Organizational, Operational, and Behavioral Causes of Product Recalls

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

George Ball

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Advisor: Dr. Rachna Shah

June, 2015

Acknowledgements

I would like to acknowledge the instrumental support of my advisor, Dr. Rachna Shah. Her mentoring, encouragement, support and friendship over the last several years have been transformational in my development. I am immensely grateful for all she has done.

I would also like to acknowledge the support of my committee members: Dr. Karen Donohue, Dr. Enno Siemsen, and Dr. Serguei Netessine. Their guidance has been an essential element of this work.

Finally, I also thank my fellow PhD students for their support and friendship in this journey, especially Suvrat Dhanorkar, Ujjal Mukherjee, Rick Hardcopf and Pettis Kent.

Dedication

I dedicate this work to my loving wife Heidi who has limitless patience and love, and to my wonderful children: John, Maria and Peter.

Abstract

Research germane to product recalls and their causes is limited. With recall rates rising in many industries, it is timely and pertinent to comprehensively investigate recalls. The focus of my dissertation is on product recalls and their causes, with the objective of recall understanding and prevention. I study three important phases in the product recall process at multiple organizational levels in the high-risk medical device industry: plant-level causes, recall decision-making, and causes and effects of firm and regulator responsiveness within the recall event.

First, I study the relationship between Food and Drug Administration (FDA) plant inspections and future recalls. Using a 7-year panel dataset and recurrent event Cox proportional hazard and propensity score matching models, I find that adverse plant inspection outcomes serve as warning signs for future recalls. I incorporate FDA investigator experience to identify reasons for, and effects of, investigator complacency in repeated plant inspections. Repeated visits to the same site by an inspector increases the recall risk and also reduces the predictability of inspection outcomes as a leading indicator of future recalls. FDA investigator rotation is shown to be an effective solution to compensate for investigator complacency.

Second, I explore behavioral factors that influence managers' decision to recall. Recall guidance provided by the FDA allows for broad managerial interpretation so it is crucial to study which factors influence managers to choose to recall. Using actual industry managers with recall experience in a controlled experiment, I find that product defects which are undetectable to physician customers pre-use are more likely to lead to a recall than detectable ones. When managers have a deeper understanding about the root cause of a defect, they are also more likely to recall. I also study individual dispositional factors unique to each manager, and surprisingly find that the level of cognitive reflection, as measured by the Cognitive Reflection Test (CRT), is the most important predictor of a recall decision in the experiment.

Finally, I study firm and regulator recall responsiveness. Responsiveness is critical in this domain: the longer a faulty medical device remains on the marketplace, the more consumers are at risk. Using an 11-year panel dataset with time-stamps for over 4,000 recalls, and multiple hazard and fixed effects panel models, I find that higher recall severity leads to slower firm and faster FDA responsiveness. However, taking longer to close a recall reduces a firm's future recalls, and this may be attributed to learning mechanisms. FDA response times also reduce future recalls.

Table of Contents

List of Figures

List of Tables

Chapter 1:

Dissertation Overview

Producing and delivering high quality products that are safe for consumers' use are critical to a firm's survival and growth. To ensure product safety, firms allocate considerable resources to develop and implement sophisticated quality control systems. Federal agencies monitor and regulate firms' design and manufacturing processes to ensure products meet quality conformance standards. Despite these efforts, product recalls have increased significantly across all regulated industries which are subject to product recall laws and requirements. In the United States, recalled products fall into five main categories: consumer products, motor vehicles, food, pharmaceuticals and medical devices. These five categories are monitored by three regulatory bodies: Consumer Product Safety Commission (CPSC), National Highway Traffic and Safety Administration (NHTSA), and the Food and Drug Administration (FDA)[1]. According to the CPSC, approximately six recalls occur per day across all types of consumer products, and the number of consumer product recalls is on the rise.[2] Similarly, NHTSA reports the average number of auto recalls per million registered U.S. vehicles has risen steadily, from 3.10 in 1980s to 8.25 in the 1990s and 11.79 between 2000 and 2010. This is significant, because each auto recall is associated with a potential economic consequence of $20 million or more[3]. Equally significant, five of the top ten largest auto recalls in U.S. history occurred in the last ten years. The medical device marketplace is not immune to this trend. Between 2003 and 2012, there was a 97% increase in the annual number of medical device recalls.[4] This number is staggering when one considers that there are thousands of devices and customers' lives that are often affected by just one medical device recall. Prior literature annotated in Table 1.1, suggests that each recall is associated with considerable costs for a company and its customers, including occasional loss of life.

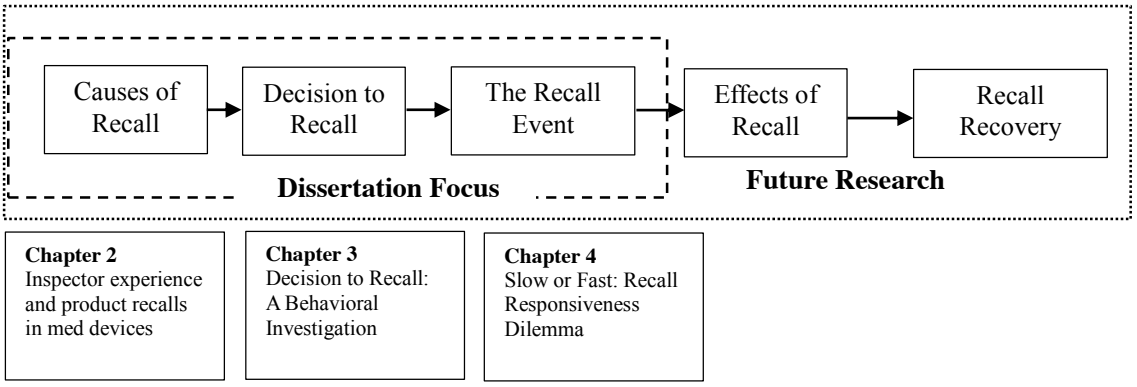
Product recalls have been explicitly studied in marketing, economics, strategy, and operations management. A comprehensive review of this literature shows that the dominant focus

[1] http://www.recalls.gov/
[2] www.cpsc.gov;http://usatoday30.usatoday.com/news/nation/story/2012-06-08/product-recall-surge-consumer-fatigue/55466398/1
[3] Jarrell and Peltzman (1985) estimate an average per car recall cost of $200. This cost includes repair, replacement, and lost sales. An average of 100,000 cars were involved in each auto recall from 1980 to 2013. http://www-odi.nhtsa.dot.gov/downloads/.
[4] http:// www.fda.gov/ downloads/AboutFDA/CentersOffices/OfficeofMedicalProductsandTobacco/ CDRH/CDRHTransparency/ UCM388442.pdf

in most of the existing research has been on the consequences of product recalls in which its impact on firms' financial and market performance are examined (Table 1.1). These studies demonstrate that product recalls are negatively associated with various measures of firm performance. At the same time, they point to the paucity of studies examining causes of product recalls. Two noteworthy exceptions to "consequence-focused" research include Haunschild and Rhee (2004) and Thirumalai and Sinha (2011), both of which model product recalls as a dependent variable. While Haunschild and Rhee (2004) examine whether the initiator of a past product recall (firm or a regulatory agency) impacts the likelihood of future recalls, Thirumalai and Sinha (2011) analyze the impact of a firm's R&D focus on future recalls. These two studies represent a first step in highlighting drivers of product recalls, but neither study focuses on operational drivers of product recalls—they conceptualize the causes at a higher level of analysis.

However, understanding ***why*** recalls occur is the necessary first step towards learning ***how*** to prevent them from occurring. The primary objective of my dissertation is to identify organizational, operational and behavioral causes of product recalls. In investigating them, I focus on two primary participants in the recall process: the government regulator and the firm manager. Regulators approve new products, inspect manufacturing facilities, and establish product quality guidelines, in an effort to protect the public from harmful product defects. Managers allocate resources to solve problems, strive to improve future product quality by learning from past mistakes, and when warranted, actually make the decision to recall. Figure 1.1 illustrates the different phases in the life cycle of a recall. It begins with the underlying causes of a recall and concludes with the eventual recovery after a recall has been undertaken and its consequence has been incurred. In my dissertation, I focus on the first three steps of the recall life cycle. Consequently, my dissertation consists of three chapters, each corresponding to the three phases of the recall process. Specifically, each chapter addresses a distinct aspect of the recall process by adopting a unique vantage point of an organizational entity. Thus, I examine 1) causes of recalls in the production plant, 2) behavioral factors influencing the managerial recall decision-making process, and 3) firm and regulatory responsiveness during the recall event. Each chapter is described briefly below.

Figure 1.1 Dissertation Structure



In chapter two, *"Inspector Experience and Product Recalls in the Medical Device Industry,"* I address an important quality control mechanism in global supply chains, that of plant inspections. In this chapter, I answer the questions: 1) How effective are Food and Drug Administration (FDA) plant quality inspections in predicting and preventing future medical device recalls and 2) how does inspector experience impact this predictability? FDA plant inspections are intended to assess the design, supplier development, manufacturing, and distribution processes associated with all products built at the plant. The FDA uses these inspections as the primary means of assessing quality in the medical device industry. It is therefore crucial to examine the validity of these inspections as it pertains to future medical device product quality. Because the FDA frequently uses the same investigator at a given plant, I also examine the effects of investigator experience.

To study these questions, I draw upon learning and complacency literature streams, compile a unique FDA inspection dataset, and a recurrent event, conditional gap-time Cox proportional hazard model. Through rigorous analyses and multiple robustness checks, I empirically demonstrate that plant inspection outcomes serve as reliable predictors of future product recalls. This is the first study, to my knowledge, to empirically associate plant inspections with future plant quality, and highlights an important mechanism to economically control quality in the context of global supply chains. However, I find that inspection predictability and future recall risks are negatively impacted when the FDA reassigns an investigator to the same plant (e.g., increasing specific experience of the investigator). Independent of the inspection outcome, the hazard of a recall at a plant increases by 48% the second time an investigator inspects a plant, and by 63% with the third visit. Repeat visits by the same investigator also reduces the ability of the inspection

outcome to predict future recalls. By the investigator's second visit to the plant, inspection outcomes no longer predict future recalls. Using a propensity score matching model to empirically demonstrate the benefits of investigator rotation, I show that rotating investigators among different plants and restricting their visit to the same plant to one time would reduce future recalls. Specifically, the rotational policy would lead to a reduction of over 900 medical device recalls in a 7-year period. While the policy is associated with an additional annual cost of approximately $800,000 to the government, it is a negligible amount relative to the increase in consumer welfare. This recommendation is currently under consideration by the FDA.

Once a product defect is released into the market place, managers are faced with the decision of whether to allow potentially harmful products to remain in the market, or to initiate an expensive product recall. In chapter three, titled, *"The Decision to Recall: A Behavioral Investigation in the Medical Device Industry"*, I study the following question: What behavioral factors influence managers' decision to recall a product? FDA's official recall policy states that companies are required to recall products that "present a risk of injury or gross deception or are otherwise defective."[5] Medical device recalls are almost always initiated by company managers, not mandated by regulators, and managers have great leeway in interpreting FDA recall regulations.[6] This makes it vital to understand what factors effect managers in this decision.

Prior behavioral research indicates the importance of examining not only factors that relate to the specific situation under consideration, but also those that are intrinsic to the decision-maker. Using a unique behavioral experiment related to the managerial recall decision, I investigate two types of managerial decision-making factors in this decision: situational and dispositional. Because the product recall decision is highly contextualized, I use actual industry managers from one of the world's largest medical device companies as subjects for the experiment. Two situational factors are significantly related to the likelihood of a recall decision: the level of detectability by the physician of the product defect, and the managerial understanding of the root cause of the product defect. Managers are 48% more likely to recall when the defect is undetectable to the physician before using the product and 65% more likely to recall when the root cause of the underlying defect is identified. A third situational factor, the product quality concern of an influential individual physician customer, does not impact the recall decision. I use several dispositional characteristics of the manager to investigate their impact on the recall decision, including a well-established

[5] http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/PostmarketRequirements/RecallsCorrectionsAndRemovals/
[6] Ibid

Cognitive Reflection Test (CRT, Frederick, 2005). The CRT measures how much a decision-maker reflects before making a decision. Higher CRT equates to more reflection, while lower CRT relates to less reflection and more intuition in decision-making. High CRT managers are 70% less likely to decide to recall than low CRT managers, independent of the experimental factors. Additionally, high CRT managers are significantly influenced by two of the situational factors in the experiment, while low CRT managers are only influenced by their individual dispositional characteristics such as gender (males are 96% less likely to recall than females) and functional area (quality department personnel are 98% less likely to recall than non-quality personnel). To my knowledge, this is the first study to investigate the recall decision-making process using actual industry managers in an experimental setting. This experiment exposes possible situational and dispositional factors that may be influential in this decision, even though the FDA does not explicitly delineate them as relevant recall criteria.

Subsequent to the recall decision, managers must expend significant resources to identify root cause and undertake corrective action related to the product defect, both to resolve the current problem and to avoid similar problems in the future. In the medical device industry, the individual steps constituting the recall process are very well-specified and clearly laid out, but the FDA does not formally mandate how much time should be spent on any one step. Intuitively, taking a long time between any two steps seems undesirable, yet anecdotal evidence suggests that there is considerable variation in the time taken to execute the recall process. In chapter four, titled: *"Slow or Fast: An Empirical Examination of the Recall Responsiveness Dilemma,"* I address causes and effects of responsiveness in the recall process, and study two related research questions: 1) What leads to quick recall response times for firms and regulators, and 2) how does such responsiveness impact future recall rates? A clear understanding of what leads to fast recall response times and how response times impact future recalls is a crucial step towards determining whether expedience is always desirable if reducing future recalls is the objective.

I decompose the medical device product recall process into three time-phases (time-to-open, time-to-classify, time-to-close) using a proprietary dataset provided by the FDA, and multiple hazard and fixed effects panel data models. Consistent with prospect theory (Kahneman and Tversky, 1979), I demonstrate that managers are risk-seeking in losses, and unfortunately delay recalling the most serious quality problems, precisely those problems which should lead to the fastest response due to their potential negative impact on customer health and safety. In other words, more severe recalls have a longer time-to-open than less severe recalls. Managers take 20% longer

to open a more severe recall (an additional 10 days) compared to a less severe recall. However, contrary to prospect theory, more severe recalls also have a longer time-to-close than less severe recalls. Managers take 23% more time-to-close more several recalls (an additional 76 days) than less severe recalls. The FDA is also observed to respond the fastest to the most serious problems, as they classify the most severe recalls over 500% faster than the least severe recalls. I also use a fixed effects negative binomial panel model to determine the effects that these different time phases (time-to-open, -close and -classify) have upon future recall occurrences at each medical device plant monitored by the FDA. I find that plants that move slower to close a recall, during which time plant personnel identify root cause and corrective action for the underlying defect, experience fewer future recalls. For every additional four months taken in the time-to-close a recall (a one half standard deviation from the mean time-to-close), plants experience one fewer recall across the time period in the panel. The slower speed may enable plants to learn from prior quality problems and apply this learning to future products. No association is observed between time-to-open and future recalls. Finally, faster FDA time-to-classify leads to fewer future recalls, but only for the least severe recall types.

As a set, this dissertation research provides a comprehensive investigation of the critical phases of the recall process using multiple methods (empirical and behavioral) and perspectives (plant, manager and regulator). The overarching goal of this research is to deepen recall understanding, and ideally lead to recall reduction and prevention. I empirically demonstrate leading indicators of product quality problems that can signal future medical device recalls, behavioral factors that contribute to the managerial recall decision, and possible learning mechanisms active within the recall event which can contribute to future recall reduction.

Table 1.1 Relevant Recall Literature

Article	Industry sector	Research Focus	Independent variables	Recall use[a]	Dependent variables	Results
Jarrell & Peltzman (1985)	Pharma; Auto	Impact of product recalls on shareholder wealth; Compare destruction of wealth to the recall costs for the firm	Recall event	IV	Stock market returns	Loss to shareholder wealth is substantially greater than the costs incurred by the firms.
Bromiley & Marcus (1989)	Auto	Stock market reaction to product recalls	Recall event	IV	Stock market returns	Stock market returns do not vary significantly with product recall announcements; they are not an effective deterrent to dubious corporate behavior.
Davidson & Worrell (1992)	Auto	Stock market reaction to product recalls	Recall event	IV	Stock market returns	Product recalls are negatively associated with stock market returns. The effect is stronger when products are replaced rather than repaired.
Archer & Weslowski (1996)	Auto	Impact of quality and service incidents upon customer loyalty	Negative quality and service incidents including recalls	IV	Customer loyalty	Product recalls do not impact customer loyalty to manufacturer
Haunschild & Rhee (2004)	Auto	Firm learning following voluntary and involuntary recalls	No. of recalls	IV, DV	# of recalls	Voluntary recalls result in more learning than mandated recalls when learning is measured as reduction in subsequent involuntary recalls.
Rhee & Haunschild (2006)	Auto	Impact of organizational reputation on market share following a recall	No. of recalls; organizational reputation; vehicle characteristics	IV	Market Share	Positive organizational reputation is associated with more negative market share reactions to product recalls. Having few substitutes for a product buffers this reaction.
Cheah et al. (2007)	Pharma	Impact of corporate social responsibility (CSR) practices on stock market reaction to product recalls	Recall event	IV	Stock market returns	Product recalls are negatively associated with stock market returns. The effect is dependent upon geography, recall severity, and CSR practices.
Chen et al. (2009)	Consumer product	Impact of recall strategy (proactive vs. passive) on stock market reaction to product recalls	Recall event	IV	Stock market returns	Proactive recalls are associated with more negative stock market reactions than passive recalls.
Thirumalai & Sinha (2011)	Medical device	Stock market reaction to product recalls and sources of recalls	Recall event; recall class & experience; product scope; R&D intensity	IV, DV	Stock market returns	Market penalties for medical device recalls are not significant. R&D focus increases likelihood of recalls.
Hora et al. (2011)	Toy	Impact of recall strategy, product defect and supply chain entity on time-to-recall	Recall strategy; source of defect; supply chain entity	DV	Time-to-recall	Preventive recall strategy, design defects & manufacturers are associated with longer time-to-recall.

a: IV = Independent Variable, DV = Dependent Variable.

Chapter 2:

Inspector Experience and Product Recalls in the Medical Device Industry

2.1 Introduction

Plant inspections are an essential instrument for managing quality in globally dispersed supply chains. Whether it is international firms auditing the quality of their own plants across the globe, or customers auditing the plants of their upstream partners, or governments auditing the plants that deliver regulated goods into distribution channels, assessing the state of quality of a plant is common practice. In supply chains with long lead times, firms cannot simply rely on inspecting incoming materials, nor can they realistically perform inspections and tests on all finished products. Firms need to assess and ensure quality performance at the source, particularly since long lead times lead to delayed feedback and tremendous uncertainty about the product already manufactured and shipped if defects are spotted among inbound supplies (Roth et al., 2008). In regulated markets, governments cannot feasibly sample every manufactured product before releasing it to customers, and instead have to rely upon the plant's quality system in order to prevent defective products from being shipped. Regulators therefore frequently assess the quality systems in place via plant inspections.

Research demonstrating the validity of plant inspections as accurate predictors of future quality performance is limited. Prior research advocates the use of plant inspections instead of product inspections (Deming, 1982) in an effort to improve quality processes and *design* and *build* quality into the product instead of *inspecting* quality into the product. This recommendation is based on the notion that plant inspections serve as reliable and more sensible, cost effective substitutes for product inspections (Concannon, 1989). In spite of this theoretical importance, there are no empirical studies to our knowledge which confirm the relationship between plant inspections and future product quality from the plant (Mayer et al., 2004; Handley and Gray, 2013). We examine this relationship in our current study.

We also examine whether inspector experience influences the predictive nature of audit outcomes by making inspections more or less informative. The literature on learning curves

indicates that general experience should lead to improved performance (Argote and Epple, 1990) and therefore more informative inspections. In contrast, research in accounting, financial, and safety audits demonstrates that inspectors can become complacent and less objective with site-specific experience (Bardach and Kagan, 2002). Given these opposing views, we empirically examine whether inspector experience creates more or less informative plant inspections. In line with the existing experience curve literature (Huckman and Pisano 2006), we differentiate between general experience (gained at all sites) and site-specific experience (gained at one site) and analyze whether these different forms of experience moderate the predictive association between inspection outcomes and future recalls, as well as whether experience directly leads to a change in the hazard of a future recall.

We study plant inspections in the context of the medical device industry. The Food and Drug Administration (FDA) regularly inspects medical device manufacturing plants, normally on a two-year cycle, and it is quite common for FDA investigators (FDA term for inspectors) to visit the same plant on multiple occasions. The intent of these inspections is twofold: to holistically assess the design, supplier development, manufacturing, and distribution processes associated with all products built at the plant, and to provide guidance to managers related to quality improvements. The FDA uses these inspections as the primary means of assessing, managing and enforcing quality in the high-risk medical device industry. Therefore, it is crucial to appraise the validity of these inspections as it pertains to future medical device product quality and to examine the possible effects of investigator experience.

One important manifestation of the absence of product quality in the medical device industry is "product recalls." Product recalls occur when a product is deemed unfit for customer use or conformance quality is lower than required (Garvin, 1987; Juran 1999). Whether the defect that leads to a recall originates from design, manufacturing, or within a supplier's processes, recalls result from a breach in the quality management system. Efforts to provide early detection or even prevention of recalls are in great need as the recall rates rise across many industries. For example, automotive recalls increased by 75% from 1980 to 2010[7], pharmaceutical recalls have increased by 77% from 2004 to 2014, and medical device recalls have risen almost 100% from 2003 to 2012.[8] If inspections consistently fail to detect the underlying problems, the efficacy of the inspection process itself is questionable.

[7] www.nhtsa.gov
[8] www.fda.gov

Consequently, we address three research questions in this study: 1) Do plant inspection outcomes serve as reliable predictors of future recall hazards, 2) Does investigator experience affect how well plant inspections predict future recalls, and 3) Independent of the inspection outcome, does investigator experience directly affect the hazard of a future recall. Using secondary data collected from several freedom of information act (FOIA) requests to the FDA over a 7-year period, and a recurrent event Cox Proportional Hazard model, we find that a favorable inspection outcome indicates a lower hazard of a future recall than an adverse inspection outcome does. While general experience has no main effect on the hazard of a recall or on the relationship between inspection outcomes and recalls, specific experience influences both whether inspection outcomes predict recalls as well as the overall hazard of recalls. The first time an investigator visits a site, their inspection outcomes are highly predictive of future recalls. In comparison to an "adverse" outcome, a "favorable" inspection outcome is associated with a 34% *decrease* in future recalls while an "adverse" inspection outcome is associated with a 52% *increase*. This relationship however diminishes very quickly. From the second inspection on, each additional visit by the same inspector increases the hazard of a recall. There is a 48% recall hazard increase on the second visit and a 63% recall hazard increase on the third visit. Additionally, from the third inspection onward, the inspection outcome does not effectively predict the future recall hazard.

These findings have practical implications for the FDA and industry managers, and inform the academic literature on quality and experience. In practice, the FDA may be able to reduce the hazard of medical device recalls by instituting investigator rotation between plants (a recommendation currently under consideration by the FDA). Similar implications may hold for practicing managers in private supply chains monitoring their upstream partners through plant inspections. Theoretically, our paper is the first, to our knowledge, to use archival data to demonstrate the relationship between plant inspections and future quality performance, indicating support for past theoretical quality research which has advocated the value of such inspections (Deming, 1982). We also bolster support for previous experience research by emphasizing the importance of distinguishing between specific and general experience (Huckman and Pisano, 2006) and demonstrating their differing impact on performance.

2.2 Research Context

A breakdown of product quality and the resulting recall in the medical device industry can

be life threatening, or can interfere with the successful treatment of patients. From 2003 to 2012, there was a 97% increase in the annual number of medical device recalls.[9] This figure is staggering when one considers that just one recall might affect thousands of devices and customers' lives. A medical device recall is defined by the FDA as, "a firm's removal or correction of a marketed product that FDA considers to be in violation of the laws it administers. A recall is a voluntary action that takes place because manufacturers and distributors carry out their responsibility to protect the public health and well-being from products that present a risk of injury or gross deception or are otherwise defective."[10]

The FDA plant inspection process is a comprehensive top-down review of the plant's quality management system and the constituent sub-systems. The review is conducted in adherence to the process described in the Quality System Inspection Techniques (QSIT) guide that an FDA investigator follows when inspecting a medical device manufacturing plant. An investigator may review any of the requirements contained within Part 820 of the U.S. Code of Federal Regulations Title 21 during the plant visit. These regulations apply to the "design, manufacture, packaging, labeling, storage, installation, and servicing of all finished devices intended for human use." (21CFR 820.1). During the inspection, the primary goal for the investigator is to "evaluate whether management with executive responsibility ensures that an adequate and effective quality system has been established and maintained."[11] It is important to note that these inspections are both intended to and capable of detecting product defects throughout the entire product realization process (e.g., from design to manufacturing to distribution). For example, an investigator may target the product design validation and change processes. In doing so, an investigator may identify a design validation step that was incorrectly implemented during product design which may serve as a warning signal for a failure in the market. Manufacturing quality control processes are also audited. Finding a gap in such a process may signal manufacturing defects that could be escaping to the customer undetected and possibly lead to a recall. Supplier development, product labeling, and distribution processes are similarly targeted. Investigators have leeway to perform the inspection as they deem appropriate, though the intent is to review aspects of each main function of a site during the inspection. The FDA publishes a frequency count of inspection concerns unearthed by inspectors during audits on their website. In 2014, the top 20 categories included

[9] http://www.fda.gov/downloads/AboutFDA/CentersOffices/OfficeofMedicalProductsandTobacco/CDRH/CDRHTransparency/UCM388442.pdf
[10] http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/PostmarketRequirements/RecallsCorrectionsAndRemovals
[11] http://www.fda.gov/ICECI/Inspections/InspectionGuides/ucm170207.htm#page1

manufacturing (lack of procedures for quality audits; procedures for training not established), purchasing (lack of purchasing controls) as well as design (lack of design control procedures, lack of procedures for design validation) related concerns, highlighting the holistic nature of this inspection process.[12]

The FDA uses three types of inspection categories: surveillance, compliance and complaint. Surveillance inspections occur regularly (approximately every two years) in order to determine the overall state of quality an establishment maintains. Compliance inspections are more narrowly targeted inspections, which normally transpire after major process changes such as a new manufacturing plant start-up, and complaint inspections occur because of one or more serious customer complaints. Upon completion, the information gathered by the FDA investigator leads to a recommended quality rating of the plant. Three quality levels can be assigned following an inspection. These are: *No Action Indicated* (NAI-highest quality score), *Voluntary Action Indicated* (VAI-moderate quality score), and *Official Action Indicated* (OAI-poorest quality score). If a plant receives a NAI, there is no official follow up from the FDA and no written FDA recommendations are given. Investigators provide voluntary written recommendations when assigning a VAI score, though these recommendations are not mandatory. An OAI, on the other hand, requires official follow up from the FDA, and denotes that significant quality system gaps were identified during the inspection. Thus, NAI is the highest quality score that a plant can receive (i.e. a favorable outcome), followed by a VAI and an OAI, which is the worst score possible (i.e. an adverse outcome).

During our extensive interviews with senior administrators at the FDA's Center for Device and Radiological Health, we learned that the timing of a plant's inspection is guided by the goal to inspect each plant once every two years, and the inspected plant is selected at random from the set of plants that need to be inspected at that time. During our conversations, we were also informed that the FDA generally reassigns investigators to plants they have inspected in the past to reduce investigator set-up costs and improve the inspection process. We were told that the FDA does not have a policy in place to rotate investigators among different plants or to systematically re-assign investigators to new plants, and most rotational changes that do occur are in response to investigator promotions, retirements, or new hires. In general, investigator assignment to plants is driven by geographic proximity between investigator and plant locations, and investigator rotation among

[12]http://www.fda.gov/ICECI/Inspections/ucm250720.htm

plants is largely due to exogenous reasons. Due to the lack of immediate feedback following an inspection outcome, it is challenging for the FDA to determine the accuracy (i.e. the information content) of their inspections. New assignments of inspectors are thus in general not driven by job performance.

2.3 Theory and Hypotheses: Inspection Outcomes and Future Recalls

Using inspections to monitor product and process standards is a well-established practice. Inspections are regularly conducted by firms both at their own manufacturing facilities as well as at their upstream supplier locations to ensure that products and processes meet established conformance standards. A key concept we use in our research is the information content of the outcomes of a plant inspection. An inspection outcome contains information if it adequately identifies deficiencies in the quality management system of a plant, or adequately assesses that no deficiencies are present. From an information theoretic perspective, information content of the inspection outcome is measured by how well the inspection outcome predicts future quality performance (e.g., recalls). Inspection outcomes that contain information should trigger improvement processes and predict future performance, whereas uninformative inspections should do neither.

Although Deming (1982) avidly advocates using supplier facility inspections as a means to improve quality at the source over wholesale supply material inspections, academic research has not systematically examined the influence of different plant inspection regimes on quality outcomes using objective and large scale archival datasets. Handley and Gray (2013) note that the connection between plant inspections and future quality outcomes is still under-investigated. This dearth of empirical research may be due to the managerial ambivalence towards plant inspections as an effective means to control quality at the upstream suppliers. In determining whether managers prefer inspections at the upstream supplier or inspections of the incoming material, Mayer et al., (2004) summarized managerial belief as "inspections (of the plant) do not guarantee delivery of the desired quality level but such inspections…increase the probability that inputs will meet the buyer's desired quality level." Using primary data collected with a survey, Handley and Gray (2013) examined the relationship between plant audits and perceived relative plant quality, and found that the use of plant inspections of a supplier's facility by the purchasing firm increases the supplier's perceived relative importance of quality.

The quality outcome of a process is a function of the underlying effectiveness of its quality management system; weak quality systems will increase the risk that poor quality product reaches the customer, whereas effective quality systems will prevent potential conformance problems and root them out before the product ships to the market. FDA plant inspections intentionally assess the effectiveness of the underlying quality management system. Intuitively, we expect that if plant inspection outcomes contain information, adverse inspection outcomes relate to a high hazard of a future recall at a plant, whereas favorable inspection outcomes indicate a low hazard of a recall. Importantly, we expect that this relationship is not causal. Inspection outcomes measure the effectiveness of the underlying quality system, but adverse inspection outcomes do not cause recalls. While a causal relationship between adverse inspection outcomes and closely following recalls may exist for a small minority of recalls which occur during or immediately following an inspection, our analysis will explicitly exclude recalls of this type, and examine only recalls that are distal from an inspection. We thus hypothesize:

> HYPOTHESIS 1*: FDA inspections contain information about future recalls such that favorable inspection outcomes are associated with a lower recall hazard than adverse inspection outcomes*.

From a counter perspective, FDA investigators who are faced with an adverse inspection outcome generate quality system improvement recommendations. Assuming that these recommendations are valuable, and that plants implement them, the future quality state of the plant should improve after such an outcome. This line of reasoning, a view espoused within the FDA, could lead to the expectation that an adverse inspection outcome actually lowers the future recall hazard. While this argument contains a kernel of truth, it is incomplete, and does not alter the logic underlying Hypothesis 1. Improvement processes take time to lower the recall risk of a weak quality management system. There is also considerable evidence to show that improvement suggestions made from parties outside the firm are not always implemented (Muthunalingam et al., 2013; Dhanorkar, et al., 2014). Further, using data from the pharmaceutical industry, Anand et al. (2012) show that while plants improve following an adverse inspection outcome, receiving a similar score in a future inspection remains highly likely. Macher et al. (2011) find similar results. As suggested by Anand et al. (2012), organizational inertia (Hannan and Freeman, 1984; Nelson and Winter, 1982) may prevent firms from immediately making significant changes to a plant's quality system. There is, however, an important relationship between inspection outcomes and paths to

improvement, which we illustrate in a stylized way in Figure 2.1.

Figure 2.1 High and Low Inspection Information

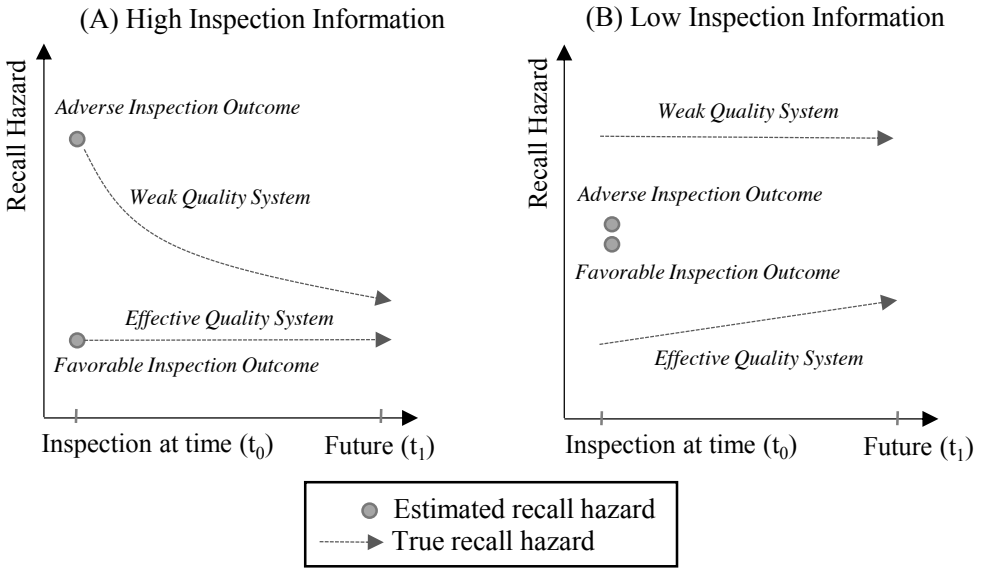


Figure 2.1A presents the evolution of a quality management system after receiving an informative inspection. The true recall hazard associated with effective and weak quality management systems is shown by the dotted lines. If the inspection is informative, its outcome measures the effectiveness of the quality management system, and thus the estimated recall hazard is similar to the true recall hazard. This correspondence is indicated by the dots in the figure at time t_0 lying on the dotted lines of the associated quality management system hazards. In the case of a favorable inspection outcome, no improvement process is kicked off by the inspection; the recall hazard of effective quality management systems is unaltered by the inspection. In the case of an adverse inspection outcome, the feedback from the inspection outcome will trigger an improvement process. For the reasons discussed in the previous paragraph, this improvement process will be incomplete and take time, but it ultimately leads to a reduction in the recall hazard of weak quality systems. Overall, the recall hazard of weak quality systems remains above the recall hazard of effective quality systems for a foreseeable period of time.

Figure 2.1B depicts the evolution of a quality system after receiving an uninformative

inspection. The estimated recall hazards of adverse and favorable inspection outcomes are tightly clustered together, and far from the true recall hazards of weak/effective quality management systems (i.e. the dots at time t_0 are far from the dotted lines). Due to the low information content of the inspection outcome, no improvement processes are triggered. If the inspection outcome is adverse, the feedback from the inspector is unlikely to lead to a real path for change, even if the underlying quality system is weak. Effective quality management systems may even deteriorate as managers may focus their attention on irrelevant tasks, fixing unbroken processes.

In summary, quality system changes following adverse inspection outcomes, for both informative and uninformative inspections, are unlikely to offset the logic underlying Hypothesis 1. However, one can see from Figure 2.1 that informative inspections lead to a much lower hazard over time than uninformative inspections, since they trigger valid improvement processes and avoid the unnecessary diversion of managerial attention. We further develop this idea in Section 2.5.

2.4 Investigator Experience and Inspection Information

Hypothesis 1 relies on the concept that plant quality inspections contain information. Whether there is a link between plant inspection outcomes and future recalls therefore depends on the degree to which investigators adequately assess the state of the quality management system of a site. Assessing the investigator efficacy at doing so is therefore paramount to better understand the relationship between plant inspections and recalls. Past FDA inspection research identifies investigator heterogeneity as an important factor in understanding plant inspection outcomes (Macher et al., 2011) and notes investigator experience as a main driver of this heterogeneity. In practice, websites have appeared that provide information on investigator characteristics and experience (www.fdazilla.com), further supporting the idea that attributes of the investigator matter and influence the inspection outcome.

We distinguish between general and site-specific investigator experience. This distinction follows Huckman and Pisano (2006) who found that cardiac surgeons who perform surgeries at one hospital had significantly higher performance (lower patient mortality) compared to surgeons who split their time between hospitals. They posited that the improved performance from hospital-specific experience results from the surgeon's familiarity with the focal hospital routines, resources and surgical team members. In the current study, FDA investigators can acquire similar levels of experience by inspecting one plant several times (specific experience) or inspecting multiple

different plants (general experience).

We further differentiate between two effects of experience: learning and complacency. That experience with a task leads to learning is a well-established phenomenon (Wright 1936), and is formalized under the *learning curve* umbrella: experience improves individual, team, and organizational performance (Argote, 2013). This phenomenon has received significant empirical support in diverse contexts. Nonetheless, empirical results from the financial auditing, regulatory and safety inspection literature suggest that learning can be dominated by complacency (Macher et al., 2011; Moore et al., 2006; Lemley and Sampat, 2012).

Learning and the Upside of Investigator Experience

The learning effects of site-specific experience observed by Huckman and Pisano (2006) are related to hospital specific familiarity, allowing the surgeon to understand specific personnel, activities, expectations, and practices that facilitate a more effective surgical procedure. This positive association between site-specific experience and performance is also seen in professional athletic teams (Berman et al., 2002), flight deck teams on aircraft carriers (Weick and Roberts, 1993), financial analysis (Groysberg et al., 2008) and software development (Huckman et al., 2009). In the literature on work teams, this phenomenon is often captured under "transactive memory systems", where individual group members become familiar with 'who knows what' within a team, making groups more effective (Lewis et al., 2005). In summary, a certain amount of tacit knowledge is honed as specific experience increases, which in turn impacts performance at the site of that experience. In our context of FDA inspections, the more familiar the investigator is with the facility, the more intimate they become with processes, products, and people which may enable a more robust ability to capture plant inadequacies and quality risks. This effect would lead to a more informative inspection outcome.

> HYPOTHESIS 2A*: Specific inspection experience increases the information content of inspections such that inspection outcomes are more predictive of future recalls with increasing specific experience*

General experience in performing plant inspections may have similar effects, though for different reasons. Costs, completion time, and quality, among other key measures, improve with repetition in manufacturing (Wright, 1936; Zimmerman, 1982; Hirschmann, 1964) healthcare (Kelsey et al., 1984; Hannan et al., 1991; Reagans et al., 2005) and other service industries (Boone

et al., 2008; Darr et al., 1995). General experience generates learning (Yelle, 1979; Argote and Epple, 1990), which enhances task outcomes. FDA investigators who develop extensive general experience are likely to observe many different plants over-time, which represent not only "best practices" when it comes to quality processes, but also "worst offenders". This experience in observing multiple plants at various stages of the quality performance spectrum may enable the investigator to more accurately identify problems at a plant and allow their inspection results to be more predictive. Further, being exposed to different plants represents a limited form of task variety for the inspector. While processes in the industry are to some degree standardized due to the regulatory nature, differences between plants do exist, and such limited variety can be beneficial for overall learning outcomes (Narayanan et al., 2009). We thus propose the following hypothesis:

> HYPOTHESIS 3A*: General inspection experience increases the information content of inspections such that inspection outcomes are more predictive of future recalls with increasing general experience*

Complacency and the Downside of Investigator Experience

Evidence for learning through experience primarily exists in the context of individuals, teams or organizations executing some specific task. Contextually, learning may not be as dominant within an auditing or inspection setting such as ours. Prior literature identifies possible negative aspects of experience in financial audits (Moore et al., 2006; Deis and Giroux, 1992; Carey and Simnett, 2006; Davis et al., 2009), plant safety audits (Short et al., 2013), and regulatory audits (Salant, 1995).

The financial auditing literature reveals that a primary source of complacency and reduced inspection objectivity is the structure of the inspection scheme; investigators are paid by the firms they audit (Moore et al., 2006). Demonstrating a reduction in inspection accuracy and objectivity with more specific experience, in the absence of financial incentives, may explain the presence of this phenomenon in other settings and reveal an important additional mechanism underlying it. In supply chain plant audits, when the auditing party has no financial incentive to bias their score towards high quality, complacency and reduced inspection performance may arise from the stale and routine nature of inspecting the same location repeatedly, from investigator fatigue or lack of interest, or from developing overly friendly relationships with auditees.

Short et al. (2013) use plant audit data to demonstrate that when an audit team includes members who have visited the factory in the past, the team is less likely to issue violations. The

authors argue that this effect may be caused not only by undue familiarity, but also by cognitive constraints. Auditors are imbued with a certain amount of tacit knowledge that bounds their ability to identify violations not in their experience history. Once an auditor has been through a facility and identified violations based on their area of expertise, they are less likely to find additional violations on return visits because they will continue to be bounded by their limited set of experiences. Repeat inspections may reduce the objectivity and information content of the inspection outcome.

"Regulatory capture" could be another possible explanation for degradation in inspection quality as investigator specific experience increases (Stigler, 1971; Salant, 1995). A stream of theoretical (and limited empirical) research signals a possible risk of an employment "revolving door" which exists between industry and government. This may lead to less objective inspections, as a regulator gains experience at a site, she/he may be incentivized to "go easier" on a site with the hope of future employment.

In sum, there is reasonable evidence to suggest that increased specific experience may hinder, not help, an investigator and reduce the information contained in the inspection outcome.

> HYPOTHESIS 2B*: Specific inspection experience decreases the information content of inspections such that inspection outcomes are less predictive of future recalls with increasing specific experience*

The previous arguments established that site-specific experience can have a downside by creating undue familiarity and complacency on the side of the investigator. Does general experience have a similar downside as well? While the effects of general experience in the inspection process are under-investigated, past studies point to such a downside. Bardach and Kagan (2002) provide a stark contrast between regulatory investigators who have general experience and those who do not. They show that investigators who are new to their role conceal their lack of self-confidence with strict legal adherence to the letter of the law, and become, potentially, more "hard-line" or accurate investigators. Conversely, experienced investigators may rely on their background and their previous inspections to apply personal filters to what they may deem acceptable, though it may not meet the true requirements.

Past research on FDA investigator heterogeneity has also identified general experience as a factor that affects performance. Macher et al. (2011) found that as FDA investigators gained more general experience, they were less likely to give a plant an OAI inspection outcome. The authors

measured experience as the total number of inspections at all plants completed by individual investigators. The authors do not assign a reason for this result. However, in light of the above literature, one possibility is that investigators slip into a complacent state of low performance after general experience increases, witnessing numerous plants, some of which operate on the extreme fringes of acceptable levels of quality. An experienced investigator may overlook more common, yet still important, quality system gaps. This would cause their inspection results to become less accurate, and less predictive of future quality.

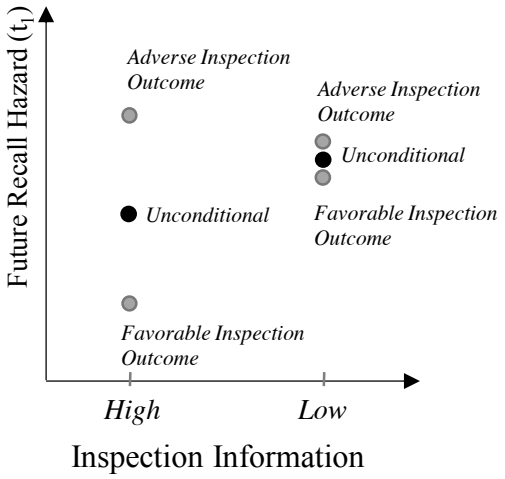
> HYPOTHESIS 3B*: General inspection experience decreases the information content of inspections such that inspection outcomes are less predictive of future recalls with increasing general experience*

2.5 Investigator Experience and Recalls

Hypothesis 1 is built on the idea that inspection outcomes contain information about the true quality state of a plant, and Hypotheses 2-3 refine this idea by suggesting that the amount of information contained in the inspection outcome depends on the experience of the investigator. As described at the end of Section 2.3, more informative inspections are better, not only for the regulator and the public, but also for the participating plant, since such information would allow a plant to react, improve their processes in time and possibly prevent future recalls.

Note that less informative inspections may lead to type I *and* type II errors which are both detrimental for quality performance. Investigators providing a plant with a 'clean bill of health' even though deep quality problems exist, point to missed opportunities for learning and improvement. If an investigator points out problems that either do not exist or have no substantive quality implications, they require plant managers to needlessly channel their effort in addressing them. Such a prioritization will crowd out their regular effort and attention on maintaining and improving their processes. In that sense, both types of errors can lead to disruptions in future quality performance. We summarize our arguments in Figure 2.2.

Figure 2.2 Unconditional Recall Hazard at Time t_1



As explained earlier, the difference between informative and uninformative inspections is the degree to which adverse/favorable inspection outcomes can differentiate the recall hazard at a plant. Figure 2.2 depicts the recall hazards of adverse/favorable outcomes (grey dots) as far apart from the unconditional recall hazard (black dot) for informative inspections. In uninformative inspections, these conditional and unconditional hazards lie in close proximity to each other. Unconditional on the outcome of an inspection, the recall hazard of a plant should be higher if inspections are not informative than if they are informative. The dotted line of the recall hazard of each type of quality management system in Figure 2.1A under high inspection information is below its counterpart under low inspection information at future time t_1 in Figure 2.1B. In other words, unconditional on the inspection outcome, the recall hazard should be lower after the inspection for an informative inspection compared to an uninformative inspection. This is shown in Figure 2.2 by the black dot under high inspection information lying below the black dot for low inspection information. More informative inspections are thus beneficial in terms of reducing the hazard of recalls.

The logical implication is that there should be a direct and causal relationship between investigator experience and future recalls. If the upside of investigator experience outweighs the downside, then specific and/or general experience among investigators should lead to more informative inspections, and a reduced hazard of future recalls, unconditional on the inspection outcome. This is, in essence, the assumption under which the FDA currently operates, and a key

reason why investigators are not intentionally rotated among different plants. If, however, the downside of investigator experience outweighs the upside, this assumption would be flawed, and investigator experience would lead to less informative inspections and would be associated with an increased hazard of recalls. In the case of specific experience, such a relationship would imply that a rotation policy might be advantageous. In the case of general experience, a relationship between experience and inspection outcomes would imply that investigators should not stay on their job for too long. We therefore test this assumption explicitly by formulating the following competing hypotheses:

HYPOTHESIS 4A*: Specific inspection experience leads to a low hazard of future recalls*.

HYPOTHESIS 4B*: Specific inspection experience leads to a high hazard of future recalls*.

HYPOTHESIS 5A*: General inspection experience leads to a low hazard of future recalls*.

HYPOTHESIS 5B*: General inspection experience leads to a high hazard of future recalls*.

Clearly Hypotheses 2-5 are formulated in a competing fashion. When there is significant support for two different relationships based upon well-established theory, competing hypotheses are not uncommon (see e.g. Jain et al., 2014). Do the positive aspects of specific experience outweigh the possible downside? Do the benefits of general experience weigh more heavily than the drawbacks? It is conceivable that the resulting effect of specific and general experience on recalls is non-linear, such that while initially learning effects prevail, complacency eventually takes over. However, it is also conceivable that the upside generally outweighs the downside (or vice versa).

While either one of our competing hypotheses is plausible, the specific context we study leads us to believe that the learning effects of experience may not be as salient as in other contexts. The FDA has more than 75 years of experience in conducting inspections; inspection routines are well documented and standardized through guidebooks and training programs for inspectors. The organization thus has accumulated a wealth of organizational knowledge and the potential learning from each additional inspection is marginal. It has also been shown that individual learning may simply be less important if the organization is highly experienced in the task at hand (Reagans et al., 2005). Additionally, while it is possible that a minority of inspections are carried out in teams, a vast majority of the inspections in our context has a single investigator. Thus, arguments of site-

specific team learning hold less sway. Further, given the regulatory nature of the industry, production processes across sites are relatively more standardized compared to other industries, making the occurrence of site-specific tacit knowledge less likely. Thus, our learning hypotheses 2A, 3A, 4A & 5A may be less plausible than our complacency hypothesis 2B, 3B, 4B & 5B. Additionally, past experience literature has demonstrated that general experience in similar "knowledge worker" settings (e.g., healthcare) has significantly less of a performance effect than specific experience (Huckman and Pisano, 2006; Huckman and Zinner, 2008). This leads us to believe that our specific experience hypotheses (2 & 4) may be more likely than our general experience hypotheses (3 & 5). Taken together, these arguments would suggest that our specific experience complacency hypotheses (2B and 4B) are more likely than our general experience learning or complacency hypotheses (2A, 3A-B, 4A, 5A-B).

It is important to emphasize that unlike Hypotheses 1-3, Hypotheses 4-5 are causal in nature. Our first three hypotheses were built on the idea that inspection outcomes simply measure the quality management system effectiveness, and thus predict future recalls without causing them directly. Our last two hypotheses postulate that based on the information contained in the inspection, inspections serve as a feedback mechanism to the organization triggering improvement processes. This creates a causal link between experience (which influences the information content of inspections) and future recalls (which are affected by the resulting improvement processes).

2.6 Research Design, Data Sources, and Variables

We investigate the relationship between FDA plant inspections and recall occurrences for medical devices in the seven years covering 2000-2006. The year 2000 was set as the starting point since recall records were not available in digital form in prior years. The year 2006 was set as the ending point since we only had identifiable plant inspection data until that point. Recalls, FDA inspection data, and plant level controls for our study come from multiple Freedom of Information Act (FOIA) requests made to the FDA. The unit of analysis for our research is the plant, while the units of observation are inspections and recalls which occurred at the plant. In practice, recalls actually occur at the product level. However, FDA inspections occur at the plant level and address the state of the quality system at that plant, which affects all products produced there. To link recalls to plants, we used a plant identifier reported by the FDA in both the recall and the inspection data sets. It is important to note that in our sample, there were no instances where a recalled product was

built in two different plants.

Dependent Variable

Recalls. The dependent variable for our research is a product recall for medical devices manufactured and sold in the U.S. between 2000 and 2006. The FDA's recall database includes the plant identifier, plant location, recall initiation date, product name, recall number, and the recall class. The recall class indicates the severity of the problem and ranges from class I (most severe) to class III (least severe). In our time frame, there are a total to 2,863 recalls (80 class I; 2,151 class II; and 632 class III). We used only unique recalls in all our analyses. In instances when one recall spawns several other related recalls in the same organization, we used the FDA recall number identification scheme to identify and remove non-unique recalls.

We made two exclusions within our sample for the main analysis: first, we exclude Class III recalls, as they represent significantly less serious quality issues (for instance labeling mistakes). Thus, our main analysis includes only Class I and II recalls. Second, we also exclude recalls that occur within a two-week window of the inspection closure date. Our goal is to estimate the hazard of a recall *following* an inspection, but sometimes, plant management may recall a product during or immediately after an FDA inspection as a gesture of good-will towards the FDA. Experts in the medical device industry also recommended to us that recalls which occur during or immediately following an FDA inspection should be excluded because these capture the state of quality at the plant *during* an inspection, and may not represent the true state of relationship between inspection outcome and future recall hazard. We reanalyze our models without these exclusions in our robustness checks.

Independent Variables

Inspection outcome. FDA inspection data includes the business unit name, location, date of the inspection, investigator ID number, type of inspection and the inspection outcome. Inspection outcomes range from the highest level of quality and compliance (No Action Indicated-NAI) to the lowest (Official Action Indicate-OAI) with a moderate score in between (Voluntary Action Indicated-VAI). Indicator variables (*No Action, Voluntary Action*) were used to measure inspection outcome in our dataset with the reference category in our model being the worst outcome (OAI). Our data contain 4,767 inspections (2,300 NAI; 1,815 VAI; and 652 OAI) covering 2,244 unique U.S. plants inspected between 2000-2006. The summary statistics associated with the number of

investigators, inspections, and the number of inspections performed by each investigator are provided in Table 2.1.

Specific experience. We measured investigator specific experience by counting the number of inspections that an investigator has performed at each plant at the time of the current inspection. We include the current inspection in the history count so that the first time an investigator inspects a facility, his/her specific experience variable equals one. The mean specific inspection experience was 1.4. While our recall and inspection outcome data cover the years 2000-2006, we obtained additional FDA data to strengthen our experience measures. Our specific and general experience measures cover all U.S. medical device inspections from 1994-2006. Though we were not able to obtain inspection history data prior to 1994, we demonstrate through robustness checks that this time window is sufficient to capture investigator experience.

General experience. Similarly, we measured general experience by counting the number of inspections performed by each investigator at all other plants at the time of the current inspection. The mean general inspection experience was 22.6. Note that we decided not to include the specific experience at a focal site in the general experience count. This choice was made to more clearly differentiate between the two concepts clearly. We tested whether an alternative scoring mechanism influences our results (by including specific experience in the general experience measure) and did not find any noteworthy differences.

Control Variables

Sales: Plant level sales values were provided by the FDA. The FDA collects annual sales from plants and groups them into ten separate categories, increasing in regular increments beginning with $0-$25,000 and ending with $50,000,000 and higher. We created indicator variables to represent each of these categories and included them as control variables. There were limited observations in the first three categories of sales ($0-$25,000; $25,000-$50,000; $50,000-$100,000), so the first three categories were grouped together to serve as the referent category.

Firm. There are 2,244 unique plants owned by 2,130 unique firms. There are 66 firms that have more than one plant in the data. To control for heterogeneity resulting from shared firms across plants, we created indicator variables for all firms which have multiple plants. Firm indicator variables are included in the analysis.

Public. An indicator variable signifies if the plant is part of a public or private firm (*Public*). A publicly traded firm may be privy to higher skilled personnel and a larger set of plant resources that

may reduce the hazard of a recall.

Percent College Degree. In our sample, manufacturing plants are dispersed throughout the United States, and certain regions have higher skilled and more educated labor than other regions. To control for the impact of geographic heterogeneity in educated labor pool availability on recall hazard, we include the percent of the population with a college degree in the zip code of the plant *(% College Degree*). We used U.S. Census data to obtain the percent of the population that have college degrees in the plant zip code.

Past Recalls. Past recalls could be indicative of future recalls. To control for the potential bias in our estimates, we computed a measure counting all prior recalls at the plant for the past three years (*Recalls (last 3 years))*. As recall data is unavailable prior to 1999, the first two years of the data (2000 and 2001) only include a one and two year window respectively. All future years in the panel (2002-2006) include a three year aggregated moving count. We also analyzed models which included a one year moving count and the results are substantively similar.

Inspection type. Our data contains all three types of inspections: surveillance, compliance and complaint inspections, which represent different levels of severity. It is possible that a complaint inspection, which is more serious, may lead to more frequent recalls. To control for this effect, an indicator variable was used to signify the type of inspection (*Surveillance*, *Compliance*). We used complaint inspections as the reference category. There are 3,536 surveillance, 1,149 compliance, and 82 complaint inspections, resulting in a total of 4,767 inspections.

Year. It is possible that FDA policies related to recall enforcement and inspection processes may change over time; becoming more or less strict in enforcing federal regulations. To control for this possibility, we added indicator variables for the year of the inspection as control variables.

Number of Products at the Plant. Plants may vary greatly by the number of products they produce. Some plants manufacture many products, while others focus on few products. It is possible that complexity ensuing from more products produced at the plant would increase the hazard of a recall. To control for this, we collected the number of products manufactured at each plant from the FDA, computed its natural log (*Ln_Number of Products*), and included it as a control variable in the analysis.

Technology at the plant. Plants may also vary in the level of technology of the products they produce. While some plants manufacture high technology products, others produce simpler devices. It is possible that plants that make technologically advanced products are more susceptible to product failures and recalls. We capture the class of products built at the plant to control for the

technology level at the plant. Medical devices are categorized into three groups (class 1, class 2, class 3) on the basis of product technology, where class 3 represents the highest technology, and class 1, the lowest technology. FDA provided us with the highest class of products built in each plant in our panel. We use this data to create three indicator variables (*Highest Class 1, Highest Class 2*, *Highest Class 3*) and treat *Highest Class 1* as the referent category.

Novelty at the plant. The novelty of a device may be associated with the likelihood of a recall. For instance, a new implantable device used to treat a previously untreated disease may experience more unexpected failures than a device that has had years of usage in the market. Highly novel devices require pre-market approval (clinical trial) while low novel devices only require demonstration of similarity to prior approved products in the market. The most novel classification of a medical device is termed a pre-market approval (PMA) device, while the least novel is a 510K Exempt (510KE). 510K is a category of medical devices which only need to demonstrate similarity to prior approved products to be approved for sale in the U.S. 510K Exempt devices are those that are of such low novelty that they do not require any prior notification by the FDA prior to marketing the device. We control for the level of novelty of product built at the plant by measuring the highest submission type of products at the plant (*Highest Submission PMA, Highest Submission 510K, Highest Submission 510KE*) and treat *Highest Submission 510KE* as the referent category. This data was also provided by the FDA.

Table 2.1 Inspection Descriptive Statistics by Year

Year	2000	2001	2002	2003	2004	2005	2006
# Investigators	238	244	226	277	239	227	244
# Inspections	426	629	639	776	759	748	790
# Plants inspected	457	654	695	831	814	799	859
Mean, Inspections per investigator	1.79	2.58	2.83	2.8	3.18	3.3	3.24
Min, Inspections per investigator	1	1	1	1	1	1	1
Max, Inspections per investigator	19	18	25	19	21	18	16

2.7 Empirical Strategy

In selecting an empirical strategy, we need to take into account the unique characteristics of our data and contextual setting. Our data consists of multiple plants that have undergone one or

more inspections over a seven-year period from 2000 to 2006. To be included in the data, a plant required at least one inspection, but not necessarily any recalls, though many plants had multiple recalls during this time period. Our research objective is to assess the hazard of a recall for a manufacturing plant following an FDA inspection while taking into account the characteristics of the previous inspection. We chose survival modeling because by using the time at which an event occurs as the dependent variable, survival models estimate the relative hazard of failure (i.e. recall) for each unit based upon different levels of covariates (inspection outcome and investigator experience), which change over time.

Within the general class of survival models, Cox-Proportional Hazard models are most frequently used because they do not require researchers to specify the underlying distributional form of the hazard, making them more flexible and less susceptible to distributional misspecification (Box-Steffensmeier and Jones, 2004). However, a critical assumption in using Cox-Proportional Hazard (CPH) model is that the hazard ratios must be constant over time. We tested for this assumption by comparing Kaplan-Meier observed survival curves with predicted Cox curves and found them close and parallel (Garrett, 1998). Quantitatively, we examined the statistical significance of Schoenfeld residuals and found that the residuals were not significant (all p-values > 0.10), indicating that hazard ratios are constant over time.

As the plants in our data experience recurrent events in the form of multiple inspections and multiple recalls, we need more specialized forms of CPH models, which can accommodate recurrent events. There are two groups of recurrent event CPH models, the shared frailty and variance-corrected models. Shared frailty models estimate the parametric distribution that account for the unobserved heterogeneity within a given plant, but the estimates are conditional upon the chosen distribution of this heterogeneity. Variance-corrected models are not constrained by a specific distribution, but are still able to correct for the shared variance that will exist when a plant has multiple events.

The next decision was to select the appropriate variance-corrected CPH model from the three frequently used options: Anderson-Gill, Marginal, and Conditional gap-time (Box-Steffensmeier and Jones, 2004). The Anderson-Gill and Marginal models both cluster on the plant to control for heterogeneity, but analyze the hazard of an event based on the initial time when the plant entered the data and not on intermittent events that occur sporadically throughout time. Conditional gap-time models fit our data well as they are able to determine the hazard of one type of event (recall) following the most previous occurrence of another event (inspection).

We use a conditional gap-time Cox Proportional Hazard model to determine the hazard of each plant experiencing a recall following an FDA inspection, which can have three distinct outcomes (NAI, VAI, and OAI) and can be conducted by investigators who vary in their levels of specific and general experience. In a CPH model, a positive beta coefficient signifies that the time to failure is lessened by the covariate and is interpreted as an increased hazard. Conversely, a negative beta coefficient indicates a longer time to failure and a decreased hazard. CPH beta coefficients are interpreted multiplicatively after exponentiation.

2.8 Results

Descriptive statistics and correlations are presented in Table 2.2. We observe that both OAI and VAI are positively and significantly correlated with recalls whereas NAI is negatively correlated with them. Additionally, while specific experience is positively related to recalls, general experience has a negative correlation.

The results of the CPH regression analysis are reported in Table 2.3. We include all our control variables, the dummy variables for the types of inspection, and the inspection outcome in column 1 (sales and firm dummies included but not shown). We then include the linear and quadratic effects of specific and general experience to capture potential main and marginal effects of experience respectively (columns 3 and 5). Finally, we include the four interaction terms between the two experience types and the two inspection outcomes (column 7). Several of our control variables are significant. A plant belonging to a publicly traded firm has a greater product recall hazard. Not surprisingly, a plant with past recall experience has a greater hazard of future recalls. The more products built at a plant, the more likely that plant is to experience a future recall.

To test Hypothesis 1, we need to establish that favorable inspection outcomes (e.g., NAI or VAI) are associated with a lower recall hazard than adverse inspection outcomes (OAI). Our results show that beta coefficients for both NAI and VAI are statistically significant and negative, suggesting that a plant receiving either an NAI or a VAI has a reduced recall hazard. Thus, we fail to reject Hypothesis 1. Compared to a plant which receives an OAI, the recall hazard decreases by 34.3% ($e^{-0.42}-1 = -0.343$) when a plant receives an NAI score and by 22.1% ($e^{-0.25}-1 = -0.221$) for VAI.

Table 2.2 Descriptive Statistics and Correlation Matrix

		Mean	Min	Max	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	Recalls[a]	1.28	0	92	20.10	1.00													
2	Public	0.38	0	1	0.49	0.32*	1.00												
3	% Coll Deg	36.95	0	88.70	17.14	0.07*	0.00	1.00											
4	Recall_3yrs	9.02	0	115	18.50	0.56*	0.52*	0.04*	1.00										
5	# Prod	50.30	1	3124	1.64	0.35*	0.30*	0.07*	0.43*	1.00									
6	High Cl 3	0.32	0	1	0.47	0.15*	0.24*	0.08*	0.24*	0.36*	1.00								
7	High Cl 2	0.58	0	1	0.49	-0.08*	-0.17*	-0.03*	-0.14*	-0.07*	-0.80*	1.00							
8	High Cl 1	0.10	0	1	0.24	-0.10*	-0.07*	-0.10*	-0.12*	-0.31*	-0.18*	-0.30*	1.00						
9	PMA	0.24	0	1	0.42	0.08*	0.19*	0.03*	0.18*	0.27*	0.81*	-0.65*	-0.14*	1.00					
10	510K	0.63	0	1	0.48	0.02	-0.08*	0.04*	-0.03*	0.11*	-0.53*	0.78*	-0.31*	-0.73*	1.00				
11	510Ke	0.13	0	1	0.29	-0.13*	-0.11*	-0.11*	-0.16*	-0.34*	-0.21*	-0.14*	0.76*	-0.18*	-0.42*	1.00			
12	Surveill.	0.70	0	1	0.46	-0.13*	-0.07*	0.07*	-0.13*	-0.01	-0.08*	0.08*	0.00	-0.07*	0.07*	-0.01	1.00		
13	Compliance	0.29	0	1	0.45	0.13*	0.07*	-0.06*	0.12*	0.00	0.08*	-0.09*	0.01	0.07*	-0.07*	0.01	-0.96*	1.00	
14	Complaint	0.01	0	1	0.13	0.00	0.01	-0.01	0.03*	0.04*	-0.01	0.03*	-0.02	0.00	0.00	0.00	-0.20*	-0.08*	1.00
15	Official Act	0.16	0	1	0.37	0.09*	0.01	-0.03*	0.09*	0.03*	-0.04*	0.04*	0.00	-0.06*	0.04*	0.01	-0.23*	0.23*	0.01
16	Volun Act	0.39	0	1	0.49	0.03*	0.00	0.03*	0.00	-0.03*	-0.02*	0.04*	-0.03*	-0.05*	0.05*	-0.02	0.02	-0.01	-0.02
17	No Act	0.45	0	1	0.50	-0.09*	-0.01	-0.01	-0.07*	0.01	0.06*	-0.07*	0.03*	0.09*	-0.08*	0.01	0.16*	-0.16*	0.01
18	Spec Exp	1.41	0	16	1.03	0.10*	0.10*	-0.41*	0.17*	0.13*	0.16*	-0.11*	-0.06*	0.20*	-0.11*	-0.06*	-0.12*	0.12*	0.00
19	Gen Exp	22.63	0	147	23.40	-0.05*	0.06*	0.07*	0.03*	0.03*	0.07*	0.04*	-0.04*	0.06*	-0.02	-0.03*	0.11*	-0.12*	0.02

		15	16	17	18	19
15	Official Act	1.00				
16	Volun Act	-0.35*	1.00			
17	No Act	-0.39*	-0.72*	1.00		
18	Spec Exp	-0.02	-0.02	0.04*	1.00	
19	Gen Exp	-0.09*	-0.04*	0.11*	0.27*	1.00

[a] Recalls per plant

The two interaction terms between specific experience and inspection outcome are positive and statistically significant (*Specific experience x Voluntary Action*, *Specific experience x No Action*), while the two interaction terms between general experience and inspection outcome are not significant. Thus, our results do not lend support to Hypotheses 2A, 3A and 3B but we fail to reject Hypothesis 2B. Our results suggest that as specific experience increases, the relationship between a favorable inspection outcome (NAI or VAI) and a low recall hazard is weakened. In other words, increase in specific experience decreases information content, and does not improve the ability of an inspection outcome to predict a future recall. In fact, increasing specific experience worsens the relationship between the inspection outcome and the recall hazard for both NAI and VAI. An interaction plot (Figure 2.3), which demonstrates the hazard of a recall at each level of specific experience (contingent upon the inspection outcome and holding all other variables at their means) confirms that as specific experience increases the recall hazard related to each inspection outcome converges (and increases with experience). This implies that the information content of inspections is reduced with an increase in specific experience.

Specific experience also has a significant and positive main effect on the hazard of a recall. Every additional visit to a plant by an investigator increases the hazard of that plant experiencing a recall by 36.3% ($e^{0.31}$-1=0.363). We reject Hypothesis 4A but fail to reject Hypothesis 4B. We do not find a significant relationship between general experience and recall hazard (reject Hypotheses 5A and 5B). We included quadratic terms to explore the possibility of a non-linear relationship between investigator experience and future recall hazard (e.g., experience could initially increase, but eventually decrease recalls). The results show that the quadratic effect of specific experience on recall hazard is negative and significant, although effect size (0.01% impact on the recall hazard) is very small. We find that the general experience quadratic term is not significant.

In sum, favorable inspection outcomes result in a lower recall hazard (H1). The more specific experience of the investigator, the lower the information content of the inspection, and the reduced ability of the inspection to predict future recalls (H2B). Finally, the more specific experience of the investigator, the higher the recall hazard, regardless of the inspection outcome (H4B).

Robustness Checks

We conduct several robustness checks for our analyses. We first replicate the main results by including data that were excluded from the main analysis. Second, we verify the findings related to

specific experience by examining it in a more nuanced manner. Finally, we try to glean the exact level of specific experience at which complacency begins. It is important to note that as the unit of analysis in this study is either an inspection or a recall, the number of observations in robustness checks change based upon the dependent variable. For example, if class III recalls are included for a robustness check, the number of observations will be higher than in the main analysis. Detailed explanations for the sample size in each analysis are included as table footnotes.

First, we analyze our model by including the class III recalls and recalls that occur within two weeks of an inspection date. We also analyzed an additional model with just class II recalls. The results of the additional CPH models are substantively similar to the full model results (Table 2.4, columns 1 and 3). The three different inspection types (surveillance, compliance, and complaint) vary greatly in potential severity. FDA experts suggest that compliance and complaint inspections are more serious, whereas surveillance inspections are routine and occur more regularly. Our main analysis included all three inspection types, but it is possible that the relationship between inspection outcomes and recalls only exists for more serious inspection types but not for the routine inspections. To verify this, we analyzed our hypothesized relationships by limiting our data to surveillance inspections only (Table 2.4, column 5). All results are in Table 2.4 are similar to the main results presented in Table 2.3.

An important conclusion from our analyses is that specific experience has both a main effect upon recalls and an interaction effect with inspection outcomes upon recalls. However, there are at least two alternate explanations that could lead to this result. First, the level of specific experience could be endogenous to plant performance: that is, the FDA may be less likely to assign a new investigator to a troubled plant. To rule out this explanation, we examined the frequency distribution of NAI, VAI, and OAI outcomes given by investigators at different levels of their specific experience. We tested the percentage of OAI inspections occurring at each level of specific experience, and found no significant difference between them (results available upon request) and can conclude that the level of specific experience is not related to plant inspection outcome.

Table 2.3 Cox Regression Analysis-Hazard of a Recall[a]

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	β	SE	β	SE	β	SE	β	SE
Public	0.26*	(0.13)	0.25+	(0.13)	0.25+	(0.13)	0.25+	(0.13)
% College Degree	0.00	(0.00)	0.00	(0.00)	0.00	(0.00)	0.00	(0.00)
Recalls (Last 3 years)	0.02**	(0.00)	0.02**	(0.00)	0.02**	(0.00)	0.02**	(0.00)
Ln_Number of Products	0.21**	(0.05)	0.20**	(0.06)	0.20**	(0.06)	0.20**	(0.06)
Highest Class 3	-0.10	(0.42)	-0.10	(0.43)	-0.11	(0.43)	-0.09	(0.43)
Highest Class 2	-0.46	(0.41)	-0.45	(0.41)	-0.45	(0.41)	-0.45	(0.41)
Highest Submission PMA	0.20	(0.42)	0.16	(0.43)	0.16	(0.43)	0.14	(0.42)
Highest Submission 510K	0.34	(0.39)	0.34	(0.39)	0.34	(0.39)	0.34	(0.39)
Surveillance	-0.21	(0.21)	-0.20	(0.21)	-0.19	(0.20)	-0.19	(0.20)
Compliance	0.23	(0.22)	0.22	(0.21)	0.21	(0.21)	0.21	(0.21)
Voluntary Action	-0.24*	(0.10)	-0.22*	(0.11)	-0.22*	(0.11)	-0.25*	(0.11)
No Action	-0.42**	(0.09)	-0.40**	(0.09)	-0.39**	(0.10)	-0.42**	(0.09)
Specific experience			0.17**	(0.03)	0.26**	(0.05)	0.31**	(0.05)
General experience			-0.00	(0.00)	-0.00	(0.00)	-0.00	(0.00)
Specific experience^2					-0.01*	(0.01)	-0.03**	(0.01)
General experience^2					0.00	(0.00)	0.00	(0.00)
Specific exp×Voluntary Action							0.14+	(0.09)
Specific exp×No Action							0.27**	(0.09)
General exp×Voluntary Action							-0.00	(0.00)
General exp×No Action							-0.00	(0.01)
Observations	6,874		6,874		6,874		6,874	
Wald χ^2	1729.18		1828.83		1945.94		2100.09	

Standard errors in parentheses + p<0.1, * p<0.05, ** p<0.01. Note: Sales, Firm and Year dummies included but not shown

[a]Excludes recalls which occur within two weeks of inspections and class III recalls. There are 632 class III recalls that occur at all times, and 124 class I and II recalls which occur within 2-week window of the inspection. 7,630 – [632 class III recalls + 124 class I and II recalls] = 6,874 inspections and recalls

Table 2.4 Robustness Checks-Time and Recall Class Exclusions

	Class I, II & III Recalls No time gap[a]		Class II Recalls[b]		Surveillance inspections only[c]	
	(1)	(2)	(3)	(4)	(5)	(6)
	β	SE	β	SE	β	SE
Public	0.20	(0.13)	0.29*	(0.13)	0.22	(0.16)
% College Degree	0.00	(0.00)	0.00	(0.00)	0.01+	(0.00)
Recalls (last 3 years)	0.02**	(0.00)	0.02**	(0.00)	0.02**	(0.00)
2000	0.56**	(0.11)	0.43**	(0.13)	0.49**	(0.16)
2001	0.31**	(0.11)	0.22+	(0.13)	0.19	(0.18)
2002	0.28**	(0.11)	0.28*	(0.12)	0.33*	(0.15)
2003	0.11	(0.11)	0.12	(0.11)	0.09	(0.13)
2004	0.14	(0.10)	0.05	(0.11)	0.09	(0.15)
2005	0.06	(0.08)	0.01	(0.09)	0.15	(0.10)
Ln_Number of Products	0.19**	(0.05)	0.20**	(0.06)	0.23**	(0.06)
Highest Class 3	-0.00	(0.36)	0.06	(0.38)	0.05	(0.38)
Highest Class 2	-0.40	(0.34)	-0.33	(0.36)	-0.33	(0.35)
Highest Submission PMA	0.13	(0.33)	0.11	(0.35)	0.04	(0.34)
Highest Submission 510K	0.32	(0.30)	0.25	(0.31)	0.51+	(0.29)
Surveillance	-0.15	(0.18)	-0.21	(0.20)		
Compliance	0.25	(0.18)	0.14	(0.21)		
Voluntary Action	-0.23*	(0.11)	-0.28**	(0.11)	-0.19	(0.13)
No Action	-0.35**	(0.10)	-0.47**	(0.09)	-0.40**	(0.14)
Specific experience	0.27**	(0.05)	0.31**	(0.05)	0.34**	(0.07)
General experience	-0.00	(0.00)	-0.00	(0.00)	-0.01+	(0.00)
Specific experience^2	-0.02**	(0.01)	-0.03**	(0.01)	-0.02**	(0.01)
General experience^2	0.00	(0.00)	0.00	(0.00)	0.00	(0.00)
Specific experience×Voluntary Action	0.13+	(0.08)	0.12	(0.08)	0.18+	(0.10)
Specific experience×No Action	0.17*	(0.07)	0.21**	(0.07)	0.20*	(0.09)
General experience×Voluntary Action	-0.00	(0.00)	-0.00	(0.00)	0.00	(0.01)
General experience×No Action	-0.00	(0.00)	-0.00	(0.00)	0.00	(0.01)
Observations	7,630		6,918		5,221	
Wald χ^2	2,346.08		2,076.18		6,406.6	

Standard errors in parentheses + p<0.1, * p<0.05, ** p<0.01

[a] All observations in the data. Includes 4,767 inspections and 2,863 recalls. 4,767 + 2,863 = 7,630 inspections and recalls. [b] Excludes class I and class III recalls. 7,630 – [80 class I recalls + 632 class III recalls] = 6,918 inspections and recalls. [c] Excludes 2,409 compliance and complaint inspections and their associated subsequent recalls. 7,630 – 2,409 =5,221 inspections and recalls. Note: Sales and Firm dummies included but not shown

Table 2.5 Propensity Score Matching Results

Specification	Treatment group	Control group	Average treatment effect on the treated	Standard Error	t-statistic
Nearest neighbor	611	470	1.76	0.41	4.31**
Stratification	611	729	1.37	0.37	4.10**
Kernel matching	611	729	1.56	0.35	4.52**
Coarsened exact matching	611	729	2.04	0.45	4.50**

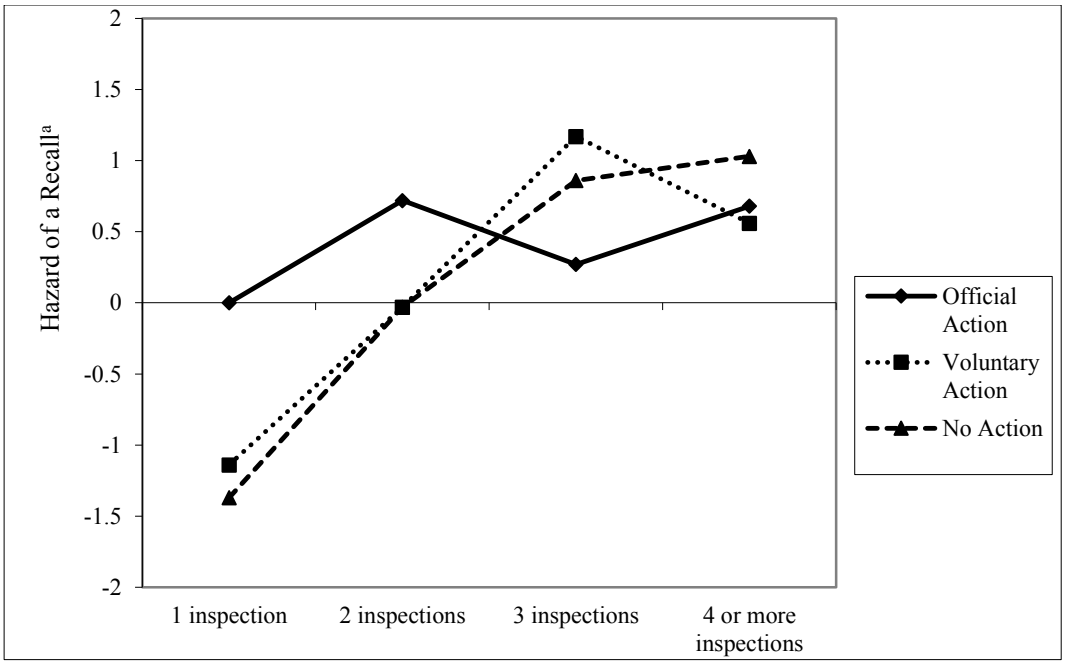
Standard errors in parentheses + p<0.1, * p<0.05, ** p<0.01

Table 2.6 Levels of Specific Experience

	(1) β	(2) SE	(3) β	(4) SE
Public	0.26*	(0.13)	0.27*	(0.13)
% College Degree	0.00	(0.00)	0.00	(0.00)
Recalls (last 3 years)	0.02**	(0.00)	0.02**	(0.00)
2000	0.38**	(0.12)	0.40**	(0.12)
2001	0.14	(0.13)	0.17	(0.13)
2002	0.17	(0.12)	0.19	(0.12)
2003	0.00	(0.10)	0.04	(0.10)
2004	-0.00	(0.11)	0.01	(0.10)
2005	-0.01	(0.09)	0.01	(0.09)
Ln_Number of Products	0.17**	(0.06)	0.18**	(0.05)
Highest Class 3	-0.03	(0.43)	0.03	(0.43)
Highest Class 2	-0.37	(0.41)	-0.36	(0.41)
Highest Submission PMA	0.16	(0.43)	0.12	(0.43)
Highest Submission 510K	0.34	(0.40)	0.34	(0.40)
Surveillance	-0.28	(0.20)	-0.23	(0.20)
Compliance	0.07	(0.21)	0.13	(0.21)
Voluntary Action	-0.28**	(0.11)	-0.35**	(0.10)
No Action	-0.44**	(0.08)	-0.58**	(0.09)
General experience	-0.00+	(0.00)	-0.00*	(0.00)
Specific experience_2 inspections	0.37**	(0.10)	0.39**	(0.10)
Specific experience_3 inspections	0.34*	(0.16)	0.49**	(0.13)
Specific experience_4 or more inspections	0.71**	(0.15)	0.79**	(0.13)
Specific experience_2 inspections x Voluntary act.			-0.07	(0.30)
Specific experience_2 inspections x No action			0.16	(0.27)
Specific experience_3 inspections x Voluntary act.			1.03**	(0.30)
Specific experience 3 inspections x No action			0.95**	(0.28)
Specific experience 4 or more inspections x Voluntary act.			0.12	(0.29)
Specific experience_4 or more inspections x No action			0.82**	(0.22)
Observations	7,630		7,630	
Wald χ^2	2,144.55		2,192.01	

Standard errors in parentheses + p<0.1, * p<0.05, ** p<0.01 Note: Sales and Firm dummies included but not shown

Figure 2.3 Recall Hazard as a Function of Inspection Outcome and Specific Experience



[a] Hazard model beta coefficients

Second, our measure for specific experience could be materially incomplete, as the experience data begins in 1994 and not at the beginning of each investigator's career. Although the investigator experience data spans 13 years, beginning in 1994, there is a possibility that our results would be different if investigator experience prior to 1994 were included. We examined this in more detail. First, we found that an investigator remains in the dataset for an average of 2.96 years. This suggests that the 13-year duration we used may be sufficiently long and most investigators were likely to have performed their first inspection within the confines of the current data. To test our conjecture, we limited our sample to investigators who had no inspections in the first five years (1994-1998) of our experience data. Our implicit assumption is that absence of inspection data in the previous five years implies that these investigators were newly assigned to plant inspections after 1998, resulting in a more accurate experience measure. In case any of these investigators had inspection experience prior to 1994, a five-year gap (equivalent to a complete re-start in the inspection process) was used. The results from the restricted sample indicate that the significance and effect sizes for substantive variables of interest are robust (results available upon request).

In Figure 2.3, we observed that as investigators gain specific experience, the overall hazard of a recall increases regardless of the inspection outcome. According to the FDA, investigator assignment to manufacturing facilities is random in nature and is guided by the FDA's goal of minimizing setup costs for the investigator, rather than any specific strategic intent. Investigators may be rotated among plants when particular regional investigators are unavailable or in cases of investigator turnover. Since these reasons for rotating investigators appear to be exogenous (a particular investigator is sick, busy or chooses to retire from the inspection process), we can, to some degree, assess whether investigator rotation leads to less recalls. To assess whether recall rates differ between those plants that see regular inspection rotation and those that see the same investigator, we estimated average treatment effects for investigator rotation. To do that, we excluded all plants which had only one inspection in our data, and categorized the remaining plants into those that see perfect rotation (i.e. a different investigator every time) and the plants that do not (i.e. investigators repeat at least once). Following Brown et al. (2013) and Iacus et al. (2012), we used our control variables to create a matched sample and conducted a propensity score analysis using four different matching methods: *nearest neighbor*, *stratification*, *kernel matching*, *and coarsened exact matching* (Table 2.5). There are a total of 1,340 plants which had more than one inspection in our dataset, 729 of these plants saw a new investigator for each inspection, and 611 saw repeat investigators. The results show that irrespective of the matching method used, there is a significant increase in recalls per plant when the plant sees the same investigator (*average treatment effect on the treated* in Table 2.5). There is an average increase of 1.7 recalls per plant, within the seven-year panel, when the treatment effect is averaged across all four methods. Hence, rotating investigators reduces the number of recalls per plant by more than one.

In our main analysis, we measured specific experience as a continuous variable by counting the number of inspections that an investigator conducted at a particular plant. We scrutinize our data further to discern the precise point at which specific experience begins to have both a main effect on recalls and a moderating effect on the relationship between inspection outcome and recalls. We use an investigator's first inspection at a plant as the reference category, and code *Specific experience_2*, *Specific experience_3*, *Specific experience_4 or more inspections* as indicator variables which equal one if the inspection was the 2^{nd}, 3^{rd}, 4^{th} or greater inspection respectively and zero otherwise. We then repeated the analysis (Table 2.6). The results show that each indicator variable is positive and significant (column 1), implying that the main effect of specific experience on product recalls begins with an investigator's 2^{nd} inspection. Moreover, the effect sizes for the

indicator variables show that the impact continues to grow as the number of inspections increase. There is a 48% recall hazard increase on the second visit ($e^{0.39}$-1=0.48), and a 63% recall hazard increase on the third visit ($e^{0.49}$-1=0.63). We then included interaction terms for each level of specific experience and inspection outcomes (column 3). The two interaction terms for the 2nd inspection at a plant (*Specific experience_2)* and the inspection outcome are not statistically significant. In contrast, the interaction terms with *Specific experience_3, and Specific experience_4 or more inspections* are statistically significant. In other words, the effect of specific experience upon the information content of the inspection outcome does not begin until the 3rd inspection, and continues with the 4th inspection onward. More specifically, the main effect of specific experience begins at the 2nd inspection, but it is not until the 3rd inspection that specific experience begins to moderate the relationship between inspection outcome and the hazard of a recall. These results are also depicted in Figure 2.3.

It is also possible that inspection outcome mediates the relationship between investigator experience and recalls. In other words, experience may lead to a certain outcome, which may then predict the recall hazard. We tested mediation with a two-stage model, where investigator experience predicts inspection outcome and the resulting recall hazard from inspection outcome. We found no evidence of mediation (detailed results are available upon request).

2.9 Discussion and Implications

This study makes three important contributions to quality management and experience literature streams. First, we validate an essential theoretical, but empirically unproven assumption (Deming, 1982) that plant inspections predict future plant quality. Favorable FDA plant inspection outcomes (VAI and NAI) indicate a significantly lower hazard of a future product recall than adverse inspection outcomes (OAI). The results from our hazard analysis show that compared to a plant that receives an OAI score, a plant that receives an NAI score has a recall hazard decrease of 34.3%. Compared to a plant that receives an NAI, a plant that receives an OAI score has a recall hazard increase of 52.2%.[13]

Second, we demonstrate that complacency effects are a significant concern in inspection and auditing settings that may outweigh benefits of learning. By using actual objective measures of

[13] Determined by analyzing our main model using NAI as the referent category and examining the coefficient for OAI.

product performance, instead of future audit findings as past studies have done, this work removes alternative explanations from previous related accounting, safety audit, and FDA inspection research that has found fewer citations result from more investigator experience (Macher et al., 2011; Short et al., 2013; Deis and Giroux, 1992; Carey and Simnett, 2006). These studies use future audit findings as their dependent variable, and find that audit findings decrease with experience. This result could be rooted in quality improvement that occurs with investigator experience, or deterioration in inspection quality. In light of our findings, it is unlikely that fewer future citations are a result of investigator learning (and improved plant quality), but are more likely attributable to investigator complacency (and reduced inspection quality). Third, by studying these questions in this context, we demonstrate that the promise of future financial gains is not the only mechanism that can create investigator complacency. The stale, routine nature of the job, and the familiarity which comes from repeat visits to a site, can lead to complacency and lower the information contained in an inspection, even when the investigator has no clear incentive to "go easier" on an inspection site. In our discussions with industry managers and FDA personnel related to these results, we learned that it is not common for medical device FDA investigators to transition into industry and that regulatory capture would not be a likely source of complacency in this setting.

From a managerial perspective, our research provides two important insights. First, our findings indicate that future recalls may be on the horizon if careful attention is not paid to heightened quality control upon receiving an adverse inspection outcome. Second, and more broadly in the context of supply chain management, our research provides some indication that adding quality inspections into the manufacturing process, as in a supplier plant inspection, may serve to more effectively guarantee quality in the product. While product quality inspections in general may be viewed in research and practice as an archaic method to control quality, regulators of industries such as medical devices, pharmaceuticals, and automotives still mandate such steps. Given such a requirement, economically allocating product inspection resources is critical. Plant inspections, which provide an accurate indicator for future quality hazard, can lead to significant potential savings for the inspecting firm by allowing them to focus limited resources more precisely. If well-trained plant investigators identify a plant's operating processes as high quality, then the resources dedicated to inspecting products from that plant might be reduced. Conversely, if a plant inspection indicates a high quality risk, valuable inspection resources can be more effectively targeted at heightening the inspection frequency and intensity of associated products and working with suppliers to root out quality problems at their source. External stakeholders may also extract

crucial guidance from these findings. Regulators, managers, investors, and supply chain partners, looking to pinpoint plants that are operating at a high hazard of a future recall can leverage publically available FDA plant inspection data as an early warning signal. Medical doctors can also use this data to improve their patient care by more carefully scanning a product's origin, which could reveal a potentiality for becoming defective. Investors may be able to use this information to adjust stock price valuations early.

The association between inspection outcomes and recalls is, however, contingent upon the experience of the investigator. We observe two effects of complacency related to specific experience. First, investigators that have previously inspected a plant, even once, are associated with an increase in the recall hazard, unconditional on the inspection outcome. The more often the same investigator visits the plant, the higher the chance that the inspection will be followed by a recall. We attribute this affect to a reduction in inspection information content as specific experience increases. Both type I and type II errors, resulting from low information inspections, may lead to a higher hazard of a recall. Additionally, if an investigator has visited the plant more than twice in the past, their inspection results provide significantly reduced information as it relates to future recalls. Inspection outcomes do not reliably predict recall hazards for this category of investigator. We also observe that unlike some past experiential learning research, we see no effect of general experience on our results. One possibility for this finding is the training process and institutional knowledge that resides within the FDA. The FDA has been performing plant audits in different evolving forms since 1938.[14] Over 75 years of institutional knowledge may create an environment in which new investigators are well prepared for their roles and apparently do not get better or worse as their inspection experience grows outside of the specific plant they are inspecting.

An apparent solution to the negative repercussions of growing specific experience may be more frequent investigator rotation among plants. Subject to travel budget constraints, the FDA could potentially rotate plant investigators among a larger set of plants to avoid repeat inspections by the same investigator. To investigate the possibility of more frequent investigator rotation, we obtained updated counts of the current-day numbers of FDA investigators and medical device plants requiring inspection. According to the publically available data on the FDA's website, listing inspections and their results from 1/1/2009 to 12/31/2012, there were 5,112 medical device plants requiring inspection as of January 2014.[15] Through an additional FOIA to the FDA, we were able

[14] http://www.fda.gov/AboutFDA/WhatWeDo/History/FOrgsHistory/ORA/ucm083663.htm

[15] http://www.fda.gov/ICECI/EnforcementActions/ucm222557.htm. Note, this file does not include investigator identification and

to identify the number of FDA medical device plant investigators every year for the past decade. Using 2013 data, the most recent year available, we found that there were 245 investigators listed. Assuming that the FDA maintains a two-year frequency for plant inspections, it would equate to 2,556 inspections per year or 10.4 inspections per investigator per year. If the FDA strategically decided to not allow investigators to repeat at a specific plant, these investigators would inspect approximately 10 new plants each year, equating to 511 years (5,112÷10) before an investigator would be required to inspect the same plant more than once. Cleary, investigator rotation while maintaining an inspection every two-years for a plant is at least a mathematical possibility. Inspecting 10 plants per year per investigator is approximately a 200% (10 compared to 3.24) increase in the average inspections completed per year per investigator, when considering 2006 inspection data, but it reduces the required work-load of the most-used investigator by about 40% (10 compared to 16) per Table 2.1. This change in strategy would lead to a more balanced inspection load for each individual investigator and simultaneously create an environment in which complacency may be reduced.

A potential negative effect of a mandatory rotation strategy could be the additional travel costs incurred. Assuming 2,556 annual inspections, all requiring the maximum U.S. travel distance across the country, the additional costs would be approximately $766,800.[16] Rotation should be able to reduce approximately 1.7 recalls per plant per seven years according to Table 2.5. However, since not all plants experience recalls, we assumed that this benefit would only be attributed to plants that had experienced at least one recall in our dataset; 532 plants. This added travel cost for FDA can be compared to a possible reduction of 904 medical device recalls (532 x 1.7). The benefit of preventing just one major recall, which can affect many people's health and safety, clearly outweighs the cost of the increased travel schedule that would be required in order to achieve a more frequent investigator rotation. Rotating only within FDA regional districts would lead to even lower travel costs, though the likelihood of repeating inspections at a plant would be higher than in the case of cross-country rotation.

An alternative explanation for our results may be related to behavioral aspects of plant management, instead of FDA investigators. In discussing our results with FDA officials, one alternative raised was the possibility that plant managers became more "savvy" as they experience

hence does not allow us to extend the time period of our main analysis.

[16] We assume an average cross-country flight cost of $600, as compared to an average regional flight cost of $300; leading to an incremental flight cost of $300 per inspection. Lodging is excluded from this analysis, as many of the current inspections already require lodging. $300 x 2,556 = $766,800.

multiple inspections by the same investigator. In essence, this argument points to manager malfeasance (learning how to hide problems from familiar investigators) as the mechanism and not investigator complacency. We have chosen to theoretically ground this study in the latter argument, which rests solidly upon results from past related literature and support from industry and regulator interviews. Fortunately, if the malfeasance argument is true in reality, the approach proposed in this paper still holds. Investigator rotation serves as a valid solution to both possible mechanisms.

Another alternative explanation may be that the more often an investigator visits a site, the better understanding he/she has of plant practices and problems and the more familiar relationship with plant personnel allows managers at the site to uncover additional quality problems leading to more recalls. A vast majority of the individuals we spoke to in leadership positions in industry and the FDA believed that investigator complacency was much more likely to be the mechanism for our results than either manager malfeasance or investigator improved plant understanding over time. Further, since recalls originating directly from an inspection are excluded from our analysis, this explanation appears implausible.

While our specific context is focused on quality inspections and product recalls, our research may apply more broadly to other inspection contexts. Could firms benefit from rotating managers responsible for performance reviews? Could investors benefit from firms rotating their auditors? Or could the public benefit from the Federal Reserve rotating the regulators associated with individual banks? We believe that rotation policies may be effective in many other contexts besides ours, and more research is clearly needed to examine the efficacy of rotation policies at improving performance outcomes. In the context of financial auditing, the practice of auditing firms rotating the partners responsible for an audit (as opposed to clients rotating auditing firms) is an established guideline. For example, the International Federation of Accountants in their international standard on quality control requires "the rotation of key audit partners after a pre-defined period"[17], and some studies of the Australian stock market seem to support this notion (Azizkhani et al., 2013). We hope our research stimulates more inquiry as to whether rotation policies within firms can generally overcome complacency effects.

In general, an efficiency and quality trade-off must be analyzed in each case. In our context, investigator efficiency may suffer somewhat in a rotation scheme, however results indicate that inspection quality improves. The quality benefit significantly outweighs the lost efficiency. In other

[17] http://www.ifac.org/sites/default/files/downloads/a007-2010-iaasb-handbook-isqc-1.pdf.

contexts, the lost efficiency may be too severe to justify a rotation strategy. This is an important consideration as additional applications of this research are pursued. More generally, our research points to a fundamental trade-off in supply chain relationship management. While we largely view experience as beneficial, supply chain settings require some form of control as well, and the complacency that comes with experience may limit the control necessary to maintain effective supply chain relations. A purchasing executive with whom we discussed this research narrated a very similar trade-off in his purchasing organization: key supplier account managers would initially benefit from the experience of managing an account, but would eventually become complacent and fail to effectively control costs within the supply chain relationship. His solution was similar to ours – he would rotate account managers to different accounts every six months to prevent complacency from taking over. This one example demonstrates the ability of managers to both leverage long-term relationships on an organizational level with key, strategic suppliers, while avoiding unwanted complacency by rotating individuals who manage these key, long-term suppliers. How prevalent is this trade-off between learning and complacency in supply chain relationships in general? Whether and when can rotation policies be used efficiently and effectively within organizations to better manage the risk of complacency? In other words, can the 'dark side of supply chain relationships' (Villena et al., 2011) be managed by effectively rotating the people responsible?

Our analysis has limitations. Additional investigator demographics and training characteristics would enable a more complete picture of the investigator. However, this information was not available. Capturing complete career inspection data for FDA investigators would improve the analysis. Further, while we examine product recalls as one quality outcome metric, other quality outcome metrics are conceivable. Specifically in our context, the FDA also collects reports on product device failures more broadly (Manufacturer and User Device Experience, MAUDE database). It would be an interesting extension of our work to examine whether inspection outcomes are predictive of MAUDE reports. Further, while our analysis shows that rotation policies may be beneficial in our context, the correlational nature of our data naturally limits the confidence we can have in these results. A field experiment in cooperation with the FDA would more firmly establish the efficacy of such rotation policies. Further, our research cannot fully establish the causal mechanism underlying the benefits of inspector rotation; behavioral lab studies could shed further light on this issue.

In this chapter, we have investigated attributes of investigators and inspections which may

help predict medical device recalls. Exposing these attributes and recommending investigator rotation at the FDA should reduce future medical device recalls. This policy change is under review by the FDA at the time of this dissertation. The next phase of the recall process, per Figure 1.1, is the managerial decision to recall. In chapter three, I investigate managerial behavioral factors that may influence the decision to recall; factors that are both situational (related to the individual product failure scenario) and dispositional (related to the disposition of the manager). Uncovering factors that influence this decision can improve the quality and objectivity of this decision.

Chapter 3:

The Decision to Recall: A Behavioral Investigation in the Medical Device Industry

3.1 Introduction

The decision to recall a product from the market is fraught with complexity. There are often multiple criteria for recalling products, many that are not clearly specified. A recall, even when it is warranted, can result in significant negative consequences for all stakeholders, including product manufacturers, managers who make the decision, and consumers who use the product. While the decision to recall may be clear cut in some cases, such as when a faulty product is shown to cause customer deaths, such situations are rare. More often, the decision is not straightforward and only taken after much deliberation and consideration. Increased customer complaints could first result in additional scrutiny of the product, leading managers to monitor the product more closely rather than recall it immediately. Whether or not to recall involves logical arguments on both sides. On the one hand, managers may be tempted to "wait out the storm" assuming that product failures will eventually abate. Such caution may be more likely if the data is noisy and no root cause has been identified for the failure. It may also be in a manager's best interest to wait and not incur the additional costs associated with a recall, as they are incentivized to increase profits. On the other hand, managers may be influenced to recall by the repercussions they perceive possible future product failures could have on a customer's health or opinion of the company.

There is anecdotal evidence to support how critical and complex these decisions can be. In some cases, it appears that recalls are not initiated when they perhaps should be. For instance, in 2010, Johnson & Johnson was sanctioned by the Food and Drug Administration (FDA) for neglecting to issue a product recall in a timely manner after receiving numerous customer complaints related to multiple faulty products.[18] In other cases, recalls are undertaken when they may not be required. Toyota's brake recall provides a recent example. Between 2008 and 2010, Toyota recalled almost ten million vehicles due to an alleged problem with the brakes. Initially,

[18] http://www.washingtonpost.com/wp-dyn/content/article/2010/05/01/AR2010050103051.html

Toyota denied having any problems, but then succumbed to intense pressure from customers, regulators and the media, and recalled all potentially faulty units.[19] After numerous investigations, including those recently completed by NASA, it was determined that the brakes were not defective and the Toyota recall was likely unwarranted.[20]

Despite the impact product recalls can have on customer health and safety, as well as firm performance, there is little empirical research examining the product recall decision. Our study addresses this research gap by examining how recall decisions are made in the medical device industry and identifying factors that are not specified in the FDA criteria and yet, they may influence the recall decision. We focus on the following factor-specific research questions: How does an individual physician concern influence the recall decision? Does the level of defect detectability contribute to the likelihood of recall? What impact does root cause understanding of defects have on the likelihood of recall? How is the recall decision influenced by individual managerial attributes, such as cognitive reflection level?

We chose the medical device industry as the setting for this study because recalls in this industry are almost always decided by managers (rather than regulators, which is common in other industries such as automotive), and these recalls often have serious public health consequences.[21] Because of the innovative nature of these products, the recall criteria provided by the FDA (the regulatory agency responsible for medical devices) are somewhat vague. FDA recall policy states that companies must recall products that "present a risk of injury or gross deception or are otherwise defective."[22] The criteria are broad and open to multiple interpretations (e.g., should managers recall every defective product, recall only products that pose a risk of injury, or recall products that are safe for use, but are grossly deceptive such as those that are mislabeled). Managers are required to decipher this policy every time they contemplate a recall. This implies that the decision outcome is dependent on an individual manager's subjective interpretation of the policy.

We test our research questions through a controlled experiment utilizing vignettes developed with our industry partners. Because of the highly contextual nature of a recall decision, we draw our subjects for the experiment from a unique subject pool of managers employed in a Fortune 500 medical device company. To inform our choice of factors to manipulate, we conducted a two phase interview process with stakeholders within the medical device industry. In the first

[19] http://www.theatlantic.com/business/archive/2011/05/who-was-really-at-fault-for-the-toyota-recalls/238076/
[20] http://www.nhtsa.gov/UA
[21] http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/PostmarketRequirements/RecallsCorrectionsAndRemovals/
[22] Ibid.

phase, we conducted interviews to identify a broad set of factors that might influence the product recall decision despite not being explicitly delineated as recall criteria by the FDA. We then mapped these factors against behavioral theory to ascertain their potential behavioral effects. In the second phase, we narrowed the list to a final set of factors based on further feedback from our industry partners. The final list of hypothesized factors include three situational factors (individual physician customer concern related to the product defect, pre-use detectability of the defect by the physician customer, and managerial understanding of root cause of the defect underlying the recall), and one dispositional factor (cognitive reflection, as measured by the Cognitive Reflection Test, Frederick (2005)). These factors are described in more detail in Section 3.2.

Results from the experiment show that, contrary to anecdotal evidence and managerial expectations, individual physician concern does not impact the decision to recall. However, a product defect that is less detectable by a physician pre-use is associated with a higher likelihood of recall. The likelihood of a recall decision increases by 48% when the defect cannot be detected by the physician. This implies that products with quality defects may be knowingly left in the marketplace if physicians can detect the defective products. The data also confirms that a better understanding of the root cause of the underlying product defect is associated with a higher likelihood of recall. The likelihood of recall increases by 65% when the root cause is understood compared to when it is not.

The manager's level of cognitive reflection (measured through the CRT test) also appears to influence an individual's likelihood to recall. Our results suggest that managers with a high CRT score are 70% less likely to recall than managers who scored low. Finally, results from post-hoc analyses show that the effects of certain situational factors on recall likelihood depends upon a manager's CRT score. Managers with high CRT scores are most affected by the root cause understanding of a defect. In contrast, managers with low CRT scores are only affected by other dispositional control variables, such as gender, experience, or their perceived relationship with the FDA (e.g., a good relationship with the FDA leads to a lower likelihood of recall).

3.2 Research Phases I and II

We conducted our research in three phases, as outlined in Table 3.1. The first two phases focused on developing a deeper understanding of the medical device industry and how the product recall process is structured, including how FDA guidelines are interpreted by practicing managers.

Our goal at the end of the first two phases was to have identified a set of unique factors that are not explicitly included in the FDA guidelines but have the potential to impact the recall decision from a behavioral perspective. These factors also needed to have face validity and be of interest to our industry partners, ensuring the study would be useful for practicing managers and regulators in future decision making.

3.2.1 Phase I: Industry & Recall Process Understanding and Initial Factor Identification

In phase I, we conducted exploratory unstructured interviews with practicing managers at two different U.S. Fortune 500 medical device companies and with FDA regulators. This helped us develop a deeper understanding of the medical device industry and the product recall decision-making process. This multi-perspective approach also allowed us to juxtapose the industry and the regulator perspectives, and understand how the intended policies were actually implemented by managers.

Table 3.1 Interview Process and Research Phases

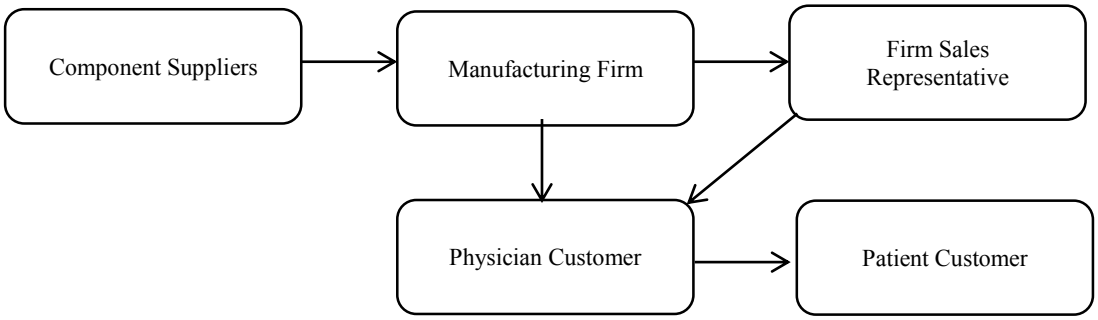
Phase (Year)	Mode	Participants	Organization	Duration	Main Objectives	Section
Phase I (2012-2013)	Exploratory unstructured interviews	- VP, Corporate Global Quality - VP, Corporate Quality - Plant Manager	Focus Firm: Fortune 500 medical device firm	3 x 3 hours	* Develop an understanding of: - Medical device industry and its supply chain - Recall policy, process and main participants - Identify broad list of potentially relevant factors	3.2.1
		Director of Quality	Fortune 500 medical device firm	4 hours		
		Director of Analytics	FDA	2 hours		
Phase II (2013-2014)	Unstructured & Structured interviews	- VP, Corporate Global Quality - VP, Corporate Quality - Plant Manager	Focus Firm: Fortune 500 medical device firm	6 hours	* Select final list of factors	3.2.3
Phase III (2014)	Experiment	- 167 managers across four functional areas from focus firm.	Focus Firm: Fortune 500 medical device firm	2 hours	* Develop subject pool list * Finalize research design, process and timing * Execute experiment	3.4

All industry participants held managerial and/or executive level positions and had significant experience in making recall decisions. Our interviews began with a focus on the medical device industry in general, and its supply chain in particular. A simplified medical device supply chain is depicted in Figure 3.1. While not all medical devices follow this exact flow, a majority do (e.g., hip or knee replacements, cardiac stents and pacemakers, hearing aids, prosthetics). Manufacturers purchase from suppliers, manufacture devices and distribute to physicians. Devices can either be transferred to sales representatives, who then sell products to physicians on behalf of the company, or sold directly to physicians. The physician normally chooses which products to use for their patients. The physician and the patient are considered important but separate customers in this supply chain. The end customer who actually uses the product is referred to internally in a medical device company as the "patient customer". This terminology emphasizes the importance of patients, each being a recipient of valuable life-saving or life-enhancing technology which has its root within the company. In contrast, the "physician customer" is also important but has a different role in the supply chain. The physician customer makes the purchasing decision, observes product performance such as product quality and features, and provides frequent feedback to company managers. The patient customer is not the source of revenue or product performance feedback, but, unlike the physician, is physically impacted by product quality.

Physician and patient customer complaints are the most common signal of product defects. Physicians report defects to the company as they occur during use or implant of the medical device, or when a patient suffers a negative health episode due to device malfunction. Quality assurance managers regularly monitor and consolidate these signals to provide an integrated view of field product performance to corporate leadership. When these signals indicate a decline in quality performance, they may initiate a product recall meeting. Product recall meetings bring together key functional area representatives (e.g., quality, manufacturing, clinical, medical) to review performance data including conditions in which product failures have occurred, impact of failures on the patient customer, and current failure rates compared to predicted failure rates. The outcome of a recall meeting is not necessarily one consolidated recommendation, but can be a list of recommendations by each functional area. Individuals can recommend no recall, which by default equates to continued monitoring of product quality, or recall. Recall recommendations by functional area are forwarded to executive leaders who make the final recall decision; their decision is heavily influenced by the recommendations from the product recall meeting. A recall decision leads to an immediate cessation of product shipments, a notification to all affected customers, a

formal communication to the FDA, and a detailed plan to retrieve or repair all affected products.

Figure 3.1 Simplified Medical Device Supply Chain



In our interviews, industry participants described the recall process in extensive detail, and shared their individual and collective experiences in making and implementing recall decisions. They gave examples of recalls that had gone well, and those that did not go as well, and conjectured on characteristics that may distinguish the two. This informal story-telling helped us hone in on an initial list of implicit factors that might exert influence on the actual recall decision. At the FDA, our main participant was the senior director of risk management at the Center for Device and Radiological Health (CDRH), the FDA division responsible for medical devices. Our in-depth interviews with the FDA director helped us understand the recall process and the main participants from a regulator's perspective. The director also outlined the stated guidelines regarding recall requirements, and potential conflicts that managers may face in making recall decisions.

At the end of phase I, we had developed an extensive understanding of the context in which the recall decision is made, the process used in making the recall decision, and the main participants and their roles. Together with our industry partners, we also developed an initial list of factors that were believed to influence the recall decision, but which were not explicitly delineated as recall criteria by the FDA. See Table 3.2, column 2 for the full list of initially considered factors.

3.2.2 Mapping Theory to Factors

To put the initial list of factors into theoretical context, we reviewed relevant behavioral literature to see whether the factors mapped to prior theories. In doing so, we recognized that the factors naturally fell into two distinct categories, situational and dispositional. We found several studies that had categorized decision-making factors similar to ours into these two groups (Kacmar et al., 2004; McNeely and Meglino, 1994; Jones, 1991; Newton and Keenan, 1991). For example, Jones (1991) studied ethical decision making in organizations, a closely related topic to product recall decisions. He concluded that situational and dispositional factors, related to a decision and a decision-maker, are pertinent and significantly influence important decisions in organizations. This distinction also fits with what we learned from the managers we interviewed. Specifically, each recall decision is made within the context of a specific product failure situation. A certain failure occurs in a specific environment and the outcome of the failure is manifested in a unique way. Although FDA criteria specifies some situational factors that could be used to make the recall decision (e.g., risk of injury, gross deception, defective product), our interest is in situational attributes that have not been explicitly specified, and yet may influence this decision.

Independent of the details surrounding the possible recall situation, each manager also brings a unique perspective that might predispose him/her in this decision. Several of the interviewees indicated it was very likely that certain dispositional criteria, unique to each manager, and separate from the situational specifics of the product failure, are influential in this decision. These dispositional decision criteria are also not specified in the FDA guidelines as relevant. Table 3.2 (column 3) shows how the initial factors identified in phase 1 breakdown between the situational versus dispositional categories.

3.2.3 Phase II: Factor Selection

The objective of phase II was to narrow the list of factors to a final set that would be used in the experiment (phase III). We conducted additional structured and unstructured interviews, analyzed each potential factor in more detail, and revisited related theory. The iteration between interviews and theory helped us discern which factors showed the highest potential to provide helpful insights for our industry partners while also having strong theoretical justification.

Through this process, we narrowed the list from fifteen to eight factors. To provide insight

into how factors were eliminated, we briefly discuss a few examples. One of the situational factors we excluded from our final list was "the risk of harm to the patient customer." Managers were fairly confident that this factor would have a strong positive impact on the likelihood to recall and so were not as interested in testing its effect. Two other excluded situational factors are related to the "opinion of others" (e.g., the FDA or a key functional area such as the quality department) in making the decision. These were also deemed less interesting to the industry representatives since it was unclear how one might overcome such a bias (or if one would want to) in an actual recall decision where FDA and quality managers play critical roles.

Two dispositional factors were also excluded. The influence of managerial incentives on the recall decision was dropped since recent recall research has already identified managerial incentives as a significant predictor of the recall decision (Wowak et al. 2014). Managers' empathy with the patient customer was excluded because our industry partners believed there was little doubt it would to lead to a higher likelihood of recall.

The final factor list consists of three situational and five dispositional factors (Table 3.2, column 5). These were each deemed to be of great interest to our industry partners and were firmly grounded in theory (see Section 3.3).

Table 3.2 Factor Development

(1)		(2)		(3)		(4)		(5)	
Phase I Interviews	➔	#	Initial List of Factors	➔	Mapping to Theory		➔ Phase II Interviews	➔	Final List of Factors
				Situational	Dispositional			Situational	Dispositional
		1	Risk of harm to patient customer	X					
		2	Recall opinion of FDA	X					
		3	Recall opinion of Quality department	X					
		4	Recall opinion of physician customer	X				X	
		5	New vs. established products	X					
		6	Root cause understanding of failure	X				X	
		7	Functional area responsible for failure	X					
		8	Physician's ability to detect defect	X				X	
		9	Firm's perceived relationship with FDA		X				X[a]
		10	Managerial incentives (e.g. stock options)		X				
		11	Gender of manager		X				X[a]
		12	Years of experience of manager		X				X[a]
		13	Manager's empathy for patient customer		X				
		14	Functional background of manager		X				X[a]
		15	Reflection in decision making		X				X[b]

[a] Dispositional factors used as control variables
[b] Cognitive reflection test used to examine this factor

3.3 Hypothesis Development

Physician Concern

We label the first situational factor *Physician Concern* because it examines the effect of an influential and outspoken physician customer on the decision to recall. The company may cater to the opinion of an influential physician because they can have significant financial impact on the company's revenue as they frequently make the purchasing decision on behalf of their patients and may have clout with other physician customers. Our partner managers believed that close relationships between the company and certain physician customers may lead to recall decisions which are not necessarily driven by data and analysis, but by close customer relationships.

Researchers have identified the tendency of managers to respond to the most salient customer feedback, and ensure that the "squeaky wheel gets the grease." (Bendoly et al., 2010; Ma et al., 2015; Mitra and Golder, 2006). For example, Ma et al. (2015) examine a large panel dataset of complaints via Twitter over the course of 10 months. They analyze thousands of individual customer complaints and find that firms take significant actions as a direct response to individual outspoken customers. This tendency to respond to one vocal customer may also be attributable to salience bias, which occurs when decision-makers respond to the most noticeable feedback, instead of aggregating all relevant feedback and taking a holistic view of the data. For example, one's perceived probability of a traffic accident is immediately raised after one observes a traffic accident (Tversky and Kahneman, 1974). Salience bias has also been observed in marketing studies which demonstrate that the prominence or salience of a brand in one's memory is significantly related to consumer product choice. (Hutchinson, 1983; Alba and Chattopadhyay, 1986; Hauser and Wernerfelt, 1990). Relatedly, a manager's perception of the need for a recall may increase following a customer complaint from one outspoken customer.

One recall example provided by a company we interviewed illustrates this relationship: A physician demanded a conference call with the manufacturing manager (a very unusual request) to ask specifically how an obvious and visible plastic coating defect on a catheter could have occurred. In the course of the conference call, the upset physician demonstrated dissatisfaction with the defect and implied that if he observed further defects, he would elevate his concerns to the FDA. Negative

opinions by physician customers such as this can escalate the significance of a product defect and a product recall. It did so in this specific case. While this example demonstrates how one vocal and influential physician can influence a recall decision, it is noteworthy that the FDA does not specify this as recall criteria. Additionally, our partner managers were interested in examining the influence of this factor because they believed that one individual physician customer's opinion should not lead to a recall, though anecdotal evidence suggested it may have a significant effect. We therefore hypothesize:

> *Hypothesis 1: Physician concern related to a product defect results in a higher likelihood of a recall*.

Defect Detectability

Defective medical devices vary in their level of detectability prior to use by the physician. The second situational factor we examine is the level of defect detectability by the physician customer prior to use of the product (*Defect Undetectable*). Certain defects, such as a cracked screen on an electronic medical device, are highly detectable, while others, such as a software coding error, are often undetectable. Our interviews revealed that managers faced an interesting conflict related to the level of pre-use physician defect detectability. On the one hand, observable defects may decrease the physician's quality opinion of the company, potentially leading managers to be more likely to recall if the defect is detectable by the physician. On the other hand, when a physician can detect a defect pre-use, the physician may serve as a type of final quality control step and not use the defective device on the patient customer; reducing the risk of patient harm. This reduced risk of harm may lead managers to be less likely to recall, even though the product is actually defective. While all the managers we interviewed were convinced of this factor's relevance in the decision, they were split on how it would influence the recall decision. We develop competing hypotheses to address this potential conflict below.

Defect Detectability and a Higher Recall Likelihood

Theoretical and empirical evidence in marketing research demonstrates that it pays to please the customer. This is especially true when the customer makes the purchasing decision for a company's products. Companies realize higher profitability (Anderson et al., 2004, 1994), stock

price (Ittner and Larcker, 1998), return on investment (Anderson and Sullivan, 1994), customer retention (Gustafsson et al., 2005), and repurchase behavior (Mittal et al., 2011), among other benefits, as customer satisfaction increases. In our context, providing physician customers with high quality products could yield both short and long-term benefits, while poor product quality could have the opposite effect. Marketing research in this domain is too vast to fully review here, but we provide a few relevant examples to support our hypothesis.

Anderson et al. (2004) study the relationship between a broad customer service index (American Customer Satisfaction Index) and performance in 200 Fortune-500 firms, and find that higher customer satisfaction is positively related to Tobin's Q, which is frequently used as a forward-looking measure of shareholder value. This relationship holds across industry type and size. Using a related measure of customer service, Fornell et al. (1996) find that product quality is one of the most influential product attributes associated with customer satisfaction, even more than product price. Mittal et al. (2001) investigate the effects of repurchase behavior for over 100,000 automotive products customers. They find that more satisfied consumers have a stronger intention to repurchase. Gustafsson et al. (2005) also study the relationship between customer satisfaction and retention, and find that in a broad telecommunications services survey, these two concepts are positively relate, satisfied customers are more likely to remain as customers. As a set, the literature on customer satisfaction is consistent on two points: satisfied customers are more likely to be retained as repeat customers, and they are a signal for broader firm success, such as increasing shareholder value. Managers may be more likely to recall a defective product that is detectable in order to increase customer satisfaction, which should lead to higher customer retention and future profits. Thus we examine the following hypothesis:

Hypothesis 2A: A detectable product defect results in a higher likelihood of a recall

Defect Detectability and a Lower Recall Likelihood

However, managers are not only concerned with the opinion of the physician customer, but also with the safety of the patient customer. Interviewees we spoke with crystallized the idea that managers may actually be less likely to recall if a defect is more detectable to the physician pre-use because there is an implied reduction of patient harm. In such a case, managers seem to view

the physician as a final quality control screen for the product, and rely on the physician to observe, and not use, a defective product, safeguarding the patient.

In an actual recall scenario, a product was accidentally packaged upside down in the sterile package. This product was equipped with external hooks to anchor in the blood vessel of the patient. If the physician inserted the device upside down, the patient would likely die from a torn artery. Additionally, this product was new to physicians, so it was unlikely they would recognize the upside down orientation, leading to a risk that the packaging error would not be detected and the patient would be harmed. The risk of this defect not being detected played an important role in this recall decision. Some of our partner managers strongly believed that when a defect is detectable to the physician pre-use, managers may be less likely to recall. Accordingly, we hypothesize:

Hypothesis 2B: A detectable product defect results in a lower likelihood of a recall.

Support for hypothesis 2A would indicate that the physician's assessment of the quality of the company's products is possibly more important to managers than protecting the safety of the patient customer. In this case, if a defect is detectable, *a recall is needed* to protect the company from a poor quality assessment by the physician, even though patient customers are possibly at a lower risk. Support for hypothesis 2A would also indicate that if the product defect is undetectable, the physician will not notice, and even though patient customers may be at risk, managers would decide that a *recall is not needed*.

Support for hypothesis 2B would indicate that the safety of the patient customer is possibly more important than the physician's assessment of the quality of the company's products. In this case, if the product defect is detectable, the patient customer is at a low risk and a *recall is not needed*, even though defective product is consciously being kept in the marketplace. Support for hypothesis 2B would also indicate that if a defect is undetectable, the physician is unlikely to screen out defective products, exposing the patient customer to possible harm; leading managers to decide that *a recall is needed*.

Defect Root Cause

Recalls can be required well before managers fully understand the cause of the failure, though it is possible that knowledge of root cause may make managers more likely to recall. The

level of understanding related to the root cause of the defect (*Root Cause Understanding*) was also deemed to be a potentially relevant situational factor for this study. While managers we interviewed recognized that they frequently waited to comprehensively understand the root cause of a defect before making a recall decision, they did not know how prevalent it was among recall decision makers. They hoped that by including this factor in the study, they would better understand its effect on the recall decision.

Understanding why something occurs is the "basic spring of action" in human life (Weiner, 1979). Attribution theory (Weiner, 1979, 1985) informs us that individuals desire to understand the cause of an event. The importance of attributing a cause to an event has been demonstrated in multiple settings including third-party logistics failures (Oflac et al., 2012), customer satisfaction related to service failures (Anderson et al., 2009), and new product designs (Schreier, 2006). Uncovering true root cause is a critical step in understanding and resolving a product defect. In the absence of a definable cause, product recalls can be challenging, though still necessary. When root cause is not understood, the "bounding" of the problem (e.g., which units should be recalled) is difficult, and can force the company to recall more products than may be necessary, leading to higher recall costs and more hassle for patient customers. When root cause is understood, the solution for overcoming the product quality problem is likely more effective and the affected units can be better isolated, leading to lower recall costs. Product engineers and other recall decision-makers search avidly for potential causes for failures, and without a well-understood root cause, managers may be more likely to choose a "wait and see" approach, hoping that product failures dissipate. It is possible that at the same failure rate, managers who understand root cause would be more likely to take action and recall than managers who do not understand root cause. Attribution theory would lead us to believe that this relationship may exist in this decision.

One applicable example from our interviews relates to an electronic cardiac device that was shorting out and experiencing premature battery depletion. A certain number of devices were failing in the field, however, engineers did not understand the reason for the failures. It was clear through data analysis that a sizable increase in field failures had occurred, but a recall was delayed until root cause could be found. Apparently some decision-makers were either not convinced that a real product quality problem existed, or they were hesitant to initiate a broad based recall that affected so many units. An engineering analysis finally pinpointed a process change at an integrated

circuit manufacturing step as the source of the failure. This discovery led to the immediate decision to recall, though the recall was limited in scope because they could bound the high-risk units to a certain process change. However, the delay between when failures first surfaced and when the root cause was identified left additional customers at risk and some customers were negatively affected. This example indicates that root cause understanding may be an important situational factor in this decision, even though it is not specified by the FDA as relevant for the recall decision. We therefore hypothesize:

> *Hypothesis 3: Root cause understanding of a product defect results in a higher likelihood of a recall*.

Individual Differences and the Recall Decision

Hypotheses 1-3 describe possible situational factors that may impact the recall decision. However, it is likely that individual differences, or dispositional factors, will also impact this decision. The decision-makers responsible for recalling a product are heterogeneous in many aspects. They have different educational and training backgrounds and represent different functional areas with different incentive structures. The factor selection process highlighted five dispositional variables that could be used in this study, per Table 3.2. Four of these variables (gender, functional area, experience and relationship with the FDA) were viewed by our partner managers as relevant control variables for the experiment, but they did not have sufficient managerial interest to be directly hypothesized. However, our partner managers had a particular interest in the effect of cognitive reflection on the likelihood of recall, and as a result, this factor is hypothesized in the study.

In complex decisions such as these, one temptation managers face is to wait for more data to accumulate. More time and more data may allow managers to make better, more informed decisions. Our interviews indicated that some managers may fall prey to the "analysis paralysis" trap, often postponing their recall decision until more data is collected and more failures occur. On the other hand, another group of managers make recall decisions very quickly and seldom voice the need to wait for additional data. These managers were described as making "knee-jerk reactions" to the recall decision. Cognitive reflection may be a lens to distinguish between these two groups

of managers; those who tend to want more data that they can reflect upon versus those who make more intuitive judgements.

The Cognitive Reflection Test (CRT) (Frederick, 2005) is one mechanism that distinguishes individual decision-makers according to their level of cognitive reflection and may be useful in understanding a subject's propensity to recall. In other words, it may help in separating the "analysis paralysis" managers from the "knee-jerk reaction" managers. The test involves three questions that have immediately apparent and seemingly intuitive answers that are all incorrect; all three questions require reflection to be answered correctly. To do so, subjects must overcome their impulsive response and reflect on what the correct answer should be. Frederick (2005) proposes the CRT as a measure of cognitive reflection that is positively correlated with similar tests such as the Scholastic Aptitude Test (SAT), the American College Test (ACT), and the Wonderlic Personnel Test (WPT) (Frederick, 2005). The CRT has been used as a means to differentiate between impulsive and reflective subjects in prior experiments (Oechssler et al., 2009; Moritz et al., 2014). The CRT has also been used to study individual differences in Newsvendor ordering behavior (Moritz et al., 2013), forecasting behavior (Moritz et al., 2014), moral decision making (Paxton et al., 2012), the tendency to fall victim to common heuristics and biases in decision making (Toplak et al., 2011), cognitive abilities for genetically inherited traits (Cesarini et al., 2012), and rulings by judges (Guthrie et al., 2007). Theory indicates that managers who score high on the CRT may be those managers who wait for more data before making recall decisions. This leads to the following prediction.

> *Hypothesis 4: Individuals with higher cognitive reflection will have a lower likelihood of recall*.

3.4 Experimental Design

To test our hypotheses, we performed a controlled experiment using managerial subjects from a leading US medical device company. We chose to study our questions in the US medical device industry for several reasons. Recalls in the medical device industry are generally voluntary in nature, such that managers make the recall decision at their own discretion unlike other industries where the regulatory body monitoring the industry mandates product recalls. For instance, the

National Highway Traffic and Safety Administration frequently mandate recalls in the automotive industry.[23] As mentioned previously, FDA recall criteria are vague, indicating a tangible benefit of establishing clarity in this context. Medical device recalls also have significant public health consequences, which makes understanding how managers reach the recall decision even more critical. Finally, this industry has a significant economic footprint as it is expected to exceed $130 billion in market size by 2016.[24]

The partner company used for our subject pool is one of the top 20 medical device companies by revenue and is in the top 200 of the U.S. Fortune 500 companies. It operates hundreds of facilities in dozens of countries, and similar to its main competitors, experiences recalls on a reasonably regular basis. It has undergone many small and large recalls and has a mature recall decision-making process.

Our study utilizes experimental subjects who are managers that have experience making recall decisions. We chose managers instead of students as experimental subjects because of the highly contextualized and unique nature of a medical device recall decision. Past experimental research indicates that while student and manager subjects have comparable performance in many tasks, certain situations may require managerial subjects (Frechette, 2012; Potters and Van Winden, 2000). In his review of 13 experiments where students and practitioners were used as subjects, Frechette (2012) compared student and manager results to each other and to the theoretically predicted responses. He concluded that using managers as subjects, when the context is unfamiliar to students can "prove very insightful in ways that studying undergraduates is not…allowing us to learn something that could not have been learned with students." (Frechette, 2012).

A team headed by the company's Corporate Global Vice-President of Quality compiled a list of potential subjects for the experiment. The guiding principle in putting the list together was to include representatives from all relevant functional areas (quality, manufacturing, clinical, medical) involved in product recall decisions. Individuals within those functional areas with significant product recall decision making experience were then randomly selected to participate. The final list consisted of 287 individuals who were each sent an email requesting participation in the study (Appendix A). The email was sent from the Corporate Global Vice-President of Quality

[23] http://www.recalls.gov/
[24] http://selectusa.commerce.gov/industry-snapshots/medical-device-industry-united-states

(the most senior quality manager at the company). This message explained the nature and objective of the study, included a link to the experiment, and clarified that all responses would be anonymous. A follow-up reminder email was sent 11 days after the initial email (Appendix A). At the end of the study duration, 167 subjects had participated in the experiment with complete responses. The experiment was conducted completely online using Qualtrics survey software, and as is typical for professional subjects, they were not compensated for their participation.

To participate, a subject first navigated to the Qualtrics website by double-clicking on the link embedded in the email. The first screen described the experiment and then asked for consent before proceeding. Once consent was given, the subject moved to the second screen, where he/she was randomly assigned to one of the treatments and was presented with the associated recall decision scenario. The second screen was visually divided into three sections: the top of the screen contained a baseline scenario that described a product failure situation in which a recall was possible but not obvious. The baseline scenario described an increasing failure rate on a cardiac medical device to ensure that the scenario presented a failure which was both critical in nature (life sustaining device) and was technologically complex (cardiac devices are highly technical). The baseline scenario was constant in all treatments, and was similar to real-life recall scenarios previously described in our interviews (see Appendix A for baseline scenario and all experimental text used). The second section on the screen below the baseline scenario contained text associated with the situational factors in that treatment. Finally, in the third section at the bottom of the screen, the subject was asked to select whether he/she would recommend recalling or not based on the information provided. After selecting one of these two options, the subject was taken to the third screen which contained questions related to the dispositional control variables and the CRT. The dispositional control variables included in the study were: gender, functional area, perceived relationship that the company maintained with the FDA, and experience at the company. These control variables are described in more detail in Section 3.6 below. The three CRT questions followed the dispositional control variables. After answering the CRT questions, subjects were invited to enter optional comments regarding the experiment in a free text section at the bottom of the screen.

The experimental design and the number of subjects per treatment are illustrated in Table 3.3. We ran a full factorial (2^3=8) set of treatments, including low and high levels for all three

situational factors. The three situational factors coincided with the main effects hypotheses (*Physician Concern, Defect Undetectable,* and *Root Cause Understanding*). We coded "Defect Undetectable" so that the high level of this factor is consistent with a defect that is not detectable to the physician customer pre-use. In this case, a negative and significant beta coefficient on "Defect Undetectable" would indicate support for Hypothesis 2A, while a positive and significant beta coefficient would indicate support for Hypothesis 2B.

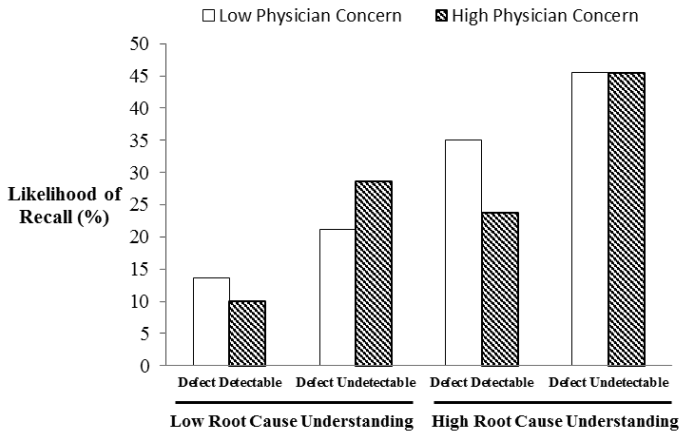
Table 3.3 Experimental Design and Number of Responses per Treatment

	Root Cause Understanding				
	Low		High		
	Defect Undetectable		Defect Undetectable		Total
Physician Concern	Low	High	Low	High	
Low	22	19	20	22	83
High	20	21	21	22	84
Total	42	40	41	44	167

3.5 T-test Results

To measure the effect of each situational factor, we computed the likelihood of recall as the percent of subjects deciding to recall within each treatment (Figure 3.2). We first discuss visual trends observed in Figure 3.2, and then present t-test results comparing the percent of subjects deciding to recall at low and high levels of each situational factor. The first factor, *Physician Concern*, does not appear to impact recall likelihood, as there is no visible trend in recall likelihood between the white and shaded bars in Figure 3.2. *Defect Undetectable* displays a possible relationship with recall likelihood, as the two sets of *Defect Detectable* bars are both lower than the corresponding *Defect Undetectable* bars, indicating that a lower percentage of subjects recalled when the defect was detectable to the physician. Finally, *Root Cause Understanding* also appears to influence recall likelihood, as the two sets of *Low Root Cause Understanding* bars are lower than the corresponding *High Root Cause Understanding* bars.

Figure 3.2 Situational Factors and Recall Likelihood



To test the statistical significance of these relationships, we conduct a t-test for each factor. The null hypothesis for these tests is that the recall likelihood at each low factor level equals the recall likelihood at each high factor level. Table 3.4 displays t-test results for each situational factor. *Physician Concern* does not impact the recall likelihood, as there is no statistical difference between the *Physician Concern* low and high groups in the likelihood to recall (p-value=0.827), providing no support for Hypothesis 1. When a defect is undetectable (*Defect Undetectable* high), subjects are more likely to recall than not recall (p-value=0.028), providing no support for Hypothesis 2A, but support for Hypothesis 2B. The more that is understood about the root cause of the defect (*Root Cause Understanding* high), the more likely the subject is to decide to recall (p-value=0.005). We therefore also support Hypothesis 3 with these results.

Table 3.4 T-test Results for Recall Likelihood and Situational Factors

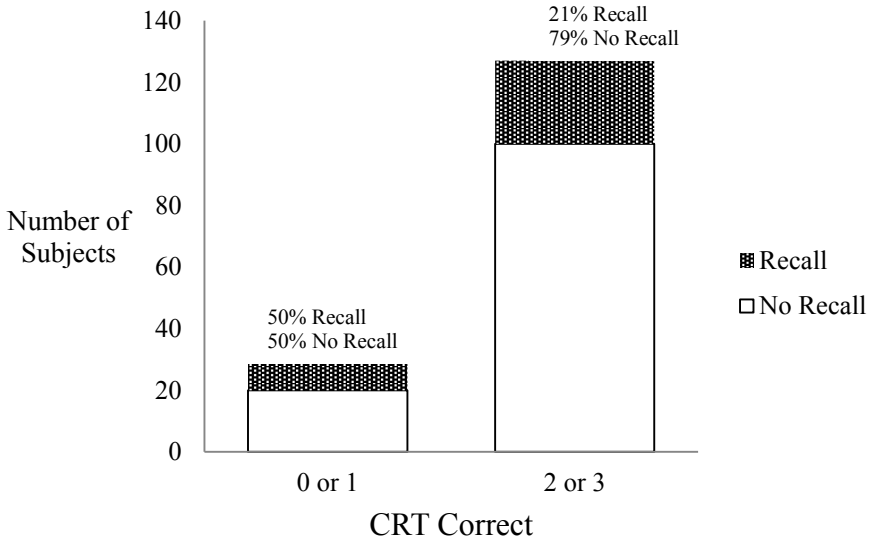
Factor	Factor Levels[a]						
	Low			High			
	Recall	N	%	Recall	N	%	p-value[b]
Physician Concern	24	83	29	23	84	27	0.827
Defect Undetectable	17	83	20	30	84	36	0.028
Root Cause Understanding	15	82	18	32	85	38	0.005

[a] Number and percentage of subjects deciding to recall in each cell
[b] Null hypothesis: Likelihood of recall at low factor level = Likelihood of recall at high factor level

To test Hypothesis 4, we pool across treatments and group subjects by their individual CRT scores. Subjects either answered zero (10.7%), one (13.2%), two (25.1%), or all three questions correctly (51%). To separate CRT scores into low and high categories, enabling us to test Hypothesis 4, we classified CRT scores of zero or one as the low CRT category and scores of two or three as the high CRT category. This categorization is consistent with how CRT scores are grouped in prior studies (Hoppe and Kusterer, 2011; Oechssler et al., 2009; Frederick, 2005). Before assessing the impact of low and high CRT, we verified that the random assignment of situational factor levels was not skewed towards low or high CRT. We conducted Chi Square tests and confirmed that there was no significant difference between the levels of each situational factor and low and high CRT.

Similar to our analysis for situational factors, we first plot the data visually and observe differences of recall likelihood across low and high CRT scores, and then confirm the visual trends statistically using a t-test. Figure 3.3 displays the likelihood of recalling at low and high CRT. Subjects who answered zero or one CRT questions correctly had an equal likelihood of recalling and not recalling, suggesting that in this category, a subject's decision to recall was primarily a random chance. In contrast, subjects who answered two or three CRT questions correctly were much less likely to recall (21% Recall vs. 79% No Recall), implying that there might be underlying causes to explain the systematic variation.

Figure 3.3 Dispositional Factor and Recall Likelihood



To confirm these observations, we present the results of a t-test in Table 3.5. The null hypothesis for the t-test is that the recall likelihood is the same across low and high CRT categories. The p-value for this t-test is p=0.000, allowing us to reject the null hypothesis. High CRT subjects were significantly less likely to recall than low CRT subjects. We support Hypothesis 4.

Table 3.5 T-test Result for Recall Likelihood and Dispositional Factor

Low CRT[a]			High CRT			
Recall	N	%	Recall	N	%	p-value[b]
20	40	50	27	127	21	0.000

[a] Number and percentage of subjects deciding to recall in each cell
[b] Null hypothesis: Likelihood of recall at low CRT = Likelihood of recall at high CRT

3.6 Logistic Regression Results

To ensure that the hypothesized relationships examined in t-tests hold after incorporating additional dispositional control variables, we regressed the likelihood of recall on the situational and dispositional factors. Regression analysis ensures that the observed relationships in the t-tests hold after including the dispositional control variables collected in the study (gender, functional

area, relationship with the FDA, and experience). Because the dependent variable (recall likelihood) is a dichotomous choice variable, we used logistic regression. Logistic regression predicts the likelihood of an outcome (a recall decision) based on changes in continuous or dichotomous predictor variables. Logistic regression results are interpreted as the change in the likelihood of the choice after exponentiation of the beta coefficient ($1\text{-}\exp^{\beta}$). We describe the dispositional control variables and how they were measured in more detail below.

We captured the gender of each respondent (*Male*) as an indicator variable with female as the reference category to control for possible differences that may exist between genders. Prior CRT studies have found a significant difference between the number of CRT questions answered correctly by males and females (Frederick, 2005; Oechssler et al., 2009; Hoppe and Kusterer, 2011). We learned in our interviews that managers in the quality department often view themselves as protectors of the customer within the company and may be more likely to decide to recall than subjects from other functional areas. We captured functional area and categorized the subjects into quality and non-quality and measured this as an indicator variable (*Quality*) with non-quality functions as the reference category. The perceived relationship that the company has with the FDA could also impact decision-makers in the recall decision. We captured this relationship as whether or not the subject perceived their company's relationship with the FDA as collaborative, average or confrontational. Because of the small percentage of responses in the confrontational category (5%), we grouped average and confrontational as the reference category and created an indicator variable for collaborative (*Collaborative FDA rel*.). It is also possible that the length of experience a subject has in the company may affect their recall decision. To control for this, we measured the number of years of experience that the manager had at the company with four categories (*0-2 years at company, 3-5 years at company, 6-10 years at company, More than 10 years at company*) and created indicator variables treating the most experience (*More than 10 years at company*) as the reference category.

We provide a descriptive summary of the dispositional control variables in Table 3.6. Approximately 70% of the subjects were male, and 60% belonged to the quality department (with the other 40% distributed between operations, clinical and the medical departments). 57% of the subjects perceived their company's relationship with the FDA to be collaborative. The number of years of experience subjects had at the company was fairly evenly distributed between new hires

(0-2 years) to very experienced personnel (more than 10 years at the company).

Table 3.6 Dispositional control variables and percentage of responses

Gender		Functional Area		Relationship with FDA		Years at the Company	
Male	71.2%	Quality	60.0%	Collaborative	56.9%	0-2	18.6%
Female	28.8%	Operations	15.6%	Average	37.7%	3-5	25.1%
		Clinical	9.0%	Confrontational	5.4%	6-10	34.1%
		Medical	4.2%			>10	22.2%
		Other	11.2%				

Finally, we controlled for the time it took a subject to answer the questions in the experiment. It is possible that subjects who rushed through the experiment were more or less likely to decide to recall, compared to those who patiently read all the text thoroughly. Qualtrics automatically recorded the time each subject took to answer the questions: it began the time when the subject opened the first screen of the experiment (consent request) and stopped the clock when the subject answered the last question and submitted the results. The mean response time in minutes was 17.65, with a minimum of 1, a maximum of 372, and a standard deviation of 33.3. We used the time taken by including the natural log of the response time (*Ln_Response_time*) as a control variable in our analysis.

We performed regression analysis in three steps. We first included situational factors (Table 3.7, column 1), then added CRT score (column 2), and finally added dispositional control variables (column 3). Per column 1, *Defect Undetectable* and *Root Cause Understanding* are both positive and significant predictors of recall likelihood while *Physician Concern* is not significant, supporting the conclusions in the t-test analyses. The likelihood of a recall increases by 48% ($\exp^{0.39}$=1.48) when the defect is undetectable (*Defect Undetectable* high) compared to when the defect is detectable. When the root cause of the defect causing the product failure is better understood (*Root Cause Understanding* high), the likelihood of a subject choosing to recall increases by 65% ($\exp^{0.50}$=1.65), compared to when the root cause is not understood.

To measure the effect of high CRT, we used an indicator variable, *CRT High (2,3),* and treated *CRT Low (0,1)* as the reference category. The results in column 2 show that compared to subjects who answered 0 or 1 CRT questions correctly, subjects who answered 2 or 3 CRT questions

correctly (*CRT High (2,3)),* were 70% less likely to recall ($exp^{-1.18}$=0.30), confirming t-test results related to CRT

In column 3, we see that none of the dispositional control variables significantly predict recall likelihood with the sole exception of *Collaborative FDA rel*., which is marginally significant (p<0.10). The result implies that when a subject perceived their company's relationship with the FDA to be collaborative, he/she was less likely to make the recall decision.

Table 3.7 Logistic Regression- Recall Likelihood

	1	2	3
Physician Concern	-0.05	-0.08	-0.09
	(0.18)	(0.19)	(0.19)
Defect Undetectable	0.39*	0.35+	0.36+
	(0.18)	(0.19)	(0.19)
Root Cause Underst.	0.50**	0.46*	0.43*
	(0.18)	(0.19)	(0.20)
CRT High (2,3)		-1.18**	-1.43**
		(0.41)	(0.45)
Male			0.27
			(0.42)
Quality			-0.32
			(0.41)
Collaborative FDA rel.			-0.67+
			(0.40)
0-2 years at company			-0.04
			(0.55)
3-5 years at company			0.64
			(0.62)
6-10 years at company			-0.26
			(0.57)
Ln_Response time			0.20
			(0.22)
Constant	-1.04***	-0.18	-0.18
	(0.19)	(0.35)	(0.85)
Observations	167	167	167
Wald Chi^2	11.77	18.38	23.51

Standard errors in parentheses + p<0.10, *p<0.05, **p<0.01
Robust standard errors used

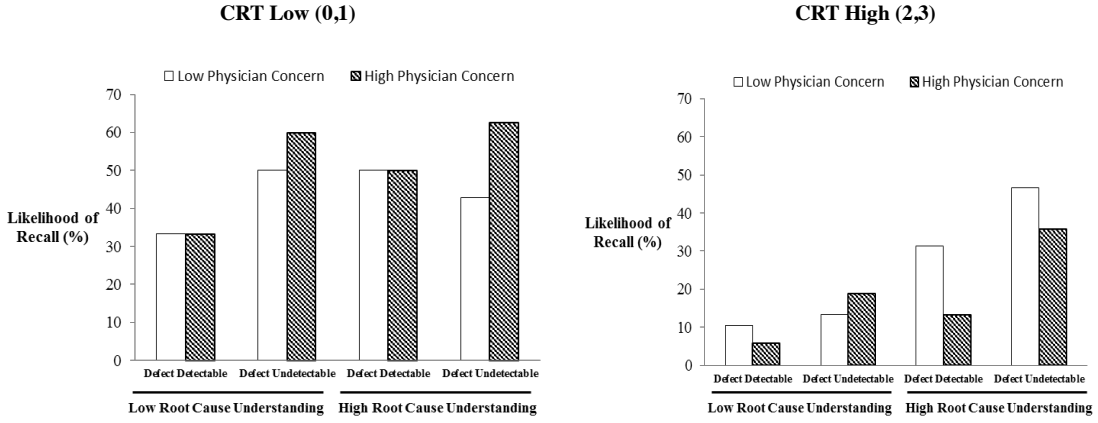
3.7 Post-Hoc Analysis: Moderating effects of CRT

It is noteworthy that the most influential factor in the t-tests and regression analyses, both in statistical significance and effect size, is a subject's CRT score. High CRT subjects, who by

definition reflect more before making a decision, more frequently decided not to recall compared to low CRT subjects. This could be because more reflective subjects have more patience and choose to delay the recall decision pending further data analysis. The unanticipated importance of CRT compared to all other factors in the experiment makes it important to perform additional analysis to delineate the role of a subject's CRT score on the relationships between the situational factors or the dispositional control variables, and the likelihood of recall. Thus, we conducted a series of post-hoc analyses by splitting our sample based on CRT scores. We reexamined the hypothesized relationships of the situational factors separately within low and high CRT categories. Similar to previous analyses, we first visually plotted these relationships, and then performed t-tests and logistics regression to statistically validate the results.

The effect of CRT on the relationship between the situational factors and the likelihood of recall is shown in Figure 3.4. We can make a few observations. First, *Physician Concern* appears to remain insignificant in both low and high CRT categories, as the difference between the white and shaded bars in likelihood of recall appears random in both low and high CRT categories. *Defect Undetectable* appears to be somewhat significant in both low and high CRT while *Root Cause Understanding* seems to be more relevant in the CRT High (2,3) category than in the CRT Low (0,1) category.

Figure 3.4 CRT Scores, Situational Factors, and the Recall Decision



We next repeated t-tests for the situational factors at low and high CRT categories. Table

3.8 displays t-test results for each situational factor after splitting the sample by low and high CRT scores. For ease of comparison, the first row for each situational factor (row a) in Table 3.8 repeats the results from the full sample t-test from Table 3.4. Rows b-c display the low and high CRT results, respectively. In the full sample, *Physician Concern* had no predictive significance on the likelihood to recall. This conclusion holds both at low and high levels of CRT. *Defect Undetectable* was a positive and significant predictor of recall likelihood in the full sample. This relationship does not persist in the CRT Low (0,1) category (p-value=0.531), and is only marginally significant in the CRT High (2,3) category (p-value=0.066). *Root Cause Understanding* was a positive and significant predictor of recall likelihood in the full sample, however this relationship only holds in the CRT High (2,3) category (p-value=0.006), and is insignificant in the CRT Low (0,1) category (p-value=0.752).

Table 3.8 T-test Results for Recall Likelihood-Split on CRT Score

Factor	Row	Sample	Factor Levels[a]						
			Low			High			
			Recall	N	%	Recall	N	%	p-value[b]
Physician Concern	a	Full sample	24	83	29	23	84	27	0.827
	b	CRT Low (0,1)	8	18	44	12	22	55	0.537
	c	CRT High (2,3)	16	65	25	11	62	18	0.348
Defect Undetectable	a	Full sample	17	83	20	30	84	36	0.028
	b	CRT Low (0,1)	7	16	44	13	24	54	0.531
	c	CRT High (2,3)	10	67	15	17	60	28	0.066
Root Cause Understanding	a	Full sample	15	82	18	32	85	38	0.005
	b	CRT Low (0,1)	7	15	47	13	25	52	0.752
	c	CRT High (2,3)	8	67	12	19	60	32	0.006

[a] Number and percentage of subjects deciding to recall in each cell
[b] Null hypothesis: Number of recall decisions at low factor level = number of recall decisions at high factor level

We also repeated the logistic regression analysis after splitting the sample by CRT scores (Table 3.9). For comparison purposes, we include the full sample results from Table 3.8 in columns 1 and 2 of Table 3.9. In the CRT Low (0,1) category, none of the situational factors are statistically significant (column 3). Only after the dispositional control variables are entered into the regression (column 4) does *Defect Undetectable* become positive and significant. Interesting relationships are observed in column 4 between the dispositional control variables the likelihood to recall for the CRT Low (0,1) category. Unlike the results in the full sample, *all* dispositional control variables

are significant in the CRT Low (0,1) category. Compared to female subjects, male subjects are 96% ($exp^{-3.17}$=0.04) less likely to decide to recall. Compared to managers in non-Quality roles, managers in the Quality department are surprisingly 98% ($exp^{-3.71}$=0.02) less likely to decide to recall. Those who perceive the company's relationship with the FDA to be collaborative are 99.9% less likely to recall than those who perceive the company's relationship with the FDA to be average or confrontational ($exp^{-8.73}$=0.0001). Less experienced subjects, in comparison to those who had more than 10 years of experience at the company, are more likely to decide to recall, as all experience indicator variables are significant in the CRT Low (0,1) category. Response time is also a positive predictor of recall likelihood in this column. For CRT Low (0,1) subjects, taking more time to respond in the experiment is positively associated with a higher likelihood to recall.

Table 3.9 Logistic Regression- Recall Likelihood-Split on CRT Score

	Full Sample		CRT Low (0,1)		CRT High (2,3)	
	1	2	3	4	5	6
Physician Concern	-0.05	-0.09	0.21	-2.01^{+}	-0.23	-0.22
	(0.18)	(0.19)	(0.33)	(1.04)	(0.23)	(0.24)
Defect Undetectable	0.39*	0.36^{+}	0.22	1.39*	0.43^{+}	0.46^{+}
	(0.18)	(0.19)	(0.33)	(0.67)	(0.23)	(0.25)
Root Cause Underst.	0.50**	0.43*	0.10	0.00	0.62**	0.59*
	(0.18)	(0.20)	(0.33)	(0.76)	(0.24)	(0.26)
CRT High (2,3)		-1.43**				
		(0.45)				
Male		0.27		-3.17^{+}		0.55
		(0.42)		(1.68)		(0.58)
Quality		-0.32		-3.71^{+}		-0.38
		(0.41)		(2.12)		(0.49)
Collaborative FDA rel.		-0.67^{+}		-8.73**		-0.35
		(0.40)		(2.79)		(0.49)
0-2 years at company		-0.04		10.31**		-0.42
		(0.55)		(3.74)		(0.63)
3-5 years at company		0.64		6.25*		0.37
		(0.62)		(2.61)		(0.66)
6-10 years at company		-0.26		14.58**		-1.41^{+}
		(0.57)		(4.81)		(0.77)
Ln_Response time		0.20		3.92**		0.04
		(0.22)		(1.24)		(0.30)
Constant	-1.04***	-0.18	-0.09	-9.30*	-1.43***	-1.24
	(0.19)	(0.85)	(0.34)	(3.75)	(0.25)	(1.04)
Observations	167	167	40	40	127	127
r^2	11.77	23.51	0.89	19.98	10.09	21.62

Standard errors in parentheses $^{+}$ p<0.10, *p<0.05, **p<0.01
Robust standard errors used

The results for CRT High (2,3) are considerably different from CRT Low (0,1). Both *Defect Undetectable* and *Root Cause Understanding* become positive predictors of recall likelihood in the CRT High (2,3) category (Table 3.9, column 5), however *Defect Undetectable* is only marginally significant, similar to the t-test results in Table 3.8. These results hold after including all dispositional control variables, and none of those variables are significant for CRT High (2,3) (Table 3.9, column 6).

In sum, by repeating our analyses in a split sample based on low and high CRT scores, we find that the two important situational factors in the full sample (*Defect Undetectable* and *Root Cause Understanding*) are only significant for subjects who scored high on the CRT. Low CRT subjects are not influenced in the recall decision by any of the situational factors. However, dispositional control variables, which were not significant in the full sample or in the high CRT category, are all significant predictors of the likelihood to recall in the low CRT category. We discuss these results and their implications in Section 3.8.

3.8 Discussion and Implications

This study has revealed several key findings that can be used by managers and regulators to improve the recall decision-making process. First, concern voiced by a single influential physician appears to have no significant impact on the likelihood to recall. This finding is contradictory to previous behavioral research which identifies the influence of salient, outspoken customer feedback on managerial decisions. Perhaps the nature of this marketplace, where the purchasing customer is separate from the customer who is most intimately affected by product quality, explains the departure from past research. Senior managers at the company where our subject pool was drawn were pleased with this result, since they believe recall decisions should be based on objective data and not influenced by any single, albeit important, customer.

Second, the likelihood of a recall appears to be significantly impacted by the technical characteristics of the potential defect, in particular whether the defect could be detected by the physician customer (and thus corrected) before use. When a potential defect is undetectable, managers are more likely to recall, perhaps in an effort to safeguard patient safety. Interestingly, this implies that the opposite is also true. When managers know that a potential defect would be

detectable by the physician customers pre-use, they are more likely to avoid recalling the product, presumably because they trust physicians to serve as a final quality screen to safeguard patient customers. Results of the physician concern and defect detectability factor indicate a general preference by managers to focus more on the patient customer's safety and less on the physician customer's opinion.

Third, the likelihood of a recall also is impacted by information surrounding the cause of a potential defect. When managers better understand the root cause of product failures, they are more likely to make a recall decision, indicating the relevance of attribution theory in the product recall decision. This result may explain popular media covered recalls in which companies failed to recall products, possibly because they did not understand root cause.

Finally, independent of these three situational factors, we also find that an individual manager's cognitive reflection level also contributes to the propensity to recall. Specifically, subjects with high cognitive reflection levels (i.e., high CRT scores) are less likely to recall compared to those with low cognitive reflection. This dispositional factor is more significant in p-value and effect size than any other factor. Post-hoc analyses further reveals that low CRT managers are not influenced by any of the situational factors. This outcome appears consistent with theory supporting the CRT (Frederick, 2005). Low CRT individuals are thought to make decisions based more on "gut" instinct, emotions and relationships and less upon data. Low CRT subjects in our setting were influenced by the dispositional control variables, such as their experience and their perceived relationship with the FDA. Defect detectability and root cause, which are both positive predictors of recall likelihood in the full analysis, only influence high CRT managers. High CRT individuals are thought to make decisions based more on analysis and reflection. Because of this, the root cause of the defect and the related elimination of randomness as a cause of the product failures matters most for this group.

Implications

Several implications for managers and regulators emerge from these findings. First, our results reveal that managers may implicitly be factoring in the possibility of using physicians as a final quality control step when making a recall decision. Senior managers from our partner company saw this as a problem that they will need to guard against through training. They felt

strongly that the decision to recall should be based on an assessment of quality of products at the time they are handed off to physician customers, rather than the predicted quality of products after intermediary steps that the company does not control. Relying on physicians as a final quality screen for defective products is something that both our partner company and the FDA deem undesirable.

Second, managers should be wary of waiting too long for additional failure analysis or root cause data before recalling. This tendency is observed in both the *Root Cause Understanding* factor and the CRT factor results. In both cases, managers demonstrated a propensity to wait for more data instead of taking action. Determining true root cause of a complex medical device failure is not a simple task. Through our interviews with our partner company, we learned that it could take months to conclusively know the cause of a failure. The FDA, in response to our findings, emphasized that root cause does not need to be determined before initiating a recall and should not necessarily serve as recall decision criteria. Managers should consider recalling products when they observe a significant increase in failure rates, even in the absence of understanding root cause. Our partner company is using this finding to calibrate their recall decision-making process, and ensure that recall decisions are made in a timely manner, not necessarily contingent upon root cause understanding.

A final managerial implication relates to a managers' cognitive reflection levels. Those managers who are more intuitive in their decision-making practices (low CRT), need to ensure they look beyond their individual predispositions and fully account for situational characteristics related to the product failure when making recall decisions, guarding against making "knee-jerk" reaction recall decisions. Managers with high CRT need to ensure they act decisively in a timely manner, and do not unnecessarily postpone the recall decision, failing victim to "analysis paralysis." Companies can use these results to expand recall decision-makers' understanding of the effect that their level of cognitive reflection or personal predispositions have upon deciding to recall, making the recall decision more objective. Managers with low CRT scores also exhibited a correlation between their perceived relationship with the FDA and likelihood to recall, which was viewed as a potential concern in our interviews with the FDA. FDA has recently invested in efforts to reduce combativeness and increase the sense of collaboration with companies. However, our results suggest that this improved relationship could lead to a lower likelihood of recall for managers with

low CRT scores. This highlights the FDA's need to strike a careful balance in their relationships with manufacturers. Relationships between regulators and companies perhaps require a certain amount of healthy animosity to reduce unwanted effects of familiarity.

Chapter 4:

Slow or Fast? An Empirical Examination of the Recall Responsiveness Dilemma

4.1 Introduction

"*General Motors doesn't know why it took 10 years to issue a recall*." [25]
- Mary Barra, CEO of General Motors. 4/1/14. Testimony before U.S. Congress

Executives at General Motors deliberated for years before deciding to recall millions of automobiles with faulty ignition switches. Although moving quicker in this case may have saved customer lives, a firm's sluggish response to obvious quality problems and a delay in recalling products is neither an isolated nor novel event. Delays in recalling products occur frequently and regularly, and across many industry sectors as diverse as food, medical device, automotive, and pharmaceuticals. In the United States, most of these industries are highly regulated and rigorously monitored by various government agencies such as Consumer Products Safety Commission (CPSC), National Highway Traffic and Safety Administration (NHTSA), and the Food and Drug Administration (FDA). Although regulatory agencies maintain strict oversight, advocate fast, timely response to quality problems as they emerge, and penalize offenders, firm responsiveness in the recall process appears to be inefficient and suboptimal as illustrated by many recent recall examples in the popular media. In 2005, Guidant Corporation, a manufacturer of implantable cardiac devices, was criticized by the Senate, federal regulators and the media for a slow recall response after receiving pressure from key physician customers to recall a certain product and after at least one patient death.[26] A recent Government Accountability Office report further corroborates the slow responsiveness by firms when recalling defective products.[27] A separate report has also indicated that regulators, like the FDA, can be slow in responding to product quality problems.[28]

The implicit rationale underlying this line of argument is that faster responsiveness in

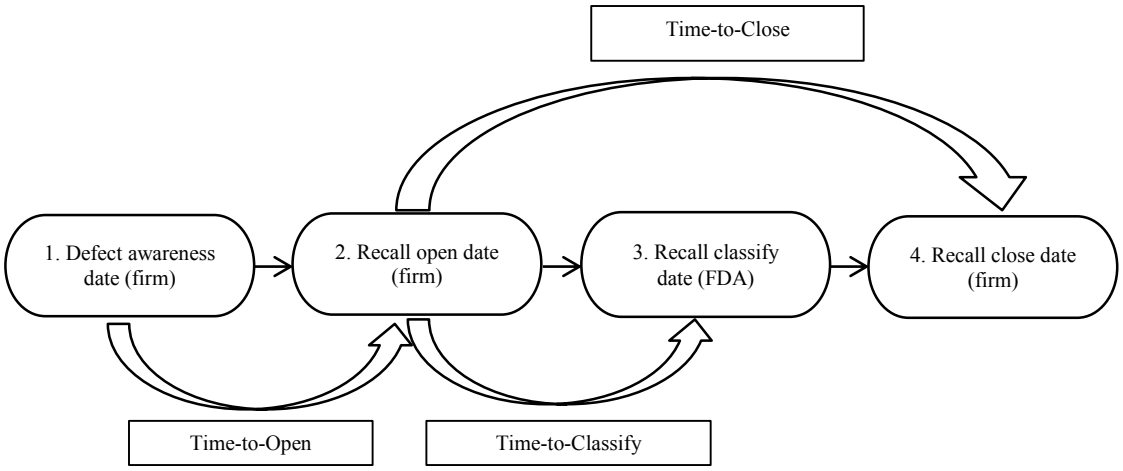
[25] http://money.cnn.com/2014/04/01/news/companies/barra-congress-testimony/
[26] http://www.nytimes.com/2005/05/24/business/24heart.html?pagewanted=all&_r=0
[27] http://www.gao.gov/products/GAO-05-51
[28] http://www.gastroendonews.com/

identifying and addressing quality problems by firms and regulators will result in better outcomes for all stakeholders. Not surprisingly, customers, manufacturers, and regulatory agencies continually advocate for speed in recalling products from the market, even though the relationship between responsiveness and future recalls has not been empirically proven. Furthermore, it is not clear whether firms and regulators respond to certain types of recalls with faster or slower speed. That is, we do not know if there is a discernable pattern in firm and regulator responsiveness to recalls.

In this paper, we address these research gaps by studying two related questions: 1) What leads to quick recall response times for firms and regulators, and 2) how does such responsiveness impact future recall rates? We study our questions by deconstructing the recall process into its constituent phases and using data compiled in collaboration with the FDA. For this study, we worked with the FDA because in the United States, it has regulatory authority over food, pharmaceutical, medical device, and cosmetic recalls; products which account for approximately $1.6 trillion of annual consumer spending.[29] Figure 4.1 shows the four most critical steps in the recall process. These steps are very well-specified and clearly laid out, but the FDA does not formally mandate how much time should be spent on any individual step. Intuitively, taking a long time between any two steps seems undesirable, yet anecdotal evidence suggests that there is considerable variation in the time taken to execute the recall process. Previous researchers have paid little attention to the underlying causes of variability in recall responsiveness and its subsequent impact on future recalls. A notable exception includes Hora et al. (2011), who showed that firm's recall strategy, defect type, and supply chain player initiating the recall, each has an impact on time-to-recall in the U.S. toy industry. Our study builds on Hora et al. (2011) but uses a finer measure of responsiveness by decomposing time-to-recall into its three separate constituent intervals using unique time-stamped data. We obtained the actual time and date stamps corresponding to the four steps for every medical device recall that occurred in the US between 2003 and 2013 from the FDA. We then computed the three different intervals representing the recall process. These are time-to-open, time-to-classify, and time-to-close a recall (Figure 4.1).

[29] http://www.bls.gov/news.release/cesan.nr0.htm

Figure 4.1 Simplified Recall Process and Critical Steps



It is important to divide time-to-recall into three separate time intervals instead of using a composite measure, because first, different entities are responsible for enacting them, and second, each of them exposes the decision maker to a different set of risks and uncertainties. Specifically, whereas a firm is responsible for opening and closing a recall, the FDA is responsible for categorizing the recall into distinct classes to indicate the potential severity of the problem. Further, the decision to open a recall compared to the decision to close one entails a different set of risks and uncertainties for the decision maker and for the firm. Arguably, the same predictor could have a dissimilar impact on the decision to open and close recalls. Thus, discriminating between the three time intervals allows us to estimate examined relationships more accurately. In this paper, we identify recall severity as a predictor of firm and FDA responsiveness because it a key characteristic differentiating recalls. Some recalls are very serious, demonstrating a risk to life if the products are not quickly recalled from the marketplace. Other recalls are more administrative, intended to correct an obscure labeling mistake that may not have ever been noticed by customers. It is likely that such diverse types of recalls lead to different responses from firms and from regulators. We draw upon prospect theory in economics (Kahneman and Tversky, 1979) to ground our relationships between severity and its impact on firm responsiveness, measured as the three time intervals. We then explore how responsiveness at the firm and the FDA effects future product recalls. We answer our questions using a unique dataset spanning 11 years and over 4000 product recalls, and rigorous econometric methods.

Our results make three important contributions which have the potential to inform current literature and impact managerial and regulatory practices related to product recalls. First, we find that the severity of a recall is associated with opening and closing recalls but not always in the hypothesized direction. Specifically, consistent with our hypothesis, firms take significantly more time-to-open more severe recalls compared to less severe recalls. Firms take 10 additional days to open high and moderately severe recalls compared to the least severe recalls, a 20% increase in time-to-recall. Unfortunately, the most severe recalls are those that put the public at the highest risk and which require the fastest response times from the firm. Contrary to our hypothesis, firms also take the longest time-to-close the most severe recalls. High severity recalls take an extra 76 days (23% increase), and moderately severe recalls an additional 250 days (72% increase) to close in comparison to the least severe recalls. We also find that consistent with federal guidelines and FDA policy objectives, the FDA classifies the most severe recalls the fastest.

Related to future recalls, our results show that closing recalls slowly reduces future recalls. A one standard deviation change in the time-to-close a recall (298 days or 10 months) leads to two fewer recalls per plant across the time of the panel, while half a standard deviation change (140 days or 4.5 months) leads to one less recall per plant. Surprisingly, the time-to-open a recall has no effect on future recalls. Apparently, while moving quickly to open a recall has significant customer upside (by quickly removing risky product from the marketplace), it has negligible firm downside (no apparent lost learning). We also find that FDA's responsiveness in classifying recalls reduces future recalls, but this effect is statistically significant only for the least severe recalls. As a set, our results suggest that the relationship between responsiveness and future recalls is nuanced and varies with recall severity. These results while counterintuitive have critical implications – conventional wisdom and economic theory suggests that firms should open and close more severe recalls faster, and such faster responsiveness should reduce future recalls. However, our results indicate that in reality, the opposite is true: higher severity slows firm responsiveness, and slower responsiveness leads to fewer future recalls.

4.2 Theory and Hypotheses

In this research, we examine predictors of recall responsiveness and investigate whether fast responsiveness reduces future recalls. Our research is informed by three distinct but related streams

of literature. Because recalls are a type of external failure cost, we review the product quality literature broadly to provide foundation and motivation to our research questions. From a managerial perspective, the decision to recall a product is inherently risky and fraught with uncertainty and unpredictability. We review literature related to prospect theory to understand how riskiness and uncertainty impact recall responsiveness. Previous researchers have also suggested that the level of employee discretion in task completion may impact the quality-speed tradeoff (Hopp et al., 2007). Therefore, we examine the literature salient to discretionary tasks in developing the relationship between responsiveness and future recalls.

Product Quality and Recalls

Product recalls occur when systemic product defects, due to flawed design or poor manufacturing, are not detected until the product is in the marketplace. In the quality literature, such quality failures are examples of external failure costs for the firm (Juran, 1999). Although quality literature is too vast to cover here, we can glean three overarching themes from reviewing the relevant literature. First, we find that a majority of empirical literature related to failure costs focuses on internal quality performance (Banker et al., 1990; Datar et al., 1993; Mukherjee et al., 2000) and qualitative measures of external quality performance (White et al., 1999; Fynes and Voss, 2001; Ahire and Dreyfus, 2000). Second, our review also found studies that have examined external failures generally, and product recalls specifically (Thirumalai and Sinha, 2011; Jarrell and Peltzman, 1985; Chen et al., 2009). A majority of these treat "product recall" as an independent variable predicting various measures of firm performance. Finally, there are only a handful of studies that have examined organizational and operational causes of product recalls using secondary data and rigorous research methods (Haunschild and Rhee, 2004; Thirumalai and Sinha, 2011; Shah et al., 2014). The paucity of research with external quality failures as the dependent variable suggests a research gap and a potential to make an important contribution to product failure literature. The rest of our review focuses on research related to product recalls.

Recently, researchers have begun to explore plant and firm level causes of recalls in order to reduce and prevent future recalls. For instance, Thirumalai and Sinha (2011) find that firms with broad product lines are more susceptible to a recall compared to firms with narrow product lines. Using automotive recall data, Haunschild and Rhee (2004) show that recalls initiated by the

automaker leads to fewer future recalls, than recalls mandated by regulators. The authors underscore the importance of learning in recall prevention, and reason that a firm's voluntary initiation of recall is associated with deeper reflection and learning. The most closely related to our study is Hora et al., (2011), who examine product and strategy drivers of a manufacturer's time-to-recall. Using data from the toy industry, the authors demonstrate that managers take more time-to-recall design related defects than manufacturing defects. Additionally, the authors find that firms which adopt preventive recall strategies take longer to recall compared to firms espousing more reactive strategies. Hora et al. (2011) measure time-to-recall as the time interval between recall date and product launch date. However, by not measuring the time-to-recall from the time when the firm first became aware of the defect, their measure represents a much longer duration. It is quite conceivable that the product may have been launched many months if not years before the defect was actually observed. Even so, this seminal study constitutes an important first step towards furthering our understanding of recall responsiveness. We build upon quality and recall literature by developing a more precise measure of recall responsiveness using actual time stamps characterizing the four phases of the recall process. Using these more rigorous measures, we substantiate conclusions from past research and uncover new relationships.

Risk, Uncertainty, and Responsiveness

In developing our hypotheses, we draw upon prospect theory (Kahneman and Tversky, 1979), which is foundational to studying decision making under risk and uncertainty. One of the most pertinent characteristics of prospect theory is that it distinguishes between risk-averse and risk-seeking behavior of decision-makers facing uncertainty. In their seminal study, Tversky and Kahneman (1981) demonstrated that when a problem with uncertainty is posed in terms of a gain, decision makers tend to be risk-averse and when the same problem is posed in terms of a loss, risk-seeking behavior is observed. They conclude that when facing an uncertain situation, decision-makers are *risk-averse in gains* and *risk-seeking in losses*. This behavior has been substantiated in multiple other studies (Bateman and Zeithaml, 1989; Fishburn and Kochenberger, 1979; Devers et al., 2007; Shefrin and Statman, 1985; Ferris et al., 1988).

From a managerial perspective, opening a recall is associated with incurring a "sure loss" today and the decision to open a recall is accompanied with inherent risk and uncertainty. The losses

are immediate and occur the instant a decision to open a recall is made, although they vary greatly in magnitude, depending upon the severity of the recall. Losses come in many forms including negative publicity generated by the recall, and its subsequent negative impact on the firm's reputation. It is not unusual for the FDA to suspend new product approvals for firms with recalls, until all existing recalls are brought under control.[30] Thus, recalls significantly affect a firm's future revenue and result in an immediate increase in costs impacting a firm's current profits. Generally Accepted Accounting Practices (GAAP) requires firms to report any known and sure future costs in the form of accruals. When a manufacturer decides to recall a product, the costs to repair or replace the affected products have to be estimated, and instead of absorbing recall costs slowly as they occur over time, manufacturers are required to accrue them immediately in their financial statements which negatively impacts earnings in the current time period.

At an individual level, recalls can have serious negative repercussions on a manager's career. Individuals involved in decision-making associated with defective products can be blamed and held responsible for the product failure, contributing to a real and perceived reduction in job security. Losses vary greatly across different recalls and increase significantly with the severity of the recall. The more serious the potential product failure and related recall, the more serious the consequences may be on the firm and the individual manager's career who is responsible for making the recall decision. Postponing the decision to open a recall to a later date may delay the loss or reduce the probability of incurring the loss to less than 100 percent. In other words, postponing the decision does not just delay the recall, it may obviate the need for a recall, if product failures abate or attention to the issue wanes. However, the delay is also likely to increase the magnitude of the loss as the impacted units are likely to increase in number and affect more customers over time. In sum, the decision to open a recall requires managers to choose between a sure loss of a known magnitude today versus a less than sure loss of an uncertain higher magnitude at a future date. As recall severity increases, the associated uncertainty and risk also rises, resulting in an increased chance of delaying the decision to open the most serious recalls.

Hypothesis 1A: Severity of a recall is negatively associated with time-to-open. Specifically, more serious recalls have longer time-to-open.

[30] http://online.wsj.com/news/articles/SB10001424052748703909804575123441387311792

In contrast, closing an open recall quickly is desirable as firms can use this to indicate in-control quality processes to customers, stock holders, and regulators. The longer a recall remains open, the greater unwanted exposure and negative publicity the firm receives from media. It may also face more intense regulatory pressure and greater potential loss of customers and market share. Generally, recalls which remain open longer have higher overall costs associated with them, and the costs increase with the severity of the problem. The higher the severity of the recall, the higher the tangible and intangible costs. These costs are likely to increase proportionally with the time that the recall remains open. Thus, in an effort to reduce the riskiest and more certain loss, decision makers are likely to close the most serious recalls fastest, demonstrating risk-seeking behavior in their decision to close a recall. It is possible that closing recalls too quickly has disadvantages because a premature closure may hinder managers from developing a comprehensive understanding of the defect leading to future recalls. However, the risk of losing potential learning associated with closing a recall too soon is less tangible and less objective than the risk of negative publicity and exposure accompanying an open recall. It is reasonable to expect managers would lean towards closing recalls quickly especially the more serious recalls because these are riskier and more costly.

Hypothesis 1B: Severity of a recall is positively associated with time-to-close. Specifically, more serious recalls have shorter time-to-close.

Hypotheses 1A and 1B are also supported by the rationale guiding a manager in a profit-maximizing firm, who should behave in a manner which enhances profits most assuredly. In deciding to open or close a recall, the profit-maximizing firm should open recalls slowly and close recalls quickly. The longer the firm waits to open the recall, the more positive effect on profits. Similarly, the faster it can close the recall, the better it is for firm's bottom line. This relationship should be further accentuated by the severity of the recall: the most serious recalls should be opened slowest and closed quickest.

Regulators such as the FDA may however be motivated to act in an opposite manner as the firm. The FDA is tasked with regulating the medical device industry and protecting the public from harmful devices. One of the responsibilities of the FDA is to classify product recalls after they are initiated by the firm. The FDA classifies the recalls according to the risk posed to the public with the objective of classifying the most severe recalls the fastest. In accomplishing this goal, the FDA

quickly alerts physicians, customers, and manufacturers, as to the seriousness of an issue. In this and in other roles, the FDA acts as an agent of the principle, which is in this instance is the taxpaying US consumer. The consumer is unable to properly monitor and track medical device product performance and risk, so the FDA performs that work on behalf of the taxpayer. Agency theory (Jensen and Meckling, 1976) informs us that if the FDA is fulfilling this responsibility in the best interests of the principle taxpayer, they should classify the most serious recalls the fastest to ensure that the consumer is alerted to this danger and can alter their behavior accordingly. We therefore hypothesize:

> *Hypothesis 1C: Severity of a recall is positively associated with FDA's time-to-classify. Specifically, more serious recalls have shorter time-to-classify.*

Responsiveness and Future Performance

In the process of opening a recall, quality and reliability engineers systematically sift through failure data to identify underlying patterns and similarities. Likewise, to close a recall, engineering personnel perform numerous tests and analyses to uncover the root cause of the problems. Although the decision to open or close a recall is based on rigorous data analysis and well-specified guidelines, it entails a considerable amount of educated guesswork and subjective assessment. That is, opening or closing a recall do not have well-established, concrete completion criteria, providing considerable flexibility to decision makers in determining when and how to end the tasks. Such *discretionary* tasks are common to many professions and frequently require knowledge workers to assess and determine when and how tasks are completed. This is in contrast to most assembly line tasks which have definite completion criteria and a strict end-point, leaving employees little discretion in deviating from them. Hopp et al., (2007) suggest that discretionary task completion allows workers to manage their workload by adjusting the quality of the output. They conclude that professional workers in discretionary jobs may consciously use quality as a lever to balance increased work-load. That is, in the absence of strict completion criteria, workers may choose to terminate their work prematurely and alter quality in order to balance an increased work load.

This trade-off between quality and speed of work in professional tasks, characterized as the *quality-speed conundrum* (Anand et al., 2011) has been empirically demonstrated in multiple settings (Tan and Netessine, 2014; Oliva and Sterman, 2001; Powell et al., 2012; Kunz et al., 2014;

Gans et al., 2003). Using panel data in a restaurant setting, Tan and Netessine (2014) find that as server workload and processing speed increases, quality measured as server sales efforts decreases – that is, the server consciously chooses to spend less time upselling pricier menu items. Powell et al. (2012) find similar relationships in a hospital setting. They show that as the physician work-load increases, the quality of their administrative work decreases significantly. This quality-speed relationship has also been demonstrated in other healthcare (Kuntz et al., 2014) and banking industry (Oliva and Sterman 2011) studies.

Faster responsiveness to open or close a recall may prevent a more thorough understanding of the product quality problem. When a firm moves too fast to open a recall, they may hinder comprehensive defect understanding by limiting the number and types of failure modes exhibited in the marketplace. Moving too fast to close a recall may prevent accurate identification of root cause of the defect. In either case, a limited understanding of the defect and its cause may exist. This limitation can hamper learning and reflection resulting from the recall. Past research has demonstrated the importance of such recall learning as a mechanism to reduce future recalls (Haunschild and Rhee, 2004).

Hypothesis 2A: Time-to-open is negatively associated with future recalls, such that plants with shorter time-to-open will have more future recalls.

Hypothesis 2B: Time-to-close is negatively associated with future recalls, such that plants with shorter time-to-close will have more future recalls.

FDA's responsiveness in classifying recalls may also have an effect on learning and future recalls. The purpose of the FDA classification is to alert the general public, physicians, and manufacturers regarding the severity of the recall. If a firm believes that a recall is not severe, and FDA classifies the recall similarly, the firm's perceptions are confirmed. If the FDA classifies differently from the firm's beliefs, the firm's perceptions are calibrated closer to regulator's perceptions. Conceivably, the sooner this classification occurs, the more likely that this calibration will resonate within the organization and lead to learning and product quality improvements. When the classification process is far removed from the time that the firm initiated the recall, the impact of this calibration is likely reduced. It is therefore possible that FDA classification speed may be associated with future recalls.

Hypothesis 2C: Time-to-classify is negatively associated with future recalls, such that when the FDA takes more time to classify a recall, plants will have more future recalls.

4.3 Research Setting, Data, and Empirical Strategy

We examine our research questions in the US-based medical device industry for two reasons. First, firms in the medical device industry represent a substantial proportion of the U.S. economy, and are diverse in size and geographic spread. By some estimates, the medical device industry is expected to exceed $130 billion in market size and 6000 firms in number by 2016.[31] Second, recalls happen regularly with great frequency, are often life-threatening to consumers, and occur because defects in increasingly complex products may go undetected after production. Thus, identifying factors that can result in reducing recalls in this industry is of critical importance to researchers, regulators, and managers on one hand, and to the overall economy on the other hand.

FDA defines a medical device recall as a "firm's removal or correction of a marketed product that FDA considers to be in violation of the laws it administers". Although FDA can mandate and force a manufacturer to recall a product from the market, a majority of recalls are voluntary in nature where a manufacturer (or a distributor) may initiate a recall on their own volition "to protect public-health and well-being from products that present a risk of injury or gross deception or are otherwise defective."[32] Because the medical device industry is highly regulated and closely monitored, the recall process is well-specified, and consists of multiple steps. A simplified representation of the recall process highlighting the most important steps shows that a recall process begins when a manufacturer decides to recall a defective product from the market (Figure 4.1).

To initiate a recall, the manufacturer is required to notify the FDA of their decision to recall the product, and also identify the date when they first became aware of the defect. This is labeled the "defect awareness date" in the official record. The manufacturer is required to contact all the customers and the FDA to notify them of the defect, and any possible corrective action. This initiates the actual recall, and is referred to as the "recall open date". FDA examines the defect notification details and classifies the recall into one of the three classes based on the severity of the

[31] http://selectusa.commerce.gov/industry-snapshots/medical-device-industry-united-states
[32] http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/PostmarketRequirements/RecallsCorrectionsAndRemovals/

defect and the potential harm to the consumer. The date associated with this step is called the "recall classify date." The classification signals FDA's perception of the seriousness of the defect to both the manufacturers and the customers. It also conveys the associated level of urgency required of the manufacturers in their response to customers. The final step in the recall process is for a manufacturer to close an open recall, and the corresponding date is called the "recall close date". Recalls can be closed when the manufacturer has identified root cause and corrective action of the problem, repaired or replaced all the affected units, and communicated this information to the FDA.[33]

The three classes of medical device recalls are labeled class I, II and III in a decreasing order of severity and potential harm to the customer. The FDA describes a class I recall as "a situation in which there is a reasonable probability that the use or exposure to an adverse product will cause serious health consequences or death."[34] A class II recall is one in which use of the product may lead to reversible or temporary customer harm, whereas a class III recall is defined as one in which such dire consequences are highly unlikely. An example of class I recall is correcting an electrical malfunction in a critical component of an implantable defibrillator whereas correcting a product labeling error, and a defective tongue depressor or examination gloves are examples of class II and III recalls respectively. Our data includes all three classes of recalls voluntarily initiated by the manufacturers. While the FDA maintains the ability to mandate a recall, it is very rare for them to do so, and they did not mandate any of the recalls in this data set.

Data for this study were obtained from the FDA in 2014 through a Freedom of Information Act (FOIA) request and covers recalls from 2003 to 2013. The beginning date for our sample is 2003 because data for recall records are not available prior to 2003; the ending date is 2013 as it was the last year for which complete recall data were available. To examine our research questions, we conduct the analysis at two different units of analyses. In our first research question, we seek to understand the relationship between recall severity and our measures of responsiveness, time-to-open, time-to-close, and time-to-classify. Hypotheses 1A-C, corresponding to research question one, examine causes of recall responsiveness and require a "recall" level unit of analysis. In the second research question and the corresponding Hypotheses 2A-C, we seek to understand the effect

[33] http://www.fda.gov/downloads/iceci/compliancemanuals/regulatoryProceduresManual/UCM074312.pdf
[34] http://www.fda.gov/safety/recalls/ucm165546.htm

of past recall responsiveness on future recalls, which requires a higher unit of analysis. To predict future recall occurrences, the analysis must be conducted at the level at which recalls occur, which is the manufacturing plant in the FDA data. To examine the second research question, we measure recall responsiveness and future recall counts at the plant in which the products are manufactured. Our data consists of 358 unique medical device manufacturing plants and 4,394 recalls across the three recall classes. The research design for research questions one and two are summarized in Table 4.1.

Table 4.1 Research Design by Research Question

	Research Question One	Research Question Two
Research Question	What effects recall responsiveness?	How does recall responsiveness effect future recalls?
Hypotheses	1A, 1B, 1C	2A, 2B, 2C
Unit-of-analysis	Recall	Plant-Year
Dependent variables	Time-To-Open,Time-To-Close, Time-To-Classify	Future Recalls at the Plant
Independent variables	Recall Severity	Ln_Process Time
Control variables	Sales	Sales[a]
	Public	Public
	Year	Year
	Ln_Units_Recalled	Lagged class I, II, III recalls
	Recall Root Cause	Ln_Average Units Per Recall

[a] Note: Sales and public are only used in robustness checks for RQ two, as FE analyses de-mean time-invariant unobservables and exclude sales and public.

DEPENDENT VARIABLES

Recall Responsiveness. *Time-To-Open*, *Time-To-Close*, and *Time-To-Classify* are the dependent variables for research question one. *Time-To-Open* is the number of days between the defect awareness date and recall open date. *Time-To-Close* is the number of days between the recall open and close dates. *Time-To-Classify* is the number of days between recall open and FDA classification dates. *Time-To-Open* and *Time-To-Close* are controlled and managed by the plant while *Time-To-Classify* is regulated by the FDA. Figures 4.2-4.3 demonstrate the trend in recall responsiveness for both the plant and the FDA in the data. While plant response times to open and close a recall increased steadily from 2003 to 2011, both demonstrate a recent decreasing trend in 2012 (Figure

3). The time-to-open a recall seems to be faster than the time-to-close, and is less variable, particularly in the most recent years. FDA response time remained consistent from 2003 to 2011, but increased both in variation and in total time in 2012 and 2013 (Figure 4.3).

Figure 4.2 Plant Recall Response Times

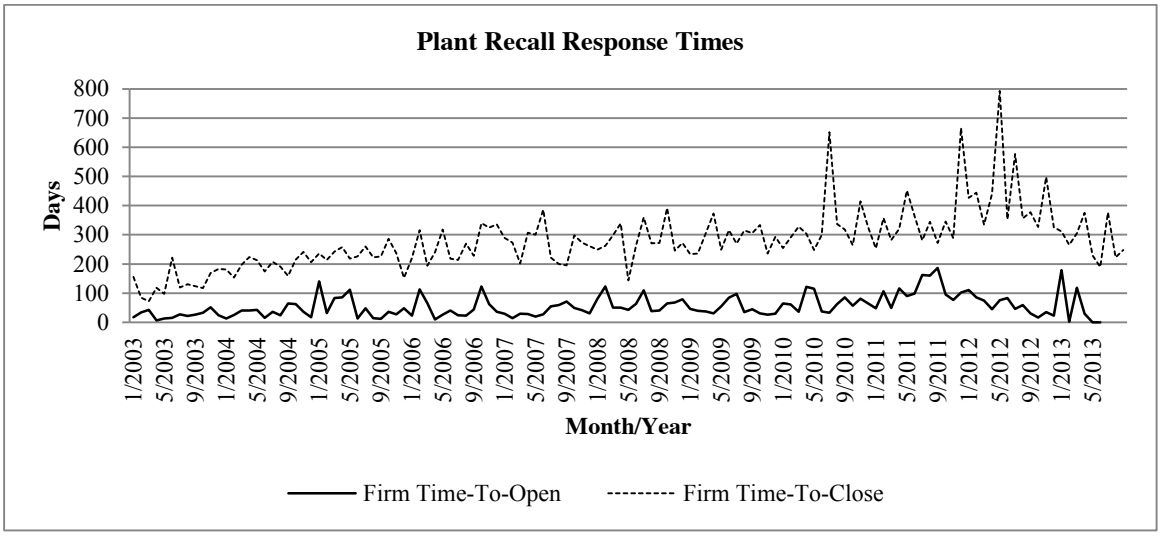
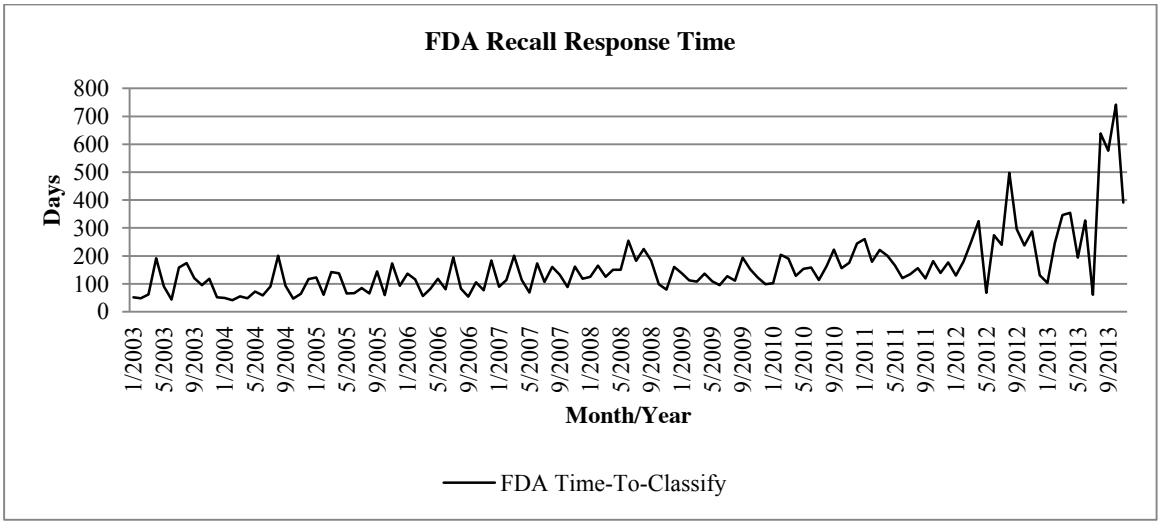
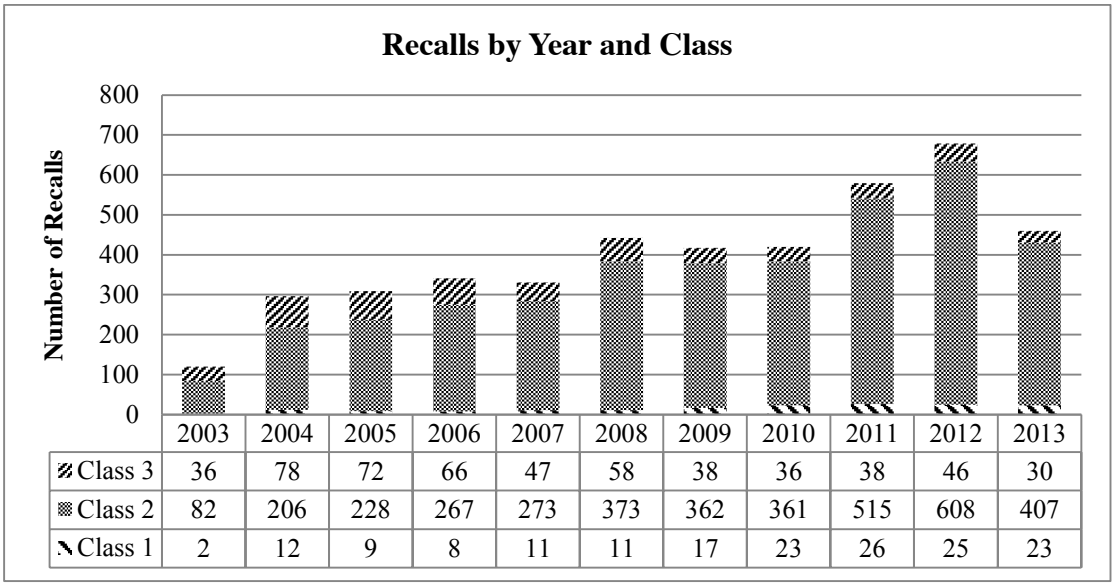


Figure 4.3 FDA Recall Response Time



Future Recalls. The dependent variable for research question two is the count of product recalls initiated on products built in each plant for every year of the panel. Because we seek to understand the relationship between recall responsiveness and subsequent recalls, this count of recalls is a lead measure, measured in the year following the year of measurement of the independent variables. The number of recalls in this industry have been increasing steadily over the past 10 years, with class II recalls accounting for a majority of this growth (Figure 4.4).

Figure 4.4 Recalls by Year and Class



	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Class 3	36	78	72	66	47	58	38	36	38	46	30
Class 2	82	206	228	267	273	373	362	361	515	608	407
Class 1	2	12	9	8	11	11	17	23	26	25	23

INDEPENDENT VARIABLES

Recall Severity. The independent variable for research question one is recall severity. The FDA classifies recalls into three classes ranging from least severe (class III-low severity) to most severe (class I=high severity) with a moderate recall class (class II=moderate severity) in between. For this measure, indicator variables for *High Severity* and *Moderate Severity* are used with *Low Severity* as the reference category.

Process Time. For research question two, the independent variables are the average time a plant takes to open and close a recall and the time the FDA takes to classify a recall in a given year. *Ln_Process Time* for the time-to-open analysis is the natural log of the time between the defect

awareness date and the recall open date. *Ln_Process Time* for the time-to-close the analysis is the natural log of the time between the recall open and close dates. *Ln_Process Time* for the time-to-classify the analysis is the natural log of the time between recall open date and FDA classification date. These variables are logged after averaging these times for all recalls that were opened and closed throughout the year.

CONTROL VARIABLES

Sales: Sales is highly correlated with plant size and is frequently used as a measure of resources available to plants including resources dedicated to quality control and thereby its ability to influence recall responsiveness. Larger sales may also imply future recalls as more products in the market may mean more opportunities for failure. FDA collects annual plant sales in ten separate categories, in regular increments beginning with $0-$25,000 and ending with $50,000,000 and higher. We acquired plant sales data from the FDA and recoded the data from one to ten corresponding to each category provided by the FDA. We use this measure (*Sales*) to control for the effect of plant size and resource availability. The mean sales were 8.52, corresponding to $18 million in annual plant sales.

Public. Publically traded plants may be under pressure to deliver more risky innovation more quickly and may be associated with a higher risk of a recall. An indicator variable is used to identify whether the plant is part of a public or private plant (*Public*).

Units Recalled. Plants are required to inform the FDA of the number of units associated with the product recall when the recall is opened. Due to the negative signals sent to customers and regulators by a large recall, plants may be less responsive in opening a recall as the number of defective units increases. The time-to-close a recall could be impacted by the size as well. The more units affected, the more units requiring repair which may lead to longer close times. The FDA may also classify recalls differently depending upon the number of units affected. To control for these effects, we measure the natural log of the units affected by each recall (*Ln_Units_Recalled*) as a control variable for research question one. For research question two, we average the units affected by all recalls at the plant in the current year to control for the impact on future recalls (*Ln_Avg Units Per Recall*).

Year. FDA policies may change over time and this might impact plant and regulator recall

responsiveness and impact the total number of recalls that occur. We use an indicator variable for the year that the recall was initiated to control for year effects and treat 2013 as the reference category.

Recall Root Cause. The root cause of the recall may affect recall responsiveness. For example, Hora et al. (2011) found that design related toy recalls led to slower time-to-recall than manufacturing recalls because of the heightened complexity associated with design problems. To control for the effect that root cause may have upon responsiveness, we utilized root cause categories assigned by the FDA. In the raw FDA data, there were 42 categories, though many of them had very few occurrence rates and were quite similar to each other in description. To streamline this measure, we classified similar categories together into four distinct groups (*Manufacturing, Software, Design,* and *Miscellaneous*). We used indicator variables for each category and treated *Design* as the reference category for research question one.

Lagged Recalls. To control for the affect that past recall tendencies may have upon future recalls, we counted the number of recalls in the current year and grouped them by recall classification (*Lagged Class I Recalls, Lagged Class II Recalls, Lagged Class III Recalls*).

Empirical Strategy

Research question one uses time-to -open, -close and -classify as three separate dependent variables. When analyzing the effects of covariates on time to an event, hazard models are commonly used (Box-Steffensmeier and Jones, 2007). The Cox Proportional Hazard (CPH) model is widely used because it is flexible and does not require researchers to specify the underlying hazard rate. However, an important assumption of the CPH is that the hazard rates (the effects of covariates on failure time) do not change across time, and when this proportionality assumption is violated, coefficients from the CPH model can become biased. To test for the appropriateness of CPH model to our data, we tested for the constancy of hazard rates over time by examining the statistical significance of Schoenfeld residuals (if residuals are significant, proportionality does not hold) (Box-Steffensmeier and Jones, 2007). Schoenfeld residuals were calculated for time-to -open, -close, and -classify for our independent variables of interest in research question one (recall severity). *Moderate Severity* for time-to-close, and *High* and *Moderate Severity* for time-to-classify demonstrated a lack of proportionality. All hazards are however proportional for time-to-open.

There are three ways to compensate for non-proportional hazards: 1) stratify the data based on specific groups which cause the non-proportionality, 2) include time interaction variables in the analysis, and 3) use a non-proportional accelerated failure time (AFT) model. As there are no clear strata for our data, we do not compensate using stratification. Instead, we include time interaction variables in the CPH analysis as necessary (*Moderate Severity* for time-to-close and *High* and *Moderate Severity* for time-to-classify). Our primary hazard model is a recurrent event CPH that accounts for shared variance that may occur when the same plant experiences multiple recalls per year (Box-Steffensmeier and Jones, 2007). We also include non-proportional AFT models in the analysis as additional verifications that our results hold, independent of the model and distribution chosen. Commonly used non-proportional AFT models are Weibull, exponential, log-normal and log-logistic (Qi, 2008; Box-Steffensmeier and Jones, 2007). For completeness, these four models are included in the hazard analysis and compared to the CPH results. It is important to note that positive beta coefficients in CPH models signify a reduced time to failure and should be interpreted as an increased hazard while positive beta coefficients in AFT models signify an increased time to failure and should be interpreted as a reduced hazard.

Research question two uses future recall counts at the plant as the dependent variable. This count variable is Poisson distributed and requires correction for over-dispersion indicating the need for a negative binomial model. Because the unit of analysis is the plant, there are recurring instances of the same plant across multiple years, necessitating a panel model approach. Secondary panel datasets with repeated occurrences over time may be susceptible to endogeneity caused by omitted variable bias, cross-sectional and temporal auto-correlation, and heteroscedasticity. Fixed effects analysis can alleviate some potential omitted variable bias by de-meaning the data and eliminating time-invariant omitted variables from the error term that could be correlated with other predictors. Chances of endogeneity caused by omitted variable bias are significantly reduced when using a fixed effect model, while a random effects model is best used when the risk of omitted variable bias is low. Our primary analysis uses fixed effects negative binomial regression with bootstrapped standard errors (1000 repetitions). Bootstrapped standard errors are one mechanism to correct for heteroscedasticity and temporal auto-correlation (Guan, 2003; Greene, 2003; Staats and Gino, 2012; Yang, Zhou and Wang, 2010). We demonstrate consistency of these results using a random effects and logistic regression model during robustness checks.

4.4 Results

We present descriptive statistics and correlation matrices in Table 4.2 and Tables 4.3-4.4 respectively. Manufacturers take 54 days on average to open a recall after becoming aware of the defect, and almost a year to close the recall once opened. *Moderate Severity* recalls are correlated with longer time-to-open and time-to-close, while least severe recalls (*Low Severity*) are negatively correlated with open and close times. *High Severity* recalls are also correlated with a faster time-to-classify. Design and software recalls are correlated with longer times to open and close while manufacturing and miscellaneous recalls are correlated with faster times. In relation to future recall occurrences, longer time-to-open and time-to-classify recalls are positively correlated with future recalls.

Table 4.2 Description of Variables and Summary Statistics

Variable	*Description*	*Mean*	*St dev.*
Research question one			
Time-To-Open	Plant time from awareness to recall initiation (days)	54.43	152.12
Time-To-Close	Plant time from initiation to recall completion (days)	329.77	365.89
Time-To-Classify	FDA time from initiation to classification (days)	142.10	169.85
Sales	Sales at the plant in 10 sales categories	8.52	1.49
Units Recalled	Number of units recalled	157,992	2,585,401
Indicator Variables			
Public	Indicator variable for public/private status of firm	0.58	0.49
Design	Design related recall	0.15	0.36
Manufacturing	Manufacturing related recall	0.32	0.46
Software	Software related recall	0.10	0.30
Misc.	Miscellaneous category of recall causes	0.43	0.50
High Severity	Class I recall	0.04	0.19
Moderator Severity	Class II recall	0.84	0.37
Low Severity	Class III recall	0.12	0.33
Research question two			
Future Recalls	Total number of recalls initiated in the next year	2.03	3.56
Lagged Class I Recalls	Total number of class I recalls at plant current year	0.10	0.37
Lagged Class II Recalls	Total number of class II recalls at plant current year	2.24	3.29
Lagged Class III Recalls	Total number of class III recalls at plant current year	0.36	0.85
Avg Units Per Recall	Average number of units recalled in the current year	205,093	1,787,273
Process Time			
Time-To-Open	Current year average time-to-open (days)	41.82	111.81
Time-To-Close	Current year average time-to-close (days)	280.40	298.11
Time-To-Classify	Current year average time-to-classify (days)	138.28	193.25

Table 4.3 Recall Level Correlation Matrix-Research Question One

		[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]
Time-To-Open	[1]	1.00												
Time-To-Close	[2]	0.18*	1.00											
Time-To-Classify	[3]	-0.07*	0.15*	1.00										
Sales	[4]	0.06*	0.04*	-0.01	1.00									
Public	[5]	0.03*	0.05*	0.02	0.27*	1.00								
Units Recalled	[6]	0.12*	0.17*	-0.05*	0.15*	0.15*	1.00							
Design	[7]	0.07*	0.13*	0.06*	-0.00	0.01	0.05*	1.00						
Manufacturing	[8]	-0.01	-0.12*	0.05*	-0.03*	-0.07*	0.00	-0.29*	1.00					
Software	[9]	0.09*	0.10*	0.11*	0.02	0.07*	-0.11*	-0.14*	-0.22*	1.00				
Misc.	[10]	-0.09*	-0.04*	-0.17*	0.01	0.01	0.03*	-0.37*	-0.59*	-0.29*	1.00			
High Severity	[11]	0.02	0.01	-0.07*	0.00	0.00	0.11*	0.12*	-0.00	0.03	-0.07*	1.00		
Moderate Severity	[12]	0.06*	0.13*	0.02	0.05*	0.02	0.02	0.03*	0.00	0.09*	-0.07*	-0.45*	1.00	
Low Severity	[13]	-0.08*	-0.16*	0.01	-0.05*	-0.02	-0.09*	-0.11*	0.00	-0.09*	0.13*	-0.08*	-0.86*	1.00

Table 4.4 Plant Level Correlation Matrix-Research Question Two

		[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
Future Recalls	[1]	1.00							
Lagged Class I	[2]	0.05*	1.00						
Lagged Class II	[3]	0.35*	0.00	1.00					
Lagged Class III	[4]	0.11*	-0.02	-0.13*	1.00				
Ln_Avg Units Recalled	[5]	0.05	0.21*	0.25*	-0.02	1.00			
Time-To-Open	[6]	0.12*	0.10*	0.29*	-0.10	0.19*	1.00		
Time-To-Close	[7]	-0.00	0.08*	0.31*	-0.11*	0.30*	0.26*	1.00	
Time-To-Classify	[8]	0.11*	-0.04	0.19*	0.03	-0.03	0.05	0.19*	1.00

The results for testing Hypotheses 1A are presented in Table 4.5. We include control variables which may affect time-to-open a recall (column 1). Compared to 2013, recalls in all previous years took a longer time-to-open, as seen by the negative and significant beta coefficient in the CPH model. Not surprisingly, the more units associated with a recall, the longer it took manufacturers to open the recall. In comparison to design recalls, manufacturing and miscellaneous recalls are opened quicker. We next incorporate the main effect of recall severity (columns 2) in a CPH model. More severe recalls took longer to open than least severe recalls, (fail to reject Hypothesis 1A), as indicated by the negative and significant coefficients for *High Severity* and *Moderate Severity* in column 2. Specifically, in comparison to the least severe recalls, moderate and high several recalls took 20% longer to open ($\exp^{-0.23}$=0.80). Using an average time-to-open of 54 days, this means that high and moderately severe recalls took 10 additional days to open than least severe recalls. Columns 3-6 in Table 4.5 substantiate these conclusions with four different AFT models which do not assume proportionality. We observe that in all but one case (log logistic, *High Severity*) every beta coefficient for *High Severity* and *Moderate Severity* in each of the AFT models are positive and significant, indicating that manufacturers take a longer time to open a recall when the recall is *High* or *Moderate Severity*, as compared to one that is the least severe.

Table 4.6 includes the results for the time-to-close analysis. We observe very similar results for time-to-close as with time-to-open. All control variables are similar in sign and significance (column 1). *High* and *Moderate Severity* recalls are both associated with a longer time-to-close when compared to the least severe recalls. Specifically, *High Severity* recalls take 23% longer to close, equating to 76 additional days, while *Moderate Severity* recalls take 72% longer to close, or 250 additional days. We reject Hypothesis 1B as contrary to our expectations and theory, manufacturing plants take longer to close a more severe recall than a less severe recall.

Table 4.5 Hazard Model: Time-To-Open

			Accelerated Failure Time Models			
	CPH Model (1)	CPH Model (2)	Log Normal (3)	Log Logistic (4)	Exponential (5)	Weibull (6)
Sales	-0.03	-0.03	0.04	0.04	0.06	0.06
	(0.02)	(0.02)	(0.04)	(0.05)	(0.05)	(0.05)
Public	-0.01	-0.01	-0.00	0.00	-0.00	0.01
	(0.10)	(0.10)	(0.20)	(0.22)	(0.21)	(0.22)
Ln_Units_Recalled	-0.32**	-0.03*	0.08**	0.09**	0.05*	0.07*
	(0.11)	(0.01)	(0.03)	(0.03)	(0.03)	(0.03)
2003	-0.41***	-0.34**	0.45+	0.34	1.12***	0.82***
	(0.11)	(0.11)	(0.25)	(0.30)	(0.25)	(0.24)
2004	-0.38**	-0.44***	0.56*	0.46	1.41***	1.04***
	(0.12)	(0.11)	(0.24)	(0.30)	(0.23)	(0.23)
2005	-0.40***	-0.39**	0.51*	0.40	1.23***	0.94***
	(0.11)	(0.12)	(0.26)	(0.31)	(0.24)	(0.25)
2006	-0.36***	-0.40***	0.61*	0.53+	1.21***	0.98***
	(0.10)	(0.11)	(0.24)	(0.29)	(0.20)	(0.23)
2007	-0.31**	-0.36***	0.53*	0.49+	1.15***	0.87***
	(0.10)	(0.10)	(0.22)	(0.26)	(0.23)	(0.21)
2008	-0.28**	-0.31**	0.42+	0.37	1.13***	0.77***
	(0.09)	(0.10)	(0.22)	(0.26)	(0.23)	(0.22)
2009	-0.43***	-0.28**	0.34+	0.25	1.06***	0.72***
	(0.10)	(0.09)	(0.19)	(0.23)	(0.20)	(0.20)
2010	-0.29**	-0.43***	0.74***	0.74**	1.22***	1.03***
	(0.09)	(0.10)	(0.20)	(0.25)	(0.22)	(0.22)
2011	-0.15+	-0.29**	0.57**	0.63*	0.71***	0.66**
	(0.09)	(0.09)	(0.21)	(0.25)	(0.21)	(0.21)
2012	-0.03*	-0.15+	0.27	0.31	0.39*	0.33+
	(0.01)	(0.08)	(0.21)	(0.25)	(0.18)	(0.19)
Manufacturing	0.25***	0.23***	-0.38***	-0.37**	-0.51***	-0.53***
	(0.05)	(0.05)	(0.10)	(0.12)	(0.12)	(0.11)
Software	-0.09	-0.09	0.25	0.36	0.09	0.17
	(0.11)	(0.11)	(0.27)	(0.33)	(0.21)	(0.24)
Misc.	0.44***	0.40***	-0.63***	-0.57**	-1.04***	-0.97***
	(0.06)	(0.07)	(0.15)	(0.20)	(0.11)	(0.12)
High Severity		-0.23*	0.40*	0.37	0.72***	0.58**
		(0.09)	(0.20)	(0.24)	(0.20)	(0.21)
Moderate Severity		-0.23***	0.35**	0.34*	0.62***	0.53***
		(0.06)	(0.13)	(0.15)	(0.13)	(0.14)
Constant			0.79+	0.65	1.89***	1.38**
			(0.44)	(0.49)	(0.45)	(0.47)
Observations	4394	4394	4394	4394	4394	4394
χ^2/R^2	140.41	167.08	75.09	61.14	236.80	154.24

Standard errors in parentheses + p<0.10, *p<0.05, **p<0.01, ***p<0.001

Finally, the hazard model results testing the relationship between recall severity and the FDA's time-to-classify a recall are presented in Table 4.7. In comparison to 2013, many previous years had a longer time-to-classify, as indicated by the negative and significant beta coefficient on years 2003 through 2012. In comparison to design related recalls, recalls that fall into a

miscellaneous category are classified faster. Moving to column 2, we observe that *High Severity* recalls are classified significantly faster than least severe recalls. If a recall is of the *High Severity* class, it will be classified 542% faster than a least severe class ($e^{1.86}$=6.42). We fail to reject Hypothesis 1C, as the FDA moves quickly to classify the most serious types of medical device recalls.

Results related to research question two are reported in Table 4.8. Fixed effects models do not include sales and public predictors as these variables do not change over time for each plant in our data and a FE analysis eliminates observed and unobserved time-invariant predictors. Two of the year indicator variables are negative predictors of future recalls, while the lagged recalls both by type of recall and units affected per recall do not affect future recalls. *Ln_Process Time* represents the average time-to -open (column 2), -close (column 3) or -classify (column 4) in Table 4.8. We reject Hypothesis 2A, as time-to-open has no relationship with future recalls but fail to reject Hypothesis 2B. We find that time-to-close is a negative and significant predictor of future recalls. Therefore, plants that take a longer time-to-close a recall have fewer future recalls.

Because the process time variable is natural log transformed within a negative binomial analysis (creating a log-log model), the beta coefficient is interpreted as the elasticity of the number of recalls per plant per year with respect to the number of days it takes to close a recall. In other words, a 1% change in the number of days to close a recall is associated with a -0.1% (β= -0.09) change in the number of recalls per plant per year. The average number of recalls per plant per year is 2.03 or 22 recalls across the 11-year panel, and the average time-to-close a recall is 280 days, with a standard deviation of 298 days. A one standard deviation change in the number of days to close a recall (298 days) would lead to two fewer recalls per plant across the 11-year panel. An increase in just 140 days (one half of one standard deviation, or approximately four months) to close a recall equates to one less medical device recalls across the 11-year panel. This is significant as even one medical device recall touches 158,000 patients on average in the US (average units recalled is 157,992 per Table 4.2). Finally, we reject Hypothesis 2C: There appears to be no relationship between FDA time-to-classify and future recalls.

Table 4.6 Hazard Model: Time-To-Close

			Accelerated Failure Time Models			
	CPH Model[a]	CPH Model	Log Normal	Log Logistic	Exponential	Weibull
	(1)	(2)	(3)	(4)	(5)	(6)
Sales	0.00	0.01	-0.00	-0.01	-0.00	-0.00
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Public	0.02	0.02	-0.01	0.04	-0.04	-0.04
	(0.07)	(0.07)	(0.08)	(0.08)	(0.07)	(0.07)
Ln_Units_Recalled	-1.00***	-1.02***	0.66***	0.62***	0.96***	0.98***
	(0.16)	(0.16)	(0.13)	(0.13)	(0.15)	(0.16)
2003	-1.00***	-1.02***	0.68***	0.63***	0.97***	0.98***
	(0.16)	(0.16)	(0.13)	(0.12)	(0.15)	(0.15)
2004	-1.01***	-1.02***	0.76***	0.66***	0.96***	0.97***
	(0.16)	(0.16)	(0.12)	(0.11)	(0.14)	(0.15)
2005	-1.17***	-1.16***	0.83***	0.76***	1.12***	1.14***
	(0.16)	(0.16)	(0.14)	(0.13)	(0.15)	(0.15)
2006	-1.06***	-1.06***	0.73***	0.67***	0.98***	0.99***
	(0.15)	(0.15)	(0.12)	(0.11)	(0.13)	(0.13)
2007	-1.06***	-1.04***	0.73***	0.73***	0.94***	0.95***
	(0.15)	(0.15)	(0.12)	(0.10)	(0.12)	(0.13)
2008	-1.09***	-1.07***	0.81***	0.81***	0.94***	0.94***
	(0.14)	(0.14)	(0.11)	(0.10)	(0.12)	(0.12)
2009	-1.00***	-0.98***	0.78***	0.79***	0.83***	0.83***
	(0.14)	(0.14)	(0.11)	(0.09)	(0.11)	(0.12)
2010	-0.82***	-0.80***	0.63***	0.65***	0.66***	0.65***
	(0.13)	(0.13)	(0.11)	(0.09)	(0.11)	(0.11)
2011	-0.43***	-0.41***	0.38***	0.36***	0.32**	0.31**
	(0.12)	(0.12)	(0.10)	(0.08)	(0.11)	(0.11)
2012	-0.04***	-0.04***	0.06***	0.06***	0.04***	0.04**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Manufacturing	0.35***	0.32***	-0.42***	-0.41***	-0.29***	-0.28***
	(0.06)	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)
Software	-0.02	-0.02	0.04	0.07	0.01	0.00
	(0.06)	(0.06)	(0.08)	(0.07)	(0.06)	(0.06)
Misc.	0.31***	0.29***	-0.33***	-0.32***	-0.29***	-0.28***
	(0.06)	(0.06)	(0.07)	(0.07)	(0.06)	(0.06)
High Severity		-0.27*	0.38**	0.32**	0.31*	0.30*
		(0.11)	(0.13)	(0.12)	(0.13)	(0.13)
Moderate Severity		-1.28***	0.45***	0.41***	0.33***	0.32***
		(0.25)	(0.09)	(0.07)	(0.08)	(0.08)
Constant			4.02***	4.21***	4.60***	4.64***
			(0.24)	(0.21)	(0.23)	(0.24)
Observations	4394	4394	4394	4394	4394	4394
χ^2/R^2	245.86	251.37	274.39	290.93	261.69	261.40

Standard errors in parentheses + p<0.10, *p<0.05, **p<0.01, ***p<0.001

[a] Time-varying covariate used for Moderate Severity. Moderate Severity interacted with ln of time

Table 4.7 Hazard Model: Time-To-Classify

			Accelerated Failure Time Models			
	CPH Model[a] (1)	CPH Model (2)	Log Normal (3)	Log Logistic (4)	Exponential (5)	Weibull (6)
Sales	-0.01	-0.01	0.02	0.02	0.01	0.01
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)
Public	-0.05	-0.04	0.05	0.03	0.05	0.05
	(0.09)	(0.08)	(0.08)	(0.08)	(0.11)	(0.11)
Ln_Units_Recalled	-0.36+	-0.32+	0.22+	0.12	0.45+	0.44+
	(0.19)	(0.19)	(0.13)	(0.12)	(0.26)	(0.25)
2003	-0.49*	-0.47*	0.25	0.12	0.66*	0.65*
	(0.22)	(0.21)	(0.17)	(0.15)	(0.30)	(0.29)
2004	-0.44*	-0.42*	0.34*	0.31*	0.42	0.43+
	(0.18)	(0.18)	(0.14)	(0.13)	(0.26)	(0.25)
2005	-0.67***	-0.65***	0.55***	0.46***	0.82**	0.81**
	(0.20)	(0.20)	(0.16)	(0.13)	(0.30)	(0.29)
2006	-0.88***	-0.87***	0.91***	0.91***	0.86***	0.87***
	(0.16)	(0.16)	(0.11)	(0.10)	(0.23)	(0.22)
2007	-1.04***	-1.04***	1.11***	1.09***	1.09***	1.10***
	(0.16)	(0.17)	(0.11)	(0.10)	(0.24)	(0.23)
2008	-1.03***	-1.03***	1.04***	1.02***	1.12***	1.13***
	(0.17)	(0.17)	(0.12)	(0.12)	(0.24)	(0.22)
2009	-0.77***	-0.77***	0.77***	0.77***	0.71**	0.73***
	(0.16)	(0.17)	(0.11)	(0.11)	(0.23)	(0.22)
2010	-0.47**	-0.47**	0.41***	0.38***	0.43+	0.44*
	(0.15)	(0.16)	(0.10)	(0.09)	(0.23)	(0.21)
2011	-0.19	-0.19	0.16+	0.19*	0.05	0.06
	(0.14)	(0.15)	(0.10)	(0.09)	(0.21)	(0.20)
2012	0.02*	0.02+	-0.02+	-0.02	-0.03+	-0.03+
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Manufacturing	0.08	0.10*	-0.08	-0.08	-0.14+	-0.13+
	(0.06)	(0.05)	(0.05)	(0.05)	(0.07)	(0.07)
Software	-0.13+	-0.11	0.17*	0.16*	0.16+	0.16+
	(0.07)	(0.07)	(0.07)	(0.08)	(0.09)	(0.09)
Misc.	0.19*	0.22*	-0.15	-0.10	-0.33**	-0.33**
	(0.09)	(0.09)	(0.10)	(0.09)	(0.12)	(0.12)
High Severity		1.86***	-0.46**	-0.50***	-0.22	-0.24
		(0.45)	(0.16)	(0.13)	(0.29)	(0.28)
Moderate Severity		0.23	-0.15*	-0.15*	-0.13	-0.13
		(0.26)	(0.07)	(0.07)	(0.08)	(0.08)
Constant			3.86***	3.81***	4.54***	4.49***
			(0.24)	(0.23)	(0.36)	(0.35)
Observations	4394	4394	4394	4394	4394	4394
χ^2/R^2	201.71	229.02	284.59	299.34	210.63	210.90

Standard errors in parentheses + p<0.10, *p<0.05, **p<0.01, ***p<0.001

[a] Time-varying covariate used for Moderate and High Severity. Moderate and High Severity interacted with ln of time

To test the robustness of our relationships, we repeat the analysis using a random effects negative binomial model and a logistic regression model (Table 4.9). For the logistic regression, we created an indicator variable which equals one if there were any recalls in the plant in a specific

year, and zero otherwise. All of our conclusions using these additional models hold.

Table 4.8 Negative Binomial Fixed Effects Regression: Future Recalls and Process Time

	Controls	Open	Close	Classify
	(1)	(2)	(3)	(4)
Lagged Class I Recalls	-0.03	-0.04	-0.03	-0.03
	(0.07)	(0.07)	(0.07)	(0.07)
Lagged Class II Recalls	0.00	0.00	0.01	0.00
	(0.01)	(0.01)	(0.01)	(0.01)
Lagged Class III Recalls	0.02	0.02	0.02	0.01
	(0.03)	(0.03)	(0.03)	(0.03)
Ln_Average Units Per Recall	-0.02+	-0.02+	-0.02	-0.02+
	(0.01)	(0.01)	(0.01)	(0.01)
2004	-0.04	-0.04	-0.02	-0.03
	(0.11)	(0.11)	(0.11)	(0.11)
2005	-0.26*	-0.26*	-0.22+	-0.26*
	(0.12)	(0.12)	(0.12)	(0.12)
2006	-0.01	-0.02	0.03	-0.02
	(0.12)	(0.11)	(0.11)	(0.12)
2007	0.14	0.13	0.19+	0.13
	(0.12)	(0.12)	(0.12)	(0.12)
2008	0.15	0.14	0.20+	0.14
	(0.11)	(0.11)	(0.11)	(0.12)
2009	0.07	0.06	0.14	0.06
	(0.12)	(0.12)	(0.12)	(0.13)
2010	0.08	0.07	0.15	0.07
	(0.14)	(0.14)	(0.14)	(0.15)
2011	-0.26	-0.28+	-0.17	-0.27+
	(0.16)	(0.16)	(0.16)	(0.16)
2012	-1.12***	-1.13***	-1.03***	-1.12***
	(0.17)	(0.17)	(0.16)	(0.17)
Ln_Process Time		0.02	-0.09*	0.02
		(0.02)	(0.05)	(0.04)
Constant	1.32***	1.29***	1.71***	1.23***
	(0.18)	(0.18)	(0.27)	(0.24)
Observations	1269	1269	1269	1269
χ^2/R^2	127.38	135.59	160.91	138.27

Standard errors in parentheses+ p<0.10, *p<0.05, **p<0.01, ***p<0.001
Bootstrapped standard errors with 1000 repetitions

It is possible that the relationship between recall responsiveness and future recalls varies based on recall severity. We perform post-hoc analyses by differentiating on recall severity as the dependent variable (Table 4.10). Interesting relationships are observed when this distinction is made. Beginning with class I recalls (columns 1-3) we observe that *Ln_Process Time* for the time-to-close a recall retains its negative and significant relationship only with future class I recalls, and that time-to-open and time-to-classify have no relationship with future class I recalls. Moving to class II recalls (columns 4-6), we observe a marginally negative and significant relationship

between time-to-close and future recalls, and no other time phases significant. For class III recalls (columns 7-9), there is a positive and significant relationship between the FDA's time-to-classify and future recalls. In other words, the longer the FDA takes to classify a recall, the more class III recalls the plant will experience in the future.

Table 4.9 Robustness Checks: Future Recalls and Process Time

	Open		Close		Classify	
	Random Effects	Logistic	Random Effects	Logistic	Random Effects	Logistic
	(1)	(2)	(3)	(4)	(5)	(6)
Sales	0.17***	0.27***	0.17***	0.27***	0.16***	0.27***
	(0.03)	(0.06)	(0.03)	(0.06)	(0.03)	(0.06)
Public	0.37**	0.64***	0.38***	0.66***	0.38***	0.67***
	(0.11)	(0.19)	(0.11)	(0.19)	(0.11)	(0.20)
Lagged Class I Recalls	0.00	0.19	0.01	0.21	0.01	0.22
	(0.07)	(0.21)	(0.07)	(0.21)	(0.07)	(0.21)
Lagged Class II Recalls	0.02*	0.18***	0.02**	0.20***	0.02*	0.18***
	(0.01)	(0.04)	(0.01)	(0.04)	(0.01)	(0.04)
Lagged Class III Recalls	0.03	0.13	0.03	0.13	0.03	0.12
	(0.03)	(0.10)	(0.03)	(0.10)	(0.03)	(0.10)
Ln_Average Units Per Recall	-0.02	-0.04	-0.01	-0.02	-0.01	-0.03
	(0.01)	(0.03)	(0.01)	(0.03)	(0.01)	(0.03)
2004	-0.08	-0.54	-0.05	-0.44	-0.07	-0.48
	(0.13)	(0.35)	(0.13)	(0.35)	(0.13)	(0.35)
2005	-0.30*	-0.74*	-0.25+	-0.59+	-0.29*	-0.69*
	(0.14)	(0.34)	(0.14)	(0.34)	(0.14)	(0.34)
2006	-0.07	-0.29	-0.02	-0.16	-0.07	-0.28
	(0.13)	(0.34)	(0.13)	(0.34)	(0.13)	(0.34)
2007	0.08	-0.25	0.14	-0.12	0.08	-0.25
	(0.13)	(0.34)	(0.13)	(0.34)	(0.13)	(0.34)
2008	0.08	-0.20	0.13	-0.06	0.06	-0.26
	(0.13)	(0.34)	(0.13)	(0.34)	(0.13)	(0.34)
2009	0.02	-0.33	0.10	-0.16	0.01	-0.39
	(0.13)	(0.34)	(0.13)	(0.34)	(0.13)	(0.34)
2010	0.03	-0.35	0.10	-0.17	0.02	-0.38
	(0.13)	(0.35)	(0.13)	(0.35)	(0.13)	(0.35)
2011	-0.34*	-1.04**	-0.23	-0.79*	-0.32*	-1.04**
	(0.14)	(0.35)	(0.14)	(0.35)	(0.14)	(0.34)
2012	-1.22***	-2.27***	-1.13***	-2.03***	-1.21***	-2.24***
	(0.17)	(0.35)	(0.17)	(0.36)	(0.17)	(0.35)
Ln_Process Time	0.03	0.05	-0.08*	-0.17*	0.03	0.13
	(0.02)	(0.04)	(0.04)	(0.08)	(0.04)	(0.08)
Constant	-0.68*	-2.18***	-0.33	-1.51*	-0.79*	-2.68***
	(0.34)	(0.57)	(0.37)	(0.62)	(0.37)	(0.66)
Observations	1491	1491	1491	1491	1491	1491
χ^2/R^2	163.33	133.70	164.60	136.78	161.88	133.94

Standard errors in parentheses + p<0.10, *p<0.05, **p<0.01, ***p<0.001

Table 4.10 Post-Hoc Analysis: Negative Binomial Fixed Effects Regression: Future Recalls of Different Recall Classes and Process Time

	Class 1 Recalls			Class 2 Recalls			Class 3 Recalls		
	Open	Close	Classify	Open	Close	Classify	Open	Close	Classify
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Lagged Class I Recalls	-0.49*	-0.49*	-0.51**	0.02	0.02	0.02	0.08	0.09	0.09
	(0.20)	(0.20)	(0.20)	(0.08)	(0.08)	(0.08)	(0.19)	(0.20)	(0.20)
Lagged Class II Recalls	0.03	0.03	0.03	-0.00	0.00	0.00	-0.03	-0.02	-0.03
	(0.03)	(0.03)	(0.03)	(0.01)	(0.01)	(0.01)	(0.03)	(0.03)	(0.03)
Lagged Class III Recalls	-0.19	-0.24	-0.21	0.02	0.02	0.02	-0.00	0.00	-0.02
	(0.17)	(0.17)	(0.17)	(0.03)	(0.03)	(0.03)	(0.05)	(0.05)	(0.05)
Ln_Average Units Per Recall	0.05	0.08	0.05	-0.03*	-0.02	-0.03*	-0.02	-0.01	-0.02
	(0.05)	(0.05)	(0.05)	(0.01)	(0.01)	(0.01)	(0.03)	(0.03)	(0.03)
2004	-0.17	-0.08	-0.14	0.06	0.08	0.06	-0.15	-0.14	-0.04
	(0.52)	(0.52)	(0.52)	(0.15)	(0.15)	(0.15)	(0.25)	(0.25)	(0.25)
2005	-1.16+	-0.92	-1.16+	-0.08	-0.04	-0.08	-0.85**	-0.80**	-0.82**
	(0.68)	(0.68)	(0.66)	(0.15)	(0.15)	(0.15)	(0.29)	(0.29)	(0.28)
2006	-0.51	-0.23	-0.55	0.17	0.21	0.18	-0.64*	-0.60*	-0.64*
	(0.52)	(0.54)	(0.52)	(0.14)	(0.15)	(0.14)	(0.27)	(0.27)	(0.27)
2007	-0.94	-0.66	-1.00+	0.37**	0.42**	0.38**	-0.78**	-0.72*	-0.79**
	(0.59)	(0.60)	(0.59)	(0.14)	(0.14)	(0.14)	(0.29)	(0.29)	(0.28)
2008	0.27	0.39	0.09	0.37**	0.42**	0.40**	-1.25***	-1.20***	-1.33***
	(0.46)	(0.46)	(0.48)	(0.14)	(0.14)	(0.14)	(0.32)	(0.32)	(0.32)
2009	-0.16	0.06	-0.32	0.33*	0.40**	0.37*	-1.36***	-1.27***	-1.46***
	(0.48)	(0.49)	(0.50)	(0.14)	(0.14)	(0.14)	(0.34)	(0.35)	(0.34)
2010	0.13	0.33	-0.08	0.32*	0.38**	0.35*	-1.35***	-1.26***	-1.48***
	(0.51)	(0.51)	(0.52)	(0.14)	(0.15)	(0.15)	(0.34)	(0.34)	(0.34)
2011	-0.44	-0.18	-0.59	-0.02	0.07	0.02	-1.51***	-1.38***	-1.55***
	(0.54)	(0.53)	(0.52)	(0.16)	(0.16)	(0.15)	(0.36)	(0.37)	(0.36)
2012	-1.64*	-1.37+	-1.79*	-0.87***	-0.79***	-0.86***	-1.98***	-1.84***	-1.90***
	(0.75)	(0.76)	(0.74)	(0.18)	(0.19)	(0.18)	(0.40)	(0.40)	(0.39)
Ln_Process Time	-0.07	-0.34*	0.09	0.02	-0.07+	-0.03	0.05	-0.09	0.25**
	(0.09)	(0.14)	(0.15)	(0.02)	(0.04)	(0.04)	(0.06)	(0.10)	(0.09)
Constant	0.85	2.09*	0.59	1.23***	1.54***	1.39***	1.07*	1.55*	0.22
	(0.89)	(1.01)	(1.10)	(0.19)	(0.24)	(0.25)	(0.46)	(0.63)	(0.60)
Observations	353	353	353	1241	1241	1241	670	670	670
χ^2/R^2	25.98	31.76	26.39	110.37	113.07	110.13	57.52	58.61	65.47

4.5 Discussion and Implications

We investigate causes and consequences of responsiveness in the medical device product recall process. We seek to understand what leads manufacturers to move quickly to respond to product problems in the marketplace and to uncover levers which can be used to increase learning and reduce future recalls. To our knowledge, this analysis is the first to predict 1) recall responsiveness using distinct time intervals computed with actual date and time stamps and 2) future recalls as a function of responsiveness. We first discuss the causes of plant and regulator responsiveness, and then the effects of this responsiveness.

Plants are slower to both open and close a recall when the problem is most severe. These findings provide empirical evidence that plants may not be responsive to the most serious problems. Not opening the most serious recalls fast may indicate resistance from the plant to expose themselves to negative publicity from the public and sanctions from regulators, while at the same time increasing risks to customers. This result is consistent with prospect theory (Kahneman and Tversky, 1979) predictions, managers are demonstrating risk-averseness with severe product recalls. While this resistance is understandable, these are precisely the types of product problems that managers should be most responsive to. These findings should spur managers to greater vigilance and reaction time when investigating the most serious product problems. Popular media coverage of recent serious recalls in multiple industries supports a case for increased vigilance. In relation to closing a recall, the results are more encouraging. More severe recalls lead to longer recall closure times. This unexpected result may indicate that once a recall has been opened, plants choose thoroughness over reputation. Possibly, the pressure to not open a severe recall is more severe than the pressure to close a severe recall quickly once opened. From a regulator perspective, the most severe recalls are classified the fastest, consistent with theory and with FDA's policy goals.[35]

In relation to future recalls, responsiveness in opening a recall does not affect future recalls but responsiveness in closing a recall does. Two conclusions can be drawn from these findings. Once the initial defect is detected, additional time taken to observe more manifestations of the defect does not seem to result in useful learning to be applied to future product quality. However,

[35] http://www.fda.gov/ICECI/ComplianceManuals/RegulatoryProceduresManual/ucm177311.htm

scrutinizing the root cause and corrective action of the recall once it is opened does seem to be an important mechanism for learning and reflection. Plants which take the most time in this phase identifying root cause and corrective action, communicating with customers and the FDA, and repairing or replacing contaminated products, have fewer future recalls. These results are consistent with research related to the quality-speed conundrum (Anand et al., 2011). Engineers and managers tasked with investigating product quality problems may cut corners and move too quickly at times, which may hinder their full understanding of the problem and its applicability to future products.

When the analysis includes a distinction of recall severity, the findings are more nuanced. Surprisingly, the relationship between the time-to-close a recall and fewer future recalls persists for the most severe recalls only. This may signify the prioritization of newly learned information within the manufacturer. As plants learn from their mistakes in the recall closure process, knowledge may only be incorporated within the most serious aspects of future products and processes. While FDA's time-to-classify was not significant in the main analysis, recall differentiation in Table 4.10 indicates that FDA responsiveness does impact future recalls, but only the least serious types of recalls (class III recalls).

We validated our findings with both senior industry and FDA personnel. The global vice-president for quality and her staff at one of the world's largest medical device companies crystallized the findings related to plant recall responsiveness. In her view, medical device manufacturers that do not prioritize quality enough are more likely to quickly push through the recall closure process and learn little from the recall. This individual informed us of her strategy to move quickly to open a recall when the problem is apparent, but to not rush through the root cause and corrective action identification state of closing a recall. Additionally, a senior director at the FDA's Center for Device and Radiological Health (CDRH) who was instrumental during data collection shared her belief that plants which rush through the recall process, especially in closing the recalls, do not learn what that they need to avoid future problems. This work provides confirmation for these senior industry and regulatory leaders and can enlighten similar individuals previously unaware of these important relationships.

Implications

There are four implications of this study. First, firms should open recalls quickly to protect

public health and safety. The possibility of lost learning is minimal when moving quickly to open a recall. Second, firms should take great care to not rush the recall closure process, extract maximal learning from the root cause and corrective action investigation process, and apply such learning to future product design and manufacturing processes. Third, they should take steps to apply this learning more broadly, as the recall reduction benefits of the recall closure process seem to only impact the most serious class I recalls, though one can imagine that the lessons learned from these experiences could be used to reduce all classes of future recalls. Finally the FDA is cautioned to not rush manufacturers to close recalls too quickly. Empirical evidence suggests that the time that the plant takes in this phase of the product recall process is associated with fewer future recalls. Instead of pushing plants to close recalls quickly, the FDA may instead use this phase of the process to aid manufacturers in deepening their understanding of the problem and ensuring that the lessons learned from each individual recall are applied as broadly as possible to future product manufacturing and design.

Our research has certain limitations. While fixed-effects analyses help mitigate significant sources of unobserved heterogeneity, some sources may still remain. For instance, collecting firm-level data and incorporating this into the analyses may provide additional insights. While preliminary analyses indicate it is not a concern, further work can be done regarding queueing effects and reverse causality in this dataset. It is possible that longer response times occur because more recalls are in queue to be closed or classified, and that this increase in the number of recalls impacts future recall occurrences and learning. Repeating these analyses with two, three, or four years after the response times are measured could also expand our understanding.

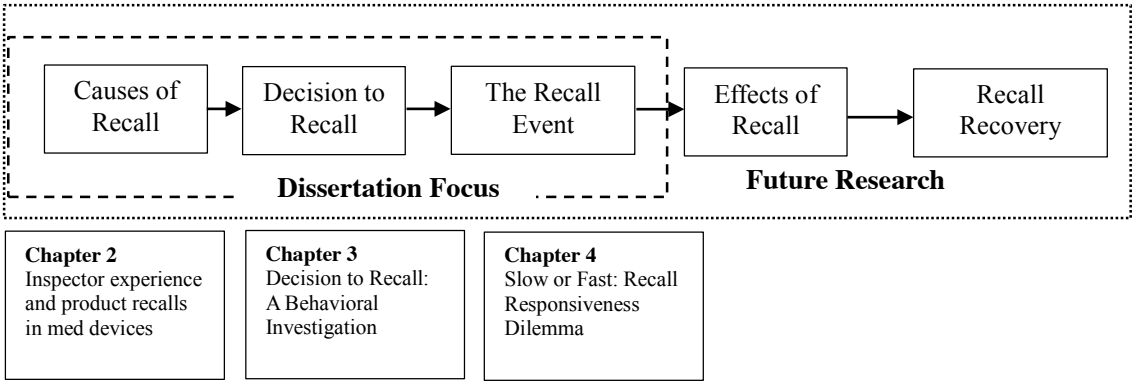
Chapter 5:

Conclusion

Product recalls have attracted sparse attention from scholars. While the current state of research identifies ramifications of recalls, leading indicators of recalls remain relatively unexplored. While studying recalls in any regulated industry would contribute to existing theory and benefit practitioners, I focus my dissertation on medical device recalls because of its potential to impact public health and safety, and the ever-increasing economic footprint of this industry. Although regulators work tirelessly to root out medical device recalls, they are still on the rise. While no one party (FDA or industry managers) may be able to completely prevent quality problems leading to recalls, this dissertation provides steps that both regulators and manufacturers can take to contain the increasing recall trend.

The dissertation consisted of three chapters, where each chapter addressed an important phase in the product recall process (Figure 5.1, repeated from Figure 1.1): Recall causes, decision-making, and responsiveness. This body of research contributes significantly to practice and theory related to product recalls. I summarize each chapter's individual contribution, areas for future research, and generalizability to other industries, below. I conclude with two important themes that can be derived from a holistic examination of the results of this dissertation.

Figure 5.1 Dissertation Structure



In chapter 2, I studied plant inspection results and the risks of using repeat inspectors to inspect a plant. Plant inspections occur on a regular basis with great frequency around the globe, and are used to manage supply chain quality, evaluate potential suppliers, and regulate product quality requirements laid out by regulatory bodies. This research demonstrated that these plant inspections serve a purpose, not only by adequately assessing the future quality risk at plants, but also by providing failing plants with an impetus for improvement. However, the investigator experience effects unearthed in this study also showed that such inspection regimes need to be carefully managed to remain effective. Using multiple Cox proportional hazard and propensity score matching models, I demonstrated that inspector rotation may serve to abate the negative effects of complacency and reduce the number of medical device recalls. Future research may identify additional criteria of inspectors that may influence their inspection accuracy, such as training, gender, experience and education. Additionally, future work may incorporate other dependent variables as measures of plant quality performance, such as medical device product complaints (MAUDE database) or internal measures of quality such as rework.

In chapter 3, I used industry managers to study behavioral factors influential in the medical device product recall decision. Through numerous hours of interviews with industry and regulatory managers, I identified potential situational and dispositional factors that may influence managers to decide to recall. I found that defect detectability and root cause understanding are both positively related to the likelihood of a recall decision. Additionally, I found that there are direct and indirect effects of a manager's CRT score. High CRT managers are significantly less likely to decide to recall, independent of any of the experimental factors. CRT also moderates several relationships in this study. Low CRT managers are only influenced by dispositional factors such as gender, experience, functional area, and their relationship with the FDA. High CRT managers are strongly influenced by the root cause understanding of the defect, and marginally influenced by defect detectability. Future work may include a student subject experiment, investigating the differences between actual industry managers and student subjects in a highly contextual setting such as product recalls. Additionally, incorporating a team-based experiment, which would simulate real-life recall decisions even more closely, may significantly add to these findings.

Finally, in chapter 4, I documented potential causes and effects of recall responsiveness for the firm and the regulator. Firms wait the longest to open the most severe recalls, while regulators

move the fastest on these types of recalls. Firms however take the longest amount of time-to-close severe recalls, apparently using extra time to study the root cause and corrective action thoroughly. While the time-to-open a recall has little relationship with future recalls, the time-to-close significantly reduces future recalls. Firms that take a longer amount of time-to-close recalls have fewer future recalls, but this is only statistically significant for the most severe types of future recalls. This may indicate learning on the part of the firm which is applied to the most important future product manufacturing and design processes. Implications of this study are clear: firms should open recalls quickly, close recalls slowly and apply recall learning more broadly. Regulators are cautioned from the results of this research to not rush firms in the recall closure process. Future avenues of research include a case-study which can track activities in the recall open and closing process to improve our knowledge of how learning is created in this process and how it is applied to future products to successfully reduce future recalls.

To highlight the similarities and differences among each chapter, I provide an overview of the dissertation in Table 5.1. While all three studies are conducted in the context of the medical device industry and involve drivers of medical device recalls, they are distinct in their units of analyses and the resulting implications. While chapters 2 and 4 are conducted at the plant level, chapter 3 is at the individual manager level. Thus, chapters 2 and 4 results can be applied to the plants in the medical device industry where as chapter 3 results are applicable to managers making the recall decision. While each chapter has implications for firms and regulators, they are distinctly different. Chapter 2 provides evidence for early warning signs for recalls at the plant, while firm implications for chapters 3 and 4 are behavioral in nature. They independently examine *how*, and *how fast*, firms' implicit decision factors and decision speed respectively, impact the decision to recall and eventually future recalls. Implications for the FDA are similar in nature. Chapters 2 and 3 indicate a downside of complacent or overly friendly relationships with firms, while chapter 4 provides procedural guidance for FDA in managing firms through their regulation of the recall process.

Table 5.1 Dissertation Summary

	Chapter #2	Chapter #3	Chapter #4
Title	Investigator Experience and Product Recalls in the Medical Device Industry	The Decision to Recall: A Behavioral Investigation in the Medical Device Industry	Slow or Fast? An Empirical Examination of the Recall Responsiveness Dilemma
Industry	Medical Device	Fortune 500 Medical Device Company	Medical Device
Unit of Analysis	Manufacturing Plant	Managerial Decision-Maker	Manufacturing Plant
Data (Years)	Secondary data (2000-2006)	Experimental data (2014)	Secondary data (2000-2013)
Research Question	* How effective are (FDA) plant inspection outcomes in predicting recalls? * How does inspector experience impact the predictability?	* Which behavioral factors influence a manager's decision to recall a product?	* What leads to quick recall response times for firms and regulators? * How does responsiveness impact future recalls?
Theoretical Lens	Learning and Complacency	Attribution Theory, Affect Heuristic, Customer Satisfaction	Prospect Theory, Quality-Speed Conundrum
Research Method	Recurrent event, Cox proportional hazard model; Propensity score matching model	Paired T-Test and Logistic Regression	Accelerated failure time and Cox proportional hazard model; Fixed Effects panel data model
Firm Implications	* FDA inspection scores signal future recalls. * Plant inspections using rotating inspectors can effectively control and monitor quality in global supply chains	* Managers should not: - expect physicians to detect their defective products - wait too long for additional data before recalling	*Firms should: - open recalls quickly, - close recalls slowly, - apply learning from recalls more broadly
Regulator Implications	* FDA should rotate inspectors across plants to avoid complacency	* FDA should guard against overly friendly relationships with firms	* FDA should allow firms to take sufficient time-to-close recalls

As a set, we can see that while the results directly apply to the medical device industry, they can be generalized to other industries that are closely regulated and monitored by federal agencies. For instance, the pharmaceutical industry and a large portion of the food industry in the

US are regulated by the FDA, and have similar processes for inspections, product recall guidelines and managerial practices. It is likely that the results of this dissertation are easily generalizable to the plants in these two industries. The automotive industry may also benefit from this research. Auto plants are inspected by NHTSA investigators, managers face multiple and complex criteria when making the recall decision, and responsiveness is pertinent in this industry as well. Finally, the consumer products industry, such as toy manufacturers, may take some lessons from this research. While consumer products companies do not normally experience federal quality inspections, their managers face difficult product recall decisions in which responsiveness is crucial to ensure consumer safety. Chapters 3 & 4 may therefore also be applicable to the consumer products industry.

Finally, there are two noteworthy themes which emerge from the research that can inform recall research and theory. First, there appears to be a very important behavioral element to product recalls that exists within each recall phase. This unexpected behavioral element is present in recall causes, recall decisions, and recall responsiveness and learning. Second, my dissertation highlights the importance of empirical analyses in demonstrating a relationship between plant-level managerial decisions and external product quality performance measures. There are only a handful of research studies using product recalls as a dependent variable. In a majority of past empirical work, studies have focused on quality programs, such as six-sigma or Total Quality Management (TQM), and a few studies have empirically analyzed causes of internal product quality measures (Banker et al., 1990; Datar et la., 1993; Fisher and Ittner, 1999; Mukherjee et al., 2000). However, very few studies use rigorous econometrics to link decisions at the production plant with external product performance, such as product recalls.

While there are clearly other things that may also matter in causing recalls and leading to the recall decision, this dissertation takes an important first step in uncovering significant sources of variation in critical phases of the recall process. My hope is that this dissertation will spur other researchers to branch out from traditional product quality research paradigms, and seek to more fully understand how managers' decisions effect actual product quality performance in the marketplace, using both empirical and behavioral research methods.

Bibliography

Ahire, S. L., & Dreyfus, P. (2000). The impact of design management and process management on quality: an empirical investigation. *Journal of Operations Management*, *18*(5), 549-575.

Alba, J. W., & Chattopadhyay, A. (1986). Salience effects in brand recall. *Journal of Marketing Research*, *23*(4), 363-369.

Anand, G., Gray, J., & Siemsen, E. (2012). Decay, shock, and renewal: Operational routines and process entropy in the pharmaceutical industry. *Organization Science*, 23(6), 1700-1716.

Anand, K. S., Pac, M. F., & Veeraraghavan, S. (2011). Quality-speed conundrum: trade-offs in customer-intensive services. *Management Science*, *57*(1), 40-56.

Anderson, E.W., Fornell, C. & Lehmann, D.R. (1994). Customer satisfaction, market share, and profitability: Findings from Sweden. *The Journal of Marketing*, 53-66.

Anderson, E. W., Fornell, C., & Mazvancheryl, S. K. (2004). Customer satisfaction and shareholder value. *Journal of marketing*, *68*(4), 172-185.

Anderson, E.W., Sullivan, M.W. (1993). The antecedents and consequences of customer satisfaction for firms, *Marketing Science*, *12*(2), 125-143.

Archer, N. P., & Wesolowsky, G. O. (1996). Consumer response to service and product quality: a study of motor vehicle owners. *Journal of Operations Management*, *14*(2), 103-118.

Argote, L., & Epple, D. (1990). Learning curves in manufacturing. *Science*, 247(4945), 920-924.

Argote, L. (2013). Organizational Learning: Creating, Retaining and Transferring Knowledge. Springer, New York NY.

Azizkhani, M., G. S. Monroe, G. Shailer. (2013). *Audit partner tenure and cost of equity capital*. Auditing: A Journal of Practice and Theory. 32(1): 183-202.

Baltagi, B. (2008). *Econometric analysis of panel data* (Vol. 1). John Wiley & Sons.

Banker, R. D., Datar, S. M., Kekre, S., & Mukhopadhyay, T. (1990). Costs of product and process complexity. *Measures for manufacturing excellence*, *2*, 269-290.

Bardach, E., & R. A. Kagan. (2002). *Going by the Book: The Problem of Regulatory Unreasonableness*. New Brunswick, NJ: Transaction.

Bateman, T. S., & Zeithaml, C. P. (1989). The psychological context of strategic decisions: A model and convergent experimental findings. *Strategic Management Journal*, *10*(1), 59-74.

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1997). Deciding advantageously before knowing the advantageous strategy. *Science*, *275*(5304), 1293-1295.

Bendoly, E., Croson, R., Goncalves, P., & Schultz, K. (2010). Bodies of knowledge for research in behavioral operations. *Production and Operations Management*, *19*(4), 434-452.

Berman, S. L., Down, J., & Hill, C. W. (2002). Tacit knowledge as a source of competitive advantage in the national basketball association. *Academy of Management Journal*, 45(1), 13-31.

Boone, T., Ganeshan, R., & Hicks, R. L. (2008). Learning and knowledge depreciation in professional services. *Management Science*, 54(7), 1231-1236.

Box-Steffensmeier, J. M., & Jones, B. S. (2004). *Event history modeling: A guide for social scientists* Cambridge University Press.

Bromiley, P. & Marcus, A. (1989). The deterrent to dubious corporate behavior: Profitability, probability and safety recalls. *Strategic Management Journal, 10*(3), 233-250.

Brown, A., Camerer, C., & Lovalla, D. (2013). Estimating structural models of equilibrium and cognitive hierarchy thinking in the field: The case of withheld movie critic reviews. *Management Science*, 59(3), 733-747.

Caplan, R. D., Cobb, S., & French, J. R. (1975). Job demands and worker health; main effects and occupational differences. *Hew Publication (NIOSH)*, 75.

Carey, P., & Simnett, R. (2006). Audit partner tenure and audit quality. *The Accounting Review*, 81(3), 653-676.

Cesarini, D., Johannesson, M., Magnusson, P. K., & Wallace, B. (2012). The behavioral genetics of behavioral anomalies. *Management science*, *58*(1), 21-34.

Cheah, E. T., Chan, W. L., & Chieng, C. L. L. (2007). The corporate social responsibility of pharmaceutical product recalls: An empirical examination of US and UK markets. *Journal of Business Ethics*, *76*(4), 427-449.

Chen, Y., Ganesan, S., & Liu, Y. (2009). Does a firm's product-recall strategy affect its financial value? An examination of strategic alternatives during product-harm crises. *Journal of Marketing*, *73*(6), 214-226.

Concannon, W.P. (1989). Managing for cost-effective purchased material quality. *Engineering Costs and Production Economics*, 18(1), 96-100.

Crossan, M. W., H. W. Lane, R. E. White. (1999). An organizational learning framework: From intuition to institution. *The Academy of Management Review*, 24(3): 522-537.

Damasio, A. R., Everitt, B. J., & Bishop, D. (1996). The somatic marker hypothesis and the possible functions of the prefrontal cortex. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *351*(1346), 1413-1420.

Darr, E. D., Argote, L., & Epple, D. (1995). The acquisition, transfer, and depreciation of knowledge in service organizations: Productivity in franchises. *Management Science*, 41(11), 1750-1762.

Datar, S. M., Kekre, S., Mukhopadhyay, T., & Srinivasan, K. (1993). Simultaneous estimation of cost drivers. *Accounting Review*, 68(3), 602-614.

Davidson, W. N., & Worrell, D. L. (1992). Research notes and communications: The effect of product recall announcements on shareholder wealth. *Strategic Management Journal, 13*(6), 467-473.

Davis, L. R., Soo, B. S., & Trompeter, G. M. (2009). Auditor tenure and the ability to meet or beat earnings forecasts. *Contemporary Accounting Research*, 26(2), 517-548.

Deis Jr, D. R., & Giroux, G. A. (1992). Determinants of audit quality in the public sector. *Accounting Review*, 67(3), 462-479.

Deming, W. E. (1982). *Quality, productivity, and competitive position*. Massachusetts Institute of Technology, Center for Advanced Engineering Study Cambridge, MA.

Devers, C. E., Wiseman, R. M., & Holmes, R. M. (2007). The effects of endowment and loss aversion in managerial stock option valuation. *Academy of Management Journal*, *50*(1), 191-208.

Dhanorkar, S., Siemsen, E., Linderman, K. (2014). Promoting change from the outside: The roles of punishment and support in the context of environmental improvements. *Working Paper*.

Dillon, R. L., & Tinsley, C. H. (2008). How near-misses influence decision making under risk: A missed opportunity for learning. *Management Science*, *54*(8), 1425-1440.

Driscoll, J. C., & Kraay, A. C. (1998). Consistent covariance matrix estimation with spatially dependent panel data. *Review of economics and statistics*, *80*(4), 549-560.

Ferris, S. P., Haugen, R. A., & Makhija, A. K. (1988). Predicting contemporary volume with historic volume at differential price levels: Evidence supporting the disposition effect. *The Journal of Finance*, *43*(3), 677-697.

Finucane, M. L., Alhakami, A., Slovic, P., & Johnson, S. M. (2000). The affect heuristic in judgments of risks and benefits. *Journal of behavioral decision making*, *13*(1), 1-17.

Fishburn, P. C., & Kochenberger, G. A. (1979). Two-Piece Von Neumann-Morgenstern Utility Functions. *Decision Sciences*, *10*(4), 503-518.

Fisher, M.L., & Ittner, C.D. (1999). The impact of product-variety on automobile assembly operations: Empirical evidence and simulation analysis. *Management Science, 45*(6), 771-786.

Frechette, G.R. (2012). Laboratory Experiments: Professionals versus Students. Frechette G, Schotter A, eds. *The Methods of Modern Experimental Economics* (Oxford University Press, Oxford, UK).

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic perspectives*, 19(4), 25-42.

Fornell, C., Johnson, M. D., Anderson, E. W., Cha, J., & Bryant, B. E. (1996). The American customer satisfaction index: nature, purpose, and findings. *The Journal of Marketing*, 7-18.

Fynes, B., & Voss, C. (2001). A path analytic model of quality practices, quality performance, and business performance. *Production and Operations Management*, *10*(4), 494-513.

Gans, N., Koole, G., & Mandelbaum, A. (2003). Telephone call centers: Tutorial, review, and research prospects. *Manufacturing & Service Operations Management*, *5*(2), 79-141.

Garrett. J.M. (1998). Predicted survival curves for the Cox proportional hazards model. *Stata technical bulletin*, 8, 285-290.

Garvin, D (1987). Competing on the 8 dimensions of quality. *Harvard Business Review,* 65(6), 101-109.

Gray, J. V., Roth, A. V., & Leiblein, M. J. (2011). Quality risk in offshore manufacturing: Evidence from the pharmaceutical industry. *Journal of Operations Management*, 29(7), 737-752.

Greene, W. H. (2003). *Econometric analysis*. Pearson Education.

Groysberg, B., Lee, L., & Nanda, A. (2008). Can they take it with them? The portability of star knowledge workers' performance. *Management Science,* 54(7), 1213-1230.

Guan, W. (2003). From the help desk: bootstrapped standard errors. *The Stata Journal*, *3*(1), 71-80.

Gustafsson, A., Johnson, M. D., & Roos, I. (2005). The effects of customer satisfaction, relationship commitment dimensions, and triggers on customer retention. *Journal of marketing*, *69*(4), 210-218.

Guthrie, C., Rachlinski, J. J., & Wistrich, A. J. (2007). Blinking on the bench: How judges decide cases. *Cornell Legal Review*, *93*, 1-40.

Hackman, J. R., & Oldham, G. R. (1976). Motivation through the design of work: Test of a theory. *Organizational Behavior and Human Performance,* 16(2), 250-279.

Handley, S. M., & Gray, J. V. (2013). Inter-organizational quality management: The use of contractual incentives and monitoring mechanisms with outsourced manufacturing. *Production and Operations Management,* 22(6), 1540-1556.

Hannan, E. L., Kilburn Jr, H., Bernard, H., O'Donnell, J. F., Lukacik, G., & Shields, E. P. (1991). Coronary artery bypass surgery: The relationship between in-hospital mortality rate and surgical volume after controlling for clinical risk factors. *Medical Care,* 29(11), 1094-1107.

Hannan, M. T., & Freeman, J. (1984). Structural inertia and organizational change. *American Sociological Review*, 49(2), 149-164.

Hardisty, D. J., & Weber, E. U. (2009). Discounting future green: money versus the environment. *Journal of Experimental Psychology: General*, *138*(3), 329-340.

Haunschild, P. R., & Rhee, M. (2004). The role of volition in organizational learning: The case of automotive product recalls. *Management Science, 50*(11), 1545-1560.

Hauser, J., & Wernerfelt, B. (1990). An evaluation cost model of consideration sets. *Journal of Consumer Research, 16*(4), 393-408.

Hirschmann, W. B. (1964). *Profit from the learning curve*. Harvard Business Review.

Hopp, W. J., Iravani, S. M., & Yuen, G. Y. (2007). Operations systems with discretionary task completion. *Management Science*, *53*(1), 61-77.

Hoppe, E. I., & Kusterer, D. J. (2011). Behavioral biases and cognitive reflection. *Economics Letters*, *110*(2), 97-100.

Hora, M., Bapuji, H., & Roth, A. V. (2011). Safety hazard and time to recall: the role of recall strategy,

product defect type, and supply chain player in the US toy industry. *Journal of Operations Management*, *29*(7), 766-777.

Huckman, R. S., & Pisano, G. P. (2006). The firm specificity of individual performance: Evidence from cardiac surgery. *Management Science*, 52(4), 473-488.

Huckman, R. S., Staats, B. R., & Upton, D. M. (2009). Team familiarity, role experience, and performance: Evidence from Indian software services. *Management Science*, 55(1), 85-100.

Huckman, R.S. & Zinner, D.E. (2008). Does focus improve operational performance? Lessons from the management of clinical trials. *Strategic Management Journal*, 29(2), 173-193.

Hutchinson, J. Wesley (1983). Expertise and the Structure of Free Recall. *Advances in Consumer Research*, *10*, 585-589.

Iacus, S.M, King, G., & Porro G. (2012). Causal inference without balance checking: Coarsened exact matching. *Political Analysis*, 20(1), 1-24.

Ittner, C.D. & Larcker, D.F. (1998). Are nonfinancial measures leading indicators of financial performance? An analysis of customer satisfaction. *Journal of Accounting Research*, *36*(1-35).

Jain, N., Girotra, K., & Netessine, S. (2014). Managing global sourcing: Inventory performance. *Management Science*, 60(5), 1202-1222.

Jarrell, G., & Peltzman, S. (1985). The impact of product recalls on the wealth of sellers. *The Journal of Political Economy*, 512-536.

Jones, T. M. (1991). Ethical decision making by individuals in organizations: An issue-contingent model. *Academy of management review*, *16*(2), 366-395.

Jenni, K., & Loewenstein, G. (1997). Explaining the identifiable victim effect. *Journal of Risk and Uncertainty*, *14*(3), 235-257.

Juran, J. M. (1999). *Quality and Income*. McGraw-Hill.

Kacmar, K.M., Carlson, D.S., & Batton, V.K. (2004). Situational and dispositional factors as antecedents of ingratiatory behaviors in organizational settings. Journal of Vocational Behavior, 65(2), 309-331.

Kahneman, D. (2011). *Thinking*, *fast and slow*. Macmillan

.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, *47*(2), 263-291.

Kahneman, D., Slovic, P., & Tversky, A. (1982). Judgment Under Uncertainty: Heuristics and Biases. Cambridge University Press.

Kelsey, S. F., Mullin, S. M., Detre, K. M., Mitchell, H., Cowley, M. J., Gruentzig, A. R., et al. (1984). Effect of investigator experience on percutaneous transluminal coronary angioplasty. *The American Journal of Cardiology*, 53(12), C56-C64.

Korniotis, G. M., & Kumar, A. (2013). State-Level Business Cycles and Local Return Predictability. *The*

Journal of Finance, *68*(3), 1037-1096.

Kuntz, L., Mennicken, R., & Scholtes, S. (2014). Stress on the ward: Evidence of safety tipping points in hospitals. *Management Science*. Articles in Advance.

Lemley, M.A. & Sampat, B. (2012). Examiner characteristics and patent office outcomes. *Review of Economics and Statistics,* 94(3), 817-827.

Lewis, K., D. Lange, L. Gillis. (2005). Transactive memory systems, learning, and learning transfer. *Organization Science*, 16(6): 581-598.

Loewenstein, G. F., Weber, E. U., Hsee, C. K., & Welch, N. (2001). Risk as feelings. *Psychological bulletin*, *127*(2), 267-286.

Ma, L., Sun, B., & Kekre, S. (2015). The Squeaky Wheel Gets the Grease—An Empirical Analysis of Customer Voice and Firm Intervention on Twitter. *Marketing Science*.

Macher, J. T., Mayo, J. W., & Nickerson, J. A. (2011). Regulator heterogeneity and endogenous efforts to close the information asymmetry gap. *Journal of Law and Economics,* 54(1), 25-54.

Mayer, K. J., Nickerson, J. A., & Owan, H. (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.

McNeely, B.L., & Meglino, B.M. (1994). The role of dispositional and situational antecedents in prosocial organizational behavior: An examination of the intended beneficiaries of prosocial behavior. *Journal of Applied Psychology, 79*(6), 836-844.

Mitra, D., & Golder, P. N. (2006). How does objective quality affect perceived quality? Short-term effects, long-term effects, and asymmetries. *Marketing Science*, *25*(3), 230-247.

Mittal, V., & Kamakura, W. A. (2001). Satisfaction, repurchase intent, and repurchase behavior: Investigating the moderating effect of customer characteristics. *Journal of marketing research*, *38*(1), 131-142.

Moore, D.A., Tetlock, P.E., Tanlu, L., & Bazerman, M.H. (2006). Conflicts of Interest and the Case of Auditor Independence: Moral Seduction and Strategic Issue Cycling. *Academy of Management Review,* 31(1), 10-29.

Moritz, B. B., Hill, A. V., & Donohue, K. L. (2013). Individual differences in the newsvendor problem: Behavior and cognitive reflection. *Journal of Operations Management*, *31*(1), 72-85.

Moritz, B., Siemsen, E., & Kremer, M. (2014). Judgmental forecasting: Cognitive reflection and decision speed. *Production and Operations Management, Forthcoming*.

Mukherjee, A., Mitchell, W., & Talbot, F. B. (2000). The impact of new manufacturing requirements on production line productivity and quality at a focused factory. *Journal of Operations Management, 18*(2), 139-168.

Murtha, T.C., Kanfer, R., & Ackerman, P.L. (1996). Toward an interactionist taxonomy of personality and situations: An intergrative situational-dispositional representation of personality traits. *Journal of*

Personality and Social Psychology, 71(1), 193.

Muthulingam, S., Corbett, C., Benartzi, S., Oppenheim, B. (2013). Energy efficiency in small and medium sized manufacturing firms: Order effects and the adoption of process improvement recommendations. *Manufacturing & Service Operations Management* 15(4): 596-615.

Narayanan, S., S. Balasubramanian, J. M. Swaminathan. (2009). A matter of balance: Specialization, task variety, and individual learning in a software maintenance environment. *Management Science*, 55(11): 1861-1876.

Nelson Richard, R., & Winter Sidney, G. (1982). An evolutionary theory of economic change. *Harvard Business School Press*, Cambridge.

Newton, T., & Keenan, T. (1991). Further analyses of the dispositional argument in organizational behavior. *Journal of Applied Psychology*, *76*(6), 781-787.

Oechssler, J., Roider, A., & Schmitz, P. W. (2009). Cognitive abilities and behavioral biases. *Journal of Economic Behavior & Organization*, *72*(1), 147-152.

Oflac, B.S., Sullivan, U.Y., & Baltacioglu, T. (2012). An attribution approach to consumer evaluations in logistics customer service failure situations. *Journal of Supply Chain Management*, *48*(4), 51-71.

Oliva, R., & Sterman, J. D. (2001). Cutting corners and working overtime: Quality erosion in the service industry. *Management Science*, *47*(7), 894-914.

Paxton, J. M., Ungar, L., & Greene, J. D. (2012). Reflection and reasoning in moral judgment. *Cognitive Science*, *36*(1), 163-177.

Potters, J., $ Van Winden, F. (2000). Professionals and students in a lobbying experiment: Professional rules of conduct and subject surrogacy. *Journal of Economic Behavior & Organization*, 43(4), 499-522.

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

Powell, A., Savin, S., & Savva, N. (2012). Physician workload and hospital reimbursement: Overworked physicians generate less revenue per patient. *Manufacturing & Service Operations Management*, *14*(4), 512-528.

Qi, J.Z. (2009). Comparison of proportional hazards and accelerated failure time models. Dissertation. University of Saskatchewan. http://ecommons.usask.ca/ bitstream/handle/ 10388/ etd-03302009-140638/JiezhiQiThesis.pdf.

Reagans, R., Argote, L., & Brooks, D. (2005). Individual experience and experience working together: Predicting learning rates from knowing who knows what and knowing how to work together. *Management Science*, 51(6), 869-881.

Roth, A.V., Tsay, A.A., Pullman, M.E., & Gray, J.V. (2008). Unraveling the Food Supply Chain: Strategic Insights from China and the 2007 Recalls. *Journal of Supply Chain Management*, 44(1), 22-39.

Salant, D.J. (1995). Behind the revolving door: a new view of public utility regulation. *The Rand Journal of economics*, 26(3), 362-377.

Schreier, M. (2006). The value increment of mass-customized products: an empirical assessment. Journal of Consumer Behavior, 5(4), 317-327.

Shah, R., Ball, G., & Netessine, S. (2014). Plant Operations and Product Recalls in the Automotive Industry: An Empirical Investigation. Working Paper.

Shefrin, H., & Statman, M. (1985). The disposition to sell winners too early and ride losers too long: Theory and evidence. *The Journal of finance*, *40*(3), 777-790.

Shiv, B., & Fedorikhin, A. (1999). Heart and mind in conflict: The interplay of affect and cognition in consumer decision making. *Journal of Consumer Research*, *26*(3), 278-292.

Short, J. L., Toffel, M. W., & Hugill, A. (2013). What shapes the gatekeepers? Evidence from global supply chain auditors. *Harvard Business School Technology & Operations Mgt.Unit Working Paper*, (14-032).

Sitkin, S. B., & Weingart, L. R. (1995). Determinants of risky decision-making behavior: A test of the mediating role of risk perceptions and propensity. *Academy of management Journal*, *38*(6), 1573-1592.

Skowronski, M. (2012). When the bored behave badly (or exceptionally). *Personnel Review*, 41(2), 143-159.

Slovic, P., Finucane, M. L., Peters, E., & MacGregor, D. G. (2007). The affect heuristic. *European Journal of Operational Research*, *177*(3), 1333-1352.

Slovic, P., Finucane, M., Peters, E., & MacGregor, D. G. (2002). Rational actors or rational fools: Implications of the affect heuristic for behavioral economics. *The Journal of Socio-Economics*, *31*(4), 329-342.

Slovic, P., Flynn, J. H., & Layman, M. (1991). Perceived risk, trust, and the politics of nuclear waste. *Science*, *254*(5038), 1603-1607.

Staats, B. R., & Gino, F. (2012). Specialization and variety in repetitive tasks: Evidence from a Japanese bank. *Management science*, *58*(6), 1141-1159.

Stanovich, K. E., & West, R. F. (2000). Advancing the rationality debate. *Behavioral and Brain Sciences*, *23*(05), 701-717.

Stigler, G. J. (1971). The theory of economic regulation. *The Bell journal of economics and management science*, 2(1), 3-21.

Tan, T. F., & Netessine, S. (2014). When does the devil make work? An empirical study of the impact of workload on worker productivity. *Management Science*, *60*(6), 1574-1593.

Terziovski, M., Samson, D., & Dow, D. (1997). The business value of quality management systems certification. Evidence from Australia and New Zealand. *Journal of Operations Management*, 15(1), 1-18.

Thirumalai, S., & Sinha, K. K. (2011). Product recalls in the medical device industry: An empirical exploration of the sources and financial consequences. *Management Science*, *57*(2), 376-392.

Tinsley, C. H., Dillon, R. L., & Cronin, M. A. (2012). How near-miss events amplify or attenuate risky decision making. *Management Science*, *58*(9), 1596-1613.

Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition*, *39*(7), 1275-1289.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *science*, *185*(4157), 1124-1131.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*(4481), 453-458.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty*, *5*(4), 297-323.

Villena, V. H., Revilla, E., Choi. T. Y. (2011). The dark side of buyer–supplier relationships: A social capital perspective. *Journal of Operations Management,* 29, 561–576.

Weick, K. E., & Roberts, K. H. (1993). Collective mind in organizations: Heedful interrelating on flight decks. *Administrative Science Quarterly,* 38(3), 357-381.

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological review*, *92*(4), 548.

Weiner, B., Russell, D., & Lerman, D. (1979). The cognition–emotion process in achievement-related contexts. *Journal of Personality and Social Psychology*, *37*(7), 1211.

White, R. E., Pearson, J. N., & Wilson, J. R. (1999). JIT manufacturing: a survey of implementations in small and large US manufacturers. *Management Science*, *45*(1), 1-15.

Wilson, T. D., Lisle, D. J., Schooler, J. W., Hodges, S. D., Klaaren, K. J., & LaFleur, S. J. (1993). Introspecting about reasons can reduce post-choice satisfaction. *Personality and Social Psychology Bulletin*, *19*(3), 331-339.

Wilson, T. D., & Schooler, J. W. (1991). Thinking too much: introspection can reduce the quality of preferences and decisions. *Journal of personality and social psychology*, *60*(2), 181-192.

Wowak, A. J., Mannor, M. J., & Wowak, K. D. (2014). Throwing caution to the wind: The effect of CEO stock option pay on the incidence of product safety problems. *Strategic Management Journal*.

Wright, T. (1936). Learning curve. *Journal of the Aeronautical Sciences,* 3(1), 122-128.

Yang, J., Zhou, Y., & Wang, Z. (2010). Conditional coskewness in stock and bond markets: time-series evidence. *Management Science*, *56*(11), 2031-2049.

Yelle, L. E. (1979). The learning curve: Historical review and comprehensive survey. *Decision Sciences,* 10(2), 302-328.

Zimmerman, M. B. (1982). Learning effects and the commercialization of new energy technologies: The case of nuclear power. *The Bell Journal of Economics,* 13(2), 297-310.

Appendix A

Experiment Communication Emails

First Email:

"[The firm name] is partnering with a research team from the [school name] to understand factors that impact product recall decisions in medical devices. A critical aspect of our research partnership is a survey-based study. The survey consists of one hypothetical scenario-based question and a few demographic questions. You have been selected to participate in this study based upon your work experience, geographical location, and functional role. The survey should not take more than 10 minutes to complete. The link below will take you to the survey…

Your participation in this research project is much needed and greatly appreciated. Please note that your responses are completely anonymous. Please complete the survey at your earliest convenience."

Reminder Email:

"This is a reminder to participate in our recall survey. If you have already responded to the survey, thank you for your time. If you have not yet responded to this important survey, we request your participation. The survey closes on [end date].

[The firm name] is partnering with a research team from [school name] to understand factors that impact product recall decisions in medical devices. A critical aspect of our research partnership is a survey-based study. The survey consists of one hypothetical scenario-based question and a few demographic questions. You have been selected to participate in this study based upon your work experience, geographical location, and functional role. The survey should not take more than 10 minutes to complete. The link below will take you to the survey…

Your participation in this research project is much needed and greatly appreciated. Please note that your responses are completely anonymous. Please complete the survey at your earliest convenience."

Experiment Text

Screen 1: Consent

The product recall decision

You are invited to participate in a study investigating factors that impact product-recall decision

making. The results of this research will be used to improve the recall decision making process. Your participation is key to the success of this study. Please note that your responses are completely anonymous and strictly confidential.

The entire survey should take no more than 10 minutes to complete. We really appreciate your taking the time to participate in the study. If you have any questions about the study, please call at [number given] or email [email given]. Thank you for your assistance.

[name]
Principal Investigator
University name

Screen 2: Recall Scenario

[Scenario background, included in every treatment]

You are a member of a cross-functional product recall team. Your team meets whenever signals occur which indicate that a product may need to be recalled. Mark Smith, the Director of Quality for your division, is the person responsible for initiating recall meetings. Mark calls the meeting to order and presents the following scenario.

A cardiac device, on the market for over two years, has had recent failures in the field. Mark reviews the failure data and compares it to predicted failure rates. The hazard analysis predicted a 0.05% failure rate for this defect type. The cumulative failure rate since product launch is 0.05%, but the failure rate has increased to 0.08% for the past 2 months. Before this, it has never been above 0.06% in any month since the product was launched; often times it is well below that number.

There are 10,000 of these devices shipped per week on average. This shipment rate has remained consistent since product launch. If the current failure rate continues at 0.08%, 156 additional devices per year will fail (.0003 x 52 x 10,000) on top of the 260 failures per year originally predicted.

[Factor specific text]

While you may need additional information before deciding whether or not to recall this product, please indicate your most likely recommendation based only on the information provided.

- o Recall. I would recommend recalling all product manufactured since the failure rate increased

- o No Recall. I would recommend that we continue to monitor and investigate, but not recall at this time.

<u>Factor Specific Text:</u>

Physician Concern Factor:

High: One of the defective products was used by Dr. Jones, a prominent physician-customer who is an advocate for your company and someone you have met on several occasions. Dr. Jones was disappointed by the product defect and has emailed the VP of Quality asking what is being done to fix this issue.

Low: *Blank*

Defect Undetectable Factor:

High: Mark also states that if the defect is present in a device, the physician is not likely to observe the defect prior to using the product.

Low: Mark explains that if the defect is present in a device, the physician is likely to observe the defect prior to using the product.

Root Cause Understanding Factor:

High: Both the manufacturing and supplier processes have been reviewed by a team of quality engineers. This team discovered that a critical manufacturing process for this product recently underwent a process change intended to improve production yields. The quality engineering team believes that the field failures are related to this process change.

Low: Both the manufacturing and supplier processes have been reviewed by a team of quality engineers, and no significant shifts in process capability or increases in defects are found. It is unclear at this point what could be causing this problem.

Screen 3. Control Variable Text

Please indicate your functional area
Operations, Quality, Clinical, Medical, or Other

How many years have you worked at your current firm?
0-2, 3-5, 5-10, More than 10

Please select gender
Male, Female

How would you describe the relationship between your company and the FDA?
Collaborative, Average, Confrontational

Please answer the next three questions to the best of your ability.
If it takes 5 machines 5 minutes to make 5 widgets, how many minutes would it take 100 machines

to make 100 widgets?
In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how many days would it take for the patch to cover half the lake?
If you have any feedback that you feel would improve this study, please enter it here. We greatly appreciate your input.