUSA, an InfoUSA database, was used to obtain credit rating scores for sample firms on a scale of 1–5. Because most plants had excellent credit (i.e., a score of 5), and only one had a rating below 4, we created a 1-0 variable, *Credit*, for which 0 indicates a top credit rating of 5, and 1 indicates lower scores. We used single imputation with logistic regression to estimate the few values missing from the ReferenceUSA database for our sample plants (Allison 2002). Also, we created a dummy variable to distinguish between plants that made almost exclusively regulated products (*Onlyreg=1*) and those that

[6] http://www.qualitydigest.com/html/iso9000.html, searched in the spring of 2006
[7] The literature shows that companies begin the work to achieve certification well before the completion of the registration process, on average 15 months prior to the certification date (Meyer 1998). We did perform robustness checks with other conceptualizations.
[8] The database was Harris' Company Reach, which is based on Dun & Bradstreet's Million Dollar Database.

produced a significant quantity of non-regulated products (*Onlyreg*=0). Classification of plants for this variable was done subjectively, using websites, news searches, and other information. Finally, we included a dummy variable based on North American Industry Classification System (NAICS)[9] codes to classify study plants based on their dominant product lines—pharmaceuticals (*Pharm* =1) and toilet goods (*Pharm*=0). We note that products classified as "toilet goods" are often regulated as drugs if they are designed to treat, cure, or prevent some illnesses (e.g., medicated skin cream); the fact that all plants in this study are in the FDA's FACTS database indicates they produce a product/product(s) regulated as drugs. Some primary classifications from D&B were modified when evidence indicated that one of these two classifications was more appropriate. Forty-nine of the internal plants and 34 of the CMs were classified as pharmaceutical manufacturing.

We included the company size (*Total Sales*) to control for company-level infrastructural support received at the plant-level. Plant size (*Employees*) is intended to directly control for issues related to scale at the plant level. Plant age (*Age*) was included to capture the fact that plants may vary with age, either through improvement or deterioration. One mechanism of deterioration over time was described by Hill (1994) as "focus regression," which can result from unmanaged increases in product and process variation over time. An examination of histograms of the continuous independent variables (i.e., *Total Sales*, *Employees*, and *Age*) and scatter plots of these variables with quality risk indicated that the independent variables required transformation due to skew and observed nonlinearities (Hair et al. 1998). In the case of *Total Sales*, *Employees*, and *Age*, the strength of the association logically decreases as the value becomes large. We chose the natural log transformation, which is commonly employed for skewed distributions (Albright et al. 2006). Also, as is recommended when including an interaction term to avoid unnecessary ill-conditioning, we mean-centered the ln*(Age)* variable for the analysis.

[9] http://www.census.gov/eos/www/naics/

3.7 Descriptive Statistics

The descriptive statistics for the variables used in this study are given in Table 2. Although the bivariate correlations depicted in Table 3 do not indicate that multicollinearity is a problem, we further checked for multicollinearity among the x-variables using PROC REG in SAS 9.1.3, and found no evidence of multicollinearity using commonly accepted rules of thumb; i.e., variance inflation factors, condition indices, and the proportion of variance in the variance-decomposition matrix (Hair et al. 1998).

Table 2: Descriptive Statistics

Name	Description	*N*	Mean	St. Dev.	Min.	Max.
$QRisk_p$	Quality risk (plant level score, from Equation (1))	154	1.08	.84	0	3.500
QRisk	Quality risk (1=low risk; 2=moderate risk; 3= high risk)	154	1.84	.77	1	3
CM	1 = Contract manufacturer	154	.50	.50	0	1
ISO 9000	1 = ISO9000 certified	154	.06	.24	0	1
TSales	Total company sales ($1,000s)	154	$9,213	$16,926	$69	$56,740,931
Emp	Plant employees	154	363	769	1	8000
Age	Age of plant, in years	154	41.3	35.6	2	158
OnlyReg	1 = Plant makes primarily regulated products	154	.40	.49	0	1
Credit	1 = Evidence of credit issues (4 or less in R-USA)	148	.14	.34	0	1
Public	1 = Public	154	.32	.47	0	1
Pharm	1=Plant makes primarily pharmaceutical products (0=primarily toilet goods)	154	.54	.50	0	1

Note. $N = 154$ plants. Note the very small plant (TSales=$69M, Emp=1). While available evidence indicates this data may be accurate, we checked our results with this plant excluded; there was no change. Note that the 158-year-old plant was Pfizer's Brooklyn, NY plant; we confirmed that production began in 1848. The plant was closed in early 2007, after our data collection.

Table 3: Pearson Bivariate Correlations ($N = 154$)

	$QRisk_p$	*QRisk*	*CM*	*ISO 9000*	*TSales*	*Emp*	*Age*	*OnlyReg*	*Credit*	*Public*	*Pharm*
$QRisk_p$	1.00										
QRisk	.87***	1.00									
CM	.29***	.25***	1.00								
ISO 9000	-.08	-.06	-.14*	1.00							
TSales	-.27***	-.17**	-.52***	-.09	1.00						
Emp	-.06	.01	-.27***	-.03	.23***	1.00					
Age	-.13	-.04	-.32***	-.06	.25***	.28**	1.00				
OnlyReg	.02	.14*	-.32***	-.15*	.22***	.11	.04	1.00			
Credit	.13	.03	.17**	-.10	-.20**	-.14*	-.20**	.06	1.00		
Public	-.17**	-.09	-.55***	-.11	.68***	.34***	.31***	.25***	-.28***	1.00	
Pharm	.06	.13	-.20**	-.21*	.16**	.10	.19**	.63***	.03	.28***	1.00

* $p < .10$; ** $p < .05$; *** $p < .01$

As shown in Table 4, simple *t*-tests (for continuous variables) and z-tests (for categorical variables) reveal that contract manufacturers and internal plants differ significantly on all control variables ($p < .01$), as well as on *quality risk* and *ISO 9000* certification. These statistical results underscore the importance of controlling for plant characteristics. We note that there should not be any systemic bias in the way the FDA inspects internal plants vs. contract manufacturers, as all drug investigations are guided by the Investigator Operations Manual, Chapter 5 (FDA 2008). The forward of this manual explicitly states that it applies "to all individuals who perform field investigations." We also confirmed via personal communication with a representative of the FDA that there is no differentiation in the auditing process between plants producing on contract for other companies and internal plants producing product under their own brand name. [10]

Table 4: Mean Differences and Differences in Proportions:
Contract Manufacturers ($N = 77$) versus Internal Plants ($N = 77$)

Name	Mean-CM	Mean-IP	Test statistic[a]
$QRisk_p$	1.32	.84	−3.73***
QRisk (Low Risk)[b]	.31	.47	1.98**
QRisk (Ave. Risk)[b]	.35	.42	.83
QRisk (High Risk)[b]	.34	.12	-3.27***
ISO9000	.025	.091	1.72*
TSales ($mil.)	499	17,900	7.44***
Emp	159	568	3.42***
Age	30	53	4.17***
OnlyReg	.247	.558	4.13***
Credit	.195	.078	-2.11**
Public	.065	.584	6.88***
Pharm	.442	.637	2.42**

[a] We used t-statistic for continuous variables (difference in means) and z-statistic for 0-1 variables (difference in proportions). [b] $QRisk_p$ was divided into three *QRisk* groups (see Table 5); dummy variables (=1 if in the specific group, 0 otherwise) were used for these tests.
*p<.10, ** p< .05, *** p <.01

[10] Further, we examined whether internal plants and contract manufacturers were systematically located in different regions of the United States. We examined the number of plants of both types occurring in the five FDA regions, the 19 FDA districts (we combined the two New York districts; and thus, controlled for 18 distinct FDA districts), the four Census regions, and the nine Census divisions in the continental United States. Chi-square tests for independence showed that there was not a significant relationship between location and plant type (p=.33 for FDA regions, p=.21 for Census regions, p=.50 for Census divisions). We also note that the internal plants in this sample are inspected more often (7.9 inspections vs. 4.8 inspections); this is driven largely by the fact that their plants tend to be larger.

4. Model Estimation, Results, and Discussion

4.1 Estimation Method

It is reasonable to expect that the true quality risk of a plants would be normally distributed if measured perfectly. At the extreme left tail of the distribution are plants for which there is almost no risk of nonconforming product being released to the trade. At the extreme right of the distribution are plants knowingly and willingly releasing product that is defective. However, our overall quality risk score ($QRisk_p$) for each plant does not capture this normal distribution for two reasons. First, our score is censored at zero. Recall that when an FDA inspector does not issue a Form 483, and the district indicates that no action is required, the quality risk for that particular inspection is "0." However, there is considerable variability of actual quality risk below this point. Reviews of inspection reports show that these plants still have some, quite possibly non-negligible, propensity to produce defective products.

Second, the distribution of the dependent variable is not continuous. A single inspection can only result in one of the seven scores listed in Table 1, and four of those seven scores are rarely obtained. There is some clustering of observations around the most common single-inspection results (0, 1.5, and 3.5). While averaging and adjusting for trends at the plant level helps alleviate this coarseness, our overall quality risk score for multiple inspections does not capture an underlying normal distribution; and thus we are left with a rather coarse measure of quality risk, particularly for plants with few inspections. We further refined our $QRisk_p$ measure using 5 logically ordered (ranked) categories of overall quality risk, as depicted in Table 5, with the consensus of the panel of experts. Following Forthofer, Lee, and Hernandez (2007), given the relatively small number of CMs in the low $QRisk_p$ range and IP plants in the top category, respectively, we further collapsed the bottom two and the top two quality risk groups for our ordered logit regression in Section 4.2. Thus, our new quality risk variable (*QRisk*) reflects the three logical, ordered groups, where 1=relatively low risk, 2=moderate risk, and 3=relatively high risk. As posited from theory, there are more internal plants than contract manufacturers in the low risk category, and the reverse pattern is observed in the high-risk category.

Table 5: Discrete Quality Risk Categories Used for Ordered Logistic Regression (N=154)

			Original 5-level Classification			*QRisk*, 3-level Classification		
Cat.	$QRisk_p$ range	Description	All	CMs	IPs	All	CMs	IPs
1	$QRisk_p = 0$	Plants with all clean inspections (lowest risk)	22	5	17			
2	$0 < QRisk_p \leq .75$	Plants with generally clean inspections (low risk)	38	19	19	60	24	36
3	$.75 < QRisk_p \leq 1.5$	Plants which, on average, receive a Form 483 but no Official Action (moderate risk)	59	27	32	59	27	32
4	$1.5 < QRisk_p \leq 2.25$	Plants with at least one Official Action (high risk)	23	15	8			
5	$QRisk_p > 2.25$	Plants that average closer to one or more Official Actions (highest risk)	12	11	1	35	26	9

4.2 Econometric Specification and Results

Our rank-order quality risk score (*QRisk*) is an imperfect measure of a plant's true, non-measurable quality risk, and these imperfections include censoring and coarseness. These factors are handled in our econometric analysis by means of ordered logit regression using STATA 9.2's "ologit" procedure. The econometric specification (following Verbeek [2004]) of the ordered logistic model for this analysis is given in Eq. 2 (where γ_i is an unknown, estimated parameter; and $QRisk_i$ is the overall quality risk score for plant i, where i = 1 to 154). The descriptions of the independent variables are given with the descriptive statistics in Table 2.

$$
\begin{aligned}
QRisk_i^* &= \beta_1 \ln(Age_i) + \beta_2 \ln(TSales_i) + \beta_3 \ln(Emp_i) + \beta_4 Onlyreg_i + \beta_5 Credit_i + \beta_6 Public_i \\
&\quad + \beta_7 Pharm_i + \beta_{10} ISO\,9000_i + \beta_{11} CM_i + \beta_{12}[CM_i * \ln(Age_i)] + \varepsilon_i \\
QRisk_i &= 1 \; if \; QRisk_i^* \leq \gamma_1 \\
QRisk_i &= 2 \; if \; \gamma_1 < QRisk_i^* \leq \gamma_2 \\
QRisk_i &= 3 \; if \; \gamma_2 > QRisk_i^*
\end{aligned}
\qquad (2)
$$

Given that there were 116 distinct companies represented by the 154 plants, we used STATA's cluster option to adjust the standard errors to account for the correlation among plants from the same company. We found no evidence that the assumption of parallel slopes was violated using the Brandt test (overall $\chi^2_{(df=10)} = 14.61$, p=.15). This non-significant result implies that we fail to reject the null hypothesis and that the two probability curves are parallel (Long and Freese 2006). The ordered logit results given in Table 6 are determined by maximum likelihood estimation. We note that these analyses do not imply

causation between the independent variables given in equation 2 and levels of quality risk. A significant finding implies that a relationship exists. Model 1 is a baseline model that excluded the two hypothesized variables (*CM* and *ISO9000*) and the interaction term ln(*Age*)**CM*. Model 2 significantly improves the fit over Model 1 (p=.0005, using STATA's "lrtest"). The significant Wald $\chi^2_{(df=10)}$ =28.4 (p=.0002) and the LR $\chi^2_{(df=10)}$ =31.5 (p=.000) in Model 2 indicates that the omnibus effect of the variables in the overall model is statistically significant from zero, indicating reasonable model fit. We note that interpretation of the odds ratio in the presence of an interaction is not straightforward. However, in Table 6, for completeness the odds ratios are given for all variables.

Table 6: Plant-Level Ordered Logit Results (N = 154)
(Nonstandardized Parameter Estimates)

	Model 1	**Odds Ratio**	**Model 2**[b]	**Odds Ratio**
Variable	*β* (*SE*)	*OR [95% CI]*	*β* (*SE*)	*OR [95% CI]*
ln(*Age*)	-.03 (.23)	.97 [.62,1.51]	.68*** (.26)	1.98 [1.19,3.30]
ln(*TSales*)	-.28*** (.09)	.76 [.63,.91]	-.17* (.09)	.84 [.70,1.01]
ln(*Emp*)	.31** (.14)	1.37 [1.03,1.81]	.23* (.14)	1.26 [.96,1.65]
Onlyreg	.86 (.55)	2.37 [.81,6.89]	1.10** (.53)	3.01 [1.06,8.52]
Credit	-.23 (.45)	.79 [.33,1.93]	-.13 (.47)	.88 [.35,2.21]
Public	.51 (.59)	1.67 [.53,5.28]	.38 (.66)	1.46 [.40,5.30]
Pharm	.05 (.53)	1.06 [.38,2.97]	.17 (.54)	1.19 [.41,3.44]
ISO9000			.50 (.71)	1.64 [.41,6.64]
CM			1.20** (.47)	3.31 [1.33,8.33]
ln(*Age*)*CM			-1.08** (.43)	.34 [.15,.79]
Wald(χ^2)	13.4*		28.4***	
LR(χ^2)	18.5**		31.5***	
Psuedo-R^2 s	5.6%, 11.3%,12.5%[a]		9.5%, 18.5%, 21.4%	
Log likelihood	-155.8		-149.3	

[a] The first pseudo-R^2 is McFadden's; the second pseudo-R^2 is Cox-Snell; and the third is the McKelvey &Zavoina R^2. The pseudo-R^2 is not comparable to that obtained in an ordinary least squares (OLS) regression and has no natural interpretation (Borooah 2002).

[b] Model 2 makes a statistically significant improvement over Model 1 (p<.01), using either the LR test or the Wald test

Note: **p*<.10, ** p< .05, *** p <.01

The overall Model 2 results indicate strong support for H1. Contract manufacturers, when other important factors are controlled for, tend to have higher quality risk than internal plants (β=1.20, p =.011). Thus, if a company were to choose to move a product from an IP to a CM with the same observable characteristics, the expected quality risk is higher, conditional on the variables in the model. While the results provide clear evidence of a quality risk in outsourcing to CMs, the interaction effect ln(Age)*CM is significant (β= -1.08, p = .012), which complicates the interpretation of the results (Hoetker 2007).

The significant interaction term indicates that the effect of age on quality risk for CMs is different than its effect on IPs. A negative interaction suggests that relative to CMs, increases in age tend to increase quality risk for IPs. To verify this, we display these interaction effects as predicted probabilities in Table 7, created using STATA's "predict" command for selected age groups. Table 7 shows that CMs pose a higher quality risk than internal plants, on average, again in support of H1, but its effect is moderated with age. At the two lower categories of age, CMs have a higher (lower) probability of being in the high (low) risk and highest (lowest) group than a comparable IP. However, Table 7 also reflects the plant age interaction. For young plants, the differences in predicted probabilities are dramatic (i.e., the probability of being in the low risk group is .20 for younger CMs versus .67 for younger IPs, on average). But, for old plants, the quality risk differences between the predicted probabilities between CMs and IPs are negligible. Closer inspection of Table 7 reveals that the predicted probabilities for the CMs are fairly stable with age, whereas those for IPs change, indicating increasing quality risk with age for IPs. In support of these observations, we computed the baseline model solely for the 77 IPs and found that the increased quality risk with age has a significant effect (p =.007). However, running the baseline model for the 77 CMs alone, the effect of age on quality risk is not significant (p=.34).

With regards to ISO 9000, the results in Table 6 lend support for H2 that ISO 9000 does not change the likelihood of quality risk, ceteris paribus. Hence we find no empirical evidence that ISO 9000 has a first-order effect on the mitigation of quality risk (β=.50, p =.49).

Table 7: Predicted probabilities of Quality Risk (*QRisk*), by plant type and plant age group**

Plant age (yrs)*QRisk	**Low Risk**	**Moderate Risk**	**High Risk**
Young (<20 yrs)			
CM (n=30)	.20	.41	.39
IP (n=12)	.67	.26	.07
Raw Diff (CM-IP)	**-.47**	**.15**	**.22**
Mid (20-49 yrs)			
CM (n=38)	.29	.43	.28
IP (n=36)	.54	.34	.13
Raw Diff (CM-IP)	**-.25**	**.09**	**.15**
Old (>=50 yrs)			
CM (n=9)	.37	.41	.22
IP (n=29)	.38	.40	.22
Raw Diff (CM-IP)	**.01**	**-.01**	**.00**

*Quality Risk (*QRisk*) categories are given in Table 5. **Ages are actual ages, not log transformed
Rows may not add to exactly 1.00 due to rounding

4.3 Robustness Checks

To check the robustness of our empirical results, we first tested the effect of several alternative measures of plant-level quality risk. H1 was generally robust ($p<.05$ for the main effect; $p<.10$ for the interaction term) to 3-level categorizations of different measures of the dependent variable, $QRisk_p$, that used a formula different from the one created by the panel of experts (Equation 1). These include the exclusion of the trend term, dividing the trend term by n instead of n+1, and the replacement of the trend term with a standard deviation term (when both the standard deviation of audit outcomes and the trend term were included, the p-values for the main effect and interaction term were .008 and .12, respectively). We also found CM and the interaction term to be robust using the five quality risk Delphi consensus group categories (Table 6) in ordered logit regression and a binary logit model with dummy coded for the high risk group vs. all others. H1 was also robust ($p<.05$ for both the CM term and the interaction term) to the inclusion of dummy variables for location (based on both FDA regions and districts and Census regions and divisions) as well as the exclusion from the analysis of the two plants whose primary classifications were R&D (included with the pharmaceutical group in the main analysis) and the five plants classified as primarily repackagers (included in the toilet goods group in the main analysis). In addition, there was one plant whose primary classification was "biologic" that we coded as pharmaceutical. The results are

robust (p<.05 for both the CM and the interaction term) to the elimination of this plant from the study. We also ran the analysis at the audit level, controlling for time. Here, H1 was partially supported. If the interaction term is excluded, CM has a moderately significant effect on quality risk (p<.10), but the interaction term dominates when it is included (p<.01), and the first-order effect of the CM variable slips to p≈.13. The non-significant ISO 9000 effect in Table 6 was also maintained in all of the robustness checks discussed above (p > .10), as well as for several different rules for classifying plants as ISO 9000 certified or not. Further details on robustness checks are available upon request.

4.4 Discussion of Results

Controlling for basic firm and plant characteristics, our results provide empirical evidence that firms in the drug sector take on significantly greater quality risk when outsourcing production to a contract manufacturer in comparison to manufacturing in an internal plant with similar observable characteristics, but that this difference dissipates with plant age (See Table 7). This overall result indicates that insights from the operations management literature on quality seem to dominate those offered by the commonly held core competence perspectives from strategic management regarding contracting out production. From OM quality management-based theory, it is difficult to observe a plant's true quality risk in a contract manufacturer. Misaligned incentives between the contract manufacturer and the buying firm, and difficulty coordinating across organizational boundaries combine with this unobservability to create a quality risk in outsourcing to contract manufacturers.

Unexpectedly, age seems to impact internal plants differently than contract manufacturers. As depicted in Table 7, increasing plant age *increases* quality risk (on average) for IPs and does not have a significant effect on CMs. It seems counterintuitive that the quality risk level of older IPs is higher than that for younger IPs. In this paragraph, we speculate why this effect may occur. For younger internal plants, it is likely that the senior company and plant management put significant effort into establishing a high quality baseline, including reapplying QM systems from other plants in the network. The necessary "unstructured technical dialogue" (Monterverde 1995) flows freely (at least, relative to outsourced

production) from corporate to the plant to allow a successful startup; quality expectations are clear and ingrained into the management team and employees. However, perhaps due to the unobservability of quality risk combined with *intra*-organizational measurement systems and incentives, quality risk creeps in over time. We suspect that as internal plants age, the attention from some corporate offices may begin wane and resemble that given to a CM. With age, some of the same multi-dimensional incentives issues that influence quality risk in outsourcing could also lead a "quality risk creep" in internal production. That is, internal managers begin to respond to their *intra*-organizational incentives of cost, delivery, and "acceptable" quality. This, combined with increasing levels of bureaucracy, possible lack of investments causing increased process variability, and complacency create a degree of entropy, which may increase quality risk in internal plants. While our assertions are plausible, an in-depth investigation of the quality risk dynamics with plant age is clearly an area for future research.

In this research, ISO 9000 does not reduce the quality risk of a plant—a result that should be interpreted with caution, as the sample contained very few ISO 9000 certified plants. The fact that ISO 9000 seems to not systematically relate to lower quality risk indicates that this commonly used method of assessing quality systems when outsourcing production to contract manufacturers may not be a sufficient indicator. The successful implementation of the easy-to-observe practices, such as those necessary for ISO 9000 certification, should not be automatically construed as assurance that the plant will operate with a low quality risk. This result underscores the challenge faced by buyers when assessing CMs—systems can appear to be in order, but quality risk may still be high

Finally, an examination of the control variables provides further insights on quality risk. Two control variables—total company sales and number of plant employees—were found to have a weak influence on the level of quality risk (β =-.17, p=.07 and β =.23, p=.10). It is not surprising that large companies and plants may be more prone to lower quality risk. Large entities generally have the resources to invest in quality management systems and infrastructure that can help create a culture of reduced risk in daily work. Similarly, it seems plausible that larger plants may have more difficulty maintaining a low quality

risk than smaller plants, ceteris paribus. We also note that when plants tend to produce regulated products exclusively, there is evidence (β=1.10, p=.04) of a greater quality risk. We leave assessment of the robustness of this result and its explanation to future research. Finally, we note that we found no systematic differences in any other control variables, including NAICS product types.

4.5 Generalizability of Results

This study has taken place in the U.S. drug manufacturing industry, examining plants in the mainland United States. It is worth commenting on how these results might be generalizable to other industries, based on the theory used to develop the hypotheses. We first note that drug products would score fairly low on the construct of "testability" (Roth et al., 2008b). Drug manufacturing is usually high-speed, and only a small percentage of products are tested. Moreover, it is neither possible to test for all possible contaminants nor for the effect of subtle process changes that may impact quality in use. Products with higher testability (such as a lower-volume, hard goods, or electronics) may have a lower quality risk in outsourcing for three related but distinct reasons. First, high testability reduces the risk of off-quality product being shipped which directly reduces quality risk. Second, high testability reduces the need for the buyer to observe the input behaviors that can lead to quality risk, thus making quality risk more observable. Third, this observability allows contracts to be written that can improve the level of cooperation across the supply chain. Thus, industries with higher testability may have less of a quality risk in outsourcing than process industries, such as drugs and food.

While testability implies that the drug industry may have a higher quality risk in outsourcing than others, the drug manufacturing industry is heavily regulated, and all manufacturers must comply with Good Manufacturing Practices (GMPs) to some degree simply to remain in business. It is plausible that other industries which share some of the characteristics of the drug sector, but are not as tightly regulated, would have an even greater quality risk in outsourcing.

5. Conclusion

In this section, we first discuss the limitations of our study and some opportunities for future research. We then review the implications for research and practice.

5.1 Limitations

This study used a collection of publicly available secondary data from multiple sources to rigorously examine the issue of quality risk in contract manufacturing, which is of mounting importance to operations, supply chain management, and business strategy. The data sources and approach both have a number of limitations, some of which can be addressed in future work. While FDA inspections are thorough, standard, and designed to assess quality risk, the combination of two inspection results (e.g., 483 yes/no and District Decision) used in this study is a coarse measure. In addition, there may be variability in FDA inspectors (Macher et al. 2006) as well as inspection purpose.

Another limitation is that the reliability of most of the control variables obtained from secondary sources cannot be readily assessed (Roth et al. 2008a). Furthermore, even if perfectly reliable, the control variable data were only available for 2006 although the inspections date back as far as 1994. Most firms and plants have presumably changed during that time. In addition, very few companies in our sample were ISO 9000 certified. Thus, our findings related to certification, while consistent with theory, should be interpreted with caution. Also, the detailed operational drivers of quality risk between CMs and internal plants were not specifically evaluated in this research. For contract manufacturers, these drivers include customer expectations and practices, as well as internal CM practices around quality.

Finally, there are alternative explanations to those provided in Section 2 that may account for the empirical results obtained. The proper interpretation of our result is this: conditional upon the variables included in the model, CMs tend to have higher quality risk than IPs. The inclusion of the control variables captures many possible alternative explanations. For example, total company sales controls for the possibility that the FDA may more aggressively pursue large companies and/or that smaller companies may lack the resources to invest in professionals who specialize in compliance. Plant size

would account for the possibility that large plants may have more potential issues for the FDA to investigate, and age would account for old plants having more disrepair and/or new plants having more start-up issues. The "only regulated control variable" may account for account for potential audit inconsistencies between plants that produce primarily unregulated versus regulated products. One potential explanation that is not controlled for in this study, and cannot be controlled for with available data, is that plant-level complexity may increase quality risk and may be systematically different between CMs and IPs. It could be possible that contract manufacturers, in serving multiple customers with diverse needs, face more product changeovers, use more general-purpose equipment, and require their employees to learn more technical and organizational processes then a comparably-sized internal plant. Further, while we have controlled for product using NAICS code, NAICS codes are a coarse measure. One may wonder whether products produced at a CM may differ in some way from products produced at an internal plant that is relevant to quality risk. To this, we note that the breadth of products produced at CMs seems similar to those produced at internal plants. CMs are often utilized to produce new products for capital avoidance, sometimes due to the possession of technology unavailable to brand-owning firms (Anscomb 2006). Further, CMs are utilized by private-label manufacturers to produce entire product lines. Practitioner surveys indicate that other factors—not the inherent importance of quality—are the drivers of outsourcing decisions (e.g., Contract Pharma 2008). And, scholarly empirical evidence indicates that the importance placed on quality is not a significant factor in an MBU's propensity to outsource (Gray et al. 2009b). Thus, we believe that there are not systematic differences regarding the importance of quality risk between products produced at IPs and CMs. However, lacking a direct measure of the importance of quality, we cannot test this assertion econometrically.

5.2 Future Research

Our research has identified areas that point toward several interesting follow-up studies. First, we note that it may be interesting to explicitly study the quality risk of plants that perform both contract manufacturing and internal production. Further, surveys of buyers and CMs could provide insights into

the nature of the internal CM systems and customer behaviors that have the potential lower the average quality risk. Clearly, there are some CMs with low quality risk, and it is important to understand what factors underlie their success and survival. Comparative analyses would provide guidance to buyers regarding differences in observable practices and behaviors between low quality-risk and high quality-risk CMs. Also, further longitudinal research is required to investigate how the causal organizational mechanisms drive quality risk dynamically, such as the entropic effect of aging facilities on quality risk that we observed in IPs. In addition, studies in other industries would enhance the generalizability of these findings. Of particular interest would be a comparison of industries with low testability and high testability. However, in industries not regulated by the FDA, new establishment-level measures of quality risk would be necessary. One option would be the creation of a latent variable based on reflective and/or formative measures (Bollen 1989) obtained with input from industry experts.

Another broad area of potential future research would be a deeper study of the dynamics of manufacturing outsourcing in the drug industry. This industry exhibits a great deal of variance in the manufacturing strategies of firms: some firms have never had internal production, a few once had such production but have since outsourced all of their production, others have maintained most production in-house, and many outsource only a portion of their volume. Moreover, as U.S. drug firms move into offshore contract manufacturing relationships, it will be increasingly important to evaluate the added quality-risk impact of geographical and cultural distances against the baseline results found here (Gray et al. 2009a). This study will hopefully spur future research regarding other difficult-to-measure implications of the outsourcing of production in this crucial industry.

5.3 Implications for Research and Practice

This paper offers several important contributions to the research community. First, we have introduced a new plant-level, objective measure of quality that is available in FDA-regulated industries and can be used in future research. Because expert judges conduct these inspections, the measure captures the difficult-to-observe aspects of quality risk. Second, this research has provided empirical evidence that in

the drug industry, outsourcing to contract manufacturers poses a quality risk, on average. Third, we surprisingly found a strong interaction effect between age and plant type, which was driven primarily by a degradation of quality risk with age among internal plants. This interaction result is intriguing and warrants further investigation. Fourth, our results lend credence to the quality management (QM) literature by indicating that easy-to-observe practices, such as those required by ISO 9000, do little to independently reduce quality risk. Fifth, this research links QM to incentives theory by explaining the underlying theoretical mechanisms by which outsourcing to CMs can pose an added quality risk.

Our study also offers managerial insights. First, despite the well-intended efforts of firms to ensure quality production when outsourcing, there still exists a systemic difference between the quality risk of IPs and CMs. The underlying mechanisms that cause the risk—including the unobservability of quality-risk lowering behaviors and the subsequent tendency to shirk in those behaviors—are difficult to eliminate. Thus, the total cost of outsourcing to CMs should specifically include the additional internal costs of prevention and quality interventions as well as an expected increased risk of quality failures. Second, managers of internal plants should beware of the apparent risk of "quality risk creep" as plants age. Third, when choosing a contract manufacturer, buying firms must thoroughly investigate potential CMs in an attempt to assess their robustness and experience relative to quality. However, because full knowledge of the quality risk is not possible to observe directly, significant effort to reduce quality risk will be necessary on an ongoing basis. Even with careful selection and monitoring, however, executives must understand that when making a decision to outsource they are likely accepting higher quality risk, *ceteris paribus*, than could be obtained internally. Our results support Fine and Whitney's (1996) call for make-buy decision competence. Arguably these make-buy skills require a deep understanding of nuanced quality risk factors described here.

In conclusion, this research reveals the tendency for contract manufacturers to operate with a higher quality risk than those of internal plants, on average. And, unfortunately, we find no evidence that documented processes provide an assurance of reduced quality risk. We believe this work adds an

important strategic operations perspective to the theory of the firm. We hope it will serve as a point of departure for future research and also highlight to managers the need to account for quality risks in their outsourcing decisions.

Acknowledgments

The authors gratefully acknowledge the support of the Joseph M. Juran Center for Leadership in Quality at the University of Minnesota and SAP AG for their support of this research. We are also grateful to Gary Koch and George Easton for their valuable insights into the model.

References

Albright, S.C., W.L. Winston, C.J. Zappe. 2006. *Data Analysis and Decision Making.* Thomson Southwestern, Mason, OH.

Alchian, A.A., H. Demsetz. 1972. Production, information costs, and economic organization. *Am. Econ. Rev.* **62**(5) 777-795.

Allison, P.D. 2002. *Missing Data.* Sage, Thousand Oaks, CA.

Anderson, E.G., A. Davis-Blake, S.S. Erzurumlu, N.R. Joglekar, G.G. Parker. 2008. The effects of outsourcing, offshoring, and distributed product development organizations on coordinating the NPD process. C.H. Loch, S. Kavadias, eds. *Handbook of New Product Development.* Elsevier, Hungary, 259-289.

Anderson, J.C., M. Rungtusanatham, R. Schroeder. 1994. A theory of quality management underlying the Deming management method. *Academy of Management Review* **19**(3) 472-509.

Anderson, S.W., J.D. Daly, M.F. Johnson. 1999. Why firms seek ISO 9000 certification: Regulatory compliance or competitive advantage? *Production and Operations Management.* **8**(1) 28-43.

Anscomb, 2006. Pharmaceutical Outsourcing Opportunities Post Launch: Contract Manufacturing, Sales and Marketing. 2nd Edition. New York: Kalorama Information.

Baiman, S., P.E. Fischer, M.V. Rajan. 2000. Information, contracting, and quality costs. *Management Science* **46**(6) 776.

Barney, J.B. 1999. How a firm's capabilities affect boundary decisions. *Sloan Manage. Rev.* **40**(3) 137-145.

Barzel, Y. 1982. Measurement cost and the organization of markets. *Journal of Law & Economics* **25**(1) 27-48.

Bollen, K. 1989. *Structural Equation Models with Latent Variables.* John Wiley & Sons, New York.

Borooah, V.K. 2002. *Logit and Probit: Ordered and Multinomial Models.* Sage, Thousand Oaks, CA.

Boyer, K.K., M. Swink, E.D. Rosenzweig. 2005. Operations strategy research in the POMS journal. *Production and Operations Management* **14**(4) 442-449.

Cachon, G.P. 2003. Supply chain coordination with contracts. A.G. de Kok, S.C. Graves, eds. *Handbooks in OR & MS.* Elsevier, 229-339.

Chesbrough, H.W., D.J. Teece. 1996. When is virtual virtuous? Organizing for innovation. *Harv. Bus. Rev.* **74**(1) 65-73.

Contract Pharma, 2008. 2008 Outsourcing Survey. May 2008. http://www.contractpharma.com/articles/2008/05/2008-annual-outsourcing-survey, accessed October 26, 2008.

Corbett, C.J., M.J. Montes-Sancho, D.A. Kirsch. 2005. The financial impact of ISO 9000 certification in the United States: An empirical analysis. *Management Science* **51**(7) 1046-1059.

Daft, R.L., R.H. Lengel. 1986. Organizational information requirements, media richness and structural design. *Management Science* **32**(5) 554-571.

Doig, S.J., R.C. Ritter, K. Speckhals, D. Woolson. 2001. Has outsourcing gone too far? *McKinsey Quarterly* **2001**(4) 24-37.

Dow, D., D. Samson, S. Ford. 1999. Exploding the myth: Do all quality management practices contribute to superior quality performance? *Production and Operations Management* **8**(1) 1-27.

D'Souza, A.G., D.J. Keeling, R.D. Phillips. 2007. Improving quality in pharma manufacturing. *McKinsey Quarterly.* September, 2007.

Economides, N. 1999. Quality choice and vertical integration. *International Journal of Industrial Organization* **17**(6) 903-914.

Edmondson, G. 2003. Look who's building Bimmers. *Business Week* (December 1, 2003) 18-19.

Eisenhardt, K.M. 1989. Agency theory: An assessment and review. *Academy of Management Review* **14**(1) 57-74.

Farley, D. 1993. Court decision strengthens FDA's regulatory power. *FDA Consumer.* **27**(7) 40-41.

FDA, 2008. Investigations Operations Manual. http://www.fda.gov/ora/inspect_ref/iom/pdftoc.html, accessed November 19, 2008.

Filer, R.K., D.L. Golbe. 2003. Debt, operating margin, and investment in workplace safety. *Journal of Industrial Economics* **51**(3) 359-381.

Fine, C., D. Whitney. 1996. Is the make-buy decision process a core competence? International Center for Research on the Management of Technology Working Paper #140-96, Massachusetts Institute of Technology, Cambridge MA.

Flynn, B.B., R.G. Schroeder, S. Sakakibara. 1995. The impact of quality management practices on performance and competitive advantage. *Decision Sciences* **26**(5) 659-691.

Forker, L.B. 1997. Factors affecting supplier quality performance. *J. Oper. Manage.* **15**(4) 243-269.

Forthofer, R.N., E.S. Lee, M. Hernandez. 2007. Biostatistics. Academic Press, St. Louis.

Gerde, V.W., J.M. Logsdon. 2001. Measuring environmental performance: Use of the Toxics Release Inventory (TRI) and other U.S. environmental databases. *Business Strategy & the Environment* **10**(5) 269-285.

Giffi, C., A.V. Roth, G.M. Seal. 1990. *Competing in World-Class Manufacturing: America's 21st Century Challenge.* Business One Irwin, Homewood, IL.

Grant, D., A.W. Jones. 2004. Do manufacturers pollute less under the regulation-through-information regime? What plant-level data tell us. *Sociological Quarterly* **45**(3) 471-486.

Grant, R.M. 1996. Toward a knowledge-based theory of the firm. *Strategic Manage. J.* **17** 109-122.

Gray, J.V., A.V. Roth, M.J. Leiblein. 2009a. Offshoring's quality risk: An explanation and empirical test. *Ohio State University Working Paper.*

Gray, J.V., A.V. Roth, B. Tomlin. 2009b. The influence of cost and quality priorities on the propensity to outsource production, *Decision Sciences* (forthcoming)

Gray, W.B., R.J. Shadbegian. 2005. When and why do plants comply? Paper mills in the 1980s. *Law & Policy* **27**(2) 238-261.

Grossman, G.M., O.D. Hart. 1986. The costs and benefits of ownership: A theory of vertical and lateral integration. *Journal of Political Economy* **94**(4) 691-719.

Gulati, R., P.R. Lawrence, P. Puranam. 2005. Adaptation in vertical relationships: Beyond incentive conflict. *Strategic Manage. J.* **26**(5) 415-440.

Hair, J.F., R.E. Anderson, R.L. Tatham, W.C. Black. 1998. *Multivariate Data Analysis.* Prentice-Hall, Upper Saddle River, NJ.

Harris, R.G., C.M. Winston. 1983. Potential benefits of rail mergers: An econometric analysis of network effects on service quality. *Rev. Econ. Stat.* **65**(1) 32-40.

Hayes, R.H., G. Pisano, D. Upton, S.C. Wheelwright. 2005. *Pursuing the Competitive Edge.* John Wiley & Sons, Inc., Hoboken, NJ.

Helberling, G. 2007. *The ISO Survey–2006.* ISO, Geneva, Switzerland.

Hill, T. 1994. *Manufacturing Strategy: Text and Cases*. McGraw-Hill, New York.

Hirschler, B. 2007. Big pharma shrinks as slowdown forces cuts. Retrieved December 13, 2007, [*Reuters* http://uk.reuters.com/article/stocksAndSharesNews/idUKNOA44257120071214?feedType=RSS&feedName=stocksAndSharesNews]

Hoetker, G. 2007. The use of logit and probit models in strategic management research: Critical issues. *Strategic Manage. J.* **28**(4) 331-343.

Holmstrom, B., P. Milgrom. 1991. Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics & Organization* **7**(2) 24-52.

Hsieh, C.M., S.G. Lazzarini, J.A. Nickerson. 2004. Outsourcing and the variability of product performance: Data from international courier services. *Working Paper* [University of Missouri—Rolla].

Hwang, I., S. Radhakrishnan, L. Su. 2006. Vendor certification and appraisal: Implications for supplier quality. *Management Science* **52**(10) 1472-1482.

Jeffries, N. 2003. Contract manufacturing revealed. *Global Cosmetic Industry,* **171** 58-61.

Kaya, M., O. Özer. 2005. Quality risk in outsourcing: Noncontractible product quality and private quality cost information. *Proceedings of the M&SOM Conference.* INFORMS, Chicago.

King, A.A., M.J. Lenox. 2001. Lean and green? An empirical examination of the relationship between lean production and environmental performance. *Production & Operations Management* **10**(3) 244-254.

Laffont, J., D. Martimort. 2002. *The Theory of Incentives: The Principle-Agent Model.* Princeton University Press, Princeton, NJ.

Laverty, K.J. 1996. Economic "short-termism": The debate, the unresolved issues, and the implications for management practice and research. *Academy of Management Review* **21**(3) 825-860.

Leiblein, M.J. 2003. The choice of organizational governance form and performance: Predictions from transaction cost, resource-based, and real options theories. *Journal of Management* **29**(6) 937.

Leiblein, M.J., J.J. Reuer, F. Dalsace. 2002. Do make or buy decisions matter? The influence of organizational governance on technological performance. *Strategic Manage. J.* **23**(9) 817-833.

Linston, H.A., M. Turoff, eds. 1975. *The Delphi Method: Techniques and Applications.* Addison-Wesley, Reading, MA.

Long, J.S., Freese, J. 2006. *Regression Models for Categorical Dependent Variables Using Stata, 2nd Edition.* Stata Press, College Station.

Macher, J.T., J.W. Mayo, J.A. Nickerson. 2006. Exploring the information asymmetry gap: Evidence from FDA regulation. *Working Paper,* Georgetown University.

Macher, J.T., B.D. Richman. 2008. Transaction cost economics: An assessment of empirical research in the social sciences. *Business & Politics* **10**(1) 1-63.

Macher, J.T., J.A. Nickerson. 2006. *Pharmaceutical Manufacturing Research Project.*

Martinez-Costa, M., T.Y. Choi, J.A. Martinez-Garcia, A.R. Martinez-Lorente. 2009. ISO 9000/1994, ISO 9001/2000 and TQM: The performance debate revisited. *Journal of Operations Management*, (forthcoming).

Mayer, K.J., J.A. Nickerson, H. Owan. 2004. Are supply and plant inspections complements or substitutes? A strategic and operational assessment of inspection practices in biotechnology. *Management Science* **50**(8) 1064-1081.

Meyer, H.R. 1998. Small firms flock to quality system--intl organization for standards' ISO 9000 certification. *Nation's Business* (March, 1998).

Michael, S.C. 2000. The effect of organizational form on quality: The case of franchising. *J. Econ. Behav. Organ.* **43**(3) 295-318.

Monteverde, K. 1995. Technical dialog as an incentive for vertical integration in the semiconductor industry. *Management Science* **41**(10) 1624-1638.

Nair, A. 2006. Meta-analysis of the relationship between quality management practices and firm performance—Implications for quality management theory development. *J. Oper. Manage.* **24**(6) 948-975.

Naveh, E., A.A. Marcus. 2004. When does the ISO 9000 quality assurance standard lead to performance improvement? Assimilation and going beyond. *IEEE Trans. Eng. Manage.* **51**(3) 352-363.

Nelson, R., S. Winter. 1982. *An Evolutional Theory of Economic Change.* Harvard University Press, Cambridge.

Ojala, M. 2004. Taming public and private company research. *Online* **28**(3) 44.

Parker, G.G., E.G. Anderson Jr. 2002. From buyer to integrator: The transformation of the supply-chain manager in the vertically disintegrating firm. *Production and Operations Management* **11**(1) 75.

Pharma Marketletter, 2007. PharmaFab agrees to stop illegal practices. April 26, 2007.

Powell, T.C. 1995. Total quality management as competitive advantage: A review and empirical study. *Strategic Manage. J.* **16**(1) 15-37.

Prahalad, C.K., G. Hamel. 1990. The core competence of the corporation. *Harv. Bus. Rev.* **68**(3) 79-91.

Quelin, B., F. Duhamel. 2003. Bringing together strategic outsourcing and corporate strategy: Outsourcing motives and risks. *European Management Journal* **21**(5) 647-661.

Quinn, J.B., F.G. Hilmer. 1994. Strategic outsourcing. *Sloan Manage. Rev.* **35**(4) 43-55.

Reyniers, D.J., C.S. Tapiero. 1995. The delivery and control of quality in supplier-producer contracts. *Management Science* **41**(10) 1581-1590.

Robinson, C.J., M.K. Malhotra. 2005. Defining the concept of supply chain quality management and its relevance to academic and industrial practice. *International Journal of Production Economics* **96**(3) 315-337.

Roth, A.V., R.G. Schroeder, X. Huang, M.M. Kristal. 2008a. *Handbook of Metrics for Research in Operations Management.* Sage Publications, Newbury Park, CA.

Roth, A.V., A.A. Tsay, M. Pullman, J.V. Gray. 2008b. Unraveling the food supply chain: Strategic insights from China and the pet food recalls. *Journal of Supply Chain Management,* forthcoming.

Roth, A.V., A.S. Marucheck, A. Kemp, D. Trimble. 1994. The knowledge factory for accelerated learning practices. *Planning Review* **22**(3) 26-33.

Samson, D., M. Terziovski. 1999. The relationship between total quality management practices and operational performance. *J. Oper. Manage.* **17**(4) 393-409.

Sharma, D.S. 2005. The association between ISO 9000 certification and financial performance. *International Journal of Accounting* **40**(2) 151-172.

Sheopuri, A., E. Zemel. 2005. Offshore & onshore contracts. *Working Paper,* [New York University]

Sosa, M.E., Eppinger, S.D., Rowles, C.M. 2004. The misalignment of product architecture and organizational structure in complex product development. *Management Science.***50** (12) 1674-1689.

Sousa, R., C.A. Voss. 2002. Quality management revisited: A reflective review and agenda for future research. *J. Oper. Manage.* **20**(1) 91-109.

Sturgeon, T.J. 2002. Modular production networks: A new American model of industrial organization. *Industrial and Corporate Change* **11**(3) 451-496.

Terlaak, A., A. King. 2006. The effect of certification with the ISO 9000 quality management standard: A signaling approach. *J. Econ. Behavior & Org.* **60**(4) 579-602

Trent, R.J., R.M. Monczka. 1999. Achieving world-class supplier quality. *Total Qual. Manage.* **10**(6) 927-938.

Tully, S. 1994. You'll never guess who really makes... *Fortune* **130**(7) 124-129.

Verbeek, M. 2004. *A Guide to Modern Econometrics.* John Wiley & Sons, Ltd., West Sussex, England.

Voss, C., K. Blackmon. 1998. Is ISO 9000 good for business? Institutional and efficiency perspectives on quality management. In *1998 Academy of Management Best Paper Proceedings.*

Wilhelmsson, M. 2004. Outsourcing patterns in the pharmaceutical industry. *IMD Case Study* **IMD-6-0263** (24.02.2004).

Williamson, O.E. 1991.Strategizing, economizing, and economic organization . *Strategic Manage. J.* **12**(8) 75-94.