

**Preventing Organizational Misconduct - An Operations
Management Perspective**

**A DISSERTATION
SUBMITTED TO THE FACULTY OF THE
UNIVERSITY OF MINNESOTA
BY**

Finn Petersen

**IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
Doctor of Philosophy**

Advisor: Rachna Shah

June, 2025

© Finn Petersen 2025
ALL RIGHTS RESERVED

Acknowledgements

Growing up, earning a Ph.D. was never something I explicitly aimed for. The first time I seriously considered it was during my Master’s studies. I’m the first in my family to attend college—which, in Germany, places me among the one percent of individuals from non-academic backgrounds who go on to earn a doctoral degree. A fact my brother likes to remind me of, which I take as his way of saying, “I’m proud of you.”

Pursuing a Ph.D. may not have been the “normal” path, but I also never doubted that I would go to college or chase my ambitions—whatever they may have been at the time. For that, I am deeply grateful to my parents and grandparents. They always supported me and created an environment in which I could grow and thrive, even when it eventually meant letting me go.

Reflecting on this journey, I’m acutely aware of the many people who have shaped me into who I am today. Now that this chapter is coming to a close, it feels only right to thank those who helped me begin—and those who helped me see it through.

The person who has had the most profound influence on my academic path is my advisor, co-author, and friend—Professor Rachna Shah. Her impact is impossible to fully express in words. She taught me to be more than just a researcher. Through her mentorship, I’ve learned the value of cultivating a research identity and leaning into my natural instincts as a community builder. I am thankful for her guidance, trust, and belief in me.

I am also incredibly grateful to my committee members for their support throughout

the years. Professor Susan Meyer Goldstein has been a constant presence across every milestone—from written prelims to dissertation defense. Professor Aseem Kaul encouraged my early exploration of strategy research during his Ph.D. seminar and provided valuable feedback on my dissertation. Professor Serguei Netessine’s guidance has also been an essential part of this work, and I look forward to collaborating on other projects in the future. Others, like Professors Anant Mishra and Marc Bellemare, do not appear on the cover of this thesis but have been part of my journey at various milestones. I deeply appreciate their support and insights over the years.

Beyond my committee, I’d like to express my heartfelt thanks to the current and former faculty of the Supply Chain & Operations Department for fostering such a supportive environment for Ph.D. students. Special thanks to Professor Kingshuk Sinha for his leadership and to Professor Kevin Linderman—now at Pennsylvania State University—for bringing me into the program.

I’m also incredibly grateful to the amazing staff who kept everything running smoothly behind the scenes. Thank you to Deborah Brashear, Diane DeBoer, Jill Johnson, and Vicki Lund in the Supply Chain & Operations office, as well as to Professor Michelle Duffy, Amy Brown, and Alex Hoppe in the Ph.D. Office for your ongoing support, kindness, and—almost most importantly—occasional free food.

Throughout my Ph.D. program, I’ve had the privilege of collaborating with outstanding co-authors, including Professors Pinar Karaca Mandic (Finance Department), Salman Arif (Accounting Department), Anna Errore (Emory University), and Ujjal Mukherjee (University of Illinois). A special thank you goes to Professors George Ball (Indiana University) and Vibhuti Dhingra (York University), who have both been generous collaborators and guiding forces in our joint work. Here’s to continued collaboration in the future.

And of course, I couldn’t have made it through this journey without my fellow Ph.D. students. A huge shoutout to Hanu Tyagi—my biggest role model and partner-in-crime

throughout the program. Special thanks to my dear friends: Arindam Debarma, for helping me stay grounded in the early years; Yeonjoo Lee, for the mutual support during our job market year; Dwaipayan Roy, for his mentorship from the very beginning; and Yi Tang, for his steady presence—my couch will always be open for your visit.

To all the other students I've had the pleasure of knowing during the program—Bhupi Juneja, Amir Karimi, Gautham Sundher, Eric Xu, Allison Murphy, Yuanyuan Ding, Yasaman Asayesh, Orchi Bhattacharyya, Andrew Shin, Minjeong Kang, Shenghan Li, and Amirhossein Salamirad—thank you for making this journey richer and more memorable.

Finally, I owe my deepest gratitude to my better half, Annkathrin, who has supported me unconditionally—whether from afar in Germany or after moving to the U.S.—even when it meant that nearly every social outing turned into a lengthy discussion about esoteric topics like “endogeneity.” She has always been my greatest cheerleader, even when I didn't always feel like being cheered on. Thank you, Annkathrin, for everything.

Dedication

To my family
who gave me the roots to grow,
the wings to fly,
and a home to return to.

Abstract

Organizational misconduct—ranging from corporate fraud to environmental disasters—poses severe risks to individuals, businesses, and society. Examples such as Theranos’ fraudulent blood testing, Ford Pinto’s flawed fuel tank design, and the Deepwater Horizon oil spill illustrate the devastating consequences of firms violating legal, ethical, or social norms. Despite growing awareness of these issues, research on effective, evidence-based prevention strategies remains limited.

This dissertation addresses this gap by exploring governance reforms and regulatory interventions as strategies to prevent misconduct. Using an Operations and Supply Chain Management (OSCM) lens, I study misconduct in three different contexts: informed trading during the medical device recall process, environmental injustice in U.S. manufacturing, and hygiene violations by restaurants. In each context, I explore why organizational misconduct occurs and how it can be effectively prevented or mitigated. The findings show that well-designed governance reforms and regulatory interventions can significantly reduce misconduct, providing valuable insights for firms, policymakers, and regulators, including the U.S. Food and Drug Administration (FDA), Securities and Exchange Commission (SEC), Environmental Protection Agency (EPA), and health departments.

The dissertation is organized as follows: Chapter 1 introduces the topic and outlines the essays. Chapter 2 frames organizational misconduct within the OSCM literature and introduces a framework to study it. Chapters 3, 4, and 5 present three empirical essays that explore different aspects of misconduct and its prevention. Finally, Chapter 6 summarizes the findings and discusses their implications for research, practice, and policy.

This research contributes to both academic debate and practical decision-making by shedding light on effective strategies to prevent organizational misconduct. As a result, we have a better understanding of how firms may operate more ethically and responsibly.

Contents

Acknowledgements	i
Dedication	iv
Abstract	v
List of Tables	xi
List of Figures	xiii
1 Introduction	1
1.1 Motivation	1
1.2 Dissertation Overview	3
1.3 Outlook	5
2 The Causes, Consequences, and Prevention of Misconduct	6
2.1 Introduction	6
2.2 Defining Organizational Misconduct	8
2.3 A Process Model to Organizational Misconduct	11
2.3.1 Theoretical Foundations	12
2.3.2 Causes	17
2.3.3 Corrective Actions and Consequences	24

2.3.4	Prevention	25
2.4	Research Opportunities	29
2.4.1	Leveraging Corrective Actions for Personal Gain	29
2.4.2	Optimizing Resource Allocation for Misconduct Prevention	30
2.4.3	Governance Reforms as a Misconduct Prevention Mechanism	31
3	Product Recall Decisions and the CEO: Informed Trading in the Medical Device Industry	33
3.1	Introduction	33
3.2	Related literature and medical device recall background	38
3.2.1	Related literature	38
3.2.2	Medical device recall process	40
3.3	Empirical setting	44
3.3.1	Data	44
3.3.2	Variables	45
3.3.3	Empirical strategy	48
3.4	Results	50
3.4.1	Presence of Informed Trading	50
3.4.2	Profitability of Sales in the Window of Opportunity	52
3.5	Supplementary Analysis	55
3.5.1	Robustness Checks	55
3.5.2	Post-hoc	58
3.6	Discussion and conclusion	62
4	Chief Sustainability Officers and Environmental Injustice: An Empirical Evaluation	69
4.1	Introduction	69
4.2	Background	74

4.2.1	Toxic Releases	75
4.2.2	Environmental Injustice in the United States	76
4.2.3	The Role of Chief Sustainability Officer	77
4.3	Hypothesis development	79
4.3.1	Linking CSO Appointments to Environmental Performance	79
4.3.2	Exploring the Moderating Role of Underserved Communities	81
4.4	Data and Empirical Strategy	83
4.4.1	Data	83
4.4.2	Variables	84
4.4.3	Primary Empirical Strategy	89
4.5	Results	92
4.5.1	Model-Free Evidence	92
4.5.2	Main Results	93
4.6	Robustness Checks and Mechanism Analyses	96
4.6.1	Robustness Checks	97
4.6.2	Alternative Explanations	100
4.6.3	Mechanism Analysis	103
4.7	Discussion and Conclusion	106

5	Improving Inspection Resource Allocation to Control Organizational Misconduct	111
5.1	Introduction	111
5.2	Literature Review	115
5.2.1	Organizational Misconduct and Routine Conformance Decay	115
5.2.2	Predicting Organizational Misconduct	116
5.2.3	Machine Learning in Operations Management	117
5.3	Empirical Setting and Data	119

5.3.1	Health Inspections in the Restaurant Industry	119
5.3.2	Data	120
5.3.3	Variable Definition	121
5.4	Machine Learning Model and Prediction	123
5.4.1	Machine Learning Model	123
5.4.2	Performance Evaluation	125
5.4.3	Prediction Results	129
5.5	Model Formulation	133
5.5.1	Single Establishment Decision-Making Model	134
5.5.2	Establishment Problem	142
5.5.3	Regulator Problem	143
5.5.4	Incorporating Machine Learning Predictions	148
5.6	Evaluation Using Simulation	150
5.6.1	Policies	150
5.6.2	Simulation Design	151
5.6.3	Results	153
5.7	Discussion and Conclusion	157
6	Conclusion and Discussion	160
6.1	Discussion	160
6.2	Key Implications	161
6.3	Future Research	163
	Bibliography	165
	Appendix A. Appendix to Introduction	180
A.1	Tables	180

Appendix B. Appendix to Chapter 3	185
B.1 Overview of the Quality and Operations Officer Roles	187
Appendix C. Appendix to Chapter 4	188
C.1 Variable Definitions	188
C.2 Computing the Demographics Around a Facility	190
C.3 Variable Balance After Matching	191
C.4 Linear Regression for Pre-Treatment Trends	193
C.5 Alternative Matching Methods	194
C.6 Alternative Measures of Underserved Communities	195
C.7 Sub-Sample Event Study	198
C.8 Placebo Analysis	199
C.9 Instrumental Variable Approach	200
C.10 Alternative Explanation: Any CSO Role Affects Toxic Releases	202
C.11 Alternative Explanation: Outsourcing of Production	204
C.12 Heterogeneity of CSO Appointment Effect	205
Appendix D. Appendix to Chapter 5	206
D.1 Food Establishment Inspection Report	208
D.2 Overview of Feature Set	210
D.3 Multinomial Inverse Regression for Review Text Variable Construction	211
D.4 Comparison of Different Machine Learning Algorithms	213
D.5 Variable Importance	214
D.6 Transition Matrix Construction	215
D.7 Simulation Results	217

List of Tables

2.1	Theoretical Foundations for Misconduct Research in OSCM	13
2.2	Selected Misconduct Prevention Strategies in OSCM Literature	26
3.1	Descriptive Statistics	50
3.2	CEO Selling Activity in Window of Opportunity for Class I Recalls	51
3.3	Profitability of CEO Sales in Window of Opportunity using Market-Adjusted 180-days Buy-and-Hold Returns	54
3.4	Robustness Check: Profitability of CEO Sales in Window of Opportunity using Size Adjusted Sales Profits	56
3.5	Robustness Check: CEO Buy-Sell-Imbalance in Window of Opportunity	57
3.6	Robustness Check: Considering Recalls without PMA or 510(k) Number in Analysis	59
3.7	CEO Selling Activity in Window of Opportunity for Class II and III Recalls	61
3.8	Quality and Operations Officer Selling Activity in Window of Opportunity	63
4.1	Descriptive Statistics and Correlations of Matched Sample	92
4.2	Model-Free Evidence of the Effect of CSO Appointments on Toxic Releases	93
4.3	TWFE Results of the Effect of CSO Appointments on Toxic Releases	94
4.4	Dynamic Treatment Effect of CSO Appointments on Toxic Releases	99
4.5	Mechanisms through which CSO appointments affect toxic releases	105
5.1	Descriptive Statistics of Raw Data	121

5.2	Performance of Gradient-Boosted Regression Tree Models for Predicting Violations	130
5.3	Transition probabilities $T_i(s, a_j, a_i, s')$ for each (a_j, a_i) combination	138
5.4	Transition probabilities $T_j(s, a_j, s')$ for each a_j	139
5.5	Simulation Setup for Establishment Problems	152
5.6	Pairwise Comparisons of Policies.	154
6.1	An Overview of the Dissertation	162
A.1	Types of Organizational Misconduct in the OSCM Literature	181
A.2	Selected Organizational Misconduct Definitions	183
B.1	Detailed Overview of the Quality and Operations Officer Roles	187
C.1	Description of Relevant Variables	189
C.2	Linear Regression Results of Pre-Treatment Trends	193
C.3	Robustness Results for the Moderation Effect Analysis with Alternative Measures for Underserved Community	197
C.4	2SLS Analysis of the Effect of CSO Presence on Toxic Releases	201
C.5	Two-Way Fixed Effects Results of the Effect of ESG Committee Creation on Toxic Releases	203
C.6	Two-Way Fixed Effects Results of the Effect of CSO Appointments on Production Ratio	204
D.1	Description of Predictors for Machine Learning Model	210
D.2	Comparing the Predictive Performance of Machine Learning Models	213
D.3	Transition Probability Matrices	216
D.4	Comparison Between Various Inspection Resource Allocation Policies	218

List of Figures

2.1	A Process Model of Organizational Misconduct	11
2.2	The Causes of Organizational Misconduct in the OSCM Literature	20
3.1	Recall Timeline and the Window of Opportunity	42
4.1	Waste Generation Process Diagram	75
4.2	Balance Plots Displaying the Likelihood to Appoint a CSO	90
4.3	Marginal Effect Plot for the Moderating Effect of <i>Black/African American</i>	96
4.4	Event Plot of the Effect of CSO Appointments on Toxic Releases	98
5.1	Framework: Integrating Predictions into Resource Allocation Decisions	113
5.2	Rolling Window Forecasting	129
5.3	Comparison of Predictions Conditional on True Outcome	131
5.5	Establishment Policy and Belief Updates	142
5.7	Discrete-Time Survival Analysis of Inspections	156
C.1	Example of 1-mile-Radius Host Neighborhood	190
C.2	Variable Balance After Matching: Covariate Balance	191
C.3	Variable Balance After Matching: Distributional Balance	192
C.4	Specification Analysis of Alternative Matching Settings	196
C.5	Event Plot of the Effect of CSO Appointments on Toxic Releases by Community Type	198
C.6	Placebo Analysis Results	199

C.7 Effect of CSO Appointments based on the CSO's Race and Gender	205
D.1 Variable Importance	214

Chapter 1

Introduction

1.1 Motivation

Researchers studying Operations and Supply Chain Management (OSCM) have historically focused on how organizations should manage “work to produce valuable results” (Browning 2020, p. 498). Yet, a growing body of work examines cases where operational decisions lead to socially undesirable outcomes such as Ford Pinto’s flawed fuel tank design (Gioia 1992), British Petroleum’s Deepwater Horizon oil rig explosion (Wang et al. 2018), or the Rana Plaza factory collapse (Liu et al. 2019, Jacobs and Singhal 2017). Inspired by such incidents, many OSCM researchers have turned their attention to organizational behavior that leads to severe negative consequences for individuals, organizations, and society such as health risks, environmental damage, human rights violations, and financial losses. I refer to this behavior as “organizational misconduct” (or just “misconduct”).

Organizational misconduct, broadly speaking, is any behavior in or by an organization that a control agent deems to contravene laws, industry or professional codes, organizational policies, social norms, or ethical principles (Greve et al. 2010). Although OSCM scholars have yet to widely adopt this conceptualization (c.f., Kistler et al. 2024, Skilton and Bernardes 2022), several established research streams within the field examine specific

misconduct-related phenomena in isolation. For instance, research has looked into environmental violations (Hardcopf et al. 2021, Mani and Muthulingam 2019, Lo et al. 2018), delayed product recalls (Darby et al. 2023, Ball et al. 2018a), misrepresentation of product specifications (Belavina et al. 2020), and workplace safety failures that result in employee injuries (Pagell et al. 2020, Liu et al. 2019). Despite their value, these studies frequently lack a unified framework, which makes it more difficult to combine their findings and create comprehensive misconduct prevention strategies.

Other disciplines—including sociology, management, and organizational theory—have taken a more unified approach to studying misconduct. Their work culminates in several review articles that offer insights into the causes and consequences of misconduct (e.g., Vaughan 1999, Tenbrunsel and Smith-Crowe 2008, Greve et al. 2010, Palmer 2012, Palmer et al. 2016b, Hersel et al. 2019, Epstein and Hanson 2021, Anand et al. 2023). Yet, despite this growing academic understanding, regulatory scrutiny, and public awareness, firms continue to engage in misconduct. This persistence emphasizes the need for a more comprehensive understanding of misconduct’s causes, consequences, and prevention strategies (Anand et al. 2023, Palmer 2012).

To help prevent organizational misconduct, my dissertation examines misconduct in three distinct contexts: informed trading during the medical device recall process, environmental injustice in U.S. manufacturing, and hygiene violations by restaurants. Building on the aforementioned organizational misconduct literature, my essays address two fundamental research questions: (1) Why does organizational misconduct occur in different operational contexts? and (2) How can it be effectively prevented or mitigated? While each context differs in the type of misconduct, they reveal common mechanisms that can help prevent misconduct, allowing for a broader understanding of how regulators and firms can prevent such behavior.

1.2 Dissertation Overview

In my dissertation, titled **Preventing Organizational Misconduct - An Operations Management Perspective**, I first provide a definition of organizational misconduct that aggregates diverse operations management concepts and present a conceptual framework that captures the causes, consequences, and prevention strategies of misconduct (Chapter 2). I then identify misconduct prevention strategies and evaluate their effectiveness in three contexts. In my first essay (Chapter 3), I investigate informed trading by CEOs during product recall decisions and propose how to prevent such behavior. My second essay (Chapter 4) examines whether Chief Sustainability Officer appointments can reduce environmental injustice, a violation of social norms. In my third essay (Chapter 5), I study how regulators can leverage machine learning to optimize inspection programs to prevent restaurant hygiene violations.

Essay I

In my first essay titled “**Product Recall Decisions and the CEO: Informed Trading in the Medical Device Industry,**” I investigate misconduct in the medical device product recall process. Specifically, I study whether CEOs engage in informed trading, which refers to selling shares based on private information for personal gain. It is illegal. The recall process provides an opportunity for such behavior because firms learn about a product defect long before they initiate the recall. CEOs, who know about most upcoming recall decisions, might exploit the opportunity by selling their shares before the market reacts to the recall news.

To test whether CEOs exploit this opportunity, I combine data on recalls, customer complaints, firm finances, stock prices, and executive stock trades for publicly traded U.S. medical device manufacturers (2003–2020). My analysis shows that CEOs increase stock sales during the recall decision process, but only for severe recalls without prior complaints.

Doing so helps them avoid financial losses of up to 20%. This presents novel evidence of illegal informed trading in an operational context. The results suggest that closer collaboration between the Food and Drug Administration (FDA) and the Security and Exchange Commission (SEC) is key to detecting and preventing such behavior.

Essay II

In my second essay titled “**Chief Sustainability Officers and Environmental Injustice: An Empirical Evaluation,**” I study the relationship between Chief Sustainability Officer (CSO) appointments and environmental injustice. Environmental injustice occurs when manufacturing facilities in low-income, Black neighborhoods (underserved communities) have higher toxic releases than those in wealthy, white neighborhoods. This results in disproportionate health risks for the underserved communities. I investigate whether appointing a CSO reduces such injustice. Drawing on the attention-based view and operations frontier theory, I hypothesize that facilities, especially those in underserved communities, reduce their toxic releases following CSO appointments.

To test this, I compile data on facility-level toxic releases, CSO appointments, and facility neighborhood demographic information. I compare the toxic releases of 165 firms that appointed a CSO between 2000 and 2020 with those of 433 firms that did not. A difference-in-differences analysis with matching shows that appointing a CSO correlates with a 12% reduction in toxic releases, driven primarily by facilities in underserved communities. This reduces environmental injustice. These findings highlight the effectiveness of CSO appointments as a new mechanism for addressing environmental injustice, which regulators have struggled to reduce since the 1980s.

Essay III

In my third essay titled “**Improving Inspection Resource Allocation to Control Organizational Misconduct,**” I explore how regulators can use machine learning to optimize the allocation of restaurant hygiene inspection resources. Hygiene violations are a well-established form of misconduct with tangible social costs, including foodborne illnesses and deaths. Inspections by health departments are essential for maintaining hygiene standards, but limited resources constrain their frequency. I study whether incorporating misconduct predictions from machine learning can improve inspection resource allocation and how it affects the effectiveness of inspection programs.

To examine this, I model hygiene violations as an interactive partially observable Markov process based on the understanding that misconduct may result from decay in adherence to operational routines. I then derive an optimal inspection resource allocation policy that incorporates machine learning predictions. Finally, I validate the model using simulations based on Yelp reviews and inspection data from over 2,000 Las Vegas restaurants from 2012 to 2019. The findings demonstrate how machine learning can improve inspection program efficiency and prevent misconduct, a novel approach in the context of regulatory interventions.

1.3 Outlook

The next chapter introduces the concept of organizational misconduct in an OSCM context. Chapters 3, 4, and 5 cover essays I, II, and III, respectively. The last chapter summarizes important findings and discusses the dissertation’s implications for OSCM research, management practices, and policymaking.

Chapter 2

The Causes, Consequences, and Prevention of Misconduct

2.1 Introduction

Before exploring the occurrence and prevention of organizational misconduct in the context of medical device recalls, environmental injustice, and restaurant hygiene violations, this chapter introduces the concept of organizational misconduct and its connection to existing OSCM phenomena. In doing so, I make the following three contributions:

First, I highlight that what may seem like isolated phenomena in the OSCM literature are, in fact, different forms of organizational misconduct. This implies that the same theoretical foundations can explain their causes or consequences and that the same mechanisms can help in their prevention. Recognizing these connections helps advance OSCM research, because it allows researchers to build on existing findings rather than addressing related issues in isolation. Consider, for instance, how competition contributes to misconduct. Theory suggests that competition creates resource scarcity and performance strain, which increases the risk of organizational misconduct (Vaughan 1999, Greve et al. 2010). Empirical findings support this hypothesis in various contexts, such as product safety

issues (Ball et al. 2018a, Lam et al. 2022), fraudulent emissions testing (Bennett et al. 2013), opioid oversupply (Skilton and Bernardes 2022), and automaker non-compliance with emissions standards (Hu et al. 2021). Thus, the relationship between competition and misconduct is well established. At the same time, less is known about how competition affects the corrective actions that firms choose after misconduct (Hersel et al. 2019), which is something that OSCM research on recall decisions (e.g., Darby et al. 2023, Mukherjee et al. 2022) could help address. Thus, by recognizing misconduct as a unifying construct to various OSCM phenomena, OSCM researchers can take advantage of existing findings (without “reinventing the wheel”) and find novel and meaningful research opportunities (Cachon 2012, Oxley et al. 2010).

Second, I create a conceptual framework that combines insights from OSCM with organizational misconduct research, providing a structured approach to examining misconduct through an OSCM lens. Even though research on misconduct is common in adjacent fields, existing OSCM studies often lack a coherent framework to systematically analyze its causes, consequences, and prevention strategies. The framework provides researchers with guidelines to study misconduct more comprehensively by synthesizing key insights from multiple literature streams. It provides a foundation for formulating more precise research questions, identifying relevant variables, and structuring empirical analyses. In doing so, I aim to give OSCM scholars the resources they need to make a significant contribution to the broader field of organizational misconduct research.

Third, I list important directions for future OSCM research on organizational misconduct. These research opportunities directly inform the three essays of my dissertation, each of which explores a different type of misconduct.

The remainder of this chapter is organized as follows: Section 2.2 defines organizational misconduct and its key components. Section 2.3 presents the conceptual framework, and Section 2.4 discusses future research opportunities.

2.2 Defining Organizational Misconduct

While OSCM researchers have begun examining undesirable behavior in organizations, their studies focus on distinct phenomena such as product safety issues, vehicle emissions non-compliance, and OSHA safety violations (see Table A.1 in the Appendix). Analyzing these phenomena reveals a common theme: all studies concern actions by organizations that harm (or potentially harm) stakeholders and are identified by regulators, media, buyers in the supply chain, or researchers. Each study contributes to this general definition by specifying: i) the harmed stakeholder, ii) the nature of harm inflicted, and iii) the entity responsible for determining the behavior as harmful. For instance, Wiengarten et al. (2017a) examine workplace safety violations that harm employees, where the violations are identified by OSHA inspectors. Scott et al. (2021) investigate road safety violations by truck drivers that harm other drivers, where the violations are identified by the FMCSA. Skilton and Bernardes (2022) and Attari et al. (2024) study opioid oversupply harming users or society, where the oversupply is identified by the researchers. These studies offer important insights. Yet they do not form a coherent body of work, even though they define undesirable behavior similar. I propose consolidating these concepts into a higher-order construct of organizational misconduct to increase their impact.

Sociology, Management, and Organizational Theory scholars have long used the concept of organizational misconduct to aggregate findings across various contexts (e.g., Greve et al. 2010, Hersel et al. 2019). Drawing on their work (see Table A.2), I propose the following definition of organizational misconduct in OSCM: *behavior in or by an organization that contravenes a line separating right from wrong; where such a line can be defined by laws, industry and professional codes, organizational rules, social norms, or ethical principles. Whether behavior contravenes a line, in turn, can be judged by a control agent or the researcher.*¹

¹I thank participants of the Professional Development Workshop “Organizational Misconduct: Developing an Operations Management-Based Research Agenda” at the 2023 Academy of Management conference for

This definition has four key components. First, contravening behavior includes not only intentional or malicious acts but also unintended adverse outcomes of well-intended actions (Vaughan 1999), lack of moral awareness (Tenbrunsel and Smith-Crowe 2008), or accidents (Greve et al. 2010). The judgment of a control agent or researcher is crucial. Control agents represent collectives (e.g., international bodies, states, professional associations) and can impose sanctions (Greve et al. 2010). In OSCM, buyers may act as de facto control agents, judging supplier behavior against a Supplier Code of Conduct (Jiang 2009) and imposing sanctions (Caro et al. 2018, Porteous et al. 2015). Other prominent control agents in OSCM include the FDA (Ball et al. 2017), OSHA (Wiengarten et al. 2017a), EPA (Muthulingam et al. 2022), and the media (Kim and Wagner 2021, Hardcopf et al. 2021). However, their judgments can be biased or limited (Greve et al. 2010) due to for example political pressures (Heitz et al. 2021) or inspection schedules (Ibanez and Toffel 2020, Ball et al. 2017, Short et al. 2016). Acknowledging these biases helps researchers understand the complex dynamics of organizational misconduct and the role of control agents in identifying and responding to it (Greve et al. 2010).

Second, I explicitly consider organizational misconduct as judged by control agents or researchers. Relying on control agents' judgments has emerged as the dominant perspective in the broader misconduct literature because it eliminates ambiguity and makes the definition empirical (Greve et al. 2010). OSCM research that follows this approach often relies on data of violations, citations, newspaper articles, or penalties to measure the occurrence of misconduct (e.g., Hardcopf et al. 2021, Scott et al. 2021, Wiengarten et al. 2017a). This approach, however, only allows researchers to consider instances of misconduct that a control agent identified. In contrast, identifying misconduct based on a researcher-defined standard is subjective but enables the exploration of unreported instances of misconduct. While rare in the broader misconduct literature, this approach is common in the OSCM literature (e.g., Attari et al. 2024, Skilton and Bernardes 2022,

their input regarding this definition.

Hu et al. 2021, Chang et al. 2021, Bennett et al. 2013, Pierce and Snyder 2008). While the empirical strategies vary, the basic approach can be described using a concept familiar to most OSCM scholars from undergraduate teaching: statistical process control charts. Researchers first use regulations or the available data to establish normal behavior (e.g., the expected emission test pass rate based on available vehicle and test characteristics). Then they compare the actual behavior (e.g., the actual emission test pass rate) against control limits established by the normal behavior to identify statistical outliers that hint at the presence of misconduct. I take this approach in my first essay (Chapter 3), where I identify the presence of informed trading based on statistically significant changes to CEO trading activity during the product recall process.

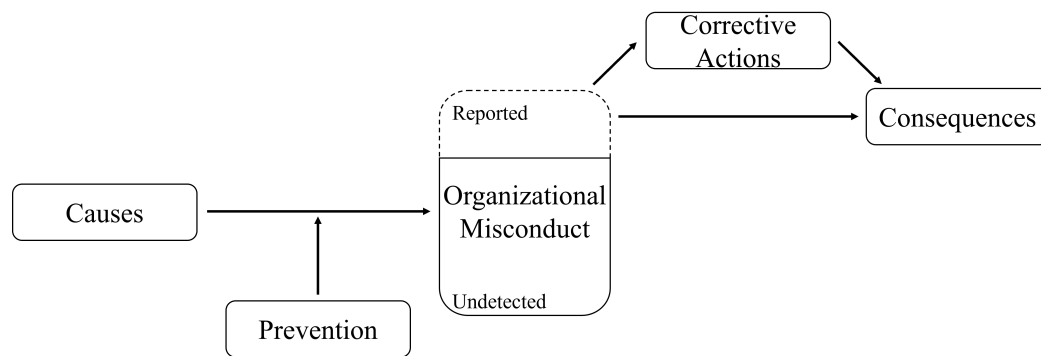
Third, the definition does not restrict misconduct to particular stakeholders or harm thresholds. Misconduct includes any behavior that harms or potentially harms stakeholders (Hersel et al. 2019), such as employees (Pagell et al. 2020), consumers (Wowak and Boone 2015), investors (Belavina et al. 2020), society (Steurer et al. 2005), buyers (Liu et al. 2019), competitors (Kim and Wagner 2021), governments (Karpoff et al. 1999), or the environment (Hardcopf et al. 2021). I do not limit misconduct to behavior that causes harm because the extent of harm can be outside the actor's control (Vaughan 1999) and because control agents frequently penalize organizations for organizational misconduct before harm occurs (e.g., Muthulingam et al. 2022, Ball et al. 2017, Gawande and Bohara 2005).

Lastly, in or by organizations refers to behavior by individuals or groups of individuals, acting in their organizational roles (Vaughan 1999). For example, a CEO who illegally demonstrates inside the Capitol in their free time (Romero 2021) is not committing misconduct. On the other hand, a CEO who illegally sells stock based on private information commits misconduct because they obtained the information in their organizational role and certain organizational prevention strategies, such as blackout periods, whistleblower protection programs, internal audits, or robust corporate governance structures, must have failed for the behavior to occur.

2.3 A Process Model to Organizational Misconduct

While existing OSCM research is clustered by the type of misconduct, I propose a framework based on the four stages of organizational misconduct: Causes, Corrective Actions, Consequences, and Prevention. Figure 2.1 illustrates the relationship between these stages.

Figure 2.1: A Process Model of Organizational Misconduct



Misconduct arises from an interplay between misconduct-causing factors and ineffective prevention strategies. First, factors that explain why individuals or groups of individuals engage in misconduct must be present. These causes include characteristics of the organization and environment that create pressure, opportunity, and predisposition for misconduct (Schnatterly et al. 2018, Vaughan 1999, Baucus 1994).

To illustrate these three categories, consider the case of fraudulent emission testing (Bennett et al. 2013, Pierce and Snyder 2008). In fraudulent emission testing, testing centers facing intense competition feel **pressure** to pass vehicles (Bennett et al. 2013). However, this pressure alone does not lead to fraud; the **opportunity**, such as easier manipulation of older car tests, must also exist (Bennett et al. 2013). Additionally, an organizational culture that normalizes misconduct **predisposes** workers to engage in such behavior, creating a situation in which independent inspectors “adjust the rate at which they pass vehicles to

the norms of those with whom they work” (Pierce and Snyder 2008).

However, these causes are necessary but not individually sufficient for organizational misconduct to occur. Instead, the occurrence of misconduct is contingent on a lack of effective prevention strategies (McKendall et al. 2002, Baucus 1994, Finney and Lesieur 1982). These strategies can either be regulatory interventions, such as inspection programs aimed at deterring misconduct by increasing the likelihood of detection (Gawande and Bohara 2005), or governance reforms, such as self-imposed blackout periods to prevent insider trading (Blackburne et al. 2021).

In the event of misconduct being discovered, firms have the opportunity to mitigate adverse outcomes through a range of corrective measures, including executive dismissals, organizational adjustments, policy reforms, and product recalls (Hersel, 2019). These actions can moderate or mediate the impact of misconduct. For instance, product recalls, a common corrective action for safety issues (Wowak et al. 2015), mediate negative consequences. It is often the recall, rather than the safety issue itself, that triggers a negative stock market reaction (Darby et al. 2023). Thus, the corrective action plays a critical role in shaping the consequences.

2.3.1 Theoretical Foundations

Before analyzing the different phases of misconduct in OSCM, it is essential to establish the theoretical underpinnings that inform this research. Table 2.1 synthesizes key theories used to examine misconduct, outlining their core principles, relevance to different phases of misconduct, and existing applications in OSCM research. While not exhaustive, this compilation captures the most widely cited theoretical perspectives in misconduct research.

Table 2.1 illustrates the diversity of theoretical approaches used to conceptualize misconduct. Some theories, such as rational choice theory and ethical decision theory, emphasize individual decision-making processes and the cognitive mechanisms underlying misconduct. Others, such as agency theory and strain theory, highlight structural and

Table 2.1: Theoretical Foundations for Misconduct Research in OSCM

Theory	Description	Phase	Examples
Agency (Eisenhardt 1989, Jensen and Meckling 1976)	Conflicts between principals (owners, regulators) and agents (employees, managers) lead to misconduct when the agent prioritizes self-interest over organizational or societal interests. Misalignment in incentives is a key driver.	Causes, Consequences	Scott et al. (2021), Steven et al. (2014), Gawande and Bohara (2005)
Conflict (Collins 1975, Coser 1967)	Different groups (e.g., labor vs. management, supplier vs. buyer) disagree about what behavior constitutes misconduct and how it is enforced. Control agents have their own interests but often represent the most powerful group's interests.	Causes	
Ethical Decision (Tenbrunsel and Smith-Crowe 2008)	Individuals assess the morality of actions before engaging in any behavior. Factors such as moral awareness, judgment, and intent influence whether misconduct occurs.	Causes	
Expectancy (Vroom 1964)	Actors engage in misconduct when they believe it will lead to desired outcomes (expected reward) and they have confidence in their ability to execute the act successfully (perceived feasibility).	Causes	Tucker (2016)

Continued on next page

Table 2.1: Theoretical Foundations for Misconduct Research in OSCM (Continued)

Theory	Description	Phase	Examples
Labeling (Becker 1963, Lemert 1951)	The response to initial wrongdoing can reinforce further misconduct, meaning that being labeled as “deviant” after an initial act of misconduct (primary deviance) can lead to further misconduct (secondary deviance):	Causes	
Network	Social networks influence behavior through shared norms, peer pressure, and the diffusion of unethical practices. Firms embedded in corrupt or non-compliant networks are more likely to engage in misconduct.	Causes	DuHadway and Narasimhan (2021), Mohliver (2019)
Normal Accident (Perrow 1984)	In highly complex and tightly coupled systems, misconduct or failures may be inevitable rather than intentional. Thus, systemic factors, rather than individual choices, lead to misconduct.	Causes	Wiengarten et al. (2017a), Wolf (2001)
Rational Choice (Arrow 1963)	Individuals are self-interested and make rational decisions based on expected benefits and costs. They may engage in misconduct if it maximizes their personal or organizational gains.	Causes	Scott and Nyaga (2019)
Regulation Compliance (Thornton et al. 2009, Genn 1993)	Compliance with laws and regulations depends on enforcement mechanisms, perceived legitimacy of rules, and the interests of stakeholders.	Causes	Hu et al. (2021)

Continued on next page

Table 2.1: Theoretical Foundations for Misconduct Research in OSCM (Continued)

Theory	Description	Phase	Examples
Strain (Greve et al. 2010, Merton 1938)	Actors are more likely to engage in misconduct if they cannot achieve their goals through legitimate means. These goals are not always objective but can be tied to aspiration levels.	Causes	Wiengarten et al. (2019)
Attribution (Lange and Washburn 2012, Kelley and Michela 1980)	Examines how stakeholders interpret and assign blame for misconduct. Attributions of internal (intentional) vs. external (situational) causes influence reputational and legal consequences.	Corrective Actions, Consequences	Mayo et al. (2022), Kim and Wagner (2021), Ball et al. (2018a)
Institutional (Meyer and Rowan 1977, DiMaggio and Powell 1983)	Organizations adhere to norms and regulations to maintain legitimacy. Misconduct damages legitimacy, leading to stakeholder withdrawal and financial losses.	Consequences	Liu et al. (2019), Lehman et al. (2014)
(Psychological)	Misconduct violates implicit or explicit agreements between firms and stakeholders, leading to trust erosion and reputational damage. Buyers and consumers adjust their behavior based on perceived contract breaches.	Consequences	Hartmann et al. (2022)
Contract (Rousseau 1995)			
Resource	Power dynamics (i.e., mutual dependence and power asymmetry) determine how misconduct is punished or tolerated between two parties.	Consequences	Beattie et al. (2021)
Dependence (Pfeffer and Salancik 1978, Emerson 1962)			

Continued on next page

Table 2.1: Theoretical Foundations for Misconduct Research in OSCM (Continued)

Theory	Description	Phase	Examples
Organizational Learning (Fiol and Lyles 1985)	Firms learn from past misconduct by adapting policies, refining decision-making processes, and institutionalizing best practices to prevent future violations. Learning can be internal or influenced by industry-wide experiences.	Corrective Actions, Prevention	Mani and Muthulingam (2019), Thirumalai and Sinha (2011)
Signaling (Spence 1973)	Firms use corrective actions and public disclosures as signals to stakeholders about their commitment to ethical behavior after misconduct is revealed. The credibility of these signals affects stakeholder reactions.	Corrective Actions	Mollenkopf et al. (2022), Ni et al. (2014)

incentive misalignment that create opportunities and pressure for misconduct. Meanwhile, perspectives like network theory and institutional theory underscore the role of social, regulatory, and organizational environments in shaping both misconduct and responses to it. The multitude of theories highlights the multitude of perspectives in how researchers frame misconduct—as a rational deviation from expectation, an outcome of systemic pressures, or a socially embedded phenomenon (Anand et al. 2023, Palmer 2012).

Integrating these theoretical lenses into OSCM research offers an opportunity for a more comprehensive understanding of misconduct. While theories such as agency theory and attribution theory have been extensively applied in the OSCM literature (e.g., Scott et al. 2021, Steven et al. 2014, Kim and Wagner 2021), others, such as labeling theory or conflict theory, remain underexplored despite their potential to highlight how firms and stakeholders interpret, respond to, and recover from misconduct.

The following sections will build on these theoretical foundations to examine the causes of misconduct, its organizational and market-level consequences, and the mechanisms to prevent and mitigate it.

2.3.2 Causes

Efforts to understand the causes of misconduct can be broadly categorized into two perspectives (Palmer 2012, Baucus and Near 1991): a **dominant perspective**, which assumes that misconduct is intentional and stems from rational decision-making, and an **alternative perspective**, which considers misconduct as potentially unintentional, arising from bounded rationality and contextual influences.

The dominant perspective is rooted in economic and decision-theoretic frameworks that conceptualize misconduct as a deliberate choice driven by incentives and constraints. This view is found in theories such as agency theory (Jensen and Meckling 1976) or rational choice theory (Arrow 1963) and reflected in models such as the fraud triangle (e.g., Schnatterly et al. 2018) and variations of the Motive, Means/Ability, Opportunity (MMO or MOA)

framework (e.g., Pendse 2012, McKendall and Wagner III 1997). The fraud triangle posits that misconduct arises when individuals face pressure or need, perceive an opportunity to act unethically, and can rationalize their behavior (Schnatterly et al. 2018). In the OSCM context, this framework has been applied to examine product safety violations and recall decisions (DuHadway and Narasimhan 2021). Similarly, the MMO/MOA framework suggests that misconduct occurs when an actor is motivated by external pressures, possesses the means to engage in misconduct due to weak oversight, and perceives an opportunity to act without detection (Pendse 2012). In OSCM research, this framework has been used to study intentional violations of road safety regulations (Scott and Nyaga 2019). The dominant perspective also underlies many analytical models designed to optimize audit and inspection mechanisms aimed at deterring misconduct (e.g., Levi et al. 2020, Caro et al. 2018, Plambeck and Taylor 2016).

While the dominant perspective provides valuable insights into misconduct, its primary limitation is its inability to account for unintentional misconduct (Palmer 2012, Baucus and Near 1991). This shortcoming is particularly relevant given my definition of misconduct (Section 2.2), which explicitly includes unintended adverse outcomes resulting from well-intended actions (Vaughan 1999), failures in moral awareness (Tenbrunsel and Smith-Crowe 2008), and accidents arising from systemic complexities (Greve et al. 2010). The alternative perspective addresses these cases by acknowledging that misconduct can emerge not only from deliberate wrongdoing but also from cognitive limitations, organizational pressures, and social influences (Palmer 2012). It is supported by theories such as ethical decisions Tenbrunsel and Smith-Crowe (2008), network (Greve et al. 2010), or labeling (Becker 1963, Lemert 1951).

From this alternative viewpoint, actors do not necessarily develop a conscious preference for misconduct before engaging in it (Tenbrunsel and Smith-Crowe 2008). Instead, they may slip into unethical behavior due to bounded rationality (Simon 1955), mindless decision-making, or social normalization of deviance (Palmer 2012). One framework

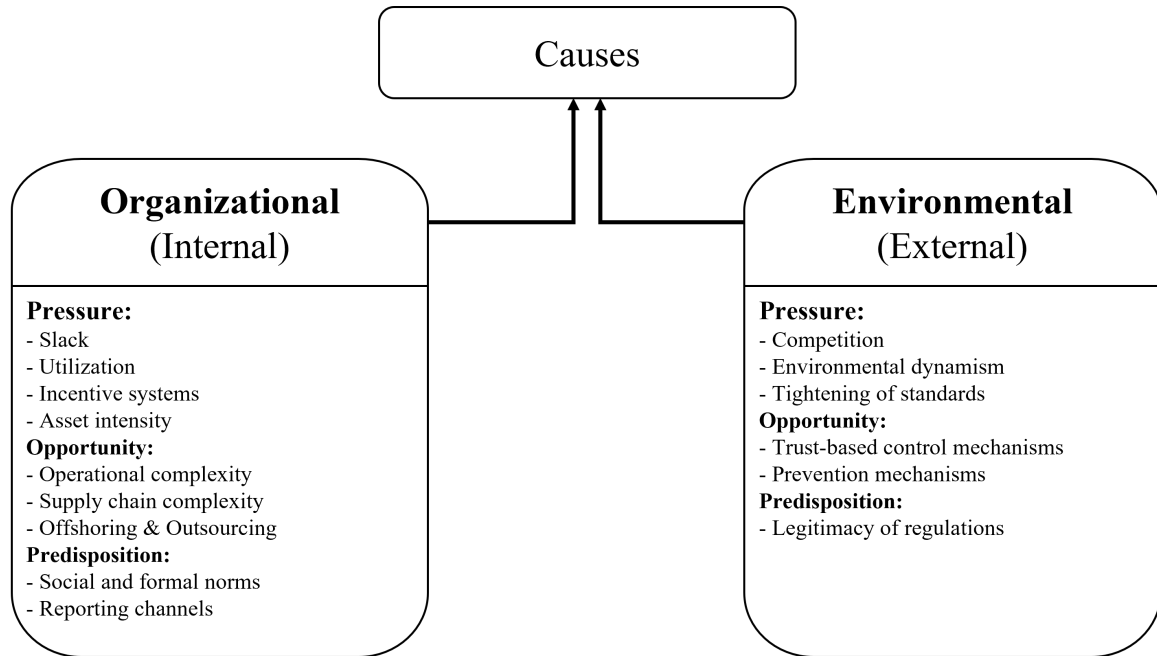
aligned with this perspective is the Pressure-Opportunity-Predisposition model proposed by Baucus (1994), which extends traditional explanations by integrating both intentional and unintentional pathways to misconduct.

Pressure arises when firms or individuals face urgent demands or constraints that compel them to act, often without full consideration of ethical implications (Greve et al. 2010, Baucus 1994). This pressure may stem from financial struggles, competitive intensity, or performance expectations (Greve et al. 2010). Opportunity refers to structural and environmental conditions that enable or fail to prevent misconduct (Baucus 1994). These conditions create the capacity for wrongdoing by reducing oversight, limiting accountability, or facilitating illicit actions through organizational complexity (Vaughan 1999, Baucus 1994). For example, regulatory gaps and weak enforcement, as well as firm size and supply chain complexity, can create opportunities for misconduct (Attari et al. 2024). Predisposition reflects organizational characteristics that shape the likelihood of engaging in misconduct, whether intentional or unintentional (Baucus 1994). It encompasses corporate culture, norms, and structural arrangements that can either encourage or inadvertently facilitate wrongdoing (Palmer 2012, Baucus 1994). For example, cultures of secrecy, the diffusion of responsibility across multiple roles, or the normalization of rule-bending in pursuit of efficiency can make misconduct more likely—even among employees who do not actively seek to break rules (Palmer 2012).

In my work, I predominantly adopt this alternative perspective. Thus, rather than viewing misconduct as an ethical lapse or managerial outlier, I conceptualize it as a process embedded in day-to-day operational decisions, routines, and incentive structures. Below, I systematically examine the causes of misconduct using the Pressure-Opportunity-Predisposition framework based on two key sources of influence (Anand et al. 2023, Schnatterly et al. 2018, Vaughan 1999, Baucus 1994): organizational characteristics (e.g., structure, processes, tasks, culture) and environmental characteristics (e.g., shareholder pressure, competition, regulatory constraints). This approach enables

a more comprehensive understanding of how internal and external factors jointly shape misconduct.

Figure 2.2: The Causes of Organizational Misconduct in the OSCM Literature



Notes. Pressure arises when firms or individuals face urgent demands or constraints that compel them to act. Opportunity refers to structural and environmental conditions that enable or fail to prevent misconduct. Predisposition reflects organizational characteristics that shape the likelihood of engaging in misconduct, whether intentional or unintentional.

Organizational (Internal) Factors

Pressure or Need

A lack of slack resources imposes constraints that generate pressure on organizations, thereby increasing the risk of misconduct (Baucus 1994). Empirical OSCM research suggests that reductions in operational slack can lead to regulatory violations. For instance, Wiengarten et al. (2017a) demonstrate that firms with fewer reserved inventory days have greater rates of OSHA violations.

High operational utilization, like few slack resources, can put pressure on organizations

and increase the likelihood of misconduct. Firms often need to speed up work under such conditions, which raises the risk of both intentional non-compliance and inadvertent errors (Perrow 1984). Supporting this claim, Shah et al. (2017) find that higher utilization rates in automotive manufacturing plants correlate with increased product safety violations and subsequent recalls.

Incentive structures within firms also influence the extent to which individuals experience pressure to engage in misconduct. Compensation schemes that link earnings to performance targets can create strong incentives for rule-breaking behaviors (Palmer 2012). According to Scott and Nyaga (2019), truck drivers are more likely to exceed legally mandated work-hour limits if their incentive to do so increases. This effect is most noticeable among drivers who own their vehicles, as the financial burden of asset ownership increases their performance pressures. This suggests that asset intensity itself may be an organizational factor contributing to misconduct.

Opportunity

Complexity within an organization can create opportunities for misconduct by preventing effective oversight and control. The management of multiple product lines, complex operational processes, and decentralized decision-making systems increases the misconduct risk. Empirical studies support this claim. For example, Shah et al. (2017) find that product variety within firms is associated with higher rates of product safety violations. Similarly, Wolf (2001) shows that more complex refinery operations have much higher rates of accidental hazardous chemical releases.

Beyond internal complexity, supplier network characteristics also shape the opportunities for misconduct. The dispersion of supply chains—measured by the extent to which firms source products from geographically distributed suppliers—can provide opportunities for misconduct. According to Levi et al. (2020), greater supply chain dispersion correlates with increased instances of food adulteration. Similarly, Attari et al.

(2024), illustrate how the vertical and horizontal complexity of pharmaceutical supply chains allowed pharmacies to evade Drug Enforcement Administration (DEA) scrutiny while oversupplying opioids.

Sourcing decisions further influence misconduct opportunities. Firms that offshore production or outsource manufacturing to third-party suppliers face higher risks of compliance failures. According to Steven et al. (2014), such strategic sourcing choices are associated with an increased likelihood of product safety violations.

Predisposition

Formal and informal norms have a considerable impact on employee actions, particularly their inclination to engage in misconduct. When unethical practices become embedded within organizational routines, they can persist and spread across different operational contexts. Pierce and Snyder (2008) find that automotive emissions inspectors working at multiple testing facilities adjust their pass rates based on local facility norms. This suggests that compliance behaviors are shaped by socialized expectations rather than universal regulatory standards. Similarly, Mohliver and Ody-Brasier (2023) find that personnel of religious nursing homes are less likely to report non-compliance, resulting in worsening violations before detection.

The availability and efficacy of internal reporting channels also influence the predisposition toward misconduct. Robust whistleblower protections and accessible reporting channels facilitate early intervention, reducing the likelihood of small violations escalating into larger issues. Tucker (2016) shows that in healthcare settings, employees are more likely to engage in policy-non-compliant workarounds when policy-compliant workarounds are difficult and they lack direct access to process owners. Similarly, Bakshi and Peura (2022) find that near-miss reporting systems reduce misconduct risks by encouraging early detection and repair of issues.

Environmental (External) Factors

Pressure or Need

Competitive intensity can increase the pressure or need to engage in misconduct. Firms operating in highly competitive markets may face short-term performance pressures that encourage rule-breaking in order to maintain profitability. Empirical studies demonstrate that geographically close competition correlates with increased instances of unethical behavior in brick-and-mortar outlets. For example, Skilton and Bernardes (2022) find that increased competition among pharmacies led to an oversupply of opioids, while Bennett et al. (2013) show that emissions testing facilities operating in competitive environments have higher rates of emission-test fraud. Similar trends can be observed across industries, with competition linked to increased product safety issues and recalls in the medical device industry (Ball et al. 2018b).

Environmental dynamism also puts strain on organizations, leading to increased pressure to engage in misconduct. Firms operating in highly dynamic markets face uncertainties that encourage risk-taking, including regulatory evasion and ethical violations, as strategies to maintain operational stability (Wiengarten et al. 2017a).

Regulatory tightening of standards is another environmental source of pressure that may drive firms toward non-compliance. Stricter legal requirements increase the financial and operational burdens on firms, which can lead to misconduct. Hu et al. (2021) demonstrate that the implementation of stricter emissions standards in the automotive industry led to more product non-compliance. Notably, the effect was stronger for firms that also faced intense competition, highlighting how various kinds of pressure can interact to increase misconduct risk.

Opportunity

The presence of trust-based control mechanisms within an industry can create environments

conducive to misconduct. When firms rely on relational governance rather than stringent enforcement mechanisms, the potential for wrongdoing increases. DuHadway and Narasimhan (2021) argue that in the automotive sector, some recalls result from firms exploiting quality assurance processes that depend on trust-based norms. In such cases, firms circumvent oversight mechanisms, falsify data, or intentionally subvert established controls.

Regulatory enforcement also has a significant impact on misconduct opportunities. The absence of effective monitoring and prevention strategies allows firms to engage in unethical behavior with minimal risk of detection or punishment (McKendall et al. 2002). Supporting this argument, Hu et al. (2021) find that in contexts where regulatory oversight is weak, tightening emissions standards increases automakers' non-compliance. Similarly, Bennett et al. (2013) demonstrate that emissions testing facilities are more likely to cheat when assessing older vehicles, which are subject to less rigorous emission test regulations.

Predisposition

Baucus (1994) identifies three environmental factors that affect the predisposition of firms to engage in misconduct: mature industry environment, prevalence of illegality in industry, and long/established relationships with regulatory agencies. However, research within the OSCM domain exploring these factors is sparse. A notable exception is work by Hu et al. (2021), who argue that the perceived legitimacy of regulations affects the likelihood of compliance.

2.3.3 Corrective Actions and Consequences

Hersel et al. (2019) provide a detailed review of the corrective actions and consequences following misconduct. Because the focus of this dissertation is on leveraging an understanding of the causes of misconduct to evaluate prevention strategies, I omit a detailed account of these post-misconduct processes and refer the interested reader to their review.

2.3.4 Prevention

Effective prevention strategies are essential for curbing the spread of misconduct (McKendall et al. 2002). While scholars of organizational misconduct have extensively examined its causes and consequences, there remains a notable gap in evidence-based mechanisms to prevent it (Anand et al. 2023, Palmer et al. 2016a, Greve et al. 2010). In contrast, the (analytical) OSCM literature provides a rich foundation for understanding and mitigating supply chain misconduct. This body of work offers valuable insights that can be leveraged to design and empirically test prevention strategies for organizational misconduct more broadly. As such, prevention represents a key avenue where OSCM scholars can contribute meaningfully to the organizational misconduct literature.

The existing OSCM literature identifies several mechanisms for curbing the spread of misconduct (see Table 2.2). I categorize these mechanisms along two dimensions. First, I assess whether they target internal or external factors causing misconduct. Second, I classify them based on their locus of control, distinguishing between organization-driven measures (i.e., governance reforms) and interventions led by external entities (i.e., regulatory interventions) (Finney and Lesieur 1982).

While the root causes of misconduct are complex, most regulatory interventions aim at reducing external factors contributing to misconduct. Often by reducing the opportunities for misconduct by increasing detection likelihood (e.g., inspections, machine learning predictions). Although this approach offers certain advantages, it also presents limitations that warrant further exploration.

Increasing the detection likelihood is effective for preventing misconduct driven by rational choice. By raising the probability of detection, these mechanisms make misconduct less appealing due to the higher expected costs (Palmer 2012, McKendall et al. 2002). For instance, Scott et al. (2021) show that electronic logging devices, which automatically record truck drivers' hours, reduce federal work rule violations. These devices increase the

Table 2.2: Selected Misconduct Prevention Strategies in OSCM Literature

Mechanism	Control	Strategy	Description
		Audits	Examination, evaluation, or observation of supplier to ensure that it complies with standards, requirements, or expectations. The threat of audits increases the likelihood of detection and allows organizations to detect supplier misconduct early. (Zhang et al. 2022, Dong et al. 2022, Dawande and Qi 2021, Caro et al. 2018, Plambeck and Taylor 2016)
	Governance reforms	Carrots and sticks	Providing incentives for desired behavior and establishing penalties for undesired behavior of suppliers (Porteous et al. 2015, Jiang 2009)
		Policy changes	Deferring payments after a funding goal is reached to reduce fraud on crowdfunding platforms (Belavina et al. 2020). More stringent test protocols for emission testing in new cars (Ben-nett et al. 2013).
Reducing external factors		Technology updates	Technologies such as electronic monitoring of truckers (Scott et al. 2021, Scott and Nyaga 2019) or reformulation of OxyCon-tin (Attari et al. 2024) can reduce misconduct opportunities.
	Regulatory interventions	Machine learning predictions	Predict the probability of violations via machine learning to increase the detection probability. (Chang et al. 2021, Campbell and Shang 2022, Mejia et al. 2019)

Continued on next page

Table 2.2: Selected Misconduct Prevention Strategies in OSCM Literature (Continued)

Mechanism	Control	Strategy	Description
		Inspections	Examine, evaluate, or observe an organization to ensure that it complies with standards, requirements, or expectations. The threat of inspections increases the likelihood of detection and allows control agents to detect misconduct early. (Ibanez and Toffel 2020, Liu et al. 2019, Mejia et al. 2019, Mani and Muthulingam 2019, Ball et al. 2017, Short et al. 2016)
		Near-miss reporting	Reduce the likelihood of accidental misconduct by implementing near-miss reporting and root-cause learning. (Bakshi and Peura 2022)
Reducing internal factors	Governance reforms	Executive appointments	Appointing a Chief Sustainability Officer focuses the organization on environmental issues, which reduces their predisposition to disadvantage communities of color in their operational decisions (Chapter 5).
	Regulatory interventions		

likelihood of uncovering violations, thereby raising the expected costs of misconduct and altering the cost-benefit analysis to favor compliance. Thus, if misconduct is the result of a rational choice between compliance and non-compliance, then increasing the detection likelihood is an effective prevention mechanism.

However, these mechanisms can be costly, making it necessary to accept some level of misconduct. For example, restaurant hygiene inspections aim to mitigate public health risks by penalizing severe violations (Ibanez and Toffel 2020). Continuous monitoring by unbiased inspectors would be ideal but is too costly and impractical. Consequently, regulators accept that some restaurants will violate hygiene regulations and focus on limiting the damage through periodic inspections (Mejia et al. 2019).

Another—often cheaper—prevention mechanism is to eliminate opportunities for misconduct through changes in routines, processes, structures, or rules. This involves identifying parts of an organization’s system that enable misconduct and altering them. For example, Belavina et al. (2020) identified that the all-or-nothing payment mechanism on most crowdfunding platforms allows for ‘funds misappropriation’ and ‘performance opacity’. They proposed a deferred payment strategy to prevent these issues, demonstrating that eliminating misconduct opportunities can be effective without the high costs of increasing detection likelihood.

Despite their benefits, detection likelihood-increasing and opportunity-elimination mechanisms are susceptible to offense displacement or hiding efforts. Offense displacement occurs when actors shift their misconduct rather than stop it (Guerette and Bowers 2009). For instance, after implementing electronic logging devices, truck drivers found other ways to violate safe driving regulations, such as speeding or ignoring traffic signals, to deliver more goods and increase earnings (Scott et al. 2021). This displacement arises because such mechanisms do not address the underlying predisposition to misconduct or the pressures that drive it. Hiding efforts occur when actors increase their efforts to mask wrongful behavior in response to detection attempts. For instance, Plambeck and Taylor (2016) show

that under certain conditions, suppliers respond to increased auditing efforts by exerting greater effort to hide information from the auditors instead of trying to reduce the occurrence of violations.

Less attention has been paid to governance reforms addressing internal factors contributing to misconduct. My second essay (Chapter 4) explores this mechanism in the context of environmental injustice. I argue that appointing a Chief Sustainability Officer raises awareness of environmental issues within the organization, reducing the likelihood of unintentionally disadvantaging communities of color, which decreases environmental injustice.

2.4 Research Opportunities

Building on the insights from the previous section, this section outlines how my dissertation essays fit within the domain of organizational misconduct.²

2.4.1 Leveraging Corrective Actions for Personal Gain

In their exploration of misconduct corrective actions and consequences, Hersel et al. (2019) conceptualize product recalls as a key corrective action in response to product safety issues. They argue that decision-makers weigh both stakeholder harm and firm-level consequences when determining how and when to respond to misconduct. This aligns with findings from the product recall literature: recall decision-makers assess patient harm alongside recall-related costs (Ball et al. 2018a), and individual characteristics, such as gender or tenure, influence recall timing and likelihood (Mayo et al. 2022, Wowak et al. 2021).

However, recall decision-makers may also face a moral hazard when initiating corrective actions—their personal incentives may conflict with firm and stakeholder interests. For example, Darby et al. (2023) show that CEO stock ownership influences recall decisions,

²See Chapter 6 for additional research opportunities.

with firms being slower to recall products when their CEOs hold high equity stakes. At the same time, research in accounting suggests that executives are predisposed to engage in misconduct. Specifically, they sometimes exploit insider knowledge about corporate processes for personal financial gain (Arif et al. 2022, Blackburne et al. 2021, Huddart et al. 2007). This raises the possibility that top management team members involved in recall decision-making could exploit the opportunity provided by their involvement by engaging in informed trading to profit from non-public recall information. This raises the following high-level research question: *Do top management teams leverage corrective action timing for personal financial gain?*

Preventing such misconduct is particularly challenging because it requires coordination between different regulatory bodies. In the U.S., the FDA oversees product safety and recall processes, while the SEC is responsible for monitoring and prosecuting illicit financial activities, including insider trading. However, these agencies operate independently and do not systematically share data, making it difficult to detect and prevent misconduct that spans both domains. In my first essay (Chapter 3), I examine whether CEOs engage in informed trading during the medical device product recall process. After establishing the likely presence of such behavior, I propose potential policy interventions, including improved cross-agency collaboration between the FDA and SEC, to mitigate this form of misconduct.

2.4.2 Optimizing Resource Allocation for Misconduct Prevention

Beyond challenges in awareness and data sharing, control agents still face a critical challenge even when misconduct is well-documented and data is abundant: how should they allocate their resources most effectively to prevent misconduct? OSCM research has used analytical models to study misconduct prevention in areas such as supplier sustainability violations (Huang et al. 2022, Caro et al. 2018) and fund misappropriation on crowdfunding platforms (Belavina et al. 2020). These studies often assume that the misconduct control agents seek to prevent is intentional. However, as highlighted in Section 2.3, misconduct frequently

emerges not from deliberate wrongdoing but from a gradual erosion of routine adherence.

A promising research direction involves modeling prevention strategies based on the “decay in routine adherence” explanation for misconduct (Mohliver 2019, Anand et al. 2012, Vaughan 1999). This perspective suggests that organizations may benefit from designing interventions that reinforce compliance habits rather than simply penalizing violations. Another crucial research question relates to how control agents can best utilize their data to detect misconduct early and improve their inspection programs (Johnson et al. 2023, Campbell and Shang 2022, Mejia et al. 2019). Integrating these two perspectives leads to the following high-level research question: *How can control agents optimize the allocation of inspection resources to prevent misconduct effectively?*

I explore this question in my third essay (Chapter 5), in which I examine how health departments can leverage machine learning predictions about restaurant hygiene violations to optimize their inspection strategies.

2.4.3 Governance Reforms as a Misconduct Prevention Mechanism

The first two research questions focus on regulatory interventions, such as SEC enforcement actions and health department inspections. However, Table 2.2 highlights another key approach: governance reforms, in which firms implement internal policies to prevent misconduct. While much of OSCM research has focused on external regulatory interventions, governance reforms remain an underexplored yet promising avenue—especially in cases where regulatory interventions have proven ineffective or too costly. This prompts the question: *Can firm governance reforms be an effective mechanism to prevent misconduct in situations where regulatory interventions are costly/have failed?*

For example, regulators have attempted to address environmental injustice—the disproportionate impact of pollution on marginalized communities—since the 1990s, but with limited success (Geltman et al. 2016, Mohai et al. 2009, O’Neil 2007). Given these setbacks, corporate governance reforms—such as appointing a Chief Sustainability

Officer (CSO)—may offer a viable alternative for improving environmental outcomes. This raises an important empirical question: “Do Chief Sustainability Officers (CSOs) reduce environmental injustice in toxic releases?” I explore this question in my second essay (Chapter 4), where I assess whether firms that adopt CSO positions demonstrate measurable improvements in their environmental performance and contribute to environmental justice.

Chapter 3

Product Recall Decisions and the CEO: Informed Trading in the Medical Device Industry

3.1 Introduction

Medical device manufacturers frequently recall faulty medical devices to protect consumers (FDA 2020). In the United States, the Food and Drug Administration (FDA) provides recall decision-making guidance for firms but leaves the actual recall decision to firm leadership (Ball et al. 2018a). Consequently, it is not uncommon that even when the defects underlying the recall are severe, recalls are initiated after periods of significant deliberation within the firm, at times resulting in consumer harm or even death during this decision-making time window (Darby et al. 2023, Sengupta et al. 2020). Exploring this important and pivotal recall decision-making process within the firm is thus an important but understudied area in the rapidly growing product recall literature (Ball et al. 2022).

Early research in this domain has taken the first steps to investigate one particular

firm leader and their relationship with the product recall decision: the CEO. For instance, Mayo et al. (2022) show that CEOs with longer tenures are more inclined to avoid recalling dangerous consumer products in an effort to avoid the reputational damage that a recall can have to the firm in general and to a long-tenured CEO in particular. Similarly, Darby et al. (2023) find that firms are significantly slower to recall the most serious defects when their CEOs have higher stock ownership in the firm. The authors suggest that one possible explanation for this finding is that CEOs with substantial firm stock holdings have more to lose from the stock market penalty that may follow a product recall. This potential loss appears to motivate firms to extend the time window between when the firm becomes aware of the defect (i.e., the defect awareness date) and when the recall is initiated (i.e., the recall initiation date) (see Section 3.2.2 for a detailed description). Both of these studies pinpoint the unique role a CEO plays in overseeing product recall decisions at the firm, and they imply a possible, though untested, moral hazard. That is, there is a possibility that CEOs behave opportunistically during this window of time to avoid personal financial losses, especially in cases of severe recalls. We refer to the time window between defect awareness date and recall initiation date as the “Window of Opportunity” and focus on one specific and potentially contentious form of opportunistic behavior that CEOs may engage in: informed trading.

Informed trading in this context occurs when a CEO sells their stock during the Window of Opportunity, that is, after the firm becomes aware of a severe defect that may require a Class I recall, but before the recall is initiated, presumably to avoid potential losses on their stock holdings.¹ Selling during the Window of Opportunity would be considered informed trading because the defect awareness date is not publicly reported, nor is it accessible to

¹The SEC refers to informed trading by corporate insiders such as CEOs as insider trading, which they define as trading by insiders based on material, non-public information (Securities and Exchange Acts of 1933 and 1934; Insider Trading Sanctions Act of 1984; Insider Trading and Securities Fraud Enforcement Act of 1988). Our use of informed trading follows the convention in the accounting literature to label trading patterns that indicate the presence of trades that are potentially based on material, non-public information but without providing legal evidence.

anyone outside the recalling firm prior to the actual recall initiation date, which normally occurs many weeks or even months after the defect awareness date.²

We focus on the most severe, Class I medical device recalls in our study because although rare, their occurrence is significant and linked to serious harm or even death for patient consumers (FDA 2020). They also elicit considerable negative attention from the public, media, investors, and the FDA (NYT 2022, Fuhr et al. 2013, Thirumalai and Sinha 2011). Given a CEO's fiduciary responsibilities related to major firm decisions, they are almost certain to be aware of, and in many cases involved in, a Class I recall decision. Published research and anecdotal reporting indicate this is the case (Darby et al. 2023, Philips 2022, Massdevice 2011). Findings in the accounting literature on informed trading (e.g., Arif et al. 2022, Blackburne et al. 2021) suggest that such events with severe negative consequences are the most likely context for informed trading. Therefore, we investigate informed trading in the product recall context to address two open questions. First, we explore whether and to what degree CEOs sell their company stock in the Window of Opportunity, that is, after defect awareness and before Class I recall initiation. Second, we examine whether CEOs avoid personal financial losses from such sales.

In examining the presence of informed trading by CEOs for Class I recalls, we segregate these most severe recalls into those with, and without, prior medical device product quality complaints that are made public on the FDA's complaint website.³ We use this approach because publicly available medical device complaints represent information leakage that might reduce a CEO's information advantage. The FDA publishes all medical device complaints through its Manufacturer and User Facility Device Experience (MAUDE) database in an effort to keep the public abreast of, and increase the transparency of, serious medical device product quality problems. Interestingly though, not all recalls are preceded

²As we explain in Section 2.2, the recalling firm must provide the defect awareness date when it initiates the recall, and this date is validated by the FDA. We gain access to the date via Freedom of Information Act (FOIA) requests for recalls that have been initiated.

³<https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfmaude/search.cfm>

by one or more complaints (FDA 2023a). A non-trivial proportion of recalls originate from internal quality signals such as production line issues or quality control trending analysis. These recalls are associated with quality defects known only within the firm, providing the firm's CEO with an information advantage. In contrast, investors, the FDA, and the general public might have some ability to foresee recalls that are preceded by MAUDE complaints, especially when those complaints involve a potential Class I recall in which serious harm or death is at play. In other words, although the FDA and the public may not know the exact timing of a forthcoming recall, they may become aware of product quality issues via MAUDE complaints reported before a recall, a view that is supported by medical device recall research (Mukherjee and Sinha 2018). This reduces the information advantage of CEOs, possibly reducing their propensity for informed trading for recalls with prior MAUDE complaints.

To examine whether CEOs engage in informed trading during the Window of Opportunity of Class I recalls, we combine data about medical device recalls, MAUDE complaints, accounting data, stock price information, and executives' stock trades. We study publicly traded medical device firms from 2003—after the Sarbanes Oxley Act changed the reporting standards of officer and director stock purchases and sales—to 2019—after which the COVID-19 pandemic started and disrupted stock markets. Our unit of analysis is the firm-trading day. We follow prior work in the accounting literature (e.g., Arif et al. 2022, Blackburne et al. 2021) to guide our empirical analysis, and compare the selling activity of CEOs on trading days during the Window of Opportunity with their selling activity on other days around the Window of Opportunity for Class I recalls. Overall, we analyze the selling activity related to 236 Class I recalls. We further categorize these recalls into 174 Class I recalls with prior MAUDE complaints and 62 without. Our empirical analysis shows that CEOs are twice as likely to sell their stock during the Window of Opportunity for Class I recalls without prior MAUDE complaints than outside it. We do not observe this pattern for Class I recalls with prior MAUDE complaints, implying that the

presence of informed trading is nuanced and appears to occur only for these most serious recalls that are most opaque to the public eye.

Next, we examine whether these sales are personally profitable for the CEOs using sale-specific abnormal returns (Arif et al. 2022, Blackburne et al. 2021). We find that CEO sales during the Window of Opportunity for Class I recalls without prior MAUDE complaints avoid losses by front-running subsequent stock price downturns. Specifically, stocks of firms whose CEOs sell during the Window of Opportunity for this subset of recalls have negative abnormal returns of about 15% over the six months following the sale. These negative abnormal returns correspond to significant profits of about \$140,000. Thus, selling during the Window of Opportunity allows CEOs to avoid substantial losses. Together, our findings suggest that CEOs may knowingly or unknowingly exploit their information advantage regarding the most serious and secret recalls by timing their stock sales and profit from the sales.

These results have significant implications for theory and practice. Towards theory, we contribute to the product recall decision-making literature as it pertains to the role of the CEO (Darby et al. 2023, Mayo et al. 2022, Wowak et al. 2015), by demonstrating that they profit from their firm-proprietary knowledge about an upcoming recall. We also contribute to the accounting literature on informed trading by executives (Arif et al. 2022, Blackburne et al. 2021, Jagolinzer et al. 2011) and the emerging operations management literature on organizational misconduct (Kistler et al. 2024, Petersen et al. 2024, Skilton and Bernardes 2022, Belavina et al. 2020). For practice and public policy, our findings reveal an opportunity to recognize and prevent informed trading through closer cooperation between the FDA and the US Security and Exchange Commission (SEC). Additionally, we identify the need for firms to implement effective corporate governance mechanisms that avert informed trading by their CEOs.

3.2 Related literature and medical device recall background

3.2.1 Related literature

Our study is related to three separate literature streams: research linking CEO characteristics to product recalls, accounting research linking information disclosure and compliance processes to informed trading, and operations management research on organizational misconduct.

Within the product recall literature that links CEO characteristics to product recalls, Mayo et al. (2022) establish that firms have a high probability of initiating a recall following the appointment of a new CEO. However, as CEO tenure lengthens, the likelihood of recalls decreases. The authors explain that this may be because long-tenured CEOs seek to prevent signaling any incompetence to the market by avoiding recalls. Other related research focuses on how stock option compensation for CEOs influences recall characteristics, indicating both a positive and negative influence. On the positive side, Liu et al. (2016) find that firms with high CEO stock option pay or short CEO tenure provide more comprehensive recall remedies to consumers, perhaps to restore consumer trust and regain firm legitimacy. On the negative side, Wowak et al. (2015) find that firms in which the CEO receives stock options as a monetary incentive initiate more recalls, likely because under their leadership, those firms undertake greater risks during product design, production, and distribution. Likewise, Darby et al. (2023) examine how the amount of CEO stock ownership impacts the speed with which a firm decides to recall defective products. Similar to our paper, those authors use the defect awareness and the recall initiation dates to bookend the recall decision-making window. The authors theorize that CEOs may be hesitant to recall a product when they have large ownership stakes in the firm, ostensibly because they expect a stock price dip once the recall is announced. Their analysis shows that firms with CEOs who have higher stock ownership tend to delay medical device recalls, and that recalls lead to negative stock market reactions. This relationship is particularly true for more severe recalls

which cause more negative stock market reactions. We extend these results in two ways. First, Darby et al. (2023) use stock ownership of CEOs to explain variation in the length of the Window of Opportunity, whereas we investigate stock trading behavior by CEOs during this window. Second, we identify the contingent nature of the stock trading behavior by showing that it is dependent upon whether or not the information about upcoming recalls is available to the general public in the form of prior MAUDE complaints.

The notion of informed trading is novel in the product recall context. However, accounting researchers have extensively examined such opportunistic behavior around key accounting-related events such as information disclosure and outcomes of compliance activities. This stream of research suggests that insiders opportunistically adjust their trading behavior after gaining non-public, material information. On the one hand, insiders sell more stock in anticipation of events that negatively affect their firm's stock price. For instance, insiders sell more heavily when they learn about SEC investigations into their firm (Blackburne et al. 2021), modified audit opinions (Arif et al. 2022), or 10-K filings that are yet to be disclosed (Huddart et al. 2007). Prior work also suggests that insiders buy more stock in anticipation of events that positively affect their firm's stock price (Kim and Oh 2023, Jagolinzer et al. 2020). This body of work suggests that insiders may use material, non-public information for personal gain. We extend this stream of work by highlighting how CEOs may exploit private information gleaned during the product recall decision-making process—an operational process designed to protect consumers—for opportunistic gain.

In recent years, operations management scholars have become increasingly interested in understanding the drivers, consequences, and prevention of organizational misconduct (Kistler et al. 2024, Petersen et al. 2024, Skilton and Bernardes 2022, Belavina et al. 2020). They often build on Greve and co-authors (Greve et al. 2010, p. 56) who define organizational misconduct as any behavior that transgresses “a line separating right from wrong; where such a line can separate legal, ethical, and socially responsible behavior from their antitheses.” Such misconduct can result from rational decision-making, where actors

consider the cost and benefits of their actions, or amoral decision-making, where actors may not realize their behavior is unethical or illegal (Palmer 2012, Tenbrunsel and Smith-Crowe 2008). We do not stipulate either explanation but merely explore whether CEOs, knowingly or unknowingly, exploit non-public firm information on the most serious product quality problems a firm faces. While the drivers of such behavior are well-established, research into effective prevention strategies is sparse (Palmer 2012). We address this problem by using our findings to suggest how regulators can identify informed trading during the window of time in which firm leaders deliberate on a product recall decision that often has life-or-death consequences for consumers.

3.2.2 Medical device recall process

In recent years, a handful of researchers have examined the product recall decision-making process, and described the critical aspects of the process (Darby et al. 2023, Wowak et al. 2021, Ball et al. 2018a). These studies primarily explore how firm characteristics or leaders' role may explain the likelihood or timing of the recall decision. Drawing from these studies and insights from our discussions with industry experts and FDA officials, we briefly outline the recall decision-making process in the medical device industry below.

A product recall at a firm begins with the emergence of a serious product quality defect, typically discovered from one of two sources: a MAUDE complaint stemming from a product failure in the market, or an internal manufacturing or quality defect (Ball et al. 2018a). Reliability engineers monitor failures in the market, while manufacturing, quality, or supplier engineers monitor internal defects. Once a defect arises, a product recall decision-making team, typically comprised of senior leadership from areas such as manufacturing, design, supplier development, legal, or regulatory, convenes to evaluate the problem (Wowak et al. 2021). This team considers the available data and conducts an analysis that examines the risk of a product failure in the market, the harm that a potential failure may cause to the consumer, and the likelihood that the defect is explained by a

systemic root cause attributable to the firm.⁴ The team may also indicate the need for further root cause analysis in an effort to make the most informed recall decision possible. If the team recommends a recall, it escalates the decision to an individual with recall decision-making authority at the firm, often a vice president of quality, a chief operations officer, or the CEO.

Regardless of who makes the final recall decision, prior research and industry sources emphasize the significant involvement of CEOs in recall decisions, especially when the recall under consideration is a Class I recall. When CEOs make the recall decision, they are directly involved in and responsible for the recall. For instance, in Medtronic’s highly publicized Sprint Fidelis lead recall, CEO Bill Hawkins “made the call that we [Medtronic] were going to voluntarily suspend the shipment [of Sprint Fidelis leads]” after carefully analyzing the data and weighing the pros and cons of such a decision (Massdevice 2011). Even when CEOs do not make the final recall decision, the firm leaders who make the recall decision regularly meet with their CEO to discuss ongoing recalls and recall-worthy defects (Darby et al. 2023). As one vice president of quality that we spoke to stated, “the CEO has to be part of that decision-making process because it directly affects how the company is running.” Furthermore, CEOs have a fiduciary duty to be involved in, or at least oversee, these decisions—in particular for the most severe recalls—because they are likely to affect a firm’s bottom line (Darby et al. 2023).

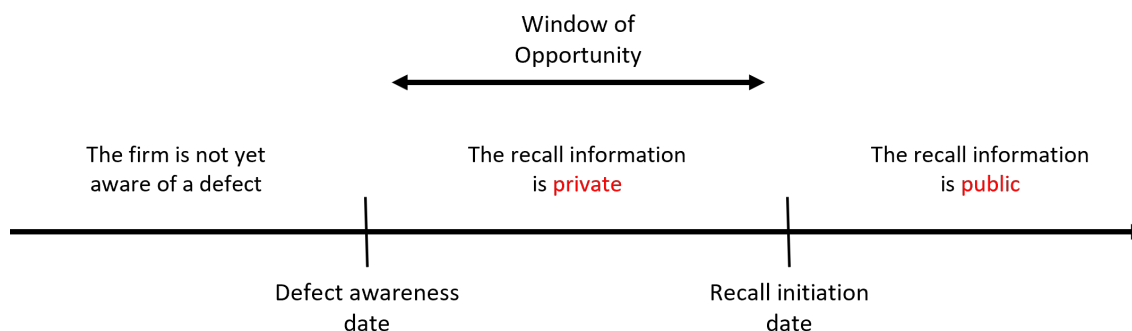
Once a firm decides to recall a product, it must immediately notify the FDA and affected customers (Wowak et al. 2021). The day on which this occurs is the recall initiation date. When notifying the FDA, the firm must also indicate the defect awareness date—the date on which they first became aware of the underlying defect (Wowak et al. 2021, FDA 2021).⁵ The FDA verifies these dates when they conduct plant inspections (Darby et al. 2023, FDA

⁴Sometimes product failures can be caused by customer misuse or they can be random and have no systemic root cause. In those cases, a recall is not warranted.

⁵See FDA Form 5072 (11/23) page 4 with the required “Firm Awareness Date” to be filled in when a firm recalls a defective product: <https://www.fda.gov/media/174165/download>.

2021). We label the time between the defect awareness and the recall initiation dates as the Window of Opportunity (see Figure 3.1).

Figure 3.1: Recall Timeline and the Window of Opportunity



Recalls can be categorized based on their severity and the presence of prior complaints. Severity is measured with the FDA-assigned recall classes. Class I recalls are the most severe, and have a reasonable probability of consumer harm or death. Class II recalls, though less severe than Class I, can still result in temporary or medically reversible adverse health effects. Class III recalls are not related to adverse health effects and involve issues such as labeling problems (FDA 2020). The categorization by severity is crucial in understanding informed trading because accounting researchers have shown that informed trading is more likely to occur for severe problems that have the most negative consequences for firms (Arif et al. 2022, Blackburne et al. 2021). In our context, the most severe Class I recalls are likely to garner the most negative attention from external stakeholders, making informed trading by CEOs most likely.

To investigate how firm-proprietary information influences stock trades by CEOs, we use the presence or absence of prior MAUDE complaints to classify recalls further. The FDA's MAUDE database makes all medical device complaint reports (e.g., death, injury, or malfunction event) publicly available. While a small portion of these reports are posted

by patients and physicians directly, the vast majority are posted by manufacturing firms based on information the firm receives from patients and physicians, in nearly real-time (Mukherjee and Sinha 2018). If a firm fails to report an adverse event that they receive from a customer, they are heavily penalized by the FDA including receiving FDA warning letters. Thus, firms are highly sensitized to accurate and timely reporting of MAUDE data (FDA 2023b). When a recall is triggered by MAUDE complaints alone or by a combination of internal signals and MAUDE complaints, the public already possesses some information about the underlying issue before the firm announces the recall. For instance, in Medtronic's Sprint Fidelis recall, the initial awareness of the defect came from a hospital report on an unusually high failure rate (Massdevice 2011), a complaint that is available in the FDA's MAUDE database. Conversely, if a recall originates solely from internal information without prior MAUDE complaints, it remains confidential within the firm. Consider, for example, Cardiovascular System's recall of their Diamondback 360 Device, a surgical instrument used to reestablish blood flow in narrowed arteries. In 2014, the firm recalled multiple devices even though there were no reported patient injuries (FDA 2014). The recall was an outcome of quality issues identified during process control in a component that could potentially fracture during use, although it had not yet happened.

While the source of information does not impact a CEO's involvement, it does affect the extent to which information about recall-worthy defects remains non-public, providing an information advantage to those who know about it. This, in turn, potentially influences the temptation to sell when firm leaders discuss a serious product quality problem. If complaints precede a recall, the stock market may anticipate it, reducing the temptation for informed trading. For instance, Kim and Wagner (2021) find that market penalties for corruption primarily stem from initial allegation reports, akin to MAUDE complaints. Conversely, if no complaints precede a recall, the secret nature of the serious quality problem may create a greater temptation to sell stock. Therefore, we analyze recalls with and without prior MAUDE complaints separately to understand their impact on informed trading by CEOs.

In summary, firms initiate product recalls after being made aware of a potentially recall-worthy defect, creating a Window of Opportunity for CEOs to engage in informed trading. Our research aims to assess the presence of informed trading in this context by investigating stock sales by CEOs during the Window of Opportunity of Class I recalls conditional on the presence of MAUDE complaints. We also examine whether such sales are profitable.

3.3 Empirical setting

3.3.1 Data

We collect and combine several data sources to investigate the relationship between medical device product recalls and informed trading. First, we compile information on product recalls via multiple Freedom of Information Act (FOIA) requests to the FDA. This data includes details such as the responsible firm’s contact information⁶, defect awareness date, recall initiation date, and severity classification. We focus on recalls with defect awareness and initiation dates between January 01, 2003, and December 31, 2019. We start from 2003, after Congress passed the Sarbanes Oxley Act, which required officers and directors to electronically file purchases and sales of their firm’s stock within two business days. Our data collection ends in 2019, after which the COVID-19 pandemic caused severe shocks for medical device firms, their recalls, and their stock price.

Next, we establish whether complaints precede recalls by linking our recall data to the FDA’s MAUDE database via the open FDA API. In the MAUDE database, the FDA publicizes adverse event reports categorizing a device using its approval number. An approval number can take on two types. Highly innovative medical devices that require clinical trials (referred to as pre-market approvals or PMAs) have a PMA approval number,

⁶We use Refinitiv’s mergers and acquisitions data to identify changes in ownership to ensure that the recall is assigned to the correct firm.

while new versions of existing devices (referred to as 510(k) approvals) have a 510(k) approval number. We use FDA-assigned approval numbers to connect recalls to MAUDE data. Our data includes 72 Class I recalls for which MAUDE reports no events before the firm initiates the recall. We consider these in the “Recalls without prior MAUDE complaints” analyses. There are 164 Class I recalls for which MAUDE reports at least one complaint (e.g., death, injury, or malfunction event) before the firm initiates the recall. We consider these in the “Recalls with prior MAUDE complaints” analyses.

We collect firm financial data from The Center for Research in Security Prices (CRSP) and Compustat for all U.S. common stocks in our sample. We also incorporate aggregate stock market data used to compute a firm’s abnormal return from the Kenneth R. French security data library.⁷

Lastly, we collect information on stock transactions by CEOs from the Thomson Reuters Insider Filings Form 4 database. We follow Jagolinzer et al. (2020) and consider only open-market purchases and sales of common stock for our analysis. Thus, we exclude option exercises, option grants, and equity gifts because CEOs have less discretion in the timing of these transactions. Combining these datasets yields a sample of 236 unique Class I recalls and 521,303 firm-trading days between 2003 and 2019. Consistent with informed trading research in the accounting literature, our unit of analysis is the firm-trading day (e.g., Huddart et al. 2007).

3.3.2 Variables

Dependent variables

CEO Selling Activity. To capture the selling activity of CEOs, we create the binary variable *CEO Selling Activity*, which equals one if the CEO at the recalling firm sells more than they buy on a given day and zero otherwise. For example, if the CEO sells stock valued

⁷http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

at \$50,000 and buys \$20,000 then we code *CEO Selling Activity* as one. If, on the other hand, they sell \$10,000 but buy \$20,000 then we code *CEO Selling Activity* as zero. This measure of selling activity is widely adopted in the accounting literature (Arif et al. 2022, Blackburne et al. 2021). Aggregating this measure across multiple days, as we do in our analyses, allows us to interpret the mean value of *CEO Selling Activity* and any coefficients that describe changes to it as the likelihood of a stock sale by a CEO on any given day (Arif et al. 2022). For instance, the mean value for *CEO Selling Activity* in our sample is 0.010, meaning that on any given day there is a 1.0% likelihood of a CEO selling their stock.

Independent variable

Window of Opportunity. To examine whether CEOs make informed sales during the Window of Opportunity, we need to identify their sales in this window. We create *Window of Opportunity*, a binary variable, that equals one if a trading day is between the defect awareness date and recall initiation date for any Class I recall of the focal firm, and zero otherwise. This approach resembles the one taken by Huddart et al. (2007) in their study of the relationship between 10-K / 10-Q reports and informed trading. To illustrate the variable construction, consider a hypothetical firm that becomes aware of a defect on July 15, 2010 and initiates the corresponding Class I recall on September 3, 2010. For this firm, *Window of Opportunity* takes the value zero for days before July 15, the value one for all days between and including July 15 and September 3, and zero for September 4 and onward (see also Figure 3.1 for a graphical representation of this coding). For our main analysis, we construct three variations of this variable: Window of Opportunity for Class I recalls without prior MAUDE complaints, Window of Opportunity for Class I recalls with prior MAUDE complaints, and Window of Opportunity for all Class I recalls combining both categories.

Other variables

We follow the accounting literature (e.g., Arif et al. 2022, Blackburne et al. 2021, Jagolinzer et al. 2011) and include seven firm-day-level control variables associated with the propensity of executives to trade. We capture the firm’s blackout period (*Blackout Period*), the firm’s market value and growth opportunities (*Firm Size*, *Book-to-Market Ratio*), as well as measures based on the firm’s stock price (*Abnormal Returns*, *Daily Abnormal Returns*, *Volatility*, *Daily Volatility*). Details on these are as follows:

Firm Size. We control for firm size because it may affect the number of recalls and a CEO’s trading behavior. We measure firm size as the natural logarithm of market capitalization, where market capitalization is the stock price multiplied by outstanding shares (Arif et al. 2022, Blackburne et al. 2021).

Book-to-Market Ratio (BM). We control for a firm’s book-to-market ratio, measured as the book value of equity scaled by market capitalization, where market capitalization is the stock price multiplied by outstanding shares (Arif et al. 2022, Blackburne et al. 2021).

Blackout Period. Most firms institute self-imposed blackout periods around earning announcement dates in which they prohibit officers from buying or selling stock. If a Window of Opportunity overlaps with such a blackout period, then CEOs would be prohibited from selling their stock in the period of overlap. We follow Blackburne et al. (2021) and set potential blackout periods as the time from 46 days before to one day after a firm’s quarterly earnings announcements. Specifically, we create *Blackout Period* as an indicator variable that takes the value of one if a firm-day falls within $[-46, +1]$ days of a firm’s earnings announcement, and zero otherwise. We obtain information on the earning announcement dates from Compustat.

Abnormal Returns. A firm’s stock price performance relative to the aggregate stock market might affect a CEO’s selling behavior. Therefore, we include abnormal returns, measured as the firm’s market-adjusted abnormal return over the fiscal year, as a control

(Arif et al. 2022, Blackburne et al. 2021).

Daily Abnormal Returns. Similar to the yearly abnormal returns, short-term swings in a firm’s stock price relative to the stock market might affect a CEO’s selling behavior. We control for this using daily abnormal returns, which we measure as a firm’s market-adjusted abnormal return over the three days preceding and including each trading day (i.e., over the [-3,0] days around a trading day). This window follows Blackburne et al. (2021) and is designed to capture the effect of contemporaneous public news that might affect *CEO Selling Activity* (Arif et al. 2022).

Volatility. The volatility of a firm’s stock returns might influence a CEO’s selling behavior. We account for this by controlling for the standard deviation of monthly stock returns over the fiscal year (Arif et al. 2022, Blackburne et al. 2021).

Daily Volatility. Similar to the yearly volatility, we control for short-term volatility in a firm’s stock returns using the standard deviation of its daily stock returns (Blackburne et al. 2021). In line with our calculation for *Daily Abnormal Returns*, we consider the three days preceding and including the trading day for this calculation.

3.3.3 Empirical strategy

Our main research question is whether CEOs engage in informed trading in the window of time between defect awareness and recall initiation. Our independent variable *Window of Opportunity* indicates whether a given day lies between the defect awareness and recall initiation dates. To empirically test whether CEOs are more likely to sell their stock during this window—a sign of potential informed trading—we follow the informed trading literature (e.g., Arif et al. 2022, Blackburne et al. 2021) and compare their selling activity within the Window of Opportunity ($Window\ of\ Opportunity_{it} = 1$) to their selling activity outside

this window. Specifically, we estimate the following equation:

$$\begin{aligned}
CEO\ Selling\ Activity_{it} = & \delta * Window\ of\ Opportunity_{it} \\
& + \beta_1 * Firm\ Size_{it} + \beta_2 * Book\ to\ Market\ Ratio_{it} \\
& + \beta_3 * Blackout\ Period_{it} + \beta_4 * Abnormal\ Returns_{it} \\
& + \beta_5 * Volatility_{it} + \beta_6 * Daily\ Abnormal\ Returns_{it} \\
& + \beta_7 * Daily\ Volatility_{it} \\
& + \gamma_i + \lambda_t + \epsilon_{it},
\end{aligned} \tag{3.1}$$

where $CEO\ Selling\ Activity_{it}$ measures the selling activity of the CEO for firm i on trading day t , $Window\ of\ Opportunity_{it}$ is our indicator variable for days between, and inclusive of, the defect awareness and recall initiation dates, γ_i and λ_t are firm and trading day fixed-effects and ϵ_{it} is the error term. Our coefficient of interest is δ , which describes the change in the likelihood of a stock sale by a CEO during the Window of Opportunity. If CEOs, on average, increase their selling activity based on information about recalls, then the coefficient for Window of Opportunity will be positive and statistically significant. This would indicate the presence of informed trading. If, however, the coefficient for Window of Opportunity is negative or not significant, then it would indicate the absence of statistically identifiable informed trading patterns.

One challenge in estimating the coefficient of *Window of Opportunity* is that underlying factors could jointly affect recalls and insider selling activity, which would lead to biased results. We address such endogeneity concerns in three ways. First, we rely on the quasi-exogenous nature of defect awareness dates. Firms constantly monitor quality signals and receive information from various third parties including doctors and patients. Thus, firms have limited control over how and when quality issues are reported. Furthermore, the FDA validates the defect awareness dates which mitigates misreporting (Darby et al. 2023, FDA 2021). Second, we include firm-day fixed effects in all our regressions, further

bolstering our identification. The firm fixed effects account for time-invariant unobservable heterogeneity between firms. The day-fixed effects account for any market-wide daily unobservable heterogeneity. Third, we follow Arif et al. (2022) and Blackburne et al. (2021) by controlling for a wide range of observable variables that can influence insider selling activity. This allows us to identify changes to a CEO’s selling activity during the Window of Opportunity—*ceteris paribus*—while holding all other observable and firm-day fixed confounders constant.

3.4 Results

Table 3.1 presents descriptive statistics for all the substantive variables. The mean *CEO Selling Activity* is 0.010 across the full sample, meaning that their average likelihood to sell on a given day is 1%.

Table 3.1: Descriptive Statistics

	Mean	SD	Min	Max	Median	N
CEO Selling Activity	0.010	0.101	0.000	1.000	0.000	523 310
Firm Size	7.221	2.349	0.152	12.958	6.994	523 310
Book-to-Market Ratio	0.402	0.464	−9.037	9.667	0.319	523 310
Blackout Period	0.525	0.499	0.000	1.000	1.000	523 310
Abnormal Returns	0.069	0.426	−2.567	3.225	0.040	523 310
Volatility	0.026	0.017	0.006	0.239	0.022	523 310
Daily Abnormal Returns	0.001	0.056	−1.052	3.665	0.000	523 310
Daily Volatility	0.020	0.023	0.000	1.942	0.013	523 310

3.4.1 Presence of Informed Trading

We examine informed trading behavior by CEOs by estimating Equation 3.1. Table 3.2 presents our results for Class I recalls without prior MAUDE complaints, Class I recalls with prior MAUDE complaints, and all Class I recalls.

Table 3.2: CEO Selling Activity in Window of Opportunity for Class I Recalls

	Recalls without prior MAUDE complaints		Recalls with prior MAUDE complaints		All Class I Recalls	
	(1)	(2)	(3)	(4)	(5)	(6)
Window of Opportunity	0.0091** (0.0033)	0.0085* (0.0033)	-0.0043*** (0.0013)	-0.0038** (0.0013)	-0.0003 (0.0014)	-0.0002 (0.0014)
Size		0.0056*** (0.0003)		0.0056*** (0.0003)		0.0056*** (0.0003)
Book-to-Market Ratio		-0.0029*** (0.0004)		-0.0029*** (0.0004)		-0.0030*** (0.0004)
Blackout Period		-0.0085*** (0.0005)		-0.0085*** (0.0005)		-0.0085*** (0.0005)
Abnormal Returns		0.0030*** (0.0004)		0.0030*** (0.0004)		0.0030*** (0.0004)
Volatility		-0.0024 (0.0129)		-0.0013 (0.0129)		-0.0016 (0.0129)
Daily Abnormal Returns		0.0122*** (0.0026)		0.0122*** (0.0026)		0.0122*** (0.0026)
Daily Volatility		0.0200** (0.0073)		0.0201** (0.0073)		0.0201** (0.0073)
Firm fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Day fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Num.Obs.	523,310	523,310	523,310	523,310	523,310	523,310
R ²	0.024	0.026	0.024	0.026	0.024	0.026

Notes: The table presents the results of a firm-day fixed effects regression with CEO selling activity as the dependent variable. The unit of analysis is firm-day. The sample is the same across all models but the computation of Window of Opportunity changes in accordance with the column header to capture the effect of different recalls. Standard errors clustered at the firm-week level in parentheses. ⁺ $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

In Section 3.2.2, we stated that CEOs might be most likely to sell their stock before Class I recalls without prior MAUDE complaints and that empirically, such behavior would be evidenced by a positive coefficient for *Window of Opportunity*. Our results support this assertion. Specifically, column (2) shows that CEOs are more likely to sell their stock in the Window of Opportunity for Class I recalls without prior MAUDE complaints (0.0085; $p < 0.05$). This means CEOs are almost twice as likely⁸ to sell stock on days during the Window of Opportunity compared to days outside the window and provides initial support for our assertion that CEOs may make informed trades during the product recall decision. At the same time, column (4) shows that CEOs are less likely to sell their stock in the Window of Opportunity for Class I recalls with prior MAUDE complaints (-0.0038 ; $p < 0.01$). We do not find significant results when considering all Class I recalls in column (6). Thus, merging these two coefficients results in a non-significant overall effect for Class I recalls. This underscores the importance of categorizing recalls by prior MAUDE complaints, a central and valuable aspect of our study. Even if the FDA and SEC collaboratively agreed to examine the risk of informed trading during the Window of Opportunity, they would need to be aware of how prior MAUDE complaints influence the likelihood of informed trading. We elaborate on this further in the discussion section.

3.4.2 Profitability of Sales in the Window of Opportunity

This section examines whether CEOs profit from selling in the Window of Opportunity. This is an important and relevant question because previous research shows that product recalls have negative consequences for firms (Cheah et al. 2007, Davidson and Worrell 1992, Jarrell and Peltzman 1985) but empirical evidence of a stock market penalty around the recall initiation date is mixed (Darby et al. 2023, Thirumalai and Sinha 2011). Similarly, previous research shows that profits of trades by “typical” insiders are economically small,

⁸The coefficient for Window of Opportunity (0.0104) means that the likelihood of a CEO sale increases by $1.85 = \frac{0.0085+0.010}{0.010}$ (i.e., twice as likely) in the Window of Opportunity compared to the sample average of 0.010 in Table 3.1.

resulting in average abnormal returns of 0.89% and mean (median) dollar profits of \$4,000 (\$141) (Cziraki and Gider 2021). Thus, whether sales in the Window of Opportunity are, in fact, profitable remains an open question.

To examine the profitability of such sales, we compute *Abnormal Sales Return* as the sale-specific buy-and-hold abnormal stock return over the 180 calendar days after the sale, multiplied by -1 (Arif et al. 2022, Blackburne et al. 2021, Ravina and Sapienza 2010).⁹ Then we compare the *Abnormal Sales Return* for CEO sales in the Window of Opportunity to three benchmarks: (i) the profitability of all sales by officers or directors, including the CEO, outside the Window of Opportunity for firms in the sample (All Other Sales); (ii) the profitability of all other sales by officers or directors, including the CEO, outside the Window of Opportunity within the same year for firms in the sample (Same Year); and (iii) the profitability of all other sales by the same CEO outside the Window of Opportunity within the preceding three years for firms in the sample (Own Sales). If sales by CEOs before product recalls are indeed informed and the information they trade on is material, then we expect them to outperform these benchmarks.

In addition to *Abnormal Sales Return*, we evaluate profits from sales (in dollars) in the Window of Opportunity to assess the financial benefit CEOs extract from recall-related information. We compute *Sales Profits* by multiplying the inflation-adjusted dollar value of the sale, expressed in 2020 US dollars, by the corresponding *Abnormal Sales Return* (Cziraki and Gider 2021).

Table 3.3 presents our results. Column 1 shows the *Abnormal Sales Return* of CEO sales during the Window of Opportunity for Class I recalls without prior MAUDE complaints, Class I recalls with prior MAUDE complaints, and all Class I recalls. Columns 2, 4, and 6 provide the *Abnormal Sales Return* for the three benchmarks. Columns 3, 5, and 7 provide the p-values from a Wilcoxon rank sum test on the difference between the *Abnormal Sales*

⁹The accounting literature generally computes abnormal returns over 180 calendar days because of the “short-swing rule” (Section 16(b) of the Securities and Exchange Act of 1934), which penalizes insiders for profits from trades within six months of each other.

Return in Column 1 and the profits of the respective benchmark.

Table 3.3: Profitability of CEO Sales in Window of Opportunity using Market-Adjusted 180-days Buy-and-Hold Returns

	Benchmarks						
	All Other Sales			Same Year		Own Sales	
	Sales Profits (1)	Sales Profits (2)	p-value (3)	Sales Profits (4)	p-value (5)	Sales Profits (6)	p-value (7)
Recalls without prior MAUDE Complaints	15.28 (5.47)	-2.36 (-16.09)	<0.001	-2.58 (-9.87)	<0.001	10.03 (4.76)	0.011
Recalls with prior MAUDE Complaints	4.11 (1.73)	-2.36 (-16.09)	0.188	-3.29 (-15.57)	<0.001	-1.65 (-0.97)	0.074
All Class I Recalls	9.54 (4.93)	-2.36 (-16.09)	<0.001	-2.99 (-18.49)	<0.001	4.75 (3.07)	8.90

Notes: The table presents sales profits for sales during the Window of Opportunity. We measure *Abnormal Returns* as the sale-specific market-adjusted buy-and-hold returns over the 180 calendar days after each sale, multiplied by -1. t-statistics appear in parentheses. The p-values are from a Wilcoxon rank sum test on the difference between the *Abnormal Returns* in the Window of Opportunity (Column 1) to the Benchmarks.

In Section 3.4.1 we showed that CEOs increase their selling activity in the Window of Opportunity for Class I recalls without prior MAUDE complaints but not for those with prior MAUDE complaints (see Table 3.2). The results in Table 3.3 provide a possible explanation for this finding. We find that CEOs who sell their stock during the Window of Opportunity of Class I recalls without prior MAUDE complaints profit significantly. These sales generate profits (i.e., they help CEOs avoid losses) of 15.28% (corresponding to *Sales Profits* of \$143,439). This is significantly larger than all three benchmarks. The magnitude of these abnormal returns is comparable to informed trades around severe SEC investigations (11.76% in Blackburne et al. 2021) and modified audit reports that lead to financial results restatements (10.98% in Arif et al. 2022).

On the other hand, CEOs who sell their stock during the Window of Opportunity of Class I recalls with prior MAUDE complaints appear to profit less (i.e., 4.11% corresponding to *Sales Profits* of \$79,981). This lower profitability might explain why we do not observe a significant increase in *CEO Selling Activity* for these recalls. Overall, the evidence in Tables 3.2 and 3.3 suggests opportunistic behavior: CEOs are more likely to sell their stock

in the Window of Opportunity for Class I recalls without prior MAUDE complaints, leading to significant personal profits.

3.5 Supplementary Analysis

We study the trading patterns that we observe more comprehensively using three robustness checks to assess the validity of our results and two post-hoc analyses to explore the boundary conditions of our findings.

3.5.1 Robustness Checks

As robustness checks, we (1) repeat the analysis with an alternative measure of the selling activity of CEOs to ensure the robustness of our findings to different dependent variables, (2) repeat the profitability analysis using an alternative market adjustment that considers company size, and (3) incorporate the relatively few recalls that did not contain a PMA or 510(k) approval number, obviating our ability to link these recalls to MAUDE data.

Alternative Measure of CEO Selling Activity

To ensure that our findings are not driven by how we operationalize our dependent variable, we use *Buy-Sell-Imbalance*, as an alternative measure. Buy-Sell-Imbalance represents the net imbalance between executives' sales and purchases and is frequently used in the accounting literature (Arif et al. 2022, Blackburne et al. 2021). We compute *Buy-Sell-Imbalance* as the value of shares bought by the CEO minus the value of shares sold by them scaled by their total trading volume. By definition, the value ranges from -1 to 1, with negative values indicating net selling activity and positive values indicating net buying activity. For example, if a CEO sells \$50,000 worth of stock and buys \$20,000 then *Buy-Sell-Imbalance* takes the value -0.43; $(\frac{20,000-50,000}{70,000})$. If, on the other hand, they sell stock worth \$10,000 but buy \$20,000 then *Buy-Sell-Imbalance* takes the value 0.33;

$(\frac{20,000-10,000}{30,000})$. As our example illustrates, higher values of *Buy-Sell-Imbalance* indicate lower relative selling activity, meaning that the algebraic sign for coefficients is opposite that in Table 3.2.

We present the results with *Buy-Sell-Imbalance* as the dependent variable in Table 3.5. Our results are consistent with our main findings. Most importantly, we continue to find evidence of informed trading with CEOs having a lower Buy-Sell-Imbalance in the Window of Opportunity for Class I recalls without prior MAUDE complaints ($\hat{\delta}_1 = -0.0077$; p - *value* < 0.05).

Alternative Adjustment For Profit Analysis

To evaluate the robustness of our insider sale profitability analysis we re-compute *Abnormal Sales Return* using size-adjusted returns (Fama and French 1995). We present our results in Table 3.4. We continue to find that CEOs who sell in the Window of Opportunity of Class I recalls without prior MAUDE complaints profit from doing so. Specifically, these sales generate profits that are significantly larger than all three benchmarks (15.12%). Thus, our findings are robust to the measure of abnormal returns.

Table 3.4: Robustness Check: Profitability of CEO Sales in Window of Opportunity using Size Adjusted Sales Profits

	Benchmarks						
	Sales Profits (1)	All Other Sales		Same Year		Own Sales	
		Sales Profits (2)	p-value (3)	Sales Profits (4)	p-value (5)	Sales Profits (6)	p-value (7)
Recalls without prior MAUDE Complaints	15.12 (5.60)	-1.41 (-3.63)	<0.001	-3.01 (-7.87)	<0.001	7.19 (2.19)	0.022
Recalls with prior MAUDE Complaints	4.40 (1.92)	-1.41 (-3.63)	0.298	-3.18 (-4.49)	<0.001	1.93 (0.95)	0.867
All Class I Recalls	9.60 (5.14)	-1.41 (-3.63)	<0.001	-3.02 (-7.63)	<0.001	5.12 (2.34)	0.158

Notes: The table presents sales profits for sales during the Window of Opportunity. We measure *Abnormal Returns* as the sale-specific cumulative size and market-adjusted returns over the 180 calendar days after each sale, multiplied by -1. t-statistics appear in parentheses. The p-values are from a Wilcoxon rank sum test on the difference between the *Abnormal Returns* in the Window of Opportunity (Column 1) to the Benchmarks.

Table 3-5: Robustness Check: CEO Buy-Sell-Imbalance in Window of Opportunity

	(1)	(2)	(3)	(4)	(5)	(6)
	Recalls without prior MAUDE complaints		Recalls with prior MAUDE complaints		All Class I Recalls	
Window of Opportunity	-0.0083* (0.0034)	-0.0077* (0.0034)	0.0039** (0.0014)	0.0033* (0.0014)	0.0004 (0.0015)	0.0001 (0.0015)
Size		-0.0058*** (0.0004)		-0.0058*** (0.0004)		-0.0058*** (0.0004)
Book-to-Market Ratio		0.0041*** (0.0006)		0.0041*** (0.0006)		0.0041*** (0.0006)
Blackout Period		0.0071*** (0.0006)		0.0071*** (0.0006)		0.0071*** (0.0006)
Abnormal Returns		-0.0037*** (0.0005)		-0.0036*** (0.0005)		-0.0036*** (0.0005)
Volatility		-0.0015 (0.0158)		-0.0025 (0.0158)		-0.0023 (0.0158)
Daily Abnormal Returns		-0.0198*** (0.0033)		-0.0198*** (0.0033)		-0.0198*** (0.0033)
Daily Volatility		0.0176* (0.0088)		0.0175* (0.0088)		0.0175* (0.0088)
Firm fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Day fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Num.Obs.	523,310	523,310	523,310	523,310	523,310	523,310
R ²	0.023	0.025	0.023	0.025	0.023	0.025

Notes: The table presents the results of a firm-day fixed effects regression with CEO Buy-Sell-Imbalance as the dependent variable. The unit of analysis is firm-day. The coefficient for Window of Opportunity, a binary indicator with value one if the trading day is between the defect awareness and recall initiation dates for a recall by the focal firm, indicates the change in the likelihood that a CEO sells more stock than they buy on a given day. Standard errors clustered at the firm-week level in parentheses. [†] $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Recalls without PMA or 510(k) number

There are 35 Class I recalls that we omitted from our main analysis because they neither have a PMA nor a 510(k) number assigned to them, preventing us from associating them with MAUDE complaint data. Here, we reintroduce these recalls to check the robustness of our findings. Because we cannot definitively classify them as recalls with or without prior MAUDE complaints, we evaluate how including them in each category separately affects our results. In columns 1 and 2 of Table 3.6, we assume these recalls had prior MAUDE complaints, while in columns 3 and 4, we assume they did not. Finally, we include them in the “All Class I recalls” category in columns 5 and 6.

We can conclude from Table 3.6 that our inability to clearly assign these recalls into one or the other complaint category does not impact our conclusions. We continue to observe an increase in selling activity in the Window of Opportunity for recalls without prior MAUDE complaints and a decrease in selling activity for those recalls with prior MAUDE complaints. There remains no detectable change in selling activity in the Window of Opportunity when combining both categories of recalls.

3.5.2 Post-hoc

As post-hoc analyses, we explore the informed trading effects we have observed more deeply. We do so along two dimensions: our focus on Class I recalls and our focus on CEOs. We first investigate whether the observed increase in the selling activity of CEOs is limited to Class I recalls or extends to Class II and III recalls as well. Then, we examine if quality and operations officers, the other senior leaders most likely to be closely associated with and informed of Class I recall decisions under deliberation (Wowak et al. 2021), also exhibit opportunistic stock trading behavior.

Table 3.6: Robustness Check: Considering Recalls without PMA or 510(k) Number in Analysis

	Recalls without PMA or 510(k) number added to...					
	Recalls without prior MAUDE complaints	Recalls without prior MAUDE complaints	Recalls with prior MAUDE complaints	All Class I Recalls		
	(1)	(2)	(3)	(4)	(5)	(6)
Window of Opportunity	0.0053* (0.0026)	0.0049+ (0.0026)	-0.0045*** (0.0012)	-0.0040*** (0.0012)	-0.0008 (0.0013)	-0.0006 (0.0013)
Size		0.0056*** (0.0003)		0.0056*** (0.0003)		0.0056*** (0.0003)
Book-to-Market Ratio		-0.0029*** (0.0004)		-0.0029*** (0.0004)		-0.0030*** (0.0004)
Blackout Period		-0.0085*** (0.0005)		-0.0085*** (0.0005)		-0.0085*** (0.0005)
Abnormal Returns		0.0030*** (0.0004)		0.0030*** (0.0004)		0.0030*** (0.0004)
Volatility		-0.0021 (0.0129)		-0.0013 (0.0129)		-0.0015 (0.0129)
Daily Abnormal Returns		0.0122*** (0.0026)		0.0122*** (0.0026)		0.0122*** (0.0026)
Daily Volatility		0.0200** (0.0073)		0.0201** (0.0073)		0.0201** (0.0073)
Firm fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Day fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Num.Obs.	523,310	523,310	523,310	523,310	523,310	523,310
R ²	0.024	0.026	0.024	0.026	0.024	0.026

Notes: The table presents the results of a firm-day fixed effects regression with recalls that do not have a PMA or 510(k) number added to the three groups (i.e., recalls without prior MAUDE complaints, recalls with prior MAUDE complaints, all Class I recalls) separately. Standard errors clustered at the firm-week level in parentheses. + $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Extension to Class II and III Recalls

We focus on Class I recalls in our main analysis because they represent the most severe cases. Consequently, they attract significant negative attention from external stakeholders and are highly likely to be within the purview of CEOs. Here, we explore whether the observed increase in the selling activity of CEOs is limited to Class I recalls or extends to Class II and III recalls. Our data includes 4,671 Class II and 465 Class III recalls. Specifically, we construct six new variations of *Window of Opportunity*: Class II or III recalls without prior MAUDE complaints, Class II or III recalls with prior MAUDE complaints, and all Class II or III recalls. We then re-estimate Equation 3.1.

Table 3.7 presents our results. The coefficients for *Window of Opportunity* are not significant in any column. This indicates that there is no statistically detectable change in the likelihood that CEOs sell their stock during the Window of Opportunity for Class II or III recalls, regardless of the presence of MAUDE complaints. This is consistent with our theorizing that CEOs are most likely to take such risks on the most serious Class I recalls only.

Extension to Quality and Operations Officers

Up to this point, our analysis focuses on the selling activity of CEOs, given their involvement in the recall decision-making process (Mayo et al. 2022) and the well-established influence of their stock compensation on the decision to recall (Darby et al. 2023, Wowak et al. 2015). However, CEOs are not the only senior managers involved in recall decisions. Senior quality and operations officers are often very involved in, or even make, the recall decision while keeping the CEO informed (Darby et al. 2023, Wowak et al. 2021). Hence, we explore whether quality and operations officers also exhibit opportunistic behavior during the Window of Opportunity.

We use BoardEx data to identify quality and operations officers for our sample firms. To

Table 3.7: CEO Selling Activity in Window of Opportunity for Class II and III Recalls

	Class II Recalls		Class III Recalls			
	Without prior MAUDE complaints (1)	With prior MAUDE complaints (2)	All Class II (3)	Without prior MAUDE complaints (4)	With prior MAUDE complaints (5)	All Class III (6)
Window of Opportunity	-0.0001 (0.0007)	-0.0006 (0.0008)	-0.0006 (0.0007)	0.0020 (0.0017)	0.0017 (0.0024)	0.0013 (0.0012)
Size	0.0056*** (0.0003)	0.0056*** (0.0003)	0.0056*** (0.0003)	0.0056*** (0.0003)	0.0056*** (0.0003)	0.0056*** (0.0003)
Book-to-Market Ratio	-0.0030*** (0.0004)	-0.0029*** (0.0004)	-0.0029*** (0.0004)	-0.0030*** (0.0004)	-0.0030*** (0.0004)	-0.0030*** (0.0004)
Blackout Period	-0.0085*** (0.0005)	-0.0085*** (0.0005)	-0.0085*** (0.0005)	-0.0085*** (0.0005)	-0.0085*** (0.0005)	-0.0085*** (0.0005)
Abnormal Returns	0.0030*** (0.0004)	0.0030*** (0.0004)	0.0030*** (0.0004)	0.0030*** (0.0004)	0.0030*** (0.0004)	0.0030*** (0.0004)
Volatility	-0.0016 (0.0129)	-0.0012 (0.0129)	-0.0013 (0.0129)	-0.0019 (0.0129)	-0.0017 (0.0129)	-0.0019 (0.0129)
Daily Abnormal Returns	0.0122*** (0.0026)	0.0122*** (0.0026)	0.0122*** (0.0026)	0.0122*** (0.0026)	0.0122*** (0.0026)	0.0122*** (0.0026)
Daily Volatility	0.0201** (0.0073)	0.0201** (0.0073)	0.0201** (0.0073)	0.0201** (0.0073)	0.0201** (0.0073)	0.0201** (0.0073)
Firm fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Day fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Num.Obs.	523,310	523,310	523,310	523,310	523,310	523,310
R ²	0.026	0.026	0.026	0.026	0.026	0.026

Notes: The table presents the results of a firm-day fixed effects regression for Class II and III recalls with CEO selling activity as the dependent variable. The unit of analysis is firm-day. The coefficient for Window of Opportunity, a binary indicator with value one if the trading day is between the defect awareness and recall initiation dates for a recall by the focal firm, indicates the change in the likelihood that a CEO sells more stock than they buy on a given day. Standard errors clustered at the firm-week level in parentheses. + $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

identify quality and operations officers, we examine the titles of individuals for keywords that suggest their involvement in the quality or operations function (e.g., “Ops,” “Operations,” “Quality”).¹⁰ This method follows prior research (Hambrick and Cannella 2004, Fu et al. 2020, Roh et al. 2016). Then we compute *Quality and Operations Officer Selling Activity* as an alternative dependent variable, which equals one if quality and operations officers at the recalling firm sell more than they buy on a given day and zero otherwise. Using this information, we replicate our previous analysis for quality and operations officers.

Table 3.8 presents our results. We observe that quality and operations officers tend to decrease their selling activity during the Window of Opportunity of Class I and II—but not Class III—recalls. In particular, the coefficients for *Window of Opportunity* are negative and significant for Class I recalls without prior MAUDE complaints, Class II recalls with prior MAUDE complaints, and all Class II recalls. While reducing selling activity aligns with an officer’s fiduciary duty to refrain from trading based on material, non-public information (Jagolinzer et al. 2011), this behavior differs from that of CEOs who increase their selling activity for Class I recalls without prior MAUDE complaints.

3.6 Discussion and conclusion

Our study explores whether CEOs engage in informed trading during the time in which the firm is deliberating on whether or not to initiate the most serious type of medical device product recall. The Window of Opportunity appears to provide them with the means and the motivation to sell their stock in anticipation of Class I recalls. However, whether they exploit this opportunity remained an open question, until now. Our findings reveal that CEOs, in fact, are more likely to sell their stock in the Window of Opportunity for Class I recalls without prior MAUDE complaints. Furthermore, we show that these sales are highly profitable for their personal fortunes. In the discussion below, we aim to achieve

¹⁰We provide a list of all BoardEx roles that we consider to be quality and operations officers in Table B.1 in the E-Companion.

Table 3.8: Quality and Operations Officer Selling Activity in Window of Opportunity

	Class I Recalls			Class II Recalls			Class III Recalls		
	Without prior MAUDE complaints (1)	With prior MAUDE complaints (2)	All Class I Recalls (3)	Without prior MAUDE complaints (4)	With prior MAUDE complaints (5)	All Class II Recalls (6)	Without prior MAUDE complaints (7)	With prior MAUDE complaints (8)	All Class III Recalls (9)
Window of Opportunity	-0.0022*** (0.0003)	-0.0001 (0.0006)	-0.0007 (0.0005)	-0.0003 (0.0003)	-0.0011*** (0.0003)	-0.0009*** (0.0002)	0.0020 (0.0017)	0.0017 (0.0024)	0.0013 (0.0012)
Size	0.0008*** (0.0001)	0.0008*** (0.0001)	0.0008*** (0.0001)	0.0008*** (0.0001)	0.0008*** (0.0001)	0.0008*** (0.0001)	0.0056*** (0.0003)	0.0056*** (0.0003)	0.0056*** (0.0003)
Book-to-Market Ratio	-0.0003* (0.0001)	-0.0003* (0.0001)	-0.0003* (0.0001)	-0.0003* (0.0001)	-0.0003* (0.0001)	-0.0003* (0.0001)	-0.0030*** (0.0004)	-0.0030*** (0.0004)	-0.0030*** (0.0004)
Blackout Period	-0.0017*** (0.0002)	-0.0017*** (0.0002)	-0.0017*** (0.0002)	-0.0017*** (0.0002)	-0.0017*** (0.0002)	-0.0017*** (0.0002)	-0.0085*** (0.0005)	-0.0085*** (0.0005)	-0.0085*** (0.0005)
Abnormal Returns	0.0007*** (0.0002)	0.0007*** (0.0002)	0.0007*** (0.0002)	0.0007*** (0.0002)	0.0007*** (0.0002)	0.0007*** (0.0002)	0.0030*** (0.0004)	0.0030*** (0.0004)	0.0030*** (0.0004)
Volatility	-0.0038 (0.0057)	-0.0041 (0.0057)	-0.0039 (0.0057)	-0.0041 (0.0057)	-0.0034 (0.0057)	-0.0036 (0.0057)	-0.0019 (0.0129)	-0.0017 (0.0129)	-0.0019 (0.0129)
Daily Abnormal Returns	0.0042*** (0.0011)	0.0042*** (0.0011)	0.0042*** (0.0011)	0.0042*** (0.0011)	0.0042*** (0.0011)	0.0042*** (0.0011)	0.0122*** (0.0026)	0.0122*** (0.0026)	0.0122*** (0.0026)
Daily Volatility	0.0065* (0.0032)	0.0065* (0.0032)	0.0065* (0.0032)	0.0064* (0.0032)	0.0065* (0.0032)	0.0065* (0.0032)	0.0201** (0.0073)	0.0201** (0.0073)	0.0201** (0.0073)
Firm fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Num.Obs.	523,310	523,310	523,310	523,310	523,310	523,310	523,310	523,310	523,310
R ²	0.019	0.019	0.019	0.019	0.019	0.019	0.026	0.026	0.026

Notes: The table presents the results of a firm-day fixed effects regression for Class I, II, and III recalls with Quality and Operations Officer selling activity as the dependent variable. The unit of analysis is firm-day. Standard errors clustered at the firm-week level in parentheses. + $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

three objectives: (1) reconcile our findings with prior research on informed trading and product recalls, (2) highlight the theoretical contributions of our study, and (3) outline its practical implications.

Our profitability analysis (Section 3.4.2) shows that sales in the Window of Opportunity for Class I recalls yield abnormal buy-and-hold returns ranging from 4.11% to 15.28%, translating to dollar profits between \$79,981 and \$143,439. These returns surpass the typical stock market reaction to high-severity recalls, which event studies estimate to range between 1% and 3% (Darby et al. 2023, Thirumalai and Sinha 2011). They also far exceed the average insider trade profits reported in the literature (e.g., \$4,000 mean and \$141 median profits in Cziraki and Gider (2021)).

To reconcile the difference with prior product recall event studies, we revisit a critical assumption of these studies: the unanticipated event assumption (McWilliams and Siegel 1997). This assumption posits that traders acquire new information from the product recall announcement itself, with no prior information leakage. However, we argue in Section 3.2.2 that information about underlying product issues can already be available to investors through complaints filed prior to the recall announcement. Such pre-existing information would undermine the unanticipated event assumption and could lead to an underestimation of the true effects. By distinguishing between recalls with and without prior complaints, our study captures variance in recall types that prior studies have not considered. Supporting this argument, we find that abnormal returns for Class I recalls without prior complaints (4.11%) align closely with the 1%-3% short-term market reactions reported in prior studies, whereas recalls with prior complaints yield substantially higher abnormal returns (15.28%). Moreover, our analysis considers a longer time frame, enabling us to capture delayed repercussions that event studies focused on the announcement date might not capture. For instance, in Philips' high-profile recall of sleep apnea and ventilator machines, the number of affected products expanded from 3 to 5 million months after the initial announcement, leading to a stock price reduction of 15%—a decline that short-term

event studies would not account for (Zipp 01/12/2022). Incorporating complaint data and using a longer observation period explains why the profitability of trades in the Window of Opportunity exceeds the typical stock market reaction to high-severity recalls.

The difference in dollar profits is similarly explained by sample variation. While Cziraki and Gider (2021) examines trades by insiders across several hundred thousand transactions—most of which are unlikely to be informed. Our study specifically focuses on potentially informed trades by CEOs. The average insider transaction in Cziraki and Gider (2021) yields a modest abnormal return of 0.89%, whereas our analysis reports abnormal returns of up to 15.28%, which is comparable to informed trading returns reported by Arif et al. (2022) and Blackburne et al. (2021). Now, because these trades generate considerably higher returns, they are also more profitable, explaining the observed difference in dollar profits.

Our findings make three important theoretical contributions. First, in examining opportunistic behavior by CEOs during the recall decision-making process, we build upon product recall studies that use CEO characteristics to predict the occurrence or delay of recalls (Darby et al. 2023, Mayo et al. 2022, Liu et al. 2016, Wowak et al. 2015). Specifically, we extend the work by Darby et al. (2023) who explain recall delays using CEO stock ownership as an indicator of financial incentive but remain silent on how the financial gains are accrued. By showing that CEOs sell stock and that these sales are profitable, we demonstrate how CEOs might accrue financial benefits for personal gain during the window of time in which the firm is deciding whether or not to initiate a serious recall.

Second, we are the first to examine and provide evidence that CEOs appear to exploit non-public information during the product recall decision-making process—an operational process that serves as a regulatory mechanism designed to manage and monitor product safety. This is particularly important because the existing literature on informed trading is focused on accounting and finance-oriented processes such as audits (Arif et al. 2022), SEC investigations (Blackburne et al. 2021), federal financial assistance (Jagolinzer et al.

2020), or 10K/Q reports (Huddart et al. 2007). Instead, our results highlight that informed trading can be present in critical operational contexts as well, paving the way for future research into other operational contexts in which there may also be a temptation to conduct informed trades, such as supplier audits or product development projects.

Finally, we contribute to the emerging literature that links operations management and organizational misconduct (Kistler et al. 2024, Petersen et al. 2024, Skilton and Bernardes 2022, Belavina et al. 2020) by demonstrating how CEOs might exploit the product recall decision-making process to engage in informed trading. While the drivers of organizational misconduct are well-established (Palmer 2012, Greve et al. 2010), research on effective prevention strategies is limited. As we discuss below, our work, which is based on an in-depth understanding of the product recall decision-making process, addresses this gap by providing regulators with a guide on how to identify and help prevent informed trading.

Our study has two key practical implications. For senior managers, particularly those concerned with corporate governance, our research highlights the importance of implementing informed trading policies that effectively limit the trading activities of executives who are involved in the product recall decision-making process. Most firms have self-imposed trading restrictions, such as blackout windows around earning announcement dates (Blackburne et al. 2021). However, these restrictions do not extend to product recall decisions. Thereby, insiders may face a moral hazard: they need to make product recall decisions in the best interest of the public health that might adversely affect them financially. Our results should prompt senior management and regulators to closely scrutinize insider sales that occur in connection with product recalls, ensuring a more robust, fair, and secure corporate environment. One possible solution is a so-called trading plan, which requires officers to schedule a trade up to 90 days before executing it (Ramonas 2022).

Concerning regulators, the SEC is primarily responsible for safeguarding the interests of individual investors, while the FDA is responsible for monitoring product safety and product recalls. However, from our discussions with senior individuals at the FDA, these

two agencies operate independently from each other. Our study identifies a collaboration opportunity between the SEC and the FDA to identify and deter informed trading. Specifically, the FDA can share defect awareness dates with the SEC, allowing the SEC to investigate abnormal trading patterns around these previously undisclosed dates, especially for those recalls that are most secret within the firm before they are initiated, those without prior MAUDE complaints. Knowing that the FDA is sharing the defect awareness dates with the SEC and that the SEC is monitoring insider sales more closely may mitigate opportunistic trading by executives. It is noteworthy as well that in discussions we have had with senior leadership at the FDA, the notion of informed trading during the recall decision-making process appears to be news to them.

Our results should be interpreted in the context of our study's limitations. First, we are unable to pinpoint which sales are definitively informed by knowledge about upcoming recalls. Instead, similar to related accounting studies, we make inferences about informed sales based on changes in average selling activity in the Window of Opportunity. This approach aligns with methods commonly found in the informed trading literature (e.g., Arif et al. 2022, Blackburne et al. 2021, Huddart et al. 2007). As a result, it is important to note that while our findings suggest the presence of informed trading, they do not provide a definitive method for identifying illegal actions. Additionally, our findings that indicate the absence of informed trading in certain contexts (e.g., during the Window of Opportunity for Class II and III recalls) mean there is no statistically identifiable pattern, but not necessarily that informed trading does not occur. Hence, our suggestions should be interpreted as a call for heightened awareness and scrutiny rather than a prosecutorial tool.

Second, our focus is restricted to one operational context, Class I medical device product recalls. Future research might be able to extend the investigation to product recalls in other industries such as automotive (Mukherjee et al. 2022) and consumer products (Mayo et al. 2022, Ni et al. 2014) or to entirely different operational contexts such as supplier audits, product development processes, or the collection of information for supply chain

transparency. Similarly to our setting, insiders might sell their stock once failures in these contexts are known internally but not yet disclosed to the public.

Trading based on private, material information about a firm is illegal in the United States. Yet, despite best efforts by policy-makers and the SEC, the practice remains prevalent because the contexts in which it happens are not always well understood (Patel and Putnins 2020). Our findings provide novel evidence that CEOs of medical device firms appear to knowingly or unknowingly benefit from the product recall decision-making process for personal gain at the expense of the firm's shareholders.

Chapter 4

Chief Sustainability Officers and Environmental Injustice: An Empirical Evaluation

Look, we all suffer from exposure to pollution in this country, but some significantly more than others, thanks to what's called environmental racism.

John Oliver (2022), *Last Week Tonight*

4.1 Introduction

Manufacturing activity is often detrimental to the environment because the production of goods generates waste and toxic releases, including harmful chemicals, gases, liquids, and metals (Muthulingam et al. 2022). These toxic releases pose serious health risks, such as cancer and chronic illnesses, particularly for communities living near industrial facilities (EPA 2023b, Mohai et al. 2009). However, the burden of industrial pollution is not shared equally across the United States. Due to systemic disparities, a disproportionate share of

environmental harm is borne by predominantly Black or African American communities, a phenomenon commonly referred to as “environmental racism” or “environmental injustice” (Kalkanci et al. 2019). The existence of environmental injustice is well-documented (e.g., Diaz 2017, Mohai et al. 2009, Chavis Jr. and Lee Charles 1987) and reflects a troubling reality: firms are more likely to build hazardous manufacturing facilities in Black or African American neighborhoods (hereafter, underserved communities) than in predominantly white neighborhoods (hereafter, better-served communities) (Pastor et al. 2001). Emerging research in Operations Management suggests that environmental injustice extends beyond these location decisions. Specifically, working papers by Shubham and Subramanian (2025) and Cole and Muthulingam (2025) suggest that facilities operating in underserved communities also have more toxic releases and adopt fewer toxic release prevention measures than otherwise similar facilities in better-served communities. In this paper, we focus on this operational dimension of environmental injustice and ask: If firm operations systematically impose higher environmental costs on underserved communities, what mechanisms—whether regulatory interventions or voluntary corporate governance reforms—can help reduce these disparities?

Regulatory efforts to address environmental injustice have had limited success. For example, President Bill Clinton’s Executive Order 12898 required federal agencies to consider environmental justice in their missions (The White House 1994), but it lacked specific action plans and enforcement mechanisms (Geltman et al. 2016, O’Neil 2007). Moreover, such policies often fail to hold manufacturing firms directly accountable for environmental harms (Bullard 1993). Recent policy shifts further complicate regulatory efforts; in February 2025, the U.S. government scaled back its commitment to environmental justice by closing the EPA’s Environmental Justice Offices and rolling back key protections (Friedman Mar 11, 2025).

These policy setbacks point to the need for an alternate approach to tackling environmental injustice, one where manufacturing firms and their facilities play a more

central role. As companies and investors face increased accountability for environmental cleanup costs at their facilities (e.g., through financial penalties, lower valuations, and reputational damage), firm governance reforms have emerged as a potential solution to bolster social and environmental performance and prevent environmental cleanup costs (Chicago Sun-Times 2023, California Environmental Protection Agency 2022, Eccles et al. 2014).

In this paper, we investigate whether a specific firm governance reform—the appointment of Chief Sustainability Officers (CSOs)—can result in the reduction of environmental injustice. Since the early 2000s, firms have increasingly appointed CSOs to lead their social and environmental initiatives and increase their credibility with sustainability-focused stakeholders (Gupta et al. 2021, Arora et al. 2020). By shifting the organizations attention towards environmental and social issues, CSOs may influence manufacturing facilities to improve their environmental performance, such as by reducing toxic releases. If such improvements are more substantial in underserved communities, CSO appointments could also serve as a mechanism for reducing environmental injustice. This premise rests on the assumption that CSO appointments improve environmental performance. However, empirical evidence towards this assumption is limited with Fu et al. (2020) finding that CSO appointments are associated with fewer socially irresponsible activities but Kanashiro and Rivera (2019) finding that CSO appointments are associated with higher toxic releases.

To address this puzzle, we first examine whether and to what extent CSO appointments affect environmental performance, measured through toxic release volumes (**RQ-1**). Based on the above arguments that CSOs shift firms' attention toward environmental issues, we hypothesize that CSO appointments lead to reduced toxic releases. We then investigate whether CSO appointments help reduce environmental injustice by analyzing if their effect is contingent on the community surrounding the facilities (**RQ-2**). Facilities in underserved communities may operate further from their operating frontier (Schmenner and Swink 1998)

due to these communities low socioeconomic power to influence environmentally conscious manufacturing operations (Diaz 2017, Taylor 1997). As a result, we hypothesize that such facilities may realize greater improvements in environmental performance following a CSO appointment, contributing to a reduction of environmental injustice.

We empirically examine these research questions using data on facility-chemical level toxic releases of U.S. manufacturing facilities between 2000 and 2020 from the Environmental Protection Agency’s (EPA) Toxic Release Inventory (TRI) program (EPA 2023b). Toxic releases are a well-established measure of environmental performance in both the operations management (Dhanorkar et al. 2023, Muthulingam et al. 2022, Fu et al. 2019) and the environmental injustice literature (Mohai and Saha 2007, Elliott et al. 2004). Our sample includes facility-chemical level observations for 598 publicly traded companies—165 (28%) of which appoint a CSO during the sample period.

To answer our first research question, we use a two-way fixed effects (TWFE) regression with matching to account for a firm’s endogenous decision to appoint a CSO. The matching step ensures common support of the sample and mitigates potential bias arising from endogenous treatment decisions (Calvo et al. 2023, Ertekin et al. 2021). Specifically, we match facility-chemical observations from firms that appoint a CSO to those from firms that do not, based on factors such as chemical type, community characteristics, total population around the facility, regulatory pressure, and the likelihood that the parent firm appoints a CSO. This likelihood is based on firm-level factors such as the firm size, the firm’s overall emissions, and the firm’s broad ESG orientation. We find that CSO appointments do indeed improve environmental performance. Specifically, our estimate indicates that toxic releases decrease by 12% following the appointment of a CSO, relative to comparable facilities without a CSO.

Next, to answer our second research question, about whether CSO appointments also reduce environmental injustice by reducing toxic releases more at facilities in underserved communities, we use a split sample analysis. Following previous research (e.g., Shubham and

Subramanian 2025, Mohai and Saha 2007) and EPA guidelines, we measure how underserved a community is by the proportion of Black or African American persons within a 1-mile radius host neighborhood of a facility. Our results show that appointing a CSO reduces toxic releases by up to 16% for facilities in underserved communities. However, we do not find evidence of a reduction in toxic releases for facilities in better-served communities. This heterogeneity in the effect of CSO appointment, in turn, reduces environmental injustice.

Our findings are robust to several additional analyses, such as accounting for the staggered appointment of CSOs (Section 4.6.1), using alternative specifications to identify underserved communities (Section 4.6.1) and considering different matching settings (Section 4.6.1). We also consider and rule out several alternative explanations for our main finding (Section 4.6.2).

Having established that CSO appointments are effective at combating environmental injustice, we next investigate two potential mechanisms driving this outcome: source reduction activities (i.e., minimizing toxic waste generation at the source) and waste management processes (i.e., preventing the release of toxic waste into the environment). Both mechanisms require significant resource investment but can effectively reduce toxic releases (Li et al. 2025, Muthulingam et al. 2022). We find that facilities increase both source reduction activities and waste management efforts following the appointment of a CSO. Notably, this effect is driven by facilities in underserved communities, which aligns with our theoretical framework and main findings. Together, these results suggest that appointing a CSO enhances overall environmental performance while also addressing environmental injustice by promoting greater adoption of resource-intensive pollution mitigation strategies in underserved communities.

Our results contribute to theory, practice, and regulation in several ways. Our primary theoretical contribution is to highlight the connection between firm governance reforms and environmental injustice. First, we empirically establish the positive effect that appointing a CSO has on environmental performance, a subject of debate in the CSO literature.

Second, we show that appointing a CSO reduces environmental injustice by reducing the toxic releases of facilities in underserved communities. In doing so, we answer the call for more operations management research that considers issues of equity and inclusion in the context of environmental performance (Kalkanci et al. 2019). We also contribute to the emerging literature on the link between operations management and organizational misconduct (Skilton and Bernardes 2022, Belavina et al. 2020) by highlighting the potential of governance reforms (such as appointing a CSO) to address misconduct in the form of environmental injustice.

For regulators, our research highlights the potential of encouraging firm governance reforms as an alternative control mechanism. This appears particularly relevant nowadays as the bodies tasked with protecting human health and the environment (e.g., the EPA) face funding challenges that limit their ability to implement regulatory interventions. For practitioners, our results suggest that social sustainability-conscious managers and stakeholders should consider pushing their firms to create a dedicated CSO role. Doing so may help direct the attention of decision-makers to social and environmental sustainability issues and ultimately lead to improved sustainability performance.

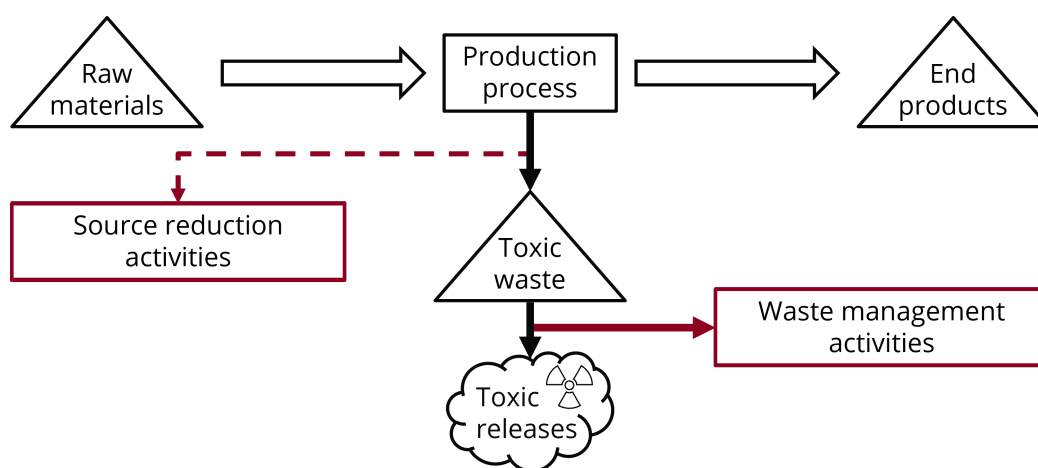
4.2 Background

To ground our theoretical model, which links the appointment of a Chief Sustainability Officer (CSO) to environmental performance and environmental injustice, we first discuss the toxic release creation process. We then describe the historical context of environmental injustice in the United States (see Mohai et al. 2009 for a comprehensive review) and expand upon the role of CSOs.

4.2.1 Toxic Releases

Manufacturing facilities generate toxic releases as an inevitable byproduct of transforming raw materials into end products (Muthulingam et al. 2022). As illustrated in Figure 4.1, these processes produce toxic waste, which, if not effectively managed, leads to toxic releases—harmful pollutants discharged into the environment. The extent of these releases depends on a facility’s operational choices regarding resource-intensive pollution reduction strategies (Li et al. 2025).

Figure 4.1: Waste Generation Process Diagram



The diagram illustrates the toxic release generation process and the two mechanisms through which facilities can prevent toxic releases. *Figure adapted from Fu et al. (2019).*

There are two main toxic release reduction strategies (Li et al. 2025, Fu et al. 2019): source reduction and waste management activities. Source reduction activities (left, dashed line in Figure 4.1) aim to minimize the generation of toxic waste at its origin. This can be achieved by adopting cleaner production technologies, improving material efficiency, or substituting hazardous inputs with safer alternatives. In contrast, waste management (right, solid red line) focuses on controlling waste after it is generated through treatment, recycling, or secure disposal methods such as containment in hazardous waste landfills. Toxic waste that is not controlled through either of these strategies is then released into the

environment. These toxic releases are the focus of our study.

4.2.2 Environmental Injustice in the United States

Environmental injustice in the United States traces back to the early 20th century, when underserved communities, particularly those inhabited by Black or African American persons, began to bear the brunt of industrial pollution and toxic releases (Taylor 1997). Because these communities had limited socioeconomic power to resist, manufacturing firms often targeted them when selecting sites for hazardous facilities (Diaz 2017, Taylor 1997). Over the latter half of the 20th century, growing awareness of environmental injustice prompted academic research on (and regulatory responses to) this situation.

One such stream of research focuses on quantifying the existence of environmental injustice, revealing stark disparities in the location and operation of hazardous manufacturing facilities. Studies show that facilities tend to operate in areas with higher percentages of Black or African American persons (Chavis Jr. and Lee Charles 1987). For example, Pastor et al. (2001) finds that for every 1% increase in the minority population of a Los Angeles census tract, there is a corresponding 3% increase in the likelihood that a toxic storage and disposal facility exists. Additionally, research shows that facilities in underserved communities are more likely to experience toxic spill accidents (Elliott et al. 2004) and have up to 30% higher negative environmental impacts during routine operations than those in better-served communities (Shubham and Subramanian 2025). Our data aligns with these findings, indicating a strong correlation between a facility's toxic releases and the presence of Black or African American persons in its host neighborhood (Table 4.2 in Section 4.5).

Another research stream establishes market incentives, regulatory systems, and power structures as the main cause of environmental injustice (Mohai et al. 2009). Specifically, research shows that environmental injustice is caused by firms actively targeting underserved communities when deciding where to operate their most harmful facilities; misguided

regulations, such as decisions to create industrial zones in or near underserved communities or unequal enforcement of regulations; and unequal socioeconomic power to resist actions unfavorable to the host community (Diaz 2017, Mohai et al. 2009, Bullard 1993). Moreover, research now overwhelmingly dismisses other possible explanations for environmental injustice such as the notion that underserved communities relocate into the host neighborhood of highly hazardous facilities because of, say, lower property prices (Mohai and Saha 2015, Pastor et al. 2001).

Even though the existence and causes of environmental injustice are well understood, attempts to control it have failed. Evidence suggests that regulatory interventions, like establishing the EPA's Office of Environmental Justice or President Bill Clinton's Executive Order 12898, which mandated considerations of environmental injustice across federal agencies, have not eliminated environmental injustice (Geltman et al. 2016, Mohai et al. 2009, O'Neil 2007).

Our study contributes to this literature by examining the effectiveness of appointing Chief Sustainability Officers—a governance reform—to reduce the disproportionate toxic releases by facilities in underserved communities as compared to their counterparts in better-served areas. Governance reforms have proven effective in addressing misconduct in other contexts (Palmer 2012), and Chief Sustainability Officers hold promise for mitigating environmental injustice due to their potential positive influence on environmental performance (Fu et al. 2020).

4.2.3 The Role of Chief Sustainability Officer

Chief Sustainability Officers (CSOs), also known as Chief Responsibility Officers or Corporate Sustainability Executives, are key figures in the modern corporate governance landscape. CSOs typically have a combination of education, experience, and skills relevant to sustainability and corporate social responsibility and come from diverse backgrounds, such as environmental functions, operations, or law (Miller and Serafeim 2014). While

their specific titles and backgrounds may vary, their core role remains consistent: driving sustainability within their organizations. Often part of a firm's top management team, CSOs are accountable for stimulating sustainable growth with limited resources (Fu et al. 2020, Unruh 2015). For example, the ESG team at Verizon—a wireless and wireline technology company with more than 130,000 employees—includes only *seven* people (Pagitsas 2022).

They have both external and internal responsibilities (Miller and Serafeim 2014). Externally, CSOs manage relationships with stakeholders through activities like non-financial reporting or engagement with non-governmental organizations (Unruh 2015), thus acting as a bridge between the organization and its stakeholders to “report on progress and tell the company's story” (Pagitsas 2022), p.258), and thereby enhancing a firm's legitimacy with sustainability-conscious stakeholders (Gupta et al. 2021). Thus, appointing a CSO signals a firm's commitment to improving social and environmental performance (Wiengarten et al. 2017b).

Internally, CSOs provide strategic direction and oversight on sustainability issues and try to improve a firm's sustainability performance (Gupta et al. 2021, Arora et al. 2020). However, due to their limited resources, CSOs cannot improve sustainability performance alone. Instead, most CSOs focus on integrating sustainability thinking into their firm's culture and promoting a collective effort toward this end (Kanashiro and Rivera 2019, Unruh 2015). For instance, at Owens Corning, various operating units manage sustainability initiatives under the guidance of the core sustainability team (Pagitsas 2022). Similarly, Boeing's CSO advances sustainability performance by incorporating environmental targets into the standard business plans of operational units (Boeing 2008). Thus, CSOs achieve sustainability improvements by fostering a culture of sustainability across the organization and involving operating units in sustainability efforts. However, it remains unclear whether most CSOs are successful in improving real and measurable sustainability metrics. Existing studies suggest that appointing a CSO can lead to better environmental performance (Ebelet et al. 2023), no change in environmental performance except for sustainability leaders

(Peters et al. 2019), or worse environmental performance (Kanashiro and Rivera 2019). Our work helps address this ambiguity by examining a more granular outcome measure (facility-level toxic releases) and using a more robust identification strategy.

4.3 Hypothesis development

Given the limited resources available to most CSOs, it is not self-evident how their appointment at the firm level affects toxic releases at the facility level or why and how this effect might vary by the type of community in a facility's host neighborhood. We argue that the attention-based view of the firm (Ocasio 1997, 2011) provides a theoretical framework that connects firm-level governance reforms to facility-level outcomes. This theory is gaining prominence in the operations management literature (Li et al. 2022, Dhanorkar et al. 2018), and management scholars use it to understand actions spanning organizational levels (Joseph and Wilson 2018, Rerup 2009).

4.3.1 Linking CSO Appointments to Environmental Performance

At its core, the attention-based view argues that the actions of an organization are the result of the attention focus of its decision-makers (Li et al. 2022, Ocasio 1997). Ample empirical evidence supports this claim. For instance, Li et al. (2022) argue that firms that focus on risk issues enhance their risk management capabilities and reduce risk exposure. Similarly, Dhanorkar et al. (2018) demonstrate that when regulatory inspections focus decision-makers' attention on environmental concerns, they respond by implementing more environmental improvement initiatives. Likewise, Muthulingam et al. (2022) show that decision-makers focused on wasteful production practices implement measures to reduce toxic releases. In short, organizations act on the issues that capture their decision-makers' attention. In our context, if decision-makers become more focused on environmental issues, they are more likely to reduce toxic releases through greater source reduction activities and

improved waste management (see Figure 4.1).

This raises the question of which issues decision-makers focus on. At any given time, decision-makers can only allocate their attention to a limited set of priorities amidst numerous competing challenges and strategic directions (Ocasio 1997, Bouquet et al. 2009). Manufacturing managers, for example, must balance issues such as cost-quality trade-offs, worker safety, technological advancements, and environmental impact. The relative importance of these priorities is shaped by the firm's external environment and internal attention structure (Ocasio 2011, 1997). For example, in Muthulingam et al. (2022), severe water shortages shift managerial attention toward reducing wasteful production practices. Similarly, in Li et al. (2022), the appointment of a Chief Risk Officer redirects attention toward risk management by altering the firm's attention structure. In our context, appointing a CSO is expected to reshape the firm's attention structure by elevating the importance of environmental issues.

The attention structure includes roles, individuals in those roles, rules governing the behavior of decision-makers, and available resources (Ocasio 1997). Appointing a CSO, represents a significant change within this structure because it introduces a new role and a new individual to the firm's top management team. This change, whether through informal signals or formal adjustments, sends a message to decision-makers at all levels of the organization. It marks sustainability as integral to the company's overarching strategy which increases the perceived importance of sustainability issues like toxic releases and makes it more likely that decision-makers focus their attention on them (Gupta et al. 2021, Arora et al. 2020). The shift may occur informally, through managerial cues, or formally, via adjustments in performance goals or incentives. A notable example is Boeing's appointment of Mary Armstrong as Vice President of Environment, Health, and Safety in 2007. Following her appointment, Boeing incorporated ambitious five-year environmental targets into its operating units' business plans, including a 25% reduction in toxic releases at major manufacturing facilities (Boeing 05/04/2007, 2008). This case illustrates how

appointing a CSO can reorient facility-level attention toward sustainability objectives, such as reducing toxic emissions.

Taken together, our theoretical discussion suggests that the importance of sustainability issues such as toxic releases might increase following the appointment of a CSO. In turn, local decision-makers might increase their focus on toxic release issues, which makes them more likely to implement measures to reduce toxic releases. We formalize this argument with the following hypothesis:

Hypothesis 1 *Appointing a Chief Sustainability Officer reduces toxic releases.*

4.3.2 Exploring the Moderating Role of Underserved Communities

Attention to sustainability issues also provides insight into why facilities in underserved communities tend to report higher toxic releases than those in better-served communities and how the impact of CSO appointments may vary depending on the type of community surrounding a facility.

The attention-based view suggests that decision-makers focus their attention on issues that they perceive as important, which is shaped not only by internal factors but also by external stakeholders, such as local communities (Ocasio 1997). Such communities can collectively take actions that benefit their well-being (Diaz 2017, Mohai and Saha 2015). This includes leveraging local facilities to reduce their toxic release volume or even persuading firms to avoid establishing facilities in certain neighborhoods (Taylor 1997). However, the effectiveness of such actions depends on socioeconomic factors closely linked to a community's economic and racial composition, with underserved communities often having less influence than better-served ones (Diaz 2017, Mohai and Saha 2015, Taylor 1997).

This disparity has long been recognized. For example, a 1984 report advised California's Waste Management Board that "middle and higher-socioeconomic strata neighborhoods

should not fall at least within the one-mile and five-mile radii” of waste-to-energy facilities because they “possess better resources to effectuate their opposition” (Powell 1984, p.26). Although this report is decades old, contemporary observations still indicate that firms see underserved communities as “the path of least resistance” for hazardous operations (Scurlock 2021, Diaz 2017). Consistent with this logic, emerging evidence suggests that facilities in underserved communities engage less in source reduction and waste management activities, likely reflecting lower managerial attention to these issues (Cole and Muthulingam 2025, Shubham and Subramanian 2025).

Importantly, this historical lack of attention may also explain how the impact of CSO appointments may vary depending on the type of community surrounding a facility. Facilities in underserved communities that have implemented fewer source reduction and waste management activities in the past may be operating further from their environmental performance frontier, making it easier for them to improve their performance once they focus on toxic release issues. According to the operating frontier theory (Schmenner and Swink 1998), a facility’s operating frontier represents its maximum achievable performance within a given set of operating choices, existing plant design, and investment constraints. Improving performance becomes increasingly difficult the closer a facility gets to this frontier. Thus, facilities with a history of attention to toxic release issues—such as those in better-served communities—may have already implemented easy source reduction activities or waste management processes, making additional improvements more resource-intensive. Conversely, facilities in underserved communities, which have previously paid less attention to these issues, may find it easier and less resource-intensive to reduce toxic releases once their attention shifts—for example, following the appointment of a CSO that redirects attention toward toxic release issues. Taken together, these arguments suggest the following hypothesis:

Hypothesis 2 *Appointing a Chief Sustainability Officer reduces toxic releases more for*

facilities in underserved communities than for those in better-served communities.

If supported, Hypothesis 2 implies that CSO appointments could help narrow disparities in toxic exposure between communities, thereby reducing environmental injustice.

4.4 Data and Empirical Strategy

4.4.1 Data

To test our hypotheses, we require information on various aspects of firms, their environmental impact, and their governance structures. Obtaining and integrating this information presents significant challenges because it is only available from diverse sources without common identifiers. This challenge may have contributed to the previously mixed findings about the environmental effect of appointing a CSO. Here, we detail our main data sources and how we connect them.

First, we need detailed data on toxic releases of individual facilities and the firms to which they belong. We obtain this data from the EPA's Toxic Release Inventory (TRI). This dataset provides information for over 650 chemicals and is a commonly used measure of the environmental impact of manufacturing firms (Muthulingam et al. 2022, Dutt and King 2014). Each observation in the data corresponds to the release quantity of a specific chemical for a specific facility, which allows us to answer our research question at the facility-chemical level. Additionally, the data includes information on facility-level Dun & Bradstreet identifiers, addresses, longitude and latitude, and the parent company's name (e.g., Boeing, Owens Corning, or Exxon Mobil). Second, we identify the presence of Chief Sustainability Officers (CSOs) in our sample firms, using the comprehensive board composition and governance information available in the BoardEx database (Fu et al. 2020). We link these datasets using a link table provided by Xiong and Png (2019) and the Dun & Bradstreet identifiers in the TRI data.

Beyond these core datasets, we also obtain data from Compustat, Thomson/Refinitiv, the U.S. Decennial Census, the U.S. Census American Community Survey, the League of Conservation Voters, and the MIT Election Data and Science Lab to enrich our analysis with additional firm-level and contextual information. We link the Compustat data to the TRI data using the methods described above. We link the TRI data to Thomson/Refinitiv using fuzzy name matching via OpenRefine (the matching protocol and log file are available upon request). Finally, we link all other datasets using facility address information.

We restrict our sample to facilities of publicly traded firms in the contiguous U.S. for which a BoardEx record exists. Our sample period is from 2000 to 2020. Our unit of analysis is the facility-chemical year, which is consistent with prior research using the TRI data (Muthulingam et al. 2022, Dutt and King 2014).

4.4.2 Variables

Below we introduce the variables used in our study (Table C.1 provides an overview).

Dependent Variable

Toxic Releases_{cit}: Consistent with the literature, our main dependent variable is the total toxic releases of chemical c for facility i in year t (Li et al. 2025, Muthulingam et al. 2022, Dutt and King 2014).¹ Toxic releases are the last and least preferable level in the EPA’s waste management hierarchy because they are most harmful to the environment and the communities surrounding a plant (EPA 2023b). To facilitate the interpretation of our coefficients, we measure *Toxic Releases_{cit}* in thousands of pounds. The higher this variable, the more toxic releases a facility produces. We winsorize the variable at 1% on the upper tail to reduce the influence of reporting errors or outliers.

¹We deviate from past research by measuring the outcome in levels and not as a logarithmic transformation. The reason is that—contrary to Muthulingam et al. (2022) and Dutt and King (2014)—we use a two-way fixed effects empirical strategy. By definition, parallel trends for this strategy only hold for one but never both functional forms, and they are more likely to hold for outcomes measured in levels (Roth et al. 2023).

Treatment Variable

After CSO Appointment_{jt}: We measure whether a firm has a CSO using an indicator variable that takes the value of one for every year t after firm j first appoints a CSO, and is zero otherwise. Following Fu et al. (2020), we identify CSO appointments from the BoardEx database based on job titles containing “sustainability,” “sustainable,” “responsibility,” “ethics,” or “environment.” Our theory suggests that creating the CSO role alters a firm’s attention structures, ultimately affecting toxic releases. Therefore, we use the first year of appointment as our treatment (Li et al. 2022). This holds even if firms subsequently change CSOs, operate for a year without a CSO, or integrate the CSO role into another functional area. Our results are robust to only considering years in which a firm has an active CSO (see Section C.9). In our sample, 165 of 598 firms created a CSO role between 2000 and 2020.

Moderating Variables

Our key moderating variable, *Underserved Community_{it}*, identifies neighborhoods with a significant proportion of Black or African American persons. We classify a neighborhood as underserved (*Underserved Community_{it}* = 1) if it falls above the 50th within-state percentile of *Black/African American_{it}*, and as better-served (*Underserved Community_{it}* = 0) otherwise. This threshold aligns with the EPA’s methodology for identifying communities that may require additional environmental policy attention in its Environmental Justice Screening Tool (EPA 2022).² Our results remain robust when applying alternative cutoffs or incorporating economic factors (see Section 4.6.1).

To compute the percentile of *Black/African American_{it}*, we first compute the percentage of persons identifying as only Black or African American within a facility’s host

²The EPA discontinued the EJScreen tool on February 5, 2025, but a restored version is currently available at <https://screening-tools.com/epa-ejscreen>.

neighborhood as:

$$\% \text{ Black/African American}_{it} = \frac{\text{Black or African American}_{it}}{\text{Total Population}_{it}}. \quad (4.1)$$

We then follow EPA guidelines (EPA 2022) and convert the raw percentages (ranging from 0% to 100%) into percentiles based on the state’s distribution of $\% \text{ Black/African American}_{it}$. This approach facilitates interpretation and reflects local demographic differences between states (Shubham and Subramanian 2025).

One challenge with computing these measures is selecting an appropriate host neighborhood (Mohai and Saha 2015). A common approach is to use the characteristics of a facility’s geographical host unit, such as state, county, zip code area, or census tract (Shubham and Subramanian 2025). However, these methods fail to account for a facility’s exact location within its host unit or its proximity to neighboring units (Mohai and Saha 2015, Chakraborty et al. 2011). To address this limitation, we use an areal apportionment method, defining the host neighborhood as a 1-mile radius around each facility (Mohai and Saha 2015, 2007). We estimate the demographic characteristics within this radius by computing the weighted average of the census block groups intersecting the area. The demographic data is sourced from the 2000 Decennial U.S. Census, the 2010-2014 American Community Survey (ACS), and the 2018-2022 ACS. A detailed explanation of this procedure is provided in Section C.2 of the E-Companion.

Control Variables

Production Ratio_{cit}: Consistent with prior research (Muthulingam et al. 2022, Dutt and King 2014), we account for changes in a facility’s production volume. Toxic releases, being a byproduct of manufacturing activity, vary following changes in the production volume of the underlying product. For instance, if the production volume of product p , which results in the release of chemical c , decreases by 10%, we expect a corresponding 10% decrease in

toxic releases of chemical c . To account for this, TRI reports the *Production Ratio* $_{cit}$ as the ratio of the production in year t to the production in year $t - 1$. We winsorize the variable at 1% on both tails to reduce the influence of reporting errors or outliers.

Toxicity Ratio $_{cit}$: Following prior work (Muthulingam et al. 2022, Dutt and King 2014), we control for changes in a facility’s toxin intensity. To compute this variable, we create a toxicity-weighted sum of all chemical releases by facility i in year t . We use the toxicity rating for each chemical from the EPA’s Risk Screening Environmental Indicators (RSEI) model as weights. Then, we compute the toxicity ratio for a given facility in year t as the natural logarithm of its toxicity-weighted chemical releases in that year divided by its toxicity-weighted chemical releases in the previous year.

Tenure Certifying Official $_{cit}$: In line with Dutt and King (2014), we account for the tenure (in years) of the certifying official as continuity in this position might help facilities reduce their toxic releases (Muthulingam et al. 2022). The certifying official is the individual who signed the TRI submission form for a specific facility and chemical, certifying its accuracy and completeness. We compute this variable by counting the number of successive years in which the same official signs the TRI submission for chemical c in facility i . A higher variable value indicates more years with the same official submitting the information.

Tenure Technical Person $_{cit}$: Following Dutt and King (2014), we control for the tenure (in years) of the technical person associated with the chemical c at facility i . The technical person is the contact person in the TRI data (Muthulingam et al. 2022). Similar to the tenure of the certifying official, we compute the variable by counting the number of subsequent years in which the TRI submission lists the same contact person.

Environmental Management System $_{it}$: Another potential influence on toxic releases is the presence of an environmental management system (EMS). The EPA defines an EMS as a framework that “helps an organization achieve its environmental goals through consistent review, evaluation, and improvement of its environmental performance.” (EPA 2023a). We code *Environmental Management System* $_{it}$ as a binary variable that takes the value of one

if facility i uses an EMS in year t . Following Muthulingam et al. (2022), we infer the usage of an EMS from reported source reduction activities in the last three years.

Regulatory Pressure_{it}: Prior research on firm governance and toxic releases suggests that regulatory pressure might influence the impact of CSOs, who become more effective if firms face strict environmental regulations (Kanashiro and Rivera 2019). Therefore, we control for regulatory pressure, measured as the extent to which U.S. House of Representatives members from the state in which facility i is located vote in favor of more stringent environmental regulations in year t . This data comes from the League of Conservation Voters (LCV), and it measures environmental regulatory pressure (Kanashiro and Rivera 2019, Delmas and Toffel 2008, Kassinis and Vafeas 2002).

Republican Vote in County_{it}: Research shows that Republican voters are generally less supportive of environmental causes (Dunlap and McCright 2008, Dunlap et al. 2001). Therefore, the political orientation of those living around and working at a facility might influence the issues that decision-makers focus on. We control for this effect using the Republican vote in a facility's county, measured as the percentage of Republican voters in presidential elections, which we collect from the MIT Election Data and Science Lab.

Total Population_{it}: The number of people living in a facility's host neighborhood might influence its toxic releases. Facilities in sparsely populated neighborhoods might have higher releases than those in densely populated neighborhoods (Shubham and Subramanian 2025). Therefore, we include the total population in the 1-mile radius host neighborhood as a control.

Black/African American_{it}: Because local communities influence toxic releases, we also account for their direct influence on toxic releases by including the percentile of Black or African American persons around facility i in year t in the analysis.

4.4.3 Primary Empirical Strategy

We use a two-way fixed effects (TWFE) regression approach with matching as our primary empirical strategy to estimate the effect that appointing a CSO has on toxic releases. In Section 4.6 we also use other empirical strategies to validate the robustness of our results.

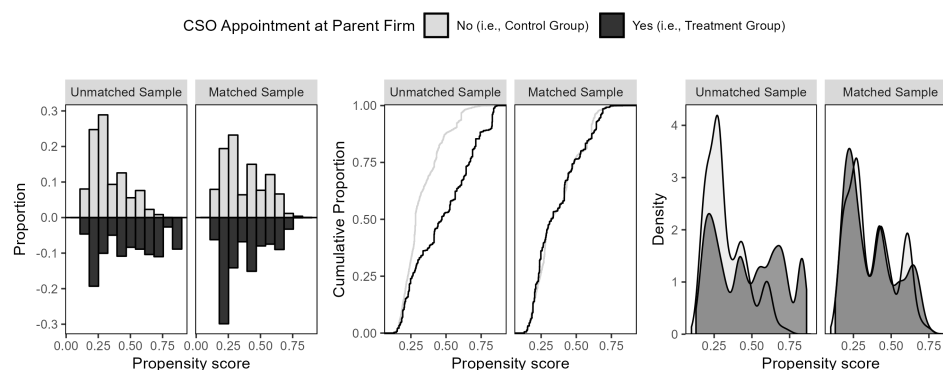
The TWFE with matching approach requires data from two groups: a treatment group and a control group. In our case, the treatment group is the facility-chemical pairs belonging to the 165 parent companies that create a CSO role between 2000 and 2020. The control group is the facility-chemical pairs belonging to the parent companies that do not create a CSO role during our sample period.

An inherent challenge in our context is the endogeneity of firm governance choices (Shaver 1998). Firms decide to create a CSO role with the expectation that the CSO will impact their environmental performance in a way that differentiates them from firms that choose not to create a CSO role. Following recent empirical operations management studies (Calvo et al. 2023, Dhanorkar et al. 2023, Ertekin et al. 2021), we account for this selection into treatment within the TWFE framework by creating a matched sample of facility-chemical pairs in the treatment and control groups. Specifically, we conduct matching to create balanced comparison groups, mitigating potential biases arising from the non-random assignment of CSOs to firms.

To implement the matching, we first compute a propensity score that predicts the conditional probability of appointing a CSO based on the parent firm’s (i) Refinitiv’s ESG score (a score from “A+” to “D-” that combines information about environmental, social, and corporate governance performance), (ii) Refinitiv’s emissions score (a score from “A+” to “D-“ that evaluates a firm’s emissions), (iii) Refinitiv’s controversies score (a score from “A+” to “D-“ that evaluates a firm’s ESG-related controversies), and (iv) size (measured as the natural logarithm of total assets). We then conduct coarsened exact matching (CEM) using this propensity score, *Total Population*, *Black/African American*,

and *Regulatory Pressure*, while exactly matching on the type of chemical (e.g., Ethylene to Ethylene and Chlorine to Chlorine) and the type of host neighborhood community. This accounts for any systematic differences between chemicals, such as their toxicity, or between facilities in different types of communities. Our results are robust to using a variety of matching settings and variables (see Section 4.6.1). Figure 4.2 compares the propensity score density distributions for the treated and control groups (see Figures C.2 and C.3 in the E-Companion for variable balances).

Figure 4.2: Balance Plots Displaying the Likelihood to Appoint a CSO



The figure plots the distribution of the likelihood (i.e., the propensity score) of parent firms appointing a CSO for facility-chemical pairs in the treatment and control groups. The left sub-figure displays a histogram of the distributional densities, the middle sub-figure displays the empirical cumulative distribution function, and the right sub-figure displays overlapping density plots. Within each sub-figure, the left display titled “Unmatched Sample” shows the balance of the sample before treatment, and the right display titled “Matched Sample” shows the balances after matching. Together, the plots indicate that facility-chemical pairs in the unmatched sample for which the firm appoints a CSO, on average, have a higher likelihood of appointing a CSO than facility-chemical pairs in the control group. These differences disappear in the “Matched Sample.”

Figure 4.2 shows that there are two strong concentrations in the unmatched sample. Many facility-chemical pairs for which the parent firm never appoints a CSO have a low propensity to do so, as evident by the spike in the white distribution around 0.25. Conversely, those facility-chemical pairs for which the parent firm eventually appoints a CSO exhibit a high propensity to do so, indicated by the spike in the dark distribution around 0.75. This imbalance disappears in our matched sample, which includes a majority

of facility-chemical pairs with a low to medium propensity to appoint a CSO. Our matched sample includes 125,878 facility-chemical year observations.

For the TWFE approach itself, we use the matched sample and compare the difference in toxic releases of facility-chemical pairs in the treatment group before and after appointing a CSO to the difference in toxic releases of the facility-chemical pairs in our control group. Specifically, we estimate the following model:

$$Toxic\ Releases_{cit} = \delta * After\ CSO\ Appointment_{jt} + \beta * X_{cit} + \alpha_{ci} + \gamma_t + \lambda_j + \epsilon_{cijt}, \quad (4.2)$$

where $Toxic\ Releases_{cit}$ is our measure of chemical-facility level environmental performance. $After\ CSO\ Appointment_{jt}$ is a treatment indicator that equals one after the appointment of a CSO for any facility-chemical pair belonging to firm j , and is zero before. X_{cit} is a vector of control variables, α_{ci} are chemical-facility fixed effects, γ_t are year fixed effects, λ_j are firm fixed effects, and ϵ_{cijt} is the error term. We cluster the standard errors at the facility level. A negative estimate of δ would support Hypothesis 1, that appointing a CSO reduces toxic releases. We test whether appointing a CSO has a disproportional effect on toxic releases at facilities in underserved communities (relative to those in better-served communities) by splitting the sample into two sub-samples, one for facilities in underserved communities ($Underserved\ Community_{it}=1$) and one for facilities in better-served communities ($Underserved\ Community_{it}=0$). A more negative estimate of δ in the underserved communities sub-sample would support Hypothesis 2, that appointing a CSO reduces toxic releases more for facilities in underserved communities.

4.5 Results

4.5.1 Model-Free Evidence

Table 4.1 presents descriptive statistics and correlations for the matched sample. Notably, *Black/African American* is positively correlated with toxic releases ($0.02; p < 0.05$), suggesting that facilities in underserved communities pollute more relative to those in better-served communities.

Table 4.1: Descriptive Statistics and Correlations of Matched Sample

Variable	Mean	SD	Min	Max	1	2	3	4	5	6	7	8	9	10	11
1. Toxic Releases (in 1,000s lb)	22.42	75.24	0.00	466.68	.										
2. After CSO Appointment	0.18	0.38	0.00	1.00	.04	.									
3. Production Ratio	1.06	0.46	0.00	4.81	-.01	-.04	.								
4. Toxicity Ratio	0.88	1.12	0.00	18.52	-.01	-.03	.09	.							
5. Tenure of Cert. Official	3.21	2.62	1.00	21.00	.00	.11	-.03	-.04	.						
6. Tenure of Tech. Person	3.92	3.39	1.00	21.00	.02	.06	-.05	-.05	.43	.					
7. EMS	0.13	0.33	0.00	1.00	-.02	-.04	-.01	-.02	-.03	-.03	.				
8. Regulatory Pressure	39.56	18.93	0.00	100.00	-.06	-.03	-.01	-.01	-.02	-.03	.03	.			
9. Republican Vote in County	52.70	14.74	0.00	90.13	.08	-.04	.00	.01	.02	.02	.00	-.38	.		
10. Total Population (in 100s)	55.85	74.47	0.01	673.18	.03	.02	-.01	-.01	-.01	-.01	.00	-.05	.01	.	
11. Black/African American	62.31	20.88	2.00	100.00	.02	.03	-.01	.00	.01	.00	.01	-.01	.03	.27	.

Note. These summary statistics are based on the matched sample used for the main analyses. Bolded values indicate a significance level of < 0.05 .

To further underscore the presence of environmental injustice and evaluate the effect of appointing a CSO, we next present model-free evidence. Table 4.2 shows the average toxic releases of facilities in underserved and better-served communities before and after the appointment of a CSO for firms that eventually appoint a CSO. Each cell in the table represents the average toxic release for a specific subgroup. As seen in Columns (7) and (8), facilities in underserved communities tend to have higher toxic releases, further pointing to the existence of environmental injustice.

Furthermore, facilities in both underserved and better-served communities have lower toxic releases following the appointment of a CSO (Columns 3 and 6), supporting Hypothesis 1. However, the relationship is stronger for facilities in underserved communities, leading to an overall decrease in the inequality of toxic releases (Column 9). This indicates

a reduction in environmental injustice and supports Hypothesis 2.

Table 4.2: Model-Free Evidence of the Effect of CSO Appointments on Toxic Releases

	Better-Served Communities			Underserved Communities			Differences between Communities		
	Pre (1)	Post (2)	Post-Pre (3)	Pre (4)	Post (5)	Post-Pre (6)	Pre (7)	Post (8)	Post-Pre (9)
Toxic Releases (in 1,000s lb)	30.91	25.24	-5.67	40.00	29.58	-10.42	9.09	4.34	-4.75

Note. The table presents model-free evidence on the effectiveness of appointing a CSO on toxic release. It reports the mean values and differences for facilities of firms that eventually appoint a CSO. We do not include statistics for facilities of firms that do not appoint a CSO because no pre- and post-period exists for these firms. The positive differences in columns (7), calculated as (4)-(1), and (8), calculated as (5)-(2), support the environmental injustice claim by highlighting that facilities in underserved communities omit more toxic releases. The negative difference of differences in column (9), calculated as (8)-(7), indicates that following the appointment of a CSO, the toxic releases of facilities in underserved communities become closer to those of facilities in better-served communities, reducing environmental injustice.

4.5.2 Main Results

While Table 4.2 provides preliminary evidence of the effectiveness of appointing a CSO, it has a notable limitation. It solely compares toxic releases before and after the appointment within facilities that eventually appoint a CSO, potentially overlooking broader trends in toxic release reduction over time. For instance, if facilities had generally reduced their toxic releases since 2000, then we would observe similar trends in Columns (1) through (6) even if appointing a CSO had no effect. Similarly, if facilities in underserved communities with high levels of toxic releases were simply ripe for improvement, then we would observe similar trends in Columns (9) even if appointing a CSO had no effect.

We address these limitations with our TWFE identification strategy, in which we use the matched sample of facilities that do not appoint a CSO as a control group. This approach allows us to control for differences in the volumes of toxic release and any general time trends in toxic release reduction. Table 4.3 displays the results of this analysis. Column (1) is our baseline model, which includes only the treatment variable, plus the facility-chemical and year fixed effects. In column (2), we add the control variables, and in column (3)

we add firm fixed effects. Columns (4) and (5) present the split sample analysis for facilities in better-served and underserved communities, respectively. In all models, the coefficient for *After CSO Appointment*_{jt} represents the effect of appointing a CSO for those facility-chemical pairs in our matched sample where the parent firm created a CSO role. We focus on the results in columns (3) to (5) in discussing the results.

Table 4.3: TWFE Results of the Effect of CSO Appointments on Toxic Releases

DV: Toxic Releases (in 1,000s lb)	Hypothesis 1			Hypothesis 2 [†]	
	(1)	(2)	(3)	(4)	(5)
After CSO Appointment	-4.601** (1.444)	-4.589** (1.425)	-4.582** (1.428)	-1.046 (1.692)	-6.352*** (1.910)
Production Ratio		1.547*** (0.359)	1.547*** (0.359)	0.941* (0.403)	1.799*** (0.474)
Toxicity Ratio		0.424*** (0.107)	0.422*** (0.107)	0.319 (0.208)	0.461*** (0.127)
Tenure of Certifying Official		0.450*** (0.119)	0.450*** (0.120)	0.248 (0.186)	0.517*** (0.148)
Tenure of Technical Person		0.005 (0.100)	0.005 (0.100)	0.037 (0.205)	-0.007 (0.110)
Environmental Management System		0.812 (0.822)	0.813 (0.824)	0.262 (1.078)	1.000 (1.086)
Regulatory Pressure		0.022 (0.040)	0.022 (0.040)	-0.026 (0.052)	0.038 (0.051)
Republican Vote in County		0.076 (0.064)	0.076 (0.064)	-0.162 (0.100)	0.160* (0.077)
Total Population (in 100s)		-0.119 (0.096)	-0.119 (0.096)	-0.012 (0.162)	-0.131 (0.106)
Black/African American		0.061 (0.083)	0.060 (0.083)	0.048 (0.087)	-0.055 (0.147)
Facility-Chemical Pair FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Firm FE	No	No	Yes	Yes	Yes
Sample	All	All	All	Better-Served Community	Underserved Community
Num.Obs.	125,878	125,878	125,878	43,412	82,466
R ²	0.852	0.852	0.852	0.876	0.843

Note. The table presents the results of a two-way fixed effects regression with toxic releases in 1000 pounds as the dependent variable. The unit of analysis is facility-chemical-year. The coefficient for After CSO Appointment, a binary treatment indicator, in column (3) corresponds to a 12.3 percent reduction in toxic releases compared to the average level of the treated group before CSO appointment. Clustered standard errors in parentheses. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

†: The coefficients for After CSO Appointment in columns (4) and (5) are significantly different with a z-score of 2.079 ($p = 0.038$).

Examining the control variables, we find that *Black/African American* is not positively

correlated with toxic releases, which contrasts with the correlation in Table 4.1. This difference arises because Table 4.3 includes facility-chemical fixed effects, which control for differences across facilities and focus on variation within facilities over time. These findings are consistent with environmental injustice research that shows that while demographic variation *between* facilities is linked to differences in toxic releases, there is little demographic change *within* a facility over time (Mohai and Saha 2015, Pastor et al. 2001).

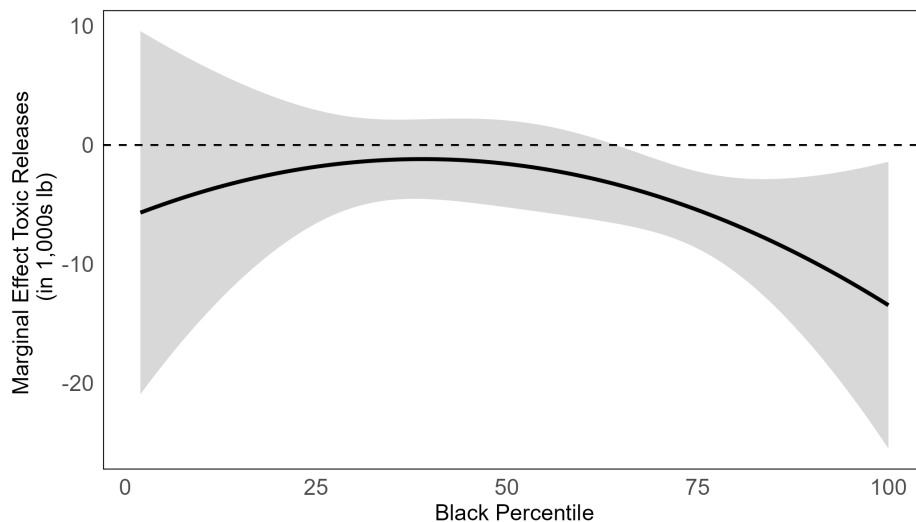
Hypothesis 1 predicts that appointing a Chief Sustainability Officer reduces toxic releases. Table 4.3 shows that facilities belonging to firms that create a CSO role subsequently reduce their toxic releases. Specifically, column (3) indicates that appointing a CSO leads to a reduction of -4.582 ($p < 0.01$) in toxic releases—a 4,582 pounds or 12.3% reduction from average toxic release levels before CSO appointment. Thus, we find support for Hypothesis 1.

Next, we examine the effect of CSO appointments on environmental injustice. Hypothesis 2 predicts that appointing a CSO decreases toxic releases more for facilities in underserved communities than for those in better-served communities, leading to reduced environmental injustice. The results in Table 4.3 support this hypothesis. The results in columns (4) and (5) show that facilities in better-served communities reduce their toxic releases to a lesser degree than those in underserved communities (i.e., $-1.046; p > 0.05$ vs. $-6.352; p < 0.001$). The difference between both coefficients is statistically significant ($p < 0.05$).

To facilitate the interpretation of this result, we plot the marginal effects of a regression in which we interact $After\ CSO\ Appointment_{jt}$ with $Black/African\ American$ and $(Black/African\ American)^2$ in Figure 4.3. This shows that the marginal effect of CSO appointments is not significantly different from zero for facilities with low percentiles for $Black/African\ American$. However, as $Black/African\ American$ increases, the decrease in toxic releases due to CSO appointments becomes more pronounced. Together, these results indicate that CSO appointments contribute to reducing environmental injustice by being

more effective for facilities in underserved communities.

Figure 4.3: Marginal Effect Plot for the Moderating Effect of *Black/African American*



The figure plots the marginal effects of appointing a CSO conditional on *Black/African American*. The regression specification is:

$$Toxic\ Releases_{cit} = \delta * After\ CSO\ Appointment_{jt} + \delta_{m1} * After\ CSO\ Appointment_{jt} * Black/AfricanAmerican_{it} + \delta_{m2} * After\ CSO\ Appointment_{jt} * (Black/AfricanAmerican_{it})^2 + (Black/AfricanAmerican_{it})^2 + \beta * X_{cit} + \alpha_{ci} + \gamma_t + \lambda_j + \epsilon_{cijt}.$$

4.6 Robustness Checks and Mechanism Analyses

We conduct additional analyses to further bolster our results. First, we conduct four robustness checks to validate our findings. Second, we considered several alternative explanations for the reduction in toxic releases following CSO appointments, and show that our results are unlikely to be an artifact of these other factors. Third, we explore two potential mechanisms through which appointing a CSO might influence toxic releases: source reduction activities and waste management efforts.

4.6.1 Robustness Checks

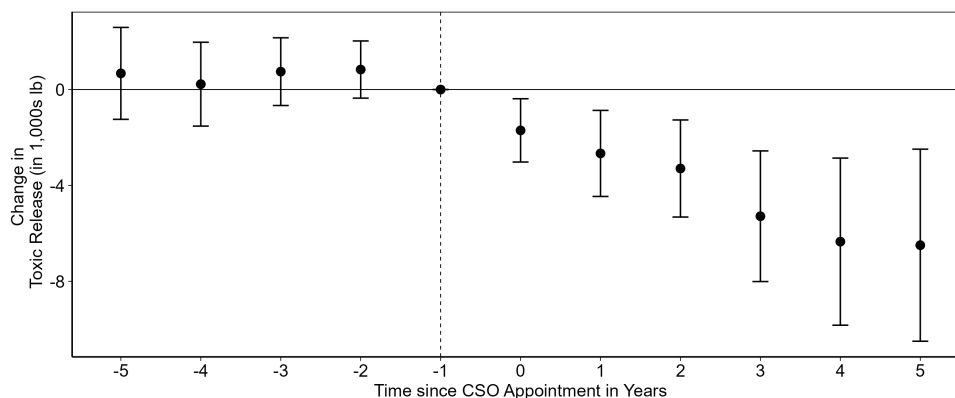
We conduct four robustness checks to validate our findings. First, we replicate our main analysis using Sun and Abraham’s (2021) estimator. This allows us to evaluate the parallel trends assumption and explicitly account for the time variation of CSO appointments across firms (i.e., staggered treatment timing). Second, we repeat our main analysis with alternative definitions of *Underserved Community*_{it}. Third, we run a specification analysis to ensure the robustness of our findings across different matching settings. Finally, we conduct a placebo analysis to determine the likelihood of our results being driven by chance.

Parallel Trends and Staggered Treatment Timing

Not all firms appoint CSOs at the same time. On the one hand, this staggered (or differential) treatment timing alleviates any concerns that a shock that happened in the same year as the treatment drives our results. On the other hand, econometricians have recently shown that the TWFE model—our main identification strategy—could be biased under staggered treatment timing (Roth et al. 2023, Sun and Abraham 2021). To address this concern, we use a dynamic treatment effects event study that corrects for staggered treatment timing (Sun and Abraham 2021). Specifically, Sun and Abraham’s (2021) estimator accounts for possible heterogeneity between different CSO appointment years. This method entails interacting CSO appointment year indicators with dummy variables that indicate the relative time before and after the appointment of a CSO. It then creates weighted averages of the treatment effects based on the sample shares of each CSO appointment year cohort. We present the results as an event plot (Figure 4.4) and regression table (Table 4.4).

Focusing on the time before the CSO appointment in Figure 4.4 allows us to evaluate the parallel trends assumption that underpins the validity of our identification strategy. The parallel trends assumption means that in the absence of a CSO appointment, the toxic

Figure 4.4: Event Plot of the Effect of CSO Appointments on Toxic Releases



This figure shows the dynamic treatment effects for five years before and five years after the appointment of a Chief Sustainability Officer (CSO). The y-axis reports coefficients and 95% confidence intervals for toxic releases (in 1,000 pounds). The x-axis shows the relative years from the CSO appointment with negative values indicating years before appointment.

releases of facility-chemical pairs in the treatment group would have followed a similar trend as those of facility-chemical pairs in the control group. While the assumption cannot be tested for the post-treatment period, we can evaluate its validity in the pre-treatment period. Consistent with the parallel trends assumption, the estimates for the years before appointing a CSO appear to develop in parallel. We also find no significant pre-treatment trends in a linear model of the pre-appointment periods (see Table C.2 in the E-Companion). Together, the visual inspection and linear regression indicate that before the appointment of a CSO, the trends for treated facility-chemical pairs where the parent firm appoints a CSO at event-time 0 are not different from those facility-chemical pairs in the control group. This instills confidence that, in the absence of a CSO appointment, the trends for the two groups would have been parallel in the post-treatment period.

For the post-treatment period, the event study plot illustrates a downward trend in toxic releases following the appointment of a CSO. This supports our theoretical arguments by indicating that the effect of CSO appointments on toxic releases is gradual rather than immediate. Table 4.4 presents the estimator's aggregated coefficients. By showing

that appointing a CSO remains associated with a significant decrease in toxic releases ($-3.933; p < 0.001$), even after explicitly accounting for the staggered treatment timing, the results support Hypothesis 1. The effect remains stronger for facilities in underserved communities ($-5.235; p < 0.001$) than for those in better-served communities ($-1.507; p > 0.05$), supporting Hypothesis 2.

Table 4.4: Dynamic Treatment Effect of CSO Appointments on Toxic Releases

DV: Toxic Releases (in 1,000s lb)	(1)	(2)	(3)
After CSO Appointment	-3.933^{***} (1.060)	-1.507 (1.111)	-5.235^{***} (1.438)
Controls	Yes	Yes	Yes
Facility-Chemical Pair FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes
Sample	All	Better-Served Community	Underserved Community
Num.Obs.	100,666	35,464	65,202
R^2	0.878	0.904	0.868

Note. The table presents the results of a dynamic treatment effects event study estimation following Sun and Abraham (2021) with toxic releases as the dependent variable. The controls include all variables from Section 4.4.2. Clustered standard errors in parentheses. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Alternative Measures of Underserved Communities

Next, we validate the robustness of our results to different definitions of underserved communities. Instead of relying solely on the share of Black or African American individuals within a 1-mile radius, we account for both racial and income characteristics. Additionally, we vary the host neighborhood radius (i.e., one and three miles) and cutoff values (i.e., 50th and 80th percentile) to classify underserved communities. Across these alternative specifications, our main results hold: appointing a CSO is associated with a reduction in toxic releases, with stronger effects observed in underserved communities (E-Companion Table C.3).

Alternative Matching Specifications

Our empirical strategy involves creating a matched sample of facility-chemical pairs in the treatment and control groups to account for the decision to appoint a CSO. While this practice follows recent empirical research in operations management, it raises the question of whether our matching settings drive our results. We therefore test the robustness of our findings using a specification analysis in which we re-run our analysis using 40 different matching settings and matching variables. Figure C.4 in the E-Companion shows that our results are robust across various matching settings, bolstering confidence in our findings.

Placebo Analysis

The staggered appointment of CSOs in our data alleviates concerns that an industry-wide shock that happened in the same year as the appointment drives our results. We implement a placebo analysis to further reduce concerns that our results might be driven by chance (Dhanorkar and Muthulingam 2020, Dhanorkar 2019). In this analysis, we assign observations to the treatment or control group by randomly allocating a CSO appointment year at the firm level. Figure C.6 in the E-Companion displays the distribution of coefficients for ‘After CSO Appointment’ that we obtain by re-estimating Equation (4.2) in 500 pseudo-samples. Our main estimate from column (3) in Table 4.3 is the 34th largest negative value, yielding a p-value of 0.068. The estimated effect of appointing a CSO for facilities in underserved communities (Column 5, Table 4.3) is the 18th largest negative value, giving a p-value of 0.036. Together, these results indicate that our results are unlikely due to chance.

4.6.2 Alternative Explanations

Our study investigates whether appointing a Chief Sustainability Officer (CSO) leads to reductions in toxic releases, particularly in underserved communities. We argue that a

CSO appointment shifts a firm’s attention toward sustainability, which, in turn, enhances environmental performance. Crucially, we do not claim that CSOs actively target reductions in environmental injustice; rather, these reductions emerge as a byproduct of (1) increased attention to the issue of toxic releases and (2) the fact that facilities in underserved communities have greater potential for emission reductions when sustainability becomes a priority because they are further from their operating frontier. However, the findings could also be explained by other explanations. We consider three such alternative explanations below.

Shifts in a firm’s environmental priority. First, toxic release reductions may result from a broader shift in a firm’s environmental priorities—potentially driven by regulatory pressures or public scrutiny—rather than the CSO appointment itself. In this scenario, appointing a CSO would be part of a larger strategic effort rather than the primary driver of improved environmental outcomes. While we account for this possibility in the matching step by including ESG score and ESG controversies as matching variables, we further test this explanation here by employing an instrumental variable approach that isolates exogenous variation in CSO appointments. Specifically, we instrument CSO appointments using the proportion of directors serving on boards of other firms with a CSO (e.g., Ambulkar et al. 2023, Wowak et al. 2021) (see Section C.9 for a detailed explanation of this procedure and the validity of the instrument). If toxic release reductions were primarily driven by broader environmental initiatives, IV estimates would differ significantly from our main results. However, as shown in Table C.4, the CSO coefficient remains negative and significant ($-9.266; p < 0.001$), with an even stronger effect in underserved communities ($-11.446; p < 0.001$). This supports the argument that CSO appointments independently contribute to toxic release reductions.

Non uniqueness of CSO effect. Second, even if the reductions are not due to a broader shift in environmental priorities, they may not be unique to CSO appointments. Other firm-level interventions designed to increase environmental awareness—such as forming

an ESG committee (Bifulco et al. 2023, Mackenzie 2007)—could yield similar results. ESG committees oversee sustainability policies, monitor compliance, and assess CSR risks (Mackenzie 2007), potentially influencing facility-level decision-making similarly to a CSO. However, ESG committees operate at a higher governance level and should have a weaker effect. To test this, we repeat our analysis using the formation of an ESG committee as the treatment condition. If ESG committees had a comparable impact, it would suggest that CSOs are not uniquely responsible for toxic release reductions. However, as shown in Table C.5 in the E-Companion, ESG committees are not associated with reductions in toxic releases—regardless of community type. These findings underscore that not all governance interventions designed to heighten sustainability awareness produce the same effects and reinforce the distinct role of CSOs.

Outsourcing production of harmful substances. Third, firms might achieve toxic release reductions by decreasing production volume or outsourcing operations rather than improving environmental practices. For example, in 2015, DuPont sold a Chloroprene-producing facility to Denka shortly after the EPA classified Chloroprene as a carcinogen (Hersher 2018). To control for production-related shifts, our models include *Production Ratio_{cit}* as a control variable. Additionally, to test whether outsourcing or facility closures drive our results, we conduct an alternative analysis using *Production Ratio_{cit}* as the dependent variable. If firms were merely shifting production elsewhere, we would expect a negative and significant coefficient for *After CSO Appointment_{jt}*. However, as shown in Table C.6, this is not the case, indicating that our findings are not driven by production displacement.

By addressing these alternative explanations, we improve our understanding of how CSO appointments influence environmental outcomes. While our main findings suggest that CSOs play a distinct role in reducing toxic releases, we take further steps to isolate their effect by accounting for alternative explanations. These additional analyses bolster our confidence that appointing a CSO contributes directly to reducing toxic releases, particularly

in underserved communities.

4.6.3 Mechanism Analysis

Next, we examine the mechanisms through which appointing a CSO affects toxic releases. Our theoretical framework suggests that a CSO appointment increases decision-makers' attention to toxic release issues, prompting them to reduce toxic releases through either source reduction activities, waste management activities, or both (see Figure 4.1).

Increase in Source Reduction Activities

Source reduction activities involve modifications to products, raw materials, or production processes that decrease the total waste generated during production. For example, Safran Landing Systems, an aircraft parts manufacturer in Kentucky, reduced its ethylene releases by optimizing pipe-cleaning processes (EPA 2018). Thus, facilities aiming to reduce their toxic releases must actively alter their production or management processes. We test this mechanism using information from the TRI Pollution Prevention reporting (Muthulingam et al. 2022, Dutt and King 2014).

Facilities can voluntarily report the types of activities they undertake to reduce toxic releases (Dutt and King 2014). Using this data, we construct a new dependent variable, *Source Reduction Activities_{cit}*, which represents the number of source reduction activities implemented by facility *i* for chemical *c* in year *t*, ranging from 0 to 4. We use this variable to test whether facilities adopt more source reduction activities after a CSO appointment.

A challenge with this data is that reporting source reduction activities is not mandatory. Consequently, missing reports may indicate either a lack of implementation (true zeros) or a failure to report (structural zeros). Despite this limitation, prior research successfully utilizes this data in toxic release studies (Muthulingam et al. 2022, Dutt and King 2014). To address this issue, we apply a two-stage Heckman selection model (Certo et al. 2016, Heckman 1976). First, we estimate the probability of a facility reporting

pollution prevention activities using a probit regression and compute the Inverse Mills Ratio. Following Muthulingam et al. (2022), we assume that facility j is likely to report source reduction activities if it has an environmental management system (i.e., if it has reported pollution prevention information for any chemical in the past three years). If so, a zero value for *Source Reduction Activities*_{cit} reflects a true zero; otherwise, it is considered a structural zero. We then include the Inverse Mills Ratio in our outcome regression to correct for selection bias (Certo et al. 2016, Heckman 1976). Table 4.5, Panel (a), presents the results of this analysis.

The results indicate that facilities within firms that appoint a CSO implement more source reduction activities (0.036; $p < 0.01$). Split-sample analysis reveals that this effect is driven by facilities in underserved communities. Specifically, in better-served communities, CSO appointments have no statistically significant impact on source reduction activities (Column 2). In contrast, facilities in underserved communities exhibit a significant increase in source reduction activities after a CSO appointment (0.045; $p < 0.01$), an effect approximately one-third larger than that observed in the full sample. These findings suggest that facilities in underserved communities become more attentive to toxic release issues following a CSO appointment.

Increase in Waste Management Activities

Beyond source reduction activities, facilities can also reduce toxic releases through waste management activities (Li et al. 2025, Fu et al. 2019). While source reduction focuses on minimizing waste at the point of generation, waste management addresses pollutants after they have been produced. These activities include treatment, recycling, or other processes designed to prevent waste from entering the environment (see Figure 4.1). Facilities employ a range of technologies and practices to achieve this, such as chemical treatments that neutralize hazardous substances, filtration systems that capture airborne pollutants, and recycling initiatives that repurpose waste materials. For example, Owens Corning reduced

Table 4.5: Mechanisms through which CSO appointments affect toxic releases

	(a)			(b)		
	Source Reduction Activities			Waste Management (in 1,000s lb)		
	(1)	(2)	(3)	(4)	(5)	(6)
After CSO Appointment	0.036** (0.013)	0.015 (0.022)	0.045** (0.016)	11.236** (4.189)	2.029 (4.135)	16.040** (5.605)
Inverse Mills Ratio [†]	-1.733*** (0.254)	-1.538*** (0.457)	-1.814*** (0.305)			
Total Waste [‡]				0.291*** (0.020)	0.285*** (0.027)	0.294*** (0.026)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Facility-Chemical Pair FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Sample	All	Better-Served Community	Underserved Community	All	Better-Served Community	Underserved Community
Num.Obs.	125,878	43,412	82,466	125,878	43,412	82,466
R ²	0.522	0.458	0.544	0.895	0.922	0.883

Note. The table presents the results of a two-way fixed effects estimation with the (a) count of implemented source reduction activities or (b) the toxic waste that is managed as the dependent variables. The unit of analysis is facility-chemical-year. The controls include: Production Ratio, Toxicity Ratio, Tenure of Certifying Official, Tenure of Technical Person, Regulatory Pressure, Republican Vote in County, Total Population, and Black/African American; plus EMS for (b). Clustered standard errors in parentheses. * p < 0.05; ** p < 0.01; *** p < 0.001.

†: An inverse mills ratio is included to account for the sample selection problem that not all facilities report source reduction activities in the TRI data. We use firm size as the excluded variable in the first stage.

‡: Building on Li et al. (2025) we account for the total waste generated in evaluating the amount of waste that is managed.

its waste by implementing a new “wet waste recycling” process that allowed the company to recover waste fibers from process water streams (Owens Corning 2024).

To empirically test whether CSO appointments lead to more waste management efforts, we construct a new dependent variable, $Waste\ Management_{cit}$, which captures the amount of waste managed by facility i for chemical c in year t (Li et al. 2025). We substitute this variable for toxic releases in Eq. (4.2) and present the results in Table 4.5, Panel (b).

The findings indicate that CSO appointments are associated with a significant increase in the quantity of managed waste (11.236; $p < 0.01$), suggesting that facilities respond to CSO appointments by investing in additional pollution control technologies and processes. This effect is stronger for facilities in underserved communities (16.040; $p < 0.001$) than for those in better-served communities (2.029; $p > 0.05$).

Together, the findings in Table 4.5 identify two plausible mechanisms through which CSO appointments influence toxic releases. Facilities within firms that appoint a CSO increase both source reduction activities and waste management efforts. Moreover, consistent with the observed stronger reductions in toxic releases among facilities in underserved communities, these effects are more pronounced in these areas. These findings underscore the role of CSOs in driving environmental improvements through both preventive and reactive pollution control mechanisms.

4.7 Discussion and Conclusion

Facilities in underserved communities have higher toxic releases than those in better-served communities (Diaz 2017, Mohai et al. 2009). In this paper, we investigated whether firm governance reforms, specifically the appointment of a Chief Sustainability Officer (CSO), can alleviate this environmental injustice. We examined whether and to what extent CSO appointments affect toxic releases (**RQ 1**) and whether the effect is contingent on the community (i.e., underserved vs. better-served) surrounding the facilities (**RQ 2**).

Our empirical findings are consistent with our theoretical arguments. Toxic releases decrease significantly following the appointment of CSOs, and this effect is more pronounced for facilities in underserved communities, which reduces environmental injustice. Two mechanisms drive these results. Following CSO appointments, facilities in underserved communities implement more source reduction activities and increase the usage of waste management activities. Taken together, our theoretical arguments and empirical evidence advance research and practice in several ways.

First, we examine the effect that appointing a top management team member like a CSO has on *facility-level* outcomes. This is especially relevant because much of the existing work on top management team appointments in general, and CSO appointments in particular, examines firm-level outcomes such as financial performance (Arora et al. 2020, Wiengarten et al. 2017b) or aggregate environmental outcomes (Fu et al. 2020, Kanashiro and Rivera 2019). Among these studies, the study by Kanashiro and Rivera (2019) is the closest to our setting. It examines the relationship between CSO appointments and firm-level toxic releases and finds that CSO appointments are associated with higher toxic releases. In contrast, our more granular outcome measures and our two-way fixed effects identification strategy allow us to reveal that, while firms with poor environmental performance might be more likely to appoint a CSO, the appointment itself improves environmental performance by reducing toxic releases.

Second, we advance research on environmental injustice by providing evidence of the power of firm governance reforms to address it. Most research in the environmental injustice literature focuses on descriptive analysis to understand its existence (Mohai and Saha 2015, Pastor et al. 2001) and applied policy analysis to understand the impact of regulatory interventions (Geltman et al. 2016, O’Neil 2007). In contrast, our analysis builds on the understanding that manufacturing firms significantly contribute to environmental injustice and should therefore be part of the solution. Our results support this argument. We find that CSO appointments not only affect overall environmental

performance but also contribute to environmental justice by being more effective in underserved communities. In doing so, we also contribute to the emerging literature that explores organizational misconduct through an operations management lens (Skilton and Bernardes 2022, Belavina et al. 2020). Specifically, we show that while regulatory interventions—one often-referenced mechanism to control organizational misconduct—seem unsuccessful in this context, governance reforms—another popular control mechanism—can effectively control organizational misconduct in the form of environmental injustice.

These findings about the effect of CSO appointments prompt an interesting question: Does the effect of appointing a CSO differ depending on the characteristics of who fills the office? Factors like CSO gender and race could both have a significant impact on the outcomes. For example, top management team member gender has been shown to influence firm outcomes, with female officers often connected with more socially desirable outcomes such as faster product recalls (Wowak et al. 2021) or a more proactive environmental agenda (Kumar and Paraskevas 2018). Figure C.7 in the E-Companion explores this question by examining toxic release reductions across CSOs of different racial and gender identities. The results suggest that appointments of non-White CSOs are associated with greater reductions in toxic releases compared to their White counterparts, while no significant difference is observed between male and female CSOs. Although this analysis is exploratory and should not be interpreted as causal, it highlights a potentially important direction for future research on the intersection of leadership diversity and environmental outcomes.

Beyond these research contributions, our study has three main practical implications. First, sustainability-conscious managers and stakeholders should consider pushing their firms to create a dedicated CSO role. Our study, together with the existing research on CSOs, reinforces the notion that appointing a CSO improves both environmental and financial performance. This is particularly pertinent since investors are becoming increasingly concerned about the regulatory and legal ramifications of the environmental practices of their portfolio companies. For example, in 2021, investors urged leading

chemical companies to discontinue the production of certain hazardous substances, estimating the potential future liabilities from such chemicals at \$25–40 billion (Jessop 2021). Since investors cannot directly monitor or control the actions of their investee companies, a CSO can serve as an intermediary safeguarding their interests.

Second, our results should encourage policymakers to shift their focus from regulatory interventions to firm governance reforms in order to address environmental injustice. While presidential executive orders have fallen short, our findings suggest that decisions such as appointing a CSO can contribute to a more environmentally just society. Thus, regulators should explore ways to encourage firms to contribute to environmental justice other than requiring government agencies to consider such issues in their actions.

Third, our results contribute to the achievement of the 2030 Agenda for U.N. Sustainable Development Goals to reduce inequality within and among countries.³ By highlighting the potential of firm governance reforms to address inequalities such as environmental injustice, our findings provide policymakers with a potential mechanism to reduce inequality within countries.

One limitation of our study is that we focus on toxic releases of publicly traded U.S. firms because of the availability of data. These firms have a greater propensity to pollute than private firms (Shive and Forster 2020), highlighting the practical significance of our findings. However, the impact of appointing a CSO might differ for private or non-U.S.-based firms. For example, CSOs might be more effective in smaller firms because they can be more involved in day-to-day operations. As such, we caution against generalizing our results to other contexts without carefully considering contextual differences. Future research could explore the relationship in other settings, if more data becomes available. Second, matching before the TWFE regression eliminates many facility-chemical pairs with a high propensity to appoint a CSO from the analysis. This might reduce the generalizability of our findings to facilities with a high likelihood of CSO appointments at the firm level. We address this

³See <https://sdgs.un.org/goals/goal10>

through our instrumental variable analysis in Section C.9, which relies on the unmatched sample and provides similar results. Nevertheless, we encourage caution when generalizing our results. Third, our theory suggests that decision-makers need to focus their attention specifically on toxic release issues to reduce toxic releases. However, focusing on a specific set of issues can come “at the expense of other[s]” (Bouquet et al. 2009), p.113). Whether the improvement in toxic releases comes at the expense of other operational outcomes, such as productivity or worker safety, is outside the scope of our study but offers a promising avenue for future research.

In conclusion, we revisit our initial overarching questions: what mechanisms can help reduce environmental injustice in the form of disproportionate toxic releases in underserved communities? Firm-governance reforms, particularly the decision to appoint a CSO, play a critical role in shaping a firm’s environmental performance. Given that environmental injustice often stems from a lack of attention to environmental issues by facilities in communities that lack the socioeconomic power for change, governance reforms that change the attention structures within firms prove effective in achieving environmental justice. Thus, in a time when institutional protections for human health and the environment are weakening, it appears that Chief Sustainability Officers can inadvertently help protect vulnerable communities against environmental harm.

Chapter 5

Improving Inspection Resource Allocation to Control Organizational Misconduct

5.1 Introduction

Recent advances in machine learning have made it possible to predict organizational outcomes using publicly available data, including retail sales (Cui et al. 2018, Steinker et al. 2017), serious drug recalls (Ball et al. 2025), and organizational misconduct (Johnson et al. 2023, Campbell and Shang 2022, Mejia et al. 2019). Yet while the accuracy of these predictions has been a major focus, an important but less explored aspect is how they can be used to improve operational decision-making (Cui et al. 2018, Steinker et al. 2017).

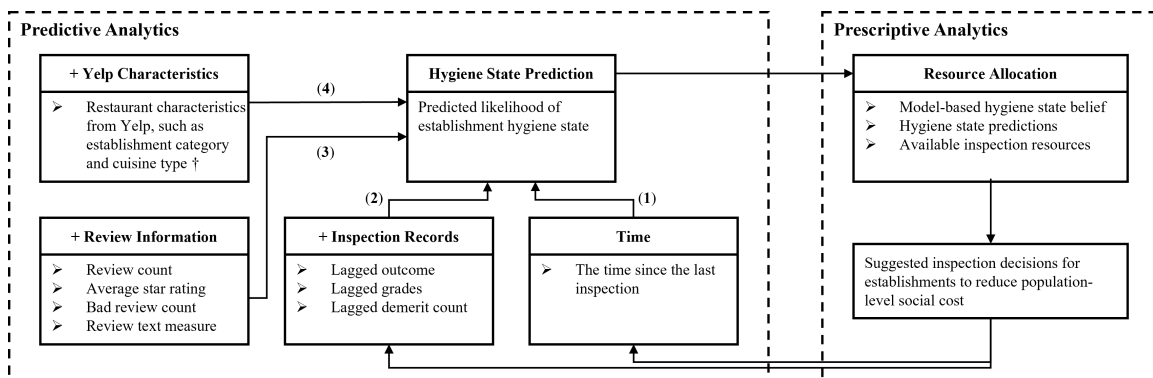
The question of how predictions might improve decision-making is especially important for regulators aiming to prevent misconduct through inspections (Johnson et al. 2023). Consider the case of restaurant hygiene violations (Mejia et al. 2019): local health departments conduct inspections to detect and deter hygiene violations, as non-conformance

with hygiene regulations contributes to tens of millions of foodborne illnesses, hundreds of thousands of hospitalizations, and hundreds of deaths annually in the U.S. (Centers for Disease Control and Prevention 2022, Mead et al. 1999). These inspections are effective because they interrupt the natural decay of adherence to regulations over time (Mohliver 2019, Anand et al. 2012). However, limited resources make continuous inspections infeasible (Mejia et al. 2019), forcing regulators to make strategic decisions about when and where to inspect—often with limited visibility into which establishments are most likely to violate. Predictions can reduce this uncertainty and have the potential “to make an inspection program even more effective” (Mejia et al. 2019, p. 1364). This promise has prompted several studies predicting the presence of restaurant hygiene violations (e.g., Farronato and Zervas 2022, Siering 2021, Mejia et al. 2019, Kang et al. 2013). Yet, how such predictions can be systematically integrated into inspection planning remains underexplored. For example, recent work by Johnson et al. (2023) shows that not all predictive models lead to improved inspection outcomes, highlighting the need to evaluate not just what we can predict, but how we can use those predictions. Because of this uncertainty, most regulators continue to make choices based on simple heuristics, such as examining establishments at regular intervals (Mohliver 2019, Farronato and Zervas 2022). This discrepancy between predictive potential and current practice prompts the following question: *How should regulators optimally allocate limited inspection resources to reduce restaurant hygiene violations based on predicted inspection outcomes?*

To answer this question, we integrate empirical machine learning (ML) predictions about the presence of hygiene violations with a prescriptive resource allocation decision model to formulate a decision framework for allocating inspection resources. Figure 5.1 displays the details of the resulting framework. The prediction step uses gradient-boosted regression trees (GBRT) (Chen and Guestrin 2016, Friedman 2001, Friedman et al. 2000) to predict the likelihood of violations based on information about the time since the last inspection, past inspection outcomes, Yelp review information, and establishment characteristics. The

prescription step builds on the fact that—in the absence of inspections—operational routines decay over time (Mohliiver 2019, Anand et al. 2012, Vaughan 1999). This allows us to model the misconduct progression of a single establishment as a Markov chain and to formulate the resource allocation problem as a capacity-constrained interactive partially observable Markov decision process (I-POMDP) (Gmytrasiewicz and Doshi 2005). We then apply Lagrangian relaxation to decouple the inspection decisions between different establishments and implement an index policy to allocate inspection resources. We evaluate the effectiveness of our approach using data from restaurant hygiene inspections conducted by the Southern Nevada Health District (SNDH) between 2012 and 2019.

Figure 5.1: Framework: Integrating Predictions into Resource Allocation Decisions



Notes. † Yelp Characteristics include: American (New), Asian Fusion, Barbecue, Bars, Breakfast & Brunch, Burgers, Cafes, Chicken Wings, Chinese, Coffee & Tea, Delis, Deserts, Event Planning & Services, Fast Food, Italian, Juice Bars & Smoothies, Mediterranean, Mexican, Nightlife, Pizza, Salad, Seafood, Specialty Food, Sports Bars, Sushi Bars, Thai.

Our findings regarding the prediction step show that it is possible to predict the unobserved hygiene-compliance state of establishments using information from past inspections and Yelp reviews. Specifically, we find that combining time since the last inspection (1 in Figure 5.1), previous inspection results (2), review text information (3), and restaurant characteristics (3) yields the best predictions. However, we also demonstrate that interpreting the predictions is not always intuitive, which may diminish their usefulness for decision-making. For instance, predicting a 30% risk of a hygiene violation implies that

the establishment is more likely to violate hygiene regulations than to comply with them. We account for this uncertainty in the I-POMDP resource allocation model.

Our findings regarding the prescription step show that machine learning predictions alone can be valuable in making inspection resource decisions but that regulators can make inspection programs substantially more effective by combining data-driven prediction with model-based planning. Specifically, we find that allocating inspection resources based on predictions alone provides a 63% relative improvement over a regular inspection interval. However, combining predictions with model-based planning in our ML INDEX policy provides an additional 48% relative improvement over a prediction-only policy.

Our study contributes to multiple streams of literature. First, we extend the research on predicting organizational misconduct by demonstrating how machine learning predictions can be operationalized to enhance regulatory decision-making. In doing so, we introduce I-POMDPs to the operations management (OM) community as a viable framework for modeling multi-agent decision problems under uncertainty. Second, we contribute to the study of organizational misconduct from an OM perspective by formalizing the relationship between routine conformance decay and misconduct progression. While prior work has recognized misconduct as routine non-conformance (Vaughan 1999) and documented the decay of routine conformance over time (Anand et al. 2012), no research has explicitly modeled these dynamics as a Markov process. This modeling approach enables simulation-based evaluation of regulatory interventions, addressing the critical gap in research on misconduct prevention strategies (Palmer 2012). Finally, we contribute to the literature on social and environmental responsibility audits (e.g., Ha et al. 2023, Zhang et al. 2022, Caro et al. 2018), where firms use audits and penalties to enforce responsible behavior among suppliers. While our data is from a restaurant context, our setting mirrors this dynamic, as regulators oversee compliance through inspections. Thus, it appears promising to model social and environmental responsibility as a Markov chain or to explore whether supplier responsibility violations can be predicted and whether these predictions can be

used to improve audit resource allocation decisions.

Regarding the practical application, our resource allocation policy can be easily implemented using open-source or widely available software, such as R, Python, or Julia. It can support regulators in optimizing their resource allocation decisions so that they can increase their level of misconduct detection and prevention with the current resources or maintain their current detection and prevention levels with reduced resources. The policy extends beyond the restaurant context to other contexts with multiple underlying misconduct severity states and limited inspection resources. Thus, the policy can have implications for all regulators that conduct inspections such as the Occupational Safety and Health Administration (OSHA), the U.S. Food and Drug Administration (FDA), the Environmental Protection Agency (EPA), the Centers for Medicare & Medicaid Services (CMS), and many others as well as their international counterparts.

5.2 Literature Review

Our research is closely related to three literature streams: organizational misconduct as routine non-conformance, prediction of organizational misconduct, and operations management studies exploring the role of machine learning in operational decision-making.

5.2.1 Organizational Misconduct and Routine Conformance Decay

Organizational misconduct refers to behavior within or by firms that a regulator judges to transgress the line separating right from wrong (Greve et al. 2010). Most empirical research on organizational misconduct has focused on its drivers and consequences, while potential mitigation strategies remain underexplored (Palmer 2012). A key aspect of misconduct is the decay of operational routines (Anand et al. 2012), which is a primary driver of non-conformance (Mohliver and Ody-Brasier 2023, Vaughan 1999).

Vaughan (1999) introduced the idea that misconduct often stems from routine non-conformance, while Anand et al. (2012) demonstrated that operational routines are prone to decay. Their findings highlight that inspections—one of the key tools regulators use to detect and prevent misconduct (Muthulingam et al. 2022)—can help renew organizational focus on routine conformance. More recently, Mohliver and Ody-Brasier (2023) provided additional evidence for routine conformance decay in the context of nursing home violations. They find that religious nursing homes experience more severe infractions during routine inspections. This occurs because religious nurses file fewer complaints to regulators, increasing the time facilities operate without oversight and allowing violations to escalate (Mohliver and Ody-Brasier 2023). Importantly, prior research indicates that the rate of decay is not uniform but varies across organizations (Mohliver and Ody-Brasier 2023, Anand et al. 2012). We build on these insights by modeling the progression of restaurant hygiene violations as a Markov chain in Section 5.5.

5.2.2 Predicting Organizational Misconduct

The prediction of inspection outcomes and organizational misconduct has been of interest in public health (e.g., Schomberg et al. 2016) and computer science (e.g., Kang et al. 2013), and it has recently gained traction in management science (e.g., Ball et al. 2025, Campbell and Shang 2022, Mejia et al. 2019, Chang and Wu 2021, Cecchini et al. 2010). Ball et al. (2025) examined whether drug review sentiment is associated with the hazard of serious drug recalls. Campbell and Shang (2022) examined whether employee reviews on Glassdoor can predict organizational misconduct, finding that these reviews are indicative of both prosecuted misconduct and whistleblower complaints. Similarly, Cecchini et al. (2010) showed that financial statement data can predict management fraud, while Chang and Wu (2021) demonstrated that waste gas data can be used to detect environmental violations. Within the restaurant industry, studies have shown that online reviews are predictive of hygiene inspection outcomes (Farronato and Zervas 2022, Siering 2021, Mejia et al. 2019),

and consumers respond to online hygiene signals (Hou et al. Forthcoming, Farronato and Zervas 2022). Many of these studies recognize the regulatory potential of predictive models. For instance, (Campbell and Shang 2022, p. 7035) suggest that corporate managers and external stakeholders could use such predictions “to detect any ongoing acts of misconduct and manage the risks of future misconduct.” However, most studies stop short of detailing how regulators can integrate these insights into operational decision-making.

A notable exception is Johnson et al. (2023), who examined how OSHA allocates inspection resources. OSHA currently selects establishments for random inspection based on injury rates from the previous two years. Johnson et al. (2023) propose using machine learning predictions of workplace injuries to refine this selection process. Their results show that targeting establishments based on predicted injuries could significantly reduce workplace safety violations compared to OSHA’s current approach. However, they also found that targeting establishments based on predicted treatment effects (i.e., expected inspection effectiveness) was less successful due to lower prediction accuracy. Our study builds on this finding by demonstrating that when machine learning predictions *replace* prior beliefs about organizational misconduct states, poor predictions can lead to suboptimal regulatory decisions. In contrast, our machine learning based index policy (ML-INDEX) allows regulators to *augment* their prior beliefs with machine learning predictions, mitigating the risk of erroneous forecasts while capitalizing on accurate ones.

5.2.3 Machine Learning in Operations Management

Operations management research has increasingly explored the role of machine learning in optimizing operational decisions (e.g., Cui et al. 2018, Ye et al. 2015). Most studies in this domain use machine learning to predict an outcome of interest, which then informs decision-making. For example, Ang et al. (2016) predicted emergency department wait times, allowing hospitals to optimize patient flow and reduce overcrowding.

The studies most relevant to our work are Steinker et al. (2017), Cui et al. (2018),

Kamalzadeh et al. (2021), and Mukherjee et al. (2025), which integrate machine learning predictions with new sources of information and evaluate the value they have on operational decisions. Steinker et al. (2017) use weather data to forecast sales, helping e-commerce firms optimize logistics and warehousing. Cui et al. (2018) leverage social media data to forecast daily sales, which allows retailers to improve their pricing and promotion decisions. Both Mukherjee et al. (2025) and Kamalzadeh et al. (2021) develop a machine learning-driven decision framework for diabetes care. They propose methods that integrate a partially observable Markov decision process and machine learning risk modeling to optimize screening policies (Kamalzadeh et al. 2021) and patient engagement (Mukherjee et al. 2025). Their results demonstrate how predictive modeling can be effectively incorporated into a formal decision-making framework to improve resource allocation. Our study builds on this literature by examining the role of machine learning in optimizing regulatory decision-making in the context of organizational misconduct. Specifically, we address a classification problem, whereas prior studies (e.g., Cui et al. 2018, Steinker et al. 2017) predominantly focused on continuous outcomes. We also incorporate multiple agents' actions, whereas prior studies (e.g., Mukherjee et al. 2025, Kamalzadeh et al. 2021) only allow for regulator actions to affect the underlying state of the world.

In summary, existing research indicates that operational routines decay over time, leading organizations to violate rules and regulations such as restaurant hygiene standards. These violations become more likely and severe as more time passes since a state of routine conformance. Regulators inspect organizations to renew their focus on compliance and detect violations. However, while regulators can observe an organization's compliance state at the last inspection and understand that compliance likely deteriorates over time, they lack real-time knowledge of an organization's current state when making inspection resource allocation decisions. Researchers have begun addressing this problem by using predictive models to estimate misconduct risk. Yet, how these predictions can be systematically integrated into operational decision-making—particularly in the context of inspection

resource allocation—remains underexplored. We build on these insights when developing a theoretical model that describes the occurrence of restaurant hygiene violations and when demonstrating how predictions can be leveraged to optimize resource allocation decisions.

5.3 Empirical Setting and Data

5.3.1 Health Inspections in the Restaurant Industry

Food safety and hygiene are critical aspects of the restaurant industry that have significant implications for public health (Jin and Leslie 2003). Health departments conduct inspections to evaluate restaurants' adherence to health and safety guidelines. In the Las Vegas area, the Southern Nevada Health District (SNHD) is responsible for inspecting food establishments and enforcing health regulations. The inspection process, grading scale, and points of emphasis of the SNHD compare favorably to other cities in which different aspects of the restaurant health inspection process have been studied, such as Los Angeles (Jin and Leslie 2003, Ho et al. 2019), Louisville (Makofske 2020), New York (Farronato and Zervas 2022, Mejia et al. 2019, Ho 2012), Seattle (Kang et al. 2013), or San Diego (Ho 2012).

In Las Vegas, each restaurant is inspected at least once annually, or more frequently when necessary, to ensure compliance with SNHD Regulations Governing the Sanitation of Food Establishments (Southern Nevada Health District 2010). Restaurants that violate hygiene regulations receive fines and must undergo re-inspection within 15 business days. This approach is consistent with practices in other jurisdictions (Farronato and Zervas 2022, Mejia et al. 2019) and similar inspection regimes, such as those for nursing homes (Mohriver 2019) and FDA plant inspections (Ball et al. 2017).

The inspection process is highly standardized. Upon arrival, inspectors conduct a visual assessment of the restaurant, identifying potential hygiene or sanitation issues. Using a Food Establishment Inspection Report (FEIR, see Appendix D.1), inspectors evaluate critical aspects such as food storage, preparation, handling procedures, general cleanliness, and the

condition of cooking equipment. Consistent with the district’s emphasis on mitigating the public health costs associated with foodborne illness, inspectors prioritize factors that may contribute to such risks (Southern Nevada Health District 2023).

Inspections result in a grade based on the number of demerits recorded in the FEIR. The demerit system ranges from 0 to 41+, with higher scores indicating poorer compliance with health regulations. Critical violations, such as spoiled or adulterated food, add 5 demerits, while major violations, such as improper sanitation of kitchenware, add 3 demerits. Restaurants with 10 or fewer demerits receive an “A” grade, those with 11 to 20 demerits receive a “B” grade, and those with 21 to 40 demerits receive a “C” grade. Restaurants accumulating more than 41 demerits are closed immediately until all issues are resolved. In this study, we use the inspection grade as our dependent variable such that anything other than “A” grades are considered violations.

5.3.2 Data

We use two primary data sources. First, we collect restaurant inspection data for Las Vegas from 2012 to 2019 (Southern Nevada Health District 2022). The dataset contains detailed inspection results, including restaurant identifiers, demerit counts, grades, inspector IDs, and inspection outcomes. It also contains inspection timing, including dates for re-inspections following routine inspections. The SNHD publishes inspection results online within five days of the inspection date.

To enrich the analysis, we combined the inspection data with Yelp restaurant reviews from the Yelp Open Dataset (Yelp Inc. 2021). This dataset includes more than 6.9 million reviews across 150,000 businesses in 11 metropolitan areas, including Las Vegas. We matched reviews to inspection records using Yelp’s business match API and restaurant details like name, address, city, and zip code. To predict hygiene violations, we use reviews posted between the most recent inspection (routine or re-inspection) and the current routine inspection date, taking into account review text and star ratings.

Table 5.1 summarizes the dataset’s descriptive statistics. These show a 63% increase in the number of restaurants from 1,728 in 2012 to 2,819 in 2019. Inspection capacity has kept up with this growth, with each restaurant being inspected twice a year on average. The average number of Yelp reviews per inspection cycle increased by 160%, from 19.56 to 50.80. This reflects the growing influence of online review platforms.

Table 5.1: Descriptive Statistics of Raw Data

	Measure	2012	2013	2014	2015	2016	2017	2018	2019
Restaurants	Count	1728	1953	2133	2319	2438	2564	2700	2819
Inspectors	Count	73	71	75	91	103	100	93	99
Inspections	Count	3225	3984	4179	4868	4923	5312	5739	5934
Inspections / Week	Count	61	75	79	92	93	102	108	112
Inspections / Restaurant	Mean	1.87	2.04	1.96	2.10	2.02	2.07	2.13	2.11
Inspections / Restaurant	SD	0.93	1.11	1.06	1.10	1.00	0.99	1.10	1.07
Reviews / Inspections	Mean	19.56	24.13	34.05	44.27	48.21	51.81	56.42	50.80
Reviews / Inspections	SD	39.17	43.07	55.72	73.13	80.52	80.18	89.36	82.17

5.3.3 Variable Definition

Outcome Variable: Hygiene violations

We operationalize hygiene violations (our outcome variable) as a binary variable indicating whether a restaurant received a grade of “B” or lower. This approach has several benefits. First, SNHD inspection records make it very transparent which activities are considered violations, providing a good measure of the underlying nefarious behavior by restaurants. Second, violations detected during health inspections are associated with possible fines and the posting of grade cards, which, if less than an “A” grade, could decrease consumer demand (Jin and Leslie 2003). Thus, restaurant owners have a vested interest in avoiding these violations.

Features

To evaluate the predictive value of various information types, we define four distinct feature sets based on the existing hygiene violation prediction literature (e.g., Farronato and Zervas 2022, Siering 2021, Mejia et al. 2019, Kang et al. 2013). The first feature set, *Time-Only* (1 in Figure 5.1), contains only the time elapsed since the last inspection. The second set, *Inspection-Record*, adds information about previous inspection outcomes, including the lagged outcome, grade, and demerit count. The third set, *Review-Information*, adds detailed Yelp review data. Finally, *Full-Model* adds restaurant characteristics obtained from Yelp, such as cuisine type (e.g., Mexican, Chinese) and establishment category (e.g., Bar, Deli). We provide an overview of these features in Table D.1 and describe the Yelp review-based features used for *Review-Information* below.

Count of Reviews For each inspection, we calculated the number of Yelp reviews posted since the previous inspection. Reviews may provide indirect signals of misconduct (Kang et al. 2013). To account for variations in the time between inspections, we scaled the number of reviews by the number of weeks since the last inspection.

Average Star Rating Yelp reviewers assign a one-to-five-star rating along with a written review of the restaurant. We computed the average star rating for reviews posted since the last inspection, as prior research suggests this metric may be a potential predictor of misconduct (Siering 2021).

Bad Reviews While the average star rating provides an overall sentiment, it does not capture variability in consumer perceptions. To address this, we calculated the number of low-star ratings (defined as three stars or fewer). This measure is based on the premise that consumers dissatisfied with a restaurant’s hygiene are more likely to leave lower ratings.

Review Text We also analyzed the text of Yelp reviews as a predictor of misconduct (Siering 2021, Kang et al. 2013). We used a lasso-penalized multinomial inverse regression framework to project review text onto our misconduct metrics, resulting in a

low-dimensional review text score for each restaurant inspection observation (Taddy 2015, 2013b). Previous studies have used this model to predict misconduct from employee reviews (Campbell and Shang 2022) and to assess the usefulness of Yelp reviews in signaling restaurant hygiene violations (Farronato and Zervas 2022). To develop this review text score, we used data from inspections conducted between 2012 and 2015. Appendix D.3 provides a detailed explanation of the construction of this measure.

5.4 Machine Learning Model and Prediction

In this section, we describe the machine learning model that we use to analyze the presence of restaurant hygiene violations. Subsection 5.4.1 describes the machine learning model. Subsection 5.4.2 discusses how we evaluate the model performance and address data imbalance issues. Finally, subsection 5.4.3 presents our results.

5.4.1 Machine Learning Model

Various machine learning techniques have been used to predict organizational misconduct, such as Naive Bayes, Gradient Boosting, Neural Networks, Random Forests, Support Vector Machines, and Super Learner (e.g., Johnson et al. 2023, Campbell and Shang 2022, Siering 2021, Mejia et al. 2019, Kang et al. 2013, Cecchini et al. 2010). In our analysis, we use a gradient-boosted regression tree (GBRT) approach (Friedman et al. 2000, Friedman 2001).¹

GBRT is an ensemble method that improves prediction performance by combining multiple weak learners, typically decision trees. The approach iteratively corrects prediction mistakes by fitting a new tree $f_t(x)$ to the residuals (or gradient) of the loss function of the previous iteration model's predictions (Mukherjee et al. 2025, Chen and Guestrin 2016).

¹Appendix D.4 compares GBRT performance with several alternative machine learning models, demonstrating that GBRT performs as well as or better than all alternatives.

For an input x_i , the model produces the following output:

$$\hat{y}_i = \phi(x_i) = \sum_{k=1}^K f_k(x_i), \quad (5.1)$$

where each f_k is a regression tree. The training process minimizes a regularized objective function:

$$\mathcal{L} = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k), \quad (5.2)$$

where $l(y_i, \hat{y}_i)$ is a differentiable convex loss function, and $\Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2$ is a regularization term that penalizes the complexity of each tree. Here, T is the number of leaves and w is the vector of leaf weights. $\Omega(f_k)$ helps prevent overfitting by discouraging overly complex trees.

The model is trained in a greedy and additive manner. At iteration t , the algorithm fits a new tree $f_t(x)$ to minimize the updated objective:

$$\mathcal{L}^{(t)} = \sum_{i=1}^n l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t), \quad (5.3)$$

where $\hat{y}_i^{(t-1)}$ represents the model's prediction for instance i at the previous iteration.

We implement the GBRT model in R using the *xgboost* and *caret* packages (Chen et al. 2025, Kuhn 2019). To capture high-order interactions among predictors, we set the tree depth to 6, allowing for five-way interactions (Campbell and Shang 2022). Following Schonlau (2005), we set the learning rate to 0.001 and train for up to 10,000 iterations. These hyperparameters were selected using 5-fold cross-validation, repeated three times, to ensure robust out-of-sample performance.

5.4.2 Performance Evaluation

Our objective is to accurately predict the presence of future restaurant hygiene violations from existing information. There are three challenges in accurately predicting organizational misconduct and evaluating predictor performance: data imbalance, model calibration, and the time dimension of the data.

Addressing the Data Imbalance Challenge

Organizational misconduct in general and restaurant hygiene violations in particular are a rare phenomenon (Palmer 2012). In our data, only 25% of all routine inspections result in a violation. This data imbalance is problematic because most standard machine learning algorithms assume that the data is balanced and that the cost of misclassification is equal (Sun et al. 2009). Machine learning models are generally trained on training data and evaluated on so-called holdout or test data (Cui et al. 2018). During the training stage, classification rules are developed that are later used to predict the class of observations in the test sample. To create these rules, most learning algorithms assume that the outcome variable is balanced, meaning there are equal observations for each outcome class. If the data is unbalanced, then classification rules to accurately predict rare events are often undiscovered or ignored. This leads to rare classes being more often misclassified than the dominant class. (Sun et al. 2009) This is especially problematic in our context, where correctly predicting violations, the rare class, is more important than correctly predicting compliance in the form of “A” grades. We address this challenge by using a synthetic minority over-sampling technique (SMOTE) (Chawla et al. 2002).²

Sampling reduces the data imbalance by selectively sampling observations from the training data to create a balanced dataset on which a model is trained (Kuhn 2019). SMOTE

²We compared the performance of SMOTE to weighting, down-sampling, and up-sampling as other imbalance correction techniques and found it to be superior. This is consistent with the findings by Chawla et al. (2002).

achieves balance by oversampling the rare class by creating synthetic observations (Chawla et al. 2002). For example, if 75 observations in the training data are for compliance and 25 observations are for violation then the technique creates a new training sample with 150 total observations by creating 50 synthetic violation observations with predictors that are similar but more general than the original observations (see Chawla et al. (2002) for a detailed description of the algorithm). The model is then trained on the new, extended training data. Importantly, the test data is not sampled in any way to accurately assess out-of-sample performance.

Addressing the Calibration Challenge

Machine learning classifiers, even those that excel at predicting the correct class, are often poorly calibrated. Calibration refers to the alignment between a model's predicted probabilities and the true probabilities observed in the data (Fernández et al. 2018). A poorly calibrated model produces predicted probabilities that do not accurately represent the actual likelihoods. Having a well-calibrated model is important because we use the predicted probabilities and not the classifications to improve the inspection resource allocation decisions.³

To illustrate, consider a binary classification scenario with two states: Compliance and Violation. Suppose a model predicts that 10 observations have a 0.6 probability of being violations. If the model is perfectly calibrated, six of these observations would indeed be violations, and four would be in compliance. In this case, the predicted probabilities align with the true distribution, and the model's precision score would be 0.6. Conversely, if nine observations were violations and only one was compliance, the predicted probability of 0.6 would underestimate the true probability, even though the precision score would increase to 0.9. This example highlights a critical issue: most models in the misconduct prediction

³We discuss the reason for using predicted probabilities and not classifications in Section 5.4.3 and how we integrate the predictions in the decision-making policy in Section 5.5.4.

literature are optimized for metrics such as precision and recall, which do not account for calibration.

Model calibration can be achieved through two primary approaches: (1) training the model using a calibration-specific metric and (2) applying post-hoc calibration techniques to the model’s predictions (Jiang et al. 2012). In this study, we implement both approaches to ensure robust calibration.

First, we use logLoss and the Brier score (Brier 1950) as our performance measures. Contrary to the more established Accuracy or Precision measures, these measures do not only value correct classification but also prediction accuracy. Specifically, logLoss measures how close the predicted probability is to the actual class. It is computed as:

$$\text{logLoss} = \frac{1}{N} \left(\sum_{i=1}^N y_i * \ln(p_i) + (1 - y_i) * \ln(1 - p_i) \right), \quad (5.4)$$

where y_i is the true value of observation i , p_i is the predicted probability for state y_i , and \ln is the natural logarithm. The measure decreases as the predictions become more accurate and better calibrated.

The Brier score is the mean squared prediction error (Brier 1950):

$$\text{Brier} = \frac{1}{N} \left(\sum_{i=1}^N (p_i - y_i)^2 \right), \quad (5.5)$$

Note that the Brier score is less sensitive to completely wrong predictions than the logLoss. For example, if a prediction is 0 but the true value is 1, then the Brier score for that observation is 1 but the logLoss is ∞ .

Second, we calibrate the model’s output using Platt scaling (Platt 2000), a post-processing technique that turns raw classifier scores into accurate probability estimates. This method involves fitting a logistic regression model on the original classifier’s outputs and true labels, thereby learning a map from scores to probabilities. The resulting adjusted

probabilities better reflect an event’s true likelihood, which makes them more appropriate for resource allocation decisions.

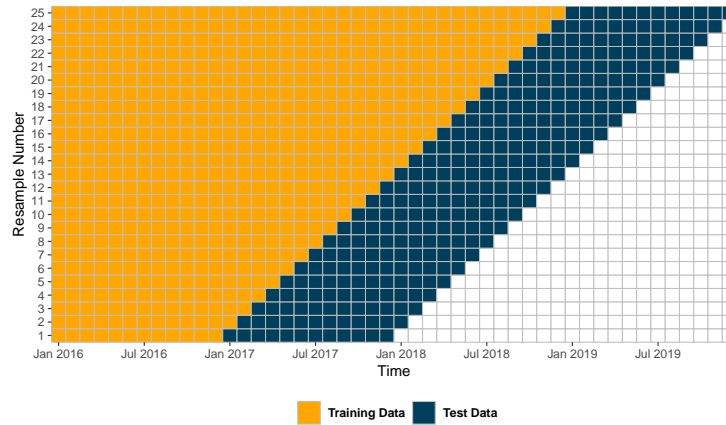
Addressing the Time-Dimension Challenge

The existing research on predicting organizational misconduct accounts for the time series feature of the data to varying degrees. Most researchers are concerned with the overall predictive ability of machine learning models and use a random train/test split with cross-validation (e.g., Mukherjee et al. 2025, Johnson et al. 2023, Campbell and Shang 2022, Siering 2021, Kang et al. 2013). However, because we are using the predictions to make resource allocation decisions, it is unrealistic to use future values to predict current violations, as done in a random train/test split. To account for this time dynamic in the data, we use rolling window forecasting to evaluate the classifier performance (Hyndman and Athanasopoulos 2018, Cui et al. 2018).⁴

Because we used the data from 2012 to 2015 to construct our review text predictor, we exclude these years from our model evaluation (Campbell and Shang 2022). This leaves the data from 2016 to 2019 to train and evaluate our models. Figure 5.2 displays our sampling scheme. The approach works as follows: We first construct an initial training set with the data from 2016 (Orange boxes for resample number 1). We then construct a test set with the data from 2017 (a one-year test horizon highlighted in blue) and evaluate the models’ out-of-sample performance. Afterwards, we extend the training period by one month while keeping the test horizon fixed. Thus, the second repetition uses training data from January 2016 through January 2017 and tests the out-of-sample performance on data from February 2017 through January 2018. We repeat this process 25 times, after which the test sample extends beyond our study period. We then use the performance values of the 25 observations to compare the predictive performance of the different feature sets.

⁴Note that Johnson et al. (2023) and Campbell and Shang (2022) conduct robustness checks using a rolling window forecasting approach.

Figure 5.2: Rolling Window Forecasting



Notes. The figure shows the rolling window forecasting approach taken in this project.

5.4.3 Prediction Results

Prediction Model Performance

Table 5.2 presents the predictive performance of gradient-boosted regression tree (GBRT) models in forecasting restaurant hygiene violations. Performance is measured using the Brier score and logLoss, where lower values indicate better predictive performance. We evaluate the improvement between models based on a paired t-test (Law 2013). We also report Accuracy and Precision measures, though these metrics are less informative in this context (see discussion in Section 5.4.2).

The results confirm that predictive performance improves as additional features are incorporated. The *Time-Only* model, which considers only the time since the last inspection, performs the worst, with the highest Brier score (0.249) and logLoss (0.691). Adding past inspection outcomes (*Inspection-Record*) significantly improves predictions, reducing the Brier score by 0.032 ($p < 0.001$) and logLoss by 0.068 ($p < 0.001$). Incorporating Yelp review data (*Review-Information*) further enhances prediction quality, with a Brier score reduction of 0.021 ($p < 0.001$) and logLoss reduction of 0.046 ($p < 0.001$).

Table 5.2: Performance of Gradient-Boosted Regression Tree Models for Predicting Violations

	Time-Only ^a (1)	Inspection-Record ^b (2)	(2)-(1) (3)	Review-Information ^c (4)	(4)-(2) (5)	Full-Model ^d (6)	(6)-(4) (7)
Brier [†]	0.249	0.217	-0.032 [<0001]	0.196	-0.021 [<0001]	0.180	-0.017 [<0001]
LogLoss [†]	0.691	0.623	-0.068 [<0001]	0.577	-0.046 [<0001]	0.539	-0.038 [<0001]
Accuracy	0.485	0.650	0.165 [<0001]	0.699	0.049 [<0001]	0.736	0.037 [<0001]
Precision	0.255	0.328	0.073 [<0001]	0.340	0.012 [0.001]	0.396	0.056 [<0001]

Notes. The table shows results from predicting the occurrence of violations (i.e., a “B” or “C” grade) during restaurant hygiene inspections using gradient-boosted regression trees (GBRT). The performance is evaluated on 25 monthly rolling window forecasts from 2016 to 2019 and after calibrating the model. p-values from a paired t-test are in brackets.

†: Lower values indicate better performance.

a: The *Time-Only* feature set includes the time since the last inspection.

b: The *Inspection-Record* includes the features from *Time-Only* and information about previous inspection outcomes (i.e., lagged outcome, lagged grades, lagged demerit count).

c: The *Review-Information* includes the features from *Inspection-Record* and data from reviews since the last inspection (i.e., review count, average star rating, bad review count, review text measure).

d: The *Full-Model* includes the features from *Review-Information* and adds and restaurant characteristics from Yelp, such as cuisine type (e.g., Mexican, Chinese) and establishment category (e.g., Bar, Deli).

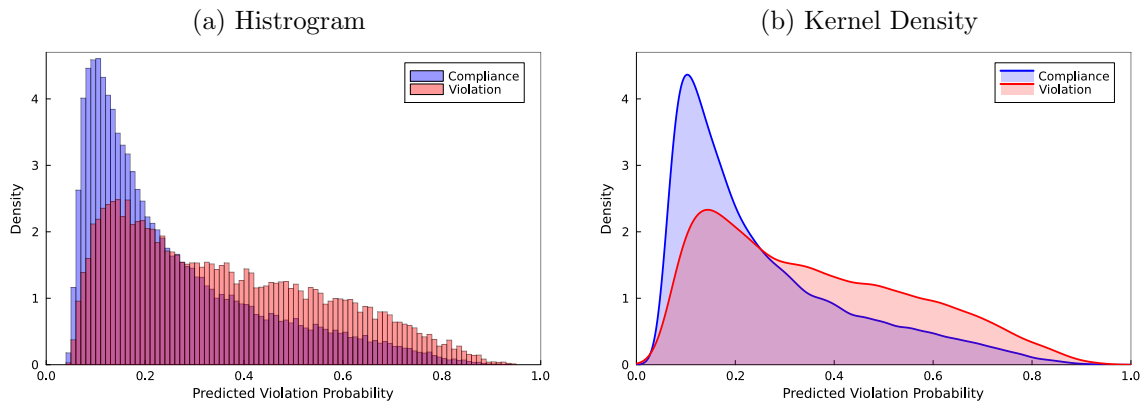
Finally, the *Full-Model*, which integrates restaurant characteristics such as cuisine type and establishment category, yields additional improvements, with the lowest Brier score (0.180) and logLoss (0.539). Appendix D.5 further examines feature importance within the *Full-Model*. The most influential predictor is the demerit count from the last inspection, highlighting the persistence of hygiene violations. Time since the last inspection remains a key factor, ranking among the top predictors. Yelp-derived attributes, such as average star ratings and review text sentiment, also contribute meaningfully, though restaurant characteristics (e.g., Fast Food, Bars) emerge as dominant features.

Prediction Model Value for Decision-Making

While the whole model greatly improves on the baseline, its ability to optimize the allocation of inspection resources remains unclear. Figure 5.3 shows a significant overlap in estimated probabilities for Violations and Compliance. The figure illustrates the density distribution

of anticipated violation probabilities, comparing true violations (red) and true compliance (blue). The x-axis represents the predicted probability of violation, while the y-axis depicts the density. This figure highlights that the predicted probabilities for violations differ depending on the true state, emphasizing the potential value of these predictions. However, the substantial overlap between the two distributions limits their usefulness. For instance, consider a prediction where the probability of violation is 0.3 and compliance is 0.7. Although this suggests a high likelihood of compliance, the likelihood ratio that this prediction is drawn from the compliance distribution (blue) vs. from the violation distribution (red) is 0.90.⁵ This implies that the prediction—despite its low predicted probability of violation—increases the belief of a Bayesian rational agent that the true state of the world is a violation. Therefore, relying solely on these predictions for resource allocation may lead to suboptimal performance.

Figure 5.3: Comparison of Predictions Conditional on True Outcome



Notes. The figure compares the distributions of the predicted probability of violation for true violations (blue) and true compliance (red) observations. Subfigure (a) presents a histogram of the prediction values corresponding to Column (6) in Table 5.2 and subfigure (b) shows the kernel density estimates used for likelihood calculations.

Recall from Section 5.4.2 that we use predicted probabilities rather than discrete classification labels to inform the regulator's decision-making. Figure 5.3 helps illustrate

⁵We use the kernel density estimates in Figure 5.4b to compute the relative likelihood of observations.

the rationale behind this choice. Traditional classification performance measures rely on a confusion matrix, which compares true and predicted class labels (Kuhn 2019). To generate these class labels from predicted probabilities, the classifier—in our case, the GBRT algorithm—applies a threshold, assigning one class label to probabilities above the threshold and another to those below it. This threshold is typically chosen to optimize classification performance metrics such as accuracy or precision. While classification labels can simplify decision-making and have been successfully used in other contexts, such as diabetes screening (Kamalzadeh et al. 2021), they discard valuable probabilistic information. Consider, for example, the previous scenario where a model predicts a 30% probability of violation and a 70% probability of compliance, resulting in a likelihood ratio of 0.9. This ratio increases a Bayesian rational agent’s belief that the true state is a violation. Now, contrast this with a prediction of a 20% probability of violation and 80% probability of compliance. Figure 5.4b shows that the density of the compliance distribution (blue) is larger than the density of the violation distribution (red), which results in a likelihood ratio of 1.15. This would shift the agent’s belief toward compliance. If we instead relied solely on classification labels, both of these cases might have been labeled as “compliance.” This would obscure the important nuance between them, treating both predictions as equally reinforcing the belief in compliance,⁶ despite their differing likelihood ratios. Retaining the full probability distribution leads to better-informed resource allocation decisions (Mukherjee et al. 2025).

In summary, our findings underscore the potential predictive value of public data but also highlight that using such predictions for operational decisions is not straightforward. In the next section, we explore how these predictive insights can inform and optimize the allocation of inspection resources.

⁶The likelihood of a predicted label, and therefore its impact on the agent’s belief state, can be derived from the confusion matrix (see Kamalzadeh et al. 2021, Section 3.3.2).

5.5 Model Formulation

After creating machine learning predictions of hygiene violations, we analyze how regulators can leverage these predictions to optimize inspection resource allocation. Building on the understanding that regulatory conformance deteriorates over time (Mohliiver and Ody-Brasier 2023, Anand et al. 2012), we define the restaurant inspection resource allocation problem as an interactive partially observable Markov decision process (I-POMDP) (Gmytrasiewicz and Doshi 2005). In this framework, the states represent establishments’ adherence to hygiene regulations, and the regulator must decide whether to inspect a given restaurant during each decision period.

Due to resource restrictions, regulators must strategically conduct inspections to reduce social costs associated with regulatory non-compliance (Mejia et al. 2019). Partially Observable Markov Decision Processes (POMDPs) provide a well-established framework for such decision-making under uncertainty and for incorporating machine learning predictions (e.g., Mukherjee et al. 2025, Kamalzadeh et al. 2021). However, they do not account for the interactive and adaptive nature of establishments, which may change their behavior in response to inspection patterns. For example, establishments may exert effort to reduce the likelihood that their hygiene compliance deteriorates or conduct self-inspection to address violations. Alternative frameworks, such as Partially Observable Markov Games (POMGs) or Partially Observable Stochastic Games (POSGs), capture this interplay between establishments and regulators but lack the belief-state representation necessary to integrate information from machine learning predictions (Kochenderfer et al. 2022). The I-POMDP framework extends POMDPs by incorporating agent models (Gmytrasiewicz and Doshi 2005), which allows regulators to make inspection decisions based on establishments’ probabilistic responses.

The core idea underlying an I-POMDP is that an agent, in this case the regulator, maintains a belief state over both the physical environment (e.g., hygiene compliance levels)

and the mental models of other agents (e.g., establishments’ effort and self-inspection strategies). Unlike traditional POMDPs, which confine state uncertainty to external factors, I-POMDPs use a recursive framework in which agents simulate the decision-making processes of others. This allows regulators to predict how establishments will change their behavior based on inspection history. A key assumption in this recursive framework is that agents have nested strategic levels such that lower-level agents do not model the higher-level agent’s decisions explicitly (Gmytrasiewicz and Doshi 2005). We model establishments as level-0 agents, which means they make self-inspection judgments based purely on previous regulatory actions without actively modeling the regulator’s strategy. In contrast, the regulator is a level-1 agent capable of developing probabilistic assumptions about establishments’ effort and self-inspection decisions.

By adopting this interactive framework, our model overcomes the limitations of static regulatory policies. The I-POMDP approach enables more granular policy decisions, allowing regulators to differentiate among establishments based on their historical compliance records and estimated risk profiles. Additionally, it enables the integration of machine learning predictions to improve decision-making efficiency.

5.5.1 Single Establishment Decision-Making Model

We formalize the inspection resource allocation problem for a single establishment as an I-POMDP, where the regulator (“she/her”, subscript i) aims to minimize social costs (i.e., maximize her reward function) by optimizing inspection policies under uncertainty, while the establishment (“he/him”, subscript j) tries to maximize his reward by timing effort and self-inspections to minimize penalties from detected violations. We use one week as a decision period, during which inspections are conducted unannounced and within a constrained geographic region, eliminating the need for additional travel arrangements. If applied to settings with longer decision periods, this time frame can be extended accordingly. The I-POMDPs of both the regulator and the establishment can be characterized as 7-tuples

$(\mathcal{IS}_{\in\{i,j\}}, \mathcal{A}, T_{\in\{i,j\}}, \mathcal{O}_{\in\{i,j\}}, O_{\in\{i,j\}}, R_{\in\{i,j\}}, \gamma)$, where $\mathcal{IS}_{\in\{i,j\}}$ is the set of interactive states, \mathcal{A} is the set of joint actions between the regulator and the establishment, $T_{\in\{i,j\}}$ are the transition probabilities between states, $\mathcal{O}_{\in\{i,j\}}$ is the set of the regulator's possible observations, $O_{\in\{i,j\}}$ are the observation probabilities, $R_{\in\{i,j\}}$ is the reward function, and γ is a discount factor (Gmytrasiewicz and Doshi 2005). We describe these components in detail below.

Sequence of Events

At the start of the establishment's lifecycle ($t = 0$), the regulator ensures that each establishment is in full compliance. This mirrors regulatory practices across industries such as healthcare and restaurants, where initial compliance inspections are mandatory before starting operations (Ball et al. 2017, Southern Nevada Health District 2010). Each decision period then follows a structured sequence of events:

- Both the establishment and the regulator update their belief about the establishment's current state based on their observations.
- The regulator generates a machine learning prediction and uses it to update her belief about the establishment's current state.
- Both the establishment and the regulator decide on their actions. The establishment decides whether to invest effort, conduct a self-inspection, or do nothing based on his current belief about his state. The regulator decides whether to inspect the establishment based on her belief about the establishment's current state, her modeling of the establishment's decision, and her capacity constraint.
- Actions occur, and the natural decay of adherence to routines takes effect. The order is: decay, establishment action, regulator action.

- The regulator incurs social costs, and the establishment receives rewards based on the state and actions.

Interactive States

$\mathcal{IS}_{\in\{i,j\}}$ is a set of interactive states defined as the combination between physical states of the environment (\mathcal{S}) and mental models of lower-level agents (Gmytrasiewicz and Doshi 2005). Formally, we define the interactive states for the regulator’s and establishment’s I-POMDP as follows (Doshi and Perez 2008):

$$\begin{aligned} \mathcal{IS}_j &= \mathcal{S}, & \Theta_j &= \left\{ \langle b_j, \hat{\theta}_j \rangle \mid b_j \in \Delta \mathcal{IS}_j \right\} \\ \mathcal{IS}_i &= \mathcal{S} \times \Theta_j. \end{aligned} \tag{5.6}$$

Here Θ_j is the set of possible intentional models of the Bayesian-rational establishment such that: $\theta_j = \langle b_j, \hat{\theta}_j \rangle$, where $\hat{\theta}_j$ is a *frame* described by the 6-tuple $(\mathcal{A}, T_j, \mathcal{O}_j, O_j, R_j, \gamma)$ that maps possible histories of the establishment’s observations to distributions over its actions.⁷ Thus, the number of models, and consequently the interactive state space, extends with the planning horizon even if the frame is considered fixed and known (Doshi and Perez 2008). In our analysis, we leverage the fact that, except for the machine learning observations, the observation space and function are the same for the establishment and the regulator. This contains the interactive state space and allows us to develop a policy.

The set \mathcal{S} consists of two states corresponding to compliance and violation, plus two inspection states that correspond to an establishment being inspected such that $\mathcal{S} = \{\text{Compliance}, \text{Violation}, \text{InspectedCompliance}, \text{InspectedViolation}\}$. The state space is partially observable as regulators only uncover the true state of an establishment during an inspection. Thus, instead of relying on the true state of establishments when making

⁷To understand this, consider the following: Let the components of the 6-tuple be fixed and known, then the different establishment models at time t are the possible combinations of the establishment’s observations. For example, if $t = 2$ and two possible observations in \mathcal{O}_j , then there are 4 possible models for every $\hat{\theta}_j$ corresponding to observing: $\{\mathcal{O}_{j,1}, \mathcal{O}_{j,1}\}$, $\{\mathcal{O}_{j,2}, \mathcal{O}_{j,1}\}$, $\{\mathcal{O}_{j,2}, \mathcal{O}_{j,1}\}$, and $\{\mathcal{O}_{j,2}, \mathcal{O}_{j,2}\}$.

inspection decisions, the regulator relies on her belief about the true state to make decisions. We use the random variable $b_{i,t}^j$ to denote the regulator’s current belief about the interactive state of establishment j at period t .

Action Space

The action space $\mathcal{A} = \mathcal{A}_j \times \mathcal{A}_i$ is the set of the joint moves of the establishment j and the regulator i . Every period, the establishment can decide to wait, to exert effort to slow down the compliance-decay, or to conduct a self-inspection (i.e., $a_{j,t} \in \mathcal{A}_j = \{\text{Wait, Effort, Self-Inspect}\}$). The regulator can decide to wait or to inspect establishment j (i.e., $a_{i,t}^j \in \mathcal{A}_i = \{\text{Wait, Inspect}\}$).

Transition Probabilities

The transition probabilities for the I-POMDP are constrained to the physical states of the environment (\mathcal{S}), meaning that one agent’s actions cannot directly control other agents’ models (Gmytrasiewicz and Doshi 2005). For the regulator, we characterize them using the transition probability function $T_i(s, a_j, a_i, s')$, which indicates the probability for moving from the current state ($s \in \mathcal{S}$) to the future state ($s' \in \mathcal{S}$) given that the establishment takes action $a_j \in \mathcal{A}_j$ and the regulator takes action $a_i \in \mathcal{A}_i$. For a given action pair (a_j, a_i) , the probabilities sum to one, and no element has a negative probability. Note that the transition probabilities for $a_j = \text{Self-Inspect}$ depend on μ_j , the probability that the establishment detects and fixes violations during self-inspection. In constructing the transition probabilities, we assume that regulator inspections perfectly identify violations but do not always fix them. This is consistent with the Las Vegas inspection data in which “only” 99 percent of re-inspection following violations lead to “A” grades. Table 5.3 presents the transition probabilities in the regulator’s I-POMDP.

In contrast, the transition probabilities of the establishment—a level 0 agent—depend only on its actions, while the actions of the regulator are incorporated as a noise factor

Table 5.3: Transition probabilities $T_i(s, a_j, a_i, s')$ for each (a_j, a_i) combination

$a_j \backslash a_i$	Wait	Inspect
Wait	$\begin{bmatrix} 0.965 & 0.035 & 0.000 & 0.000 \\ 0.000 & 1.000 & 0.000 & 0.000 \\ 0.965 & 0.035 & 0.000 & 0.000 \\ 0.955 & 0.045 & 0.000 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 & 0.965 & 0.035 \\ 0.000 & 0.000 & 0.000 & 1.000 \\ 0.000 & 0.000 & 0.965 & 0.035 \\ 0.000 & 0.000 & 0.955 & 0.045 \end{bmatrix}$
Effort	$\begin{bmatrix} 0.975 & 0.025 & 0.000 & 0.000 \\ 0.000 & 1.000 & 0.000 & 0.000 \\ 0.975 & 0.025 & 0.000 & 0.000 \\ 0.965 & 0.035 & 0.000 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 & 0.975 & 0.025 \\ 0.000 & 0.000 & 0.000 & 1.000 \\ 0.000 & 0.000 & 0.975 & 0.025 \\ 0.000 & 0.000 & 0.965 & 0.035 \end{bmatrix}$
Self-Inspect	$\begin{bmatrix} a & b & 0.000 & 0.000 \\ \mu_j & 1 - \mu_j & 0.000 & 0.000 \\ a & b & 0.000 & 0.000 \\ c & d & 0.000 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 & a & b \\ 0.000 & 0.000 & \mu_j & 1 - \mu_j \\ 0.000 & 0.000 & a & b \\ 0.000 & 0.000 & c & d \end{bmatrix}$

Notes. Appendix D.6 provides detailed information about the computation of these transition probabilities. a: $0.965 + 0.035 * \mu_j$; b: $0.035 * (1 - \mu_j)$; c: $0.955 + 0.045 * \mu_j$; d: $0.045 * (1 - \mu_j)$

(Gmytrasiewicz and Doshi 2005). The interval between inspections for the SNHD follows a geometric distribution, implying that the per-period inspection likelihood from the establishment's perspective follows a Bernoulli distribution with probability 0.05. Thus, $T(s, a_j, s')$ is characterized as in Table 5.4.

Observations and Observation Probabilities

In each period, an observation (i.e., $o_i \in \mathcal{O}_i = o_j \in \mathcal{O}_j$) provides information about the state of the system. For now, we limit the observations to signal whether a regulator inspection happened and what the outcome of the inspection was such that $\mathcal{O}_i = \mathcal{O}_j = \{\text{Nothing, InspectedCompliance, InspectedViolation}\}$. We will consider machine learning predictions about the true state for the regulator separately (see Section 5.5.4). The function $O(s')$ is independent of the agents' actions and characterizes how the observations are

Table 5.4: Transition probabilities $T_j(s, a_j, s')$ for each a_j

a_j	Wait	Effort	Self-Inspect
	$\begin{bmatrix} 0.917 & 0.033 & 0.048 & 0.002 \\ 0.000 & 0.950 & 0.000 & 0.050 \\ 0.917 & 0.033 & 0.048 & 0.002 \\ 0.908 & 0.042 & 0.048 & 0.002 \end{bmatrix}$	$\begin{bmatrix} 0.926 & 0.024 & 0.049 & 0.001 \\ 0.000 & 0.950 & 0.000 & 0.050 \\ 0.926 & 0.024 & 0.049 & 0.001 \\ 0.917 & 0.033 & 0.048 & 0.002 \end{bmatrix}$	$\begin{bmatrix} 0.95 a & 0.95 b & 0.05 a & 0.05 b \\ 0.95 \mu_j & 0.95 \dot{\mu}_j & 0.05 \mu_j & 0.05 \dot{\mu}_j \\ 0.95 a & 0.95 b & 0.05 a & 0.05 b \\ 0.95 c & 0.95 d & 0.05 c & 0.05 d \end{bmatrix}$

Notes. Appendix D.6 provides detailed information about the computation of these transition probabilities. $\dot{\mu}_j$: $1 - \mu_j$; a: $0.965 + 0.035 * \mu_j$; b: $0.035 * (1 - \mu_j)$; c: $0.955 + 0.045 * \mu_j$; d: $0.045 * (1 - \mu_j)$

generated based on the future state. We assume perfect observability of inspections and no observability otherwise such that:

$$O(s') = \begin{cases} \text{Nothing} & \text{if } s' = \{\text{Compliance, Violation}\} \\ \text{InspectedCompliance} & \text{if } s' = \{\text{InspectedCompliance}\} \\ \text{InspectedViolation} & \text{if } s' = \{\text{InspectedViolation}\}. \end{cases} \quad (5.7)$$

Reward

Rewards depend on the physical states of the environment (\mathcal{S}) and the actions taken. The regulator incurs social costs ($SC_{\text{Violation}}$) when the establishment is in a state of violation, and a cost (C_{Inspect}) for each inspection. Otherwise, the regulator does not “earn” anything from the restaurant’s operation.⁸

We assume that restaurants in the compliance state do not impose social costs. This assumption is based on regulatory logic: if completely complying restaurants continued to produce considerable social costs, the regulator would raise the compliance standard. This view aligns with the interpretation in Jin and Leslie (2003), who argue that while “A”-graded restaurants do not eliminate all foodborne illness risks, their risk level is no greater than the baseline risk of eating at home. Therefore, social costs from violations

⁸It is easy to account for the social benefit of a restaurant by adding a positive value SB to the reward formulation. However, since SB would apply uniformly across all state-action combinations, it does not affect the policy and can be omitted without loss of generality.

should be interpreted as relative to this outside option. We define the regulator's reward function as:

$$R_i(s', a_i) = \begin{cases} 0 & \text{if } s = \text{Compliance}, a_i = \text{Wait} \\ -C_{Inspect} & \text{if } s = \text{Compliance}, a_i = \text{Inspect} \\ -SC_{Violation} & \text{if } s = \text{Violation}, a_i = \text{Wait} \\ -(SC_{Violation} + C_{Inspect}) & \text{if } s = \text{Violation}, a_i = \text{Inspect} \\ 0 & \text{if } s = \text{InspectedCompliance}, a_i = \text{Wait} \\ -C_{Inspect} & \text{if } s = \text{InspectedCompliance}, a_i = \text{Inspect} \\ -SC_{Violation} & \text{if } s = \text{InspectedViolation}, a_i = \text{Wait} \\ -C_{Inspect} & \text{if } s = \text{InspectedViolation}, a_i = \text{Inspect}. \end{cases} \quad (5.8)$$

In contrast, the restaurant's reward function can be formulated as:

$$R_j(s', a_j) = \begin{cases} Profit & \text{if } s = \text{Compliance}, a_j = \text{Wait} \\ Profit - C_{Effort} & \text{if } s = \text{Compliance}, a_j = \text{Effort} \\ Profit - C_{SelfInspect} & \text{if } s = \text{Compliance}, a_j = \text{Inspect} \\ Profit & \text{if } s = \text{Violation}, a_j = \text{Wait} \\ Profit - C_{Effort} & \text{if } s = \text{Violation}, a_j = \text{Effort} \\ Profit - C_{SelfInspect} & \text{if } s = \text{Violation}, a_j = \text{Inspect} \\ Profit & \text{if } s = \text{InspectedCompliance}, a_j = \text{Wait} \\ Profit - C_{Effort} & \text{if } s = \text{InspectedCompliance}, a_j = \text{Effort} \\ Profit - C_{SelfInspect} & \text{if } s = \text{InspectedCompliance}, a_j = \text{Inspect} \\ Profit - Penalty & \text{if } s = \text{InspectedViolation}, a_j = \text{Wait} \\ Profit - Penalty - C_{Effort} & \text{if } s = \text{InspectedViolation}, a_j = \text{Effort} \\ Profit - Penalty - C_{SelfInspect} & \text{if } s = \text{InspectedViolation}, a_j = \text{Inspect}, \end{cases} \quad (5.9)$$

where the *Profit* is the difference between revenue and costs from operating the restaurant, C_{Effort} is the cost associated with effort, $C_{SelfInspect}$ is the cost associated with self-inspecting, and *Penalty* captures the loss of revenue associated with receiving a downgrade plus potential fees.⁹

⁹In reality, the revenue loss from a downgrade is not contained to the period of detection. However, explicitly modeling this temporal effect would introduce non-stationarity. In following Deo et al. (2013), we therefore assume that the penalty occurs immediately.

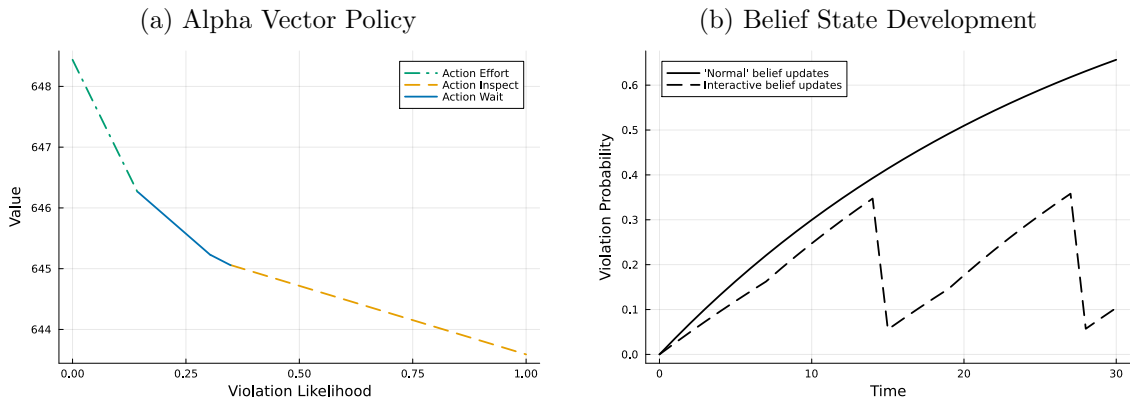
5.5.2 Establishment Problem

The establishment’s objective is to maximize the reward associated with operating the restaurant for some feasible policy π . Formally:

$$\max_{\pi} \mathbb{E} \left[\sum_{t=1}^T \gamma^t R(b'_{j,t}, a_{j,t}) \right]. \quad (5.10)$$

Because the establishment is a 0-level agent, meaning it does not model the regulator’s behavior but incorporates their behavior as noise, its problem reduces to a POMDP (Gmytrasiewicz and Doshi 2005). While POMDPs remain computationally hard to solve, several exact and approximate value iteration approaches exist that compute policies in reasonable times. Examples include the Two-Pass algorithm (Sodnik 1971), Witness algorithm (Littman et al. 1995), point-based value iteration (PBVI) (Shani et al. 2013, Pineau et al. 2003), Successive Approximations of the Reachable Space under Optimal Policies (SARSOP) (Kurniawati et al. 2009)). We use PBVI implemented via the Julia package “PointBasedValueIteration” to compute the establishment’s policy (see Figure 5.6a).

Figure 5.5: Establishment Policy and Belief Updates



Notes. Subfigure (a) shows an establishment’s alpha vector policy and subfigure (b) shows the belief state development considering “normal” and interactive belief updates. The figures are computed for the following values: $\mu_j = 0.85$, $C_{Effort} = 0.1$, $C_{SelfInspect} = 5.0$, $Profit = 20.0$, $Penalty = 40.0$, $\gamma = 0.99$.

The policy has two thresholds: \bar{b}_{Wait} and $\bar{b}_{Violation}$. If the violation likelihood is low, establishment's will exert effort to reduce their transition to a violation state (i.e., $a_j = \text{Effort}$). After the violation likelihood increases above \bar{b}_{Wait} , the establishment will stop to invest effort and instead wait (i.e., $a_j = \text{Wait}$). After the violation likelihood increases above the second threshold $\bar{b}_{Violation}$, the establishment will conduct a self-inspection (i.e., $a_j = \text{Wait}$) to identify and fix violations. While the exact policy in Figure 5.6a depends on the parameters of the problem, we observe the following relationships: \bar{b}_{Wait} depends on the cost of effort such that it decreases (increases) as the cost of effort increases (decreases). $\bar{b}_{Violation}$ decreases as the penalty for violations increases. $\bar{b}_{Violation}$ increases as the self-inspection cost increases. $\bar{b}_{Violation}$ is independent of the profit. $\bar{b}_{Violation}$ decreases as the likelihood of detecting violations during self-inspection (μ_j) increases.

Figure 5.6b highlights the development of the belief state $b_{i,t}^j$ assuming no inspections from the regulator. It shows that accounting for the establishment's actions (i.e., $a_{j,t} \in \{\text{Wait}, \text{Effort}, \text{Self-Inspect}\}$) changes the belief about the establishment's true state compared to assuming a static establishment (i.e., $a_{j,t} \in \{\text{Wait}\}$).

5.5.3 Regulator Problem

After establishing the optimal policy for a single establishment, we now focus on the regulator's decision problem.

Capacity Constraint

Thus far, we have considered the relationship between a regulator and a single establishment. However, regulators oversee many establishments and have limited inspection resources (Mejia et al. 2019). Thus, the regulator must make the inspection decisions while considering all establishments jointly, subject to the constraint that the inspection cost cannot exceed a given capacity constraint K .

Let N be the number of establishments, each following the dynamics discussed in the

preceding subsections, $\mathbf{a}_t = (a_{i,t}^1, a_{i,t}^2, \dots, a_{i,t}^N)$ be a vector capturing the regulator's decisions in period t , and $\mathbf{B}_t = (b_{i,t}^1, b_{i,t}^2, \dots, b_{i,t}^N)$ be a vector of vectors indicating her belief across the states of all N decision problems at period t . Formally, the regulator must choose \mathbf{a}_t such that:

$$\sum_{j=1}^N 1_{\{a_{i,t}^j = \text{Inspect}\}} \leq K, \forall t \in \{1, 2, \dots, T\}. \quad (5.11)$$

This setup implies that we have N independent decision problems that are weakly linked by the joint capacity constraint (Kamalzadeh et al. 2021, Hawkins 2003).

Objective

The regulator aims to minimize the total social costs associated with hygiene violations by strategically allocating inspection resources. This is equivalent to maximizing the total discounted reward. The regulator seeks to find an optimal policy π that governs inspection decisions over time, balancing immediate costs with long-term benefits. Formally, the objective function is given by:

$$\max_{\pi} \mathbb{E} \left[\sum_{t=1}^T \sum_{j=1}^N \gamma^t R(b'_{i,t}{}^j, a_{i,t}^j) \right], \quad (5.12)$$

where $b'_{i,t}{}^j$ represents the belief about the hygiene compliance state of establishment j at the end of time t given the corresponding inspection action $a_{i,t}^j$. $R(\bullet, \bullet)$ denotes the reward function capturing the impact of compliance and inspection decisions, and $\gamma \in (0, 1)$ is the discount factor reflecting the regulator's preference for current versus future costs.

Value Function and Optimality Equations

To model the regulator's decision-making process as a dynamic optimization problem, we define a value function that quantifies the expected total discounted reward from a given system state onward, assuming optimal decisions in all future periods. The regulator's

goal is to allocate inspection resources in a way that maximizes this value function while adhering to capacity constraints.

Let $J(\mathbf{B}_t)$ represent the maximum expected discounted reward from period t onward, given that the belief vector at the start of period t is \mathbf{B}_t . The optimal policy follows from solving the following constrained Bellman equation (Puterman 2005, Hawkins 2003):

$$J(\mathbf{B}_t) = \max_{\mathbf{a}_t} \left\{ \sum_{j=1}^N R(b'_{i,t}{}^j, a_{i,t}^j) + \gamma \mathbb{E}[J(\mathbf{B}_{t+1}) \mid \mathbf{B}_t, \mathbf{a}_t] \mid \sum_{j=1}^N 1_{\{a_{i,t}^j = \text{Inspect}\}} \leq K \right\}. \quad (5.13)$$

The capacity constraint ensures that inspections are limited to at most K establishments in each period. The optimization problem described above is a weakly coupled dynamic optimization problem (Hawkins 2003), where individual establishment-level decisions are connected through the shared inspection resource constraint (Kamalzadeh et al. 2021). To address the problem, we apply Lagrangian relaxation, which transforms the constrained optimization problem into an unconstrained one by introducing a dual variable λ that penalizes violations of the capacity constraint (Kamalzadeh et al. 2021, Hawkins 2003, Bertsekas 1976). The Lagrangian function for period t is given by:

$$\begin{aligned} L(\mathbf{B}_t, \lambda) &= \max_{\mathbf{a}_t} \left\{ \sum_{j=1}^N R(b'_{i,t}{}^j, a_{i,t}^j) + \gamma \mathbb{E}[L(\mathbf{B}_{t+1}) \mid \mathbf{B}_t, \mathbf{a}_t] - \lambda \left(\sum_{j=1}^N 1_{\{a_{i,t}^j = \text{Inspect}\}} - K \right) \right\} \\ &= \max_{\mathbf{a}_t} \left\{ \sum_{j=1}^N \left(R(b'_{i,t}{}^j, a_{i,t}^j) - \lambda 1_{\{a_{i,t}^j = \text{Inspect}\}} \right) + \gamma \mathbb{E}[L(\mathbf{B}_{t+1}) \mid \mathbf{B}_t, \mathbf{a}_t] + \lambda K \right\}. \end{aligned} \quad (5.14)$$

Here, the term $\lambda 1_{\{a_{i,t}^j = \text{Inspect}\}}$ effectively reduces the reward for selecting the inspection action, discouraging inspections when resources are constrained. Because λ is identical across establishments and time periods, the problem decomposes into independent subproblems for each establishment (Hawkins 2003). Specifically, we can rewrite Equation

(5.14) as:

$$L(\mathbf{B}_t, \lambda) = \sum_{j=1}^N L(b'_{i,t}{}^j, \lambda) + \frac{\lambda K}{1 - \gamma}, \quad (5.15)$$

where each establishment-specific function is given by:

$$L(b'_{i,t}{}^j, \lambda) = \max_{a_{i,t}^j} \left\{ \left(R(b'_{i,t}{}^j, a_{i,t}^j) - \lambda 1_{\{a_{i,t}^j = \text{Inspect}\}} \right) + \gamma \mathbb{E} \left[L(B_{j,t+1}) \mid b'_{i,t}{}^j, a_{i,t}^j \right] \right\}. \quad (5.16)$$

This reformulation enables solving for each establishment's optimal policy independently while accounting for the global constraint through the dual variable λ . Further, we can interpret the term $R(b'_{i,t}{}^j, a_{i,t}^j) - \lambda 1_{\{a_{i,t}^j = \text{Inspect}\}}$ as a modified reward function that incorporates a penalty for inspections:

$$R'_i(s', a_i) = \begin{cases} 0 & \text{if } s = \text{Compliance}, a_i = \text{Wait} \\ -(C_{\text{Inspect}} + \lambda) & \text{if } s = \text{Compliance}, a_i = \text{Inspect} \\ -SC_{\text{Violation}} & \text{if } s = \text{Violation}, a_i = \text{Wait} \\ -(SC_{\text{Violation}} + C_{\text{Inspect}} + \lambda) & \text{if } s = \text{Violation}, a_i = \text{Inspect} \\ 0 & \text{if } s = \text{InspectedCompliance}, a_i = \text{Wait} \\ -(C_{\text{Inspect}} + \lambda) & \text{if } s = \text{InspectedCompliance}, a_i = \text{Inspect} \\ -SC_{\text{Violation}} & \text{if } s = \text{InspectedViolation}, a_i = \text{Wait} \\ -(SC_{\text{Violation}} + C_{\text{Inspect}} + \lambda) & \text{if } s = \text{InspectedViolation}, a_i = \text{Inspect}. \end{cases}$$

Using this modified reward function, equation (5.16) can be rewritten as:

$$L(b'_{i,t}{}^j, \lambda) = \max_{a_{i,t}^j} \left\{ R'(b'_{i,t}{}^j, a_{i,t}^j) + \gamma \mathbb{E} \left[L(B_{j,t+1}) \mid b'_{i,t}{}^j, a_{i,t}^j \right] \right\}. \quad (5.17)$$

This allows us to solve the optimization problem for a fixed λ using established solving techniques such as Interactive Point Based Value Iteration (I-PBVI) (Doshi and Perez 2008).

INDEX Policy

We then choose the value for λ that satisfies the capacity constraint in Equation (5.11) using an adaption of Hawkins's (2003) INDEX heuristic. Specifically, we use Algorithm 1, which uses binary search to find a λ for which $a_{i,t}^j = \text{Inspect}$ for exactly K establishments. To restrict the runtime of the algorithm and break ties, we use Algorithm 2 if λ converges to a small enough region.

Algorithm 1 INDEX (Adapted from Hawkins (2003))

Input: $\mathcal{S} = \emptyset$, ϵ , $\lambda_L = \Lambda$, $\lambda_U = -\Lambda$, $\Lambda = R_{min}$

- 1: **while** $|\mathcal{S}| \neq K$ **do**
- 2: $\lambda_M = \frac{\lambda_L - \lambda_U}{2}$
- 3: **for** $j = 1, \dots, N$ **do**
- 4: **if** $L_{a_{i,t}^j = \text{Inspect}}(b_{i,t}^j, \lambda) \geq L_{a_{i,t}^j = \text{Wait}}(b_{i,t}^j, \lambda)$ **then**
- 5: $\mathcal{S} \leftarrow \mathcal{S} \cup \{j\}$
- 6: **end if**
- 7: **end for**
- 8: **if** $|\mathcal{S}| < K$ **then**
- 9: $\lambda_U = \lambda_M$
- 10: **else if** $|\mathcal{S}| > K$ **then**
- 11: $\lambda_L = \lambda_M$
- 12: **end if**
- 13: **if** $\lambda_U - \lambda_L < \epsilon$ **then**
- 14: $\lambda_M = \frac{\lambda_L - \lambda_U}{2}$
- 15: $\mathcal{S} \leftarrow \text{RANK}()$
- 16: **end if**
- 17: **end while**
- 18: **for** $j = 1, \dots, N$ **do**
- 19: **if** $j \in \mathcal{S}$ **then**
- 20: $a_{i,t}^j = \text{Inspect}$
- 21: **else**
- 22: $a_{i,t}^j = \text{Wait}$
- 23: **end if**
- 24: **end for**

Algorithm 2 RANK (Adapted from Hawkins (2003))

Input: $\mathcal{S} = \emptyset$, $\mathcal{N} = 1, \dots, N$, λ_M

1: **while** $\mathcal{S} \neq K$ **do**

2: $j = \operatorname{argmax} \left\{ L_{a_{i,t}^j = \text{Inspect}}(b_{i,t}^j, \lambda) - L_{a_{i,t}^j = \text{Wait}}(b_{i,t}^j, \lambda) \mid j \in \mathcal{N} \setminus \mathcal{S} \right\}$

3: $\mathcal{S} \leftarrow \mathcal{S} \cup \{j\}$

4: **end while**

5: **return** \mathcal{S}

5.5.4 Incorporating Machine Learning Predictions

As outlined in Section 5.4, regulators can leverage machine learning techniques to enhance their assessment of an establishment’s compliance status. These predictive models refine the belief vector by incorporating probabilistic updates, rather than relying solely on direct observations, such as inspection outcomes. Unlike traditional observations, machine learning predictions aggregate information across multiple decision periods, requiring a distinct treatment within our framework.

To illustrate this distinction, consider the role of the *Fast Food* category in predicting regulatory violations. As demonstrated in Section 5.4.3, this category significantly influences predictions. However, because *Fast Food* is a static, time-invariant feature, treating machine learning predictions as direct observations—where the belief vector is updated in each period—would artificially inflate the probability of a violation. This issue arises because the same static information would be repeatedly incorporated into belief updates, distorting the regulator’s decision-making process. To mitigate this effect, we update only the current belief vector $b_{i,t}^j$ at the time of decision-making.

Let $p_{i,t}^j$ denote the regulator’s machine learning prediction for the compliance state of establishment j at time t . Suppose the model assigns probabilities of 95% to *Compliance*, 5% to *Violation*, and 0% to both *InspectedCompliance* and *InspectedViolation*. Then, the prediction vector is:

$$p_{i,t}^j = [0.95, 0.05, 0.00, 0.00].$$

From this we compute the likelihood of the predicted probabilities under each possible compliance state using the Kernel density estimates from Figure 5.4b. Specifically, the likelihood of observing $p_{i,t}^j(\text{Compliance})$ given that the establishment is in the *Compliance* state is:

$$Pr(p_{i,t}^j(\text{Compliance})|\text{Compliance}) = f_C(p_{i,t}^j(\text{Compliance})),$$

where f_C represents the probability density function of the true-compliance distribution. Similarly, the likelihood of observing $p_{i,t}^j(\text{Violation})$ given that the establishment is in the *Violation* state is:

$$P(p_{i,t}^j(\text{Violation})|\text{Violation}) = f_V(p_{i,t}^j(\text{Violation})),$$

where f_V represents the probability density function of the true-violation distribution. The likelihood of observing $p_{i,t}^j(\text{InspectedCompliance})$ is 0 if $p_{i,t}^j(\text{InspectedCompliance}) = 0$ and 1 if $p_{i,t}^j(\text{InspectedCompliance}) = 1$, with an analogous condition applying to $p_{i,t}^j(\text{InspectedViolation})$. To formally integrate these predictions into the belief update process, we employ Bayes' theorem in its odds form (Downey 2013):

$$o(H | E) = o(H) \frac{P(E | H)}{P(E | \neg H)}, \quad (5.18)$$

where $o(H) = \frac{P(H)}{P(\neg H)}$ represents the prior odds in favor of hypothesis H , characterizing the prior belief before new evidence is incorporated. The term $\frac{P(E|H)}{P(E|\neg H)}$ quantifies the relative impact of the observed evidence E on updating the belief in H . A factor greater than 1 strengthens belief in H , while a value less than 1 weakens it.

Applying this framework to the compliance monitoring problem, we update the belief

for each state is as follows:

$$o(is | p_{i,t}^j) = o(is) \frac{Pr(p_{i,t}^j(is) | is)}{\sum_{\substack{is' \in \mathcal{IS} \\ is' \neq is}} Pr(p_{i,t}^j(is') | is')}, \quad \forall is \in \mathcal{IS}, \quad (5.19)$$

where $o(is) = \frac{b_{i,t}^j(is)}{1 - b_{i,t}^j(is)}$ and $Pr(p_{i,t}^j(is) | is)$ comes from the likelihood functions derived from the machine learning model's predictive distributions. The posterior odds are subsequently transformed to probabilities and normalized to yield the updated belief vector $\tilde{b}_{i,t}^j$. To integrate these updated beliefs into the decision-making framework, we redefine the belief vector \mathbf{B}_t in the optimization problem:

$$\tilde{\mathbf{B}}_t = [\tilde{b}_{i,t}^1, \tilde{b}_{i,t}^2, \dots, \tilde{b}_{i,t}^N]. \quad (5.20)$$

We then use this updated belief vector in line (4) of the INDEX algorithm, and label the resulting policy: *ML-INDEX*.

5.6 Evaluation Using Simulation

In this section, we use simulations to analyze how incorporating the machine learning predictions from Section 5.4 in the inspection resource allocation policy from Section 5.5 affects the social welfare costs. We first summarize the different policies, then describe the simulation setup, and finally discuss the results.

5.6.1 Policies

We evaluate the performance of five policies: Random, Regular, Predictions, INDEX, and ML-INDEX. The *Random* policy is inspired by Kim and Xu (2024), who compare the performance of random and periodic policies to prevent the occurrence of risk events in banking. In every period, we randomly choose K —the capacity limit—establishments

to inspect. Each establishment is equally likely to be inspected in every period. The *Regular* policy is inspired by the current practice in restaurant inspections. In every period, we choose to inspect the K establishments with the longest time since the last inspection. This results in a regular inspection interval. The *Predictions* policy is inspired by Johnson et al. (2023), who rely exclusively on Machine Learning predictions in assigning OSHA inspections. In every period, we first order the establishments by their predicted Violation likelihood.¹⁰ We then choose to inspect the K establishments with the highest violation likelihood without considering any other factors. The *INDEX* policy is described in Section 5.5.3. In every period, we choose to inspect a subset \mathcal{S} of establishments which we identify using Algorithm 1. The *ML-INDEX* policy integrates Machine Learning predictions in the inspection resource allocation decision as described in Section 5.5.4. In every period, we first update the belief state based on the prediction and then use Algorithm 1 to select which establishments to inspect.

5.6.2 Simulation Design

For our simulation, we consider a finite horizon consisting of 50 one-week periods. A one-week look-ahead schedule is appropriate for restaurant inspections because they are conducted unannounced and all within close proximity. No travel arrangements are needed. If, in other settings, the look-ahead schedule is longer, then the estimation period can be extended. We consider four contexts for our simulation: a homogeneous establishment population with low inspection capacity, a homogeneous establishment population with high inspection capacity, a heterogeneous establishment population with low inspection capacity, and a heterogeneous establishment population with high inspection capacity. We

¹⁰We generate the predictions by sampling from the Kernel density estimate corresponding to the true state of the establishment. To ensure that repeated predictions for the same establishment are correlated, we use a quantile transformation approach. Each establishment is assigned a seed value from a uniform random variable upon initialization and after inspections (i.e., the seed resets). In each period, we then augment the seed with Gaussian noise ($\mu = 0, \sigma = 0.2$) before creating a prediction based on mapping the seeds quantile to the corresponding value of the prediction distribution. This method introduces controlled randomness while preserving a consistent but non-deterministic relationship between related predictions.

repeat each setting eight times with different random seeds to allow statistical inference.

Table 5.5 summarizes the simulation parameters.

Table 5.5: Simulation Setup for Establishment Problems

Context	Parameter	Case	Values
Homogeneous	Establishments	–	160
	Time since last inspection at beginning	–	$\mathcal{U}(5, 20)$
	Profit	–	20
	Penalty	–	40
	Self inspection cost ($C_{SelfInspect}$)	–	3
	Cost of effort (C_{Effort})	–	0.1
	Self-inspection correction probability (μ_j)	–	0.85
	Social costs ($SC_{Violation}$)	–	16
	Self inspection cost ($C_{SelfInspect}$)	–	3
	Capacity (K)	Low High	3 9
Heterogeneous	Establishments	–	160
	Time since last inspection at beginning	–	$\mathcal{U}(5, 20)$
	Profit	–	$20 + \mathcal{U}(-5, 5)$
	Penalty	–	$40 + \mathcal{U}(-5, 5)$
	Self inspection cost ($C_{SelfInspect}$)	–	$3 + \mathcal{U}(-1, 1)$
	Cost of effort (C_{Effort})	–	$0.1 + \mathcal{U}(-0.1, 0.1)$
	Self-inspection correction probability (μ_j)	–	$0.85 + \mathcal{U}(-0.01, 0.01)$
	Social costs ($SC_{Violation}$)	–	$16 + \mathcal{U}(-1, 1)$
	Capacity (K)	Low High	3 9

Notes. The social cost of non-compliance ($SC_{Violation} = \$16$ per restaurant per week) is based on the empirical effects of Los Angeles’ restaurant grading policy (Jin and Leslie 2003). Following the introduction of hygiene cards, approximately 3,500 restaurants improved their compliance, leading to a 20% reduction in 400 annual hospitalizations. Combining hospitalization costs (estimated at \$17,500 per case (Dhaliwal et al. 2021)) and productivity losses from non-hospitalized foodborne illnesses (based on 0.5 days of missed work at \$60,000 GDP per capita and a 234:1 ratio of illnesses to hospitalizations (Mead et al. 1999)), the weekly social cost per non-compliant restaurant is estimated as the sum of hospitalization cost (i.e., $\$17,500 \times 0.00044 = \7.70) and productivity loss (i.e., $\$60,000 \div 365 \times 0.5 \times 0.1 = \8.30).

Each simulation consists of 160 establishments. In both contexts, the time since the last inspection is randomly assigned based on a discrete uniform distribution with integer values between five and 20. The true establishment state is sampled randomly with state probabilities based on the time since the last inspection and the transition probabilities for

$a_i = Wait$ in Section 5.5.1. In the homogeneous context, all establishments share the same parameters for profit, penalty, self-inspection correction probability (μ_j), self-inspection cost ($C_{SelfInspection}$), effort cost (C_{Effort}), and social costs ($SC_{Violation}$). In the heterogeneous context, the parameter values vary between establishments based on a noise factor from a uniform distribution. In both contexts, we evaluate a low- and high-capacity case that allows for three and nine inspections per period, respectively. This is consistent with the inspection capacity to establishment ratio in Las Vegas.

5.6.3 Results

Policy Comparison

We use two metrics to evaluate the performance of the five policies. First, we report *Reward* as the cumulative discounted reward of employing policy m , computed as:

$$\Phi^m = \sum_{t=1}^T \sum_{j=1}^N \gamma^t R(s'_{j,t}, a_{i,t}^{j*}), \quad (5.21)$$

where $s'_{j,t} \in \mathcal{S}$ is the true state of establishment j at the end of period t and $a_{i,t}^{j*}$ is the inspection decision according to policy m for j in period t . Second, we report *Relative Improvement* as the relative improvement in net benefit of policy m compared to policy n (Cui et al. 2018, Deo et al. 2013). We compute *Relative Improvement* as:

$$\delta_{m,n} = \frac{(\Phi^m - \Phi^{Baseline}) - (\Phi^n - \Phi^{Baseline})}{\Phi^n - \Phi^{Baseline}} = \frac{\Phi^m - \Phi^n}{\Phi^n - \Phi^{Baseline}}, \quad (5.22)$$

where $\Phi^{Baseline}$ is the cumulative discounted reward of a no-inspection baseline (i.e., running the simulation without conducting any inspections). Table 5.6 provides the pairwise comparisons of all policies using these two measures.

The results demonstrate that incorporating machine learning predictions can substantially improve inspection resource allocation. In Panel A, we observe significant

Table 5.6: Pairwise Comparisons of Policies.

	Regular	ML	INDEX	ML INDEX
Panel A: Cumulative discounted reward				
Random	-320.78 ± 486.70	$-1988.46 \pm 390.17^{***}$	$-3824.93 \pm 392.71^{***}$	$-4217.61 \pm 442.26^{***}$
Regular		$-1667.68 \pm 570.97^{***}$	$-3504.15 \pm 594.92^{***}$	$-3896.83 \pm 689.98^{***}$
Predictions			$-1836.47 \pm 450.19^{***}$	$-2229.15 \pm 442.88^{***}$
INDEX				$-392.68 \pm 209.64^{***}$
Panel B: Relative improvement over baseline				
Random	0.03 ± 0.13	$0.67 \pm 0.25^{***}$	$1.29 \pm 0.44^{***}$	$1.46 \pm 0.52^{***}$
Regular		$0.78 \pm 0.41^{***}$	$1.46 \pm 0.61^{***}$	$1.66 \pm 0.72^{***}$
Predictions			$0.36 \pm 0.13^{***}$	$0.45 \pm 0.14^{***}$
INDEX				$0.06 \pm 0.04^{***}$

Notes. The table presents individual 99% confidence intervals for all pairwise comparisons of policies based on the raw simulation results (see Table D.4). Since we conduct $5(5 - 1)/2 = 10$ comparisons, we use a 99% confidence interval for each comparison to achieve 90% overall confidence (Law 2013). Significance levels from a paired-t test (Panel A) and t-test (Panel B) are indicated as: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

differences in cumulative discounted rewards across policies. The only exception is the comparison between the *Random* and *Regular* policies, which shows no statistically significant difference. Each successive policy yields a statistically significant improvement over the previous one, establishing a clear performance hierarchy: *Random/Regular* < *Predictions* < *INDEX* < *ML-INDEX*.

The *Predictions* policy—based solely on predicted violation likelihoods—performs better than both the *Random* and *Regular* policies, yet it is consistently outperformed by the *INDEX*-based approaches. The *ML-INDEX* policy achieves the best outcomes overall, reducing average social costs by more than \$4,000 compared to random inspections, over \$2,000 relative to the *Predictions* policy, and more than \$390 relative to the *INDEX* policy. These differences are statistically significant at the 0.001% level, underscoring the robustness of the improvements.

Panel B reports relative improvements over a no-inspection baseline, which is less sensitive to specific parameter settings in the simulation (Deo et al. 2013). These results reinforce the findings in Panel A in that the *ML-INDEX* policy consistently delivers the

highest relative improvement. On average, it reduces social costs by 166 percent relative to the *Regular* policy and by 6 percent relative to the *INDEX* policy. The differences are again statistically significant.

Together, the results highlight that while machine learning predictions can be valuable in informing inspection resource allocation decisions, their greatest value emerges when they are combined with a dynamic optimization framework. The consistent pattern of statistical significance across both panels highlights the strength of the proposed *ML-INDEX* policy compared to existing alternatives. These findings suggest that regulators can realize substantial efficiency gains by combining predictive analytics with model-based planning in their inspection strategies.

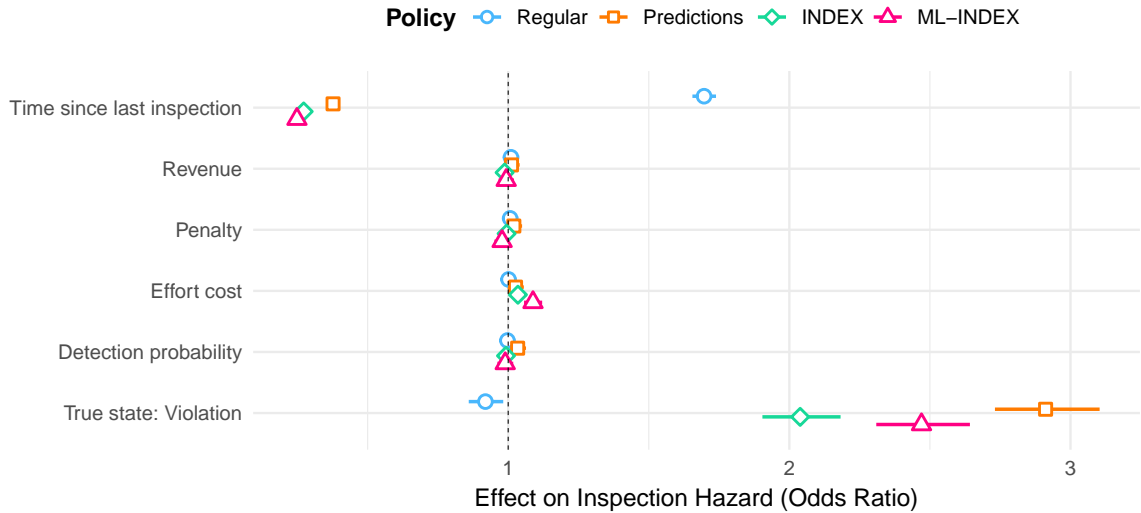
Additional Observations

To understand the drivers underlying the improvements in performance among the different policies, we investigate how the *Regular*, *Predictions*, *INDEX*, and *ML-INDEX* policies make inspection decisions based on the establishment characteristics. Figure 5.7 presents the results of a discrete-time survival analysis linking the inspection likelihood to the I-POMDP characteristics.

First, the time since the last inspection is positively associated with inspection probability under the *Regular* policy, reflecting a basic heuristic that prioritizes establishments that have not been inspected recently. In contrast, this association is negative for the *Predictions*, *INDEX*, and *ML-INDEX* policies. These policies do not follow a regular inspection interval. Instead, they focus inspections on higher-risk facilities, resulting in some establishments being systematically inspected less frequently, which increases their time since the last inspection even as their inspection probability remains low.

Second, among all establishment-level parameters varied in the simulation, effort cost has the strongest positive association with inspection likelihood. Establishments with higher

Figure 5.7: Discrete-Time Survival Analysis of Inspections



Notes. The figure presents the coefficients of a Discrete-Time Survival Analysis with inspection decisions as the dependent variable based on the heterogeneous simulations. The effect sizes are exponentiated and standardized estimates, meaning that a one standard deviation increase in the parameter is associated with an increase of $effect - 1$ in the probability of an inspection.

effort costs are less incentivized to invest in maintaining compliance, leading to a faster decay in hygiene compliance. This decay increases the probability that these establishments enter violation states, thereby raising their likelihood of inspection under risk-based policies such as *Predictions*, *INDEX*, and *ML-INDEX*.

Third, the inspection probability under the *Predictions*, *INDEX*, and *ML-INDEX* policies is strongly positively associated with an establishment's true violation state, but the mechanisms behind this relationship differ across policies. The *Predictions* policy relies on machine learning outputs, which are directly influenced by changes in the underlying state, leading to responsive increases in inspection probability when an establishment violates. In contrast, the *INDEX* policy uses model-based inference of compliance decay to target establishments in violation without observing the true situation. The *ML-INDEX* policy combines the model-based structure of *INDEX* with machine learning predictions, resulting in an intermediate effect size between the two extremes, as illustrated in Figure 5.7.

5.7 Discussion and Conclusion

This study explores how regulators can leverage machine learning to improve inspection resource allocation, particularly in the context of restaurant hygiene inspections. We propose an integrated framework that combines predictive analytics—using gradient-boosted regression trees (GBRT)—with prescriptive decision-making, formulated as a capacity-constrained interactive partially observable Markov decision process (I-POMDP). The resulting *ML-INDEX* policy effectively combines model-based planning with data-driven predictions.

Applying this framework to Las Vegas restaurant inspection data and Yelp reviews (2012–2019), we show that it is not only possible to predict inspection outcomes, but also how embedding these predictions in a resource allocation model improves regulatory performance. The *ML-INDEX* policy significantly increases inspection program effectiveness by up to 160% compared to standard inspection programs. This improvement is achieved by tailoring inspection intervals to each establishment’s predicted risk and modeled conformance decay.

Our research makes several contributions across disciplinary boundaries. First, we extend the literature on predicting organizational misconduct (e.g., Ball et al. 2025, Campbell and Shang 2022, Mejia et al. 2019, Chang and Wu 2021, Cecchini et al. 2010) by demonstrating how machine learning predictions can be operationalized through a structured decision framework. Second, we introduce I-POMDPs to the operations management (OM) community as a promising framework for modeling multi-agent decision-making under uncertainty—especially relevant in regulatory contexts. Third, we contribute to the literature on organizational misconduct by formalizing the dynamics of routine conformance decay as a Markov process (Anand et al. 2012, Vaughan 1999), enabling simulation-based evaluations of inspection strategies and addressing long-standing calls to study misconduct prevention strategies more rigorously (Palmer 2012, Anand et al. 2023).

Finally, we inform the literature on social and environmental responsibility audits (e.g., Ha et al. 2023, Zhang et al. 2022, Caro et al. 2018) by demonstrating that risk predictions can be effectively used to allocate inspection resources in oversight relationships, such as those between regulators and firms or buyers and suppliers.

From a practical standpoint, our ML-INDEX policy is readily implementable using open-source tools such as R, Python, or Julia. It provides a scalable and adaptive strategy that allows regulators to improve misconduct detection and prevention with existing resources or to retain current levels of inspection effectiveness even with reduced personnel. This last point is especially important as governments are trying to reduce their overhead costs by reducing government-agency personnel. Moreover, while our study focuses on restaurant hygiene, the policy generalizes to other regulatory contexts. Potential applications include inspection programs run by the Occupational Safety and Health Administration (OSHA), the U.S. Food and Drug Administration (FDA), the Environmental Protection Agency (EPA), the Centers for Medicare & Medicaid Services (CMS), the National Highway Traffic Safety Administration (NHTSA), and their international counterparts.

While our study offers valuable contributions, it is not without limitations. First, while our simulation-based evaluation provides a robust assessment of the ML-INDEX policy’s potential, it does not fully capture the complexity of real-world inspection dynamics. Future empirical validation through field experiments would be important to substantiate the real-world impact of ML-enhanced inspection resource allocation policies. Second, our model restricts the strategy level of establishments, meaning that they do not explicitly model the regulator’s behavior but base their decisions on past inspection patterns. Extending the model to higher strategic interaction levels thus provides an interesting future research opportunity. Third, using public data to inform inspection programs is not without risks. Establishments might “game the system” and use negative Yelp reviews to increase inspection scrutiny for their competitors. Such Yelp review fraud (Luca and Zervas 2016)

would undermine the effectiveness of our policy if it can't be detected in the prediction step.

In conclusion, this paper demonstrates the power of integrating predictive analytics into regulatory decision-making. By formalizing misconduct dynamics and embedding predictions into a principled resource allocation framework, regulators can move toward more proactive, data-informed enforcement strategies. As regulatory challenges become more complex and data-rich, such tools will be essential for building effective, efficient, and adaptable oversight systems.

Chapter 6

Conclusion and Discussion

6.1 Discussion

This dissertation sought to prevent organizational misconduct by framing it as a high-level concept encompassing several OSCM phenomena (see Figure 2.1 in Chapter 2) and investigating its causes and potential prevention strategies in three contexts: informed trading during the medical device recall process (Chapter 3), environmental injustice in US manufacturing (Chapter 4), and restaurant hygiene violations (Chapter 5). Collectively, these essays addressed two central research questions: (1) Why does organizational misconduct occur in different operational contexts? and (2) How can it be effectively prevented or mitigated?

While each essay focuses on a distinct type of misconduct, they all highlight common preventative strategies, providing a more complete picture of how regulators and businesses might lower the risk of misconduct. Rather than viewing misconduct as an ethical lapse or managerial outlier, I conceptualize it as a process embedded in day-to-day operational decisions, routines, and incentive structures (see also Palmer 2012). From this perspective, misconduct can be prevented both by reshaping organizational systems to reduce the pressures and predispositions that lead to misconduct (e.g., Chapter 4) and by improving

the efficiency and effectiveness of regulatory interventions (e.g., Chapters 3 and 5).

The dissertation's findings not only offer new avenues for OSCM research but also provide practical guidance for firms and regulators looking to prevent and mitigate misconduct. Next, I go over the essays' key findings and discuss their implications for both research and practice. Finally, I outline future research directions that expand on this work and investigate other applications of misconduct prevention in operations management.

6.2 Key Implications

This dissertation offers several key implications for Operations and Supply Chain Management (OSCM) research, management practice, and policymaking. To highlight these implications, I present an overview of my essays in Table 6.1.

In Essay I, I analyzed informed trading by CEOs during the medical device recall. My analysis shows that CEOs increase stock sales during the recall decision process, but only for severe recalls without prior complaints. Doing so helps them avoid financial losses of up to 20%. This presents novel evidence of illegal informed trading in an operational context. These findings show that information asymmetry in operational decisions can be used for personal gain, even in highly regulated environments. This emphasizes the significance of balancing decision-making transparency and executive accountability. For policymakers, the findings highlight an opportunity for regulators like the FDA and SEC to work more closely together. For firms, the findings highlight the necessity of internal controls that minimize opportunities for opportunistic behavior during sensitive operational procedures.

Essay II changed the focus to environmental injustice in the manufacturing sector, as well as firm governance reforms. I find that appointing Chief Sustainability Officers (CSOs) significantly reduces toxic releases, with facilities in underserved communities causing the majority of the reduction. As such, appointing a CSO reduces environmental injustice. This offers practical guidance to firms looking to balance environmental performance with equity

Table 6.1: An Overview of the Dissertation

	Essay I	Essay II	Essay III
Title	Product Recall Decisions and the CEO: Informed Trading in the Medical Device Industry	Chief Sustainability Officers and Environmental Injustice: An Empirical Evaluation	Improving Inspection Resource Allocation to Control Organizational Misconduct
Misconduct Type	Informed trading by CEOs during medical device recalls	Environmental injustice by manufacturing firms	Hygiene regulation violations by restaurants
Research Question	Do CEOs engage in informed trading during the medical device product recall process?	Do Chief Sustainability Officers (CSOs) reduce environmental injustice in toxic releases?	Can machine learning misconduct predictions improve inspection resource allocation?
Data	Product recalls and trading information (2003-2019)	Toxic releases and CSO appointments (2000-2020)	Las Vegas inspection outcomes and YELP reviews (2012-2019)
Methods	Firm-day fixed effects analysis, long-window event study	Difference-in-Differences with matching	Machine learning, analytical model, simulation
Theoretical Contributions	Demonstrates how information asymmetry enables misconduct in operational decision-making	Identifies CSO appointments as a new strategy for addressing environmental injustice	Introduces a predictive and prescriptive approach to regulatory inspections
Practical Implications	Highlights the need for better coordination between the FDA and SEC to detect informed trading during recall decisions	Incentivizes firms to appoint CSOs because they can improve environmental performance in addition to financial benefits	Shows how regulators can integrate machine learning into inspection planning to improve compliance and public health outcomes

issues, particularly in light of increased ESG scrutiny. For regulators, the findings show how firm-level governance models might supplement traditional environmental regulatory interventions to overcome persistent disparities.

In Essay III, I explored how regulators can use machine learning to better allocate hygiene inspection resources in the restaurant industry. The findings show that, when properly integrated into the decision-making process, predictive analytics can significantly increase inspection program efficiency and reduce hygiene violations. This adds to the growing literature on AI-supported operational decision-making and gives regulators a framework for data-driven enforcement planning. The essay highlights how digital footprints (such as customer reviews) can provide early warning signs of potential non-compliance, allowing for more proactive responses via inspections. Furthermore, our proposed policy is not restricted to the restaurant industry but extends to a wide range of contexts where authorities regulate misconduct through inspections, such as OSHA for labor violations or supplier audits for code of conduct violations.

Taken together, the three essays reveal a set of mechanisms—regulatory interventions and firm governance reforms—that can be used to prevent or mitigate organizational misconduct. These mechanisms are not mutually exclusive; instead, they can be included in larger plans that connect operational decisions with legal, ethical, and social norms. This suggests that OSCM scholars should incorporate more explicit consideration of misconduct and prevention into their research agenda. For practitioners and regulators, the dissertation gives practical insights for developing prevention strategies to reduce the occurrence of organizational misconduct.

6.3 Future Research

This dissertation offers a novel operational perspective on organizational misconduct, demonstrating how misaligned incentives, information asymmetries, and decay of routine

adherence can give rise to harmful behaviors within firms. Building on this foundation, my future research will continue to deepen and expand our understanding of misconduct in operational contexts, with the ultimate aim of informing more effective prevention strategies.

Several natural extensions emerge from the dissertation essays. For example, Essay I examines informed trading, a concern that regulators have responded to by requiring CEOs to pre-schedule their trades up to 90 days in advance (Ramonas 2022). This regulatory intervention prompts an important follow-up question: do trade schedules inadvertently incentivize executives to delay bad news—such as product recalls—until after their scheduled trade dates? I plan to study this by combining a stylized model of managerial behavior with data on recall timing and executive stock trades.

A second promising direction concerns the role of supply chain transparency in shaping public reactions to ESG violations. While experimental research indicates that transparency can increase consumer willingness to pay and reduce negative reactions (e.g., Mollenkopf et al. 2022, Kraft et al. 2022), I propose a more nuanced relationship: Transparent firms may face stronger consumer penalties following violations, precisely because their customers expect better behavior. I plan to empirically test this hypothesis using a regression discontinuity in time design and high-frequency sales data.

Beyond these projects, I remain committed to the broader agenda of understanding and preventing misconduct in various operational domains, including product safety issues, employee mistreatment, and environmental violations. My current research conceptualizes misconduct as a unifying concept underlying a wide range of operational misbehaviors (see Chapter 2), and I plan to extend this theory through empirical tests across different settings and industries. In doing so, I intend to contribute not only to the OSCM discipline but also to the broader management and sociology literatures on organizational misconduct. Ultimately, my goal is to provide insights that are practical, offering concrete solutions to societal and economic challenges, and to contribute to a more responsible, ethical, and sustainable business world.

Bibliography

- Adams RB, Ferreira D (2009) Women in the boardroom and their impact on governance and performance. *Journal of financial economics* 94(2):291–309.
- Ambulkar S, Arunachalam S, Bommaraju R, Ramaswami S (2023) Should a firm bring a supplier into the boardroom? *Production and Operations Management* 32(1):28–44.
- Anand A, Rottig D, Parameswar N, Zwerg-Villegas AM (2023) Diving deep into the dark side: A review and examination of research on organizational misconduct in emerging markets. *Business Ethics, the Environment & Responsibility* 32(2):612–637.
- 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.
- Ang E, Kwasnick S, Bayati M, Plambeck EL, Aratow M (2016) Accurate Emergency Department Wait Time Prediction. *Manufacturing & Service Operations Management* 18(1):141–156.
- Arif S, Kepler JD, Schroeder J, Taylor D (2022) Audit process, private information, and insider trading. *Review of Accounting Studies* 27(3):1125–1156.
- Arora P, Hora M, Singhal V, Subramanian R (2020) When do appointments of corporate sustainability executives affect shareholder value? *Journal of Operations Management* 66(4):464–487.
- Arrow KJ (1963) Uncertainty and the Welfare Economics of Medical Care. *American Economic Review* 53(5):941–973.
- Attari I, Helm JE, Mejia J (2024) Hiding Behind Complexity: Supply Chain, Oversight, Race, and the Opioid Crisis. *Production and Operations Management* .
- Bakshi N, Peura H (2022) Prevent or Report? Managing Near Misses for Safer Operations. *Manufacturing & Service Operations Management* 24(4):2064–2080.
- Ball G, Shah R, Donohue K (2018a) The decision to recall: A behavioral investigation in the medical device industry. *Journal of Operations Management* 62(1):1–15.
- Ball G, Siemsen E, Shah R (2017) Do Plant Inspections Predict Future Quality? The Role of Investigator Experience. *Manufacturing & Service Operations Management* 19(4):534–550.
- Ball GP, Bavafa H, Blanco CC, Park H, Wowak KD (2025) Gender and Serious Drug Recalls: A Textual Sentiment Analysis of Drug Reviews on WebMD. *Production and Operations Management* 34(4):698–710.
- Ball GP, Shah R, Wowak KD (2018b) Product competition, managerial discretion, and manufacturing recalls in the U.S. pharmaceutical industry. *Journal of Operations Management* 58-59(1):59–72.
- Ball GP, Wowak KD, Mukherjee UK (2022) Product Recall Research: Dimensions, Methods, and Regulator Implications. Chou M, Gibson H, Staats B, Shier D, Greenberg HJ, eds., *Tutorials*

- in Operations Research: Emerging and Impactful Topics in Operations*, 116–132 (INFORMS), ISBN 978-0-9906153-7-8, <http://dx.doi.org/10.1287/educ.2022.0240>.
- Baucus MS (1994) Pressure, Opportunity and Predisposition: A Multivariate Model of Corporate Illegality. *Journal of Management* 20(4):699–721.
- Baucus MS, Near JP (1991) Can Illegal Corporate Behavior Be Predicted? An Event History Analysis. *Academy of Management Journal* 34(1):9–36.
- Beattie G, Durante R, Knight B, Sen A (2021) Advertising Spending and Media Bias: Evidence from News Coverage of Car Safety Recalls. *Management Science* 67(2):698–719.
- Becker HS (1963) *Outsiders: Studies in the Sociology of Deviance* (Free Press of Glencoe), 1 edition.
- Belavina E, Marinesi S, Tsoukalas G (2020) Rethinking Crowdfunding Platform Design: Mechanisms to Deter Misconduct and Improve Efficiency. *Management Science* 66(11):4980–4997.
- Bennett VM, Pierce L, Snyder JA, Toffel MW (2013) Customer-Driven Misconduct: How Competition Corrupts Business Practices. *Management Science* 59(8):1725–1742.
- Bertsekas DP (1976) On Penalty and Multiplier Methods for Constrained Minimization. *SIAM Journal on Control and Optimization* 14(2):216–235.
- Bifulco GM, Savio R, Paolone F, Tiscini R (2023) The CSR committee as moderator for the ESG score and market value. *Corporate Social Responsibility and Environmental Management* 30(6):3231–3241.
- Blackburne T, Kepler JD, Quinn PJ, Taylor D (2021) Undisclosed SEC Investigations. *Management Science* 67(6):3403–3418.
- Boeing (05/04/2007) Boeing Names Armstrong to Lead New Enterprise Environmental Organization. <https://boeing.mediaroom.com/2007-05-04-Boeing-Names-Armstrong-to-Lead-New-Enterprise-Environmental-Organization>.
- Boeing (2008) 2008 Environment Report. http://www.mdc.com/aboutus/environment/environmental_report/media/pdf/boeing-2008-environment-report.pdf.
- Bouquet C, Morrison A, Birkinshaw J (2009) International attention and multinational enterprise performance. *Journal of International Business Studies* 40(1):108–131.
- Brier GW (1950) VERIFICATION OF FORECASTS EXPRESSED IN TERMS OF PROBABILITY. *Monthly Weather Review* 78(1):1–3.
- Browning TR (2020) Operations management writ large. *Journal of Operations Management* 66(5):494–500.
- Bullard RD (1993) *Confronting environmental racism: Voices from the grassroots* (Boston Mass.: South End Press), 1st ed. edition, ISBN 0896084469.
- Cachon GP (2012) What Is Interesting in Operations Management? *Manufacturing & Service Operations Management* 14(2):166–169.
- California Environmental Protection Agency (2022) Press Release: California holds Boeing accountable for cleanup at toxic Santa Susana Field Laboratory. <https://calepa.ca.gov/2022/05/09/press-release-california-holds-boeing-accountable-for-cleanup-at-toxic-santa-susana-field-laboratory/>.
- Calvo E, Cui R, Wagner L (2023) Disclosing Product Availability in Online Retail. *Manufacturing & Service Operations Management* 25(2):427–447.
- Campbell DW, Shang R (2022) Tone at the Bottom: Measuring Corporate Misconduct Risk from the Text of Employee Reviews. *Management Science* 68(9):7034–7053.

- Caro F, Chintapalli P, Rajaram K, Tang CS (2018) Improving Supplier Compliance Through Joint and Shared Audits with Collective Penalty. *Manufacturing & Service Operations Management* 20(2):363–380.
- Cecchini M, Aytug H, Koehler GJ, Pathak P (2010) Detecting Management Fraud in Public Companies. *Management Science* 56(7):1146–1160.
- Centers for Disease Control and Prevention (2022) Foodborne Illnesses and Germs. <https://www.cdc.gov/foodsafety/foodborne-germs.html#print>.
- Certo ST, Busenbark JR, Woo Hs, Semadeni M (2016) Sample selection bias and Heckman models in strategic management research. *Strategic Management Journal* 37(13):2639–2657.
- Chakraborty J, Maantay JA, Brender JD (2011) Disproportionate proximity to environmental health hazards: methods, models, and measurement. *American journal of public health* 101 Suppl 1(Suppl 1):S27–36.
- Chang CH, Wu Q (2021) Board Networks and Corporate Innovation. *Management Science* 67(6):3618–3654.
- Chang X, Huang Y, Li M, Bo X, Kumar S (2021) Efficient Detection of Environmental Violators: A Big Data Approach. *Production and Operations Management* 30(5):1246–1270.
- Chavis Jr BF, Lee Charles (1987) *Toxic Wastes and Race in the United States* (New York: United Church Christ).
- Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP (2002) SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research* 16:321–357.
- Cheah ET, Chan WL, Chieng CLL (2007) The Corporate Social Responsibility of Pharmaceutical Product Recalls: An Empirical Examination of U.S. and U.K. Markets. *Journal of Business Ethics* 76(4):427–449.
- Chen T, Guestrin C (2016) XGBoost. Krishnapuram B, Shah M, Smola A, Aggarwal C, Shen D, Rastogi R, eds., *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794 (New York, NY, USA: ACM), ISBN 9781450342322, <http://dx.doi.org/10.1145/2939672.2939785>.
- Chen T, He T, Benesty M, Khotilovich V, Tang Y, Cho H, Chen K, Mitchell R, Cano I, Zhou T, Li M, Xie J, Lin M, Geng Y, Li Y, Yuan J (2025) xgboost: Extreme Gradient Boosting. <http://dx.doi.org/10.32614/CRAN.package.xgboost>.
- Chicago Sun-Times (2023) BP refinery hit with \$40 million fine, agrees to spend \$200 million on pollution control. <https://chicago.suntimes.com/2023/5/18/23729343/bp-refinery-hit-40-million-fine-agrees-spend-200-million-pollution-control>.
- Cole D, Muthulingam S (2025) Barriers to Pollution Reduction in Minority Communities — Evidence from Manufacturing Facilities in the United States.
- Collins R (1975) *Conflict Sociology: Toward an explanatory science* (New York: Academic Press), ISBN 0121813509.
- Coser LA (1967) *Continuities in the Study of Social Conflict* (New York: Free Press), ISBN 978-0029067802.
- Craig BA, Sendi PP (2002) Estimation of the transition matrix of a discrete-time Markov chain. *Health economics* 11(1):33–42.
- Cui R, Gallino S, Moreno A, Zhang DJ (2018) The Operational Value of Social Media Information. *Production and Operations Management* 27(10):1749–1769.
- Cziraki P, Gider J (2021) The Dollar Profits to Insider Trading. *Review of Finance* 25(5):1547–1580.

- Darby JL, Ketchen DJ, Ball GP, Mukherjee U (2023) CEO Stock Ownership, Recall Timing, and Stock Market Penalties. *Manufacturing & Service Operations Management* 25(5):1909–1930.
- Davidson WN, Worrell DL (1992) Research notes and communications: The effect of product recall announcements on shareholder wealth. *Strategic Management Journal* 13(6):467–473.
- Dawande M, Qi A (2021) Auditing, Inspections, and Testing for Social Responsibility in Supply Networks. Swaminathan JM, Deshpande V, eds., *Responsible business operations*, volume 10 of *Springer Series in Supply Chain Management*, 243–259 (Cham, Switzerland: Springer), ISBN 9783030519568, http://dx.doi.org/10.1007/978-3-030-51957-5_{_}12.
- Delmas MA, Toffel MW (2008) Organizational responses to environmental demands: opening the black box. *Strategic Management Journal* 29(10):1027–1055.
- Deo S, Iravani S, Jiang T, Smilowitz K, Samuelson S (2013) Improving Health Outcomes Through Better Capacity Allocation in a Community-Based Chronic Care Model. *Operations Research* 61(6):1277–1294.
- Dhaliwal S, Hoffmann S, White A, Ahn JW, McQueen RB, Scallan Walter E (2021) Cost of Hospitalizations for Leading Foodborne Pathogens in the United States: Identification by International Classification of Disease Coding and Variation by Pathogen. *Foodborne pathogens and disease* 18(11):812–821.
- Dhanorkar S (2019) Environmental Benefits of Internet-Enabled C2C Closed-Loop Supply Chains: A Quasi-Experimental Study of Craigslist. *Management Science* 65(2):660–680.
- Dhanorkar S, Muthulingam S (2020) Do E-Waste Laws Create Behavioral Spillovers? Quasi-Experimental Evidence from California. *Production and Operations Management* 29(7):1738–1766.
- Dhanorkar S, Muthulingam S, Noh IJ (2023) Does Legalizing Marijuana Increase Toxic Waste? Evidence from Manufacturing Facilities in the United States. *SSRN Electronic Journal* .
- Dhanorkar SS, Siemsen E, Linderman KW (2018) Promoting Change from the Outside: Directing Managerial Attention in the Implementation of Environmental Improvements. *Management Science* 64(6):2535–2556.
- Diaz RS (2017) Getting to the Root of Environmental Injustice: Evaluating Claims, Causes, and Solutions. *Georgetown Environmental Law Review* 29(4).
- DiMaggio PJ, Powell WW (1983) The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields. *American Sociological Review* 48(2):147.
- Dong L, Rashkova I, Shi D (2022) Food Safety Audits in Developing Economies: Decentralization vs. Centralization. *Manufacturing & Service Operations Management* 24(6):2863–2881.
- Doshi P, Perez D (2008) Generalized point based value iteration for interactive POMDPs. *AAAI'08: Proceedings of the 23rd national conference on Artificial intelligence* 63–68.
- Downey A (2013) *Think Bayes* (Sebastopol CA: O'Reilly), first edition edition, ISBN 978-1449370787.
- DuHadway S, Narasimhan R (2021) Subverting Process-Based Controls: Oscillation in Automotive Recalls and a Simulation on Opportunism within a Network. *Decision Sciences* 52(6):1326–1363.
- Dunlap RE, McCright AM (2008) A Widening Gap: Republican and Democratic Views on Climate Change. *Environment: Science and Policy for Sustainable Development* 50(5):26–35.
- Dunlap RE, Xiao C, McCright AM (2001) Politics and Environment in America: Partisan and Ideological Cleavages in Public Support for Environmentalism. *Environmental Politics* 10(4):23–48.

- Dutt N, King AA (2014) The Judgment of Garbage: End-of-Pipe Treatment and Waste Reduction. *Management Science* 60(7):1812–1828.
- Ebele NE, Bae SM, Kim JD (2023) The Impact of Chief Sustainability Officers on Environmental Performance of Korean Listed Companies: The Mediating Role of Corporate Sustainability Practices. *Sustainability* 15(20):14819.
- Eccles RG, Ioannou I, Serafeim G (2014) The Impact of Corporate Sustainability on Organizational Processes and Performance. *Management Science* 60(11):2835–2857.
- Eisenhardt KM (1989) Agency Theory: An Assessment and Review. *Academy of Management Review* 14(1):57–74.
- Elliott MR, Wang Y, Lowe RA, Kleindorfer PR (2004) Environmental justice: frequency and severity of US chemical industry accidents and the socioeconomic status of surrounding communities. *Journal of epidemiology and community health* 58(1):24–30.
- Emerson RM (1962) Power-Dependence Relations. *American Sociological Review* 27(1):31.
- EPA (2018) Source Reduction Activities — US EPA. <https://www.epa.gov/trinationalanalysis/source-reduction-activities>.
- EPA (2022) EJScreen Technical Documentation. https://www.epa.gov/sites/default/files/2021-04/documents/ejscreen_technical_document.pdf.
- EPA (2023a) What is an EMS? <https://www.epa.gov/ems/what-ems>.
- EPA (2023b) What is the Toxics Release Inventory? — US EPA. <https://www.epa.gov/toxics-release-inventory-tri-program/what-toxics-release-inventory>.
- Epstein MJ, Hanson KO (2021) *Rotten: Why corporate misconduct continues and what to do about it* (Los Altos, California: Lanark Press), 1st ed. edition, ISBN 978-1735336114.
- Ertekin N, Gümüş M, Nikoofal ME (2021) Online-Exclusive or Hybrid? Channel Merchandising Strategies for Ship-to-Store Implementation. *Management Science* .
- Fama EF, French KR (1995) Size and Book-to-Market Factors in Earnings and Returns. *The Journal of Finance* 50(1):131.
- Farronato C, Zervas G (2022) Consumer Reviews and Regulation: Evidence from NYC Restaurants. <http://dx.doi.org/10.3386/w29715>.
- FDA (2014) Cardiovascular Systems Diamondback 360 Peripheral Orbital Atherectomy System - Sheath May Fracture During Use. <https://www.fdanews.com/ext/resources/files/08/08-20-14-DiamondbackRecall.pdf>.
- FDA (2020) Recalls, Corrections and Removals (Devices). <https://www.fda.gov/medical-devices/postmarket-requirements-devices/recalls-corrections-and-removals-devices>.
- FDA (2021) Regulatory Procedures Manual: Chapter 7 Recall Procedures .
- FDA (2023a) About Manufacturer and User Facility Device Experience (MAUDE). <https://www.fda.gov/medical-devices/mandatory-reporting-requirements-manufacturers-importers-and-device-user-facilities/about-manufacturer-and-user-facility-device-experience-maude>.
- FDA (2023b) Reporting Serious Problems to FDA. <https://www.fda.gov/safety/medwatch-fda-safety-information-and-adverse-event-reporting-program/reporting-serious-problems-fda>.
- Fernández A, García S, Galar M, Prati RC, Krawczyk B, Herrera F (2018) *Learning from imbalanced data sets* (Cham, Switzerland: Springer), ISBN 978-3-319-98073-7.

- Finney HC, Lesieur HR (1982) A Contingency Theory of Organizational Crime. Bacharach SB, ed., *Research in Sociology of Organizations*, 255–299 (Greenwich CT: JAI Press), ISBN 978-0892321704.
- Fiol CM, Lyles MA (1985) Organizational Learning. *The Academy of Management Review* 10(4):803.
- Friedman J, Hastie T, Tibshirani R (2000) Additive logistic regression: a statistical view of boosting (With discussion and a rejoinder by the authors). *The Annals of Statistics* 28(2).
- Friedman JH (2001) Greedy function approximation: A gradient boosting machine. *The Annals of Statistics* 29(5).
- Friedman L (Mar 11, 2025) E.P.A. Plans to Close All Environmental Justice Offices. *New York Times (Online)*; New York .
- Fu R, Tang Y, Chen G (2020) Chief sustainability officers and corporate social (Ir)responsibility. *Strategic Management Journal* 41(4):656–680.
- Fu W, Kalkanci B, Subramanian R (2019) Are Hazardous Substance Rankings Effective? An Empirical Investigation of Information Dissemination About the Relative Hazards of Chemicals and Emissions Reductions. *Manufacturing & Service Operations Management* 21(3):602–619.
- Fuhr T, George K, Pai J (2013) The Business Case for Medical Device Quality. https://www.mckinsey.com/~media/mckinsey/dotcom/client_service/public%20sector/regulatory%20excellence/the_business_case_for_medical_device_quality.ashx.
- Gawande K, Bohara AK (2005) Agency Problems in Law Enforcement: Theory and Application to the U.S. Coast Guard. *Management Science* 51(11):1593–1609.
- Geltman EG, Gill G, Jovanovic M (2016) Beyond Baby Steps: An Empirical Study of the Impact of Environmental Justice Executive Order 12898. *Family & community health* 39(3):143–150.
- Genn H (1993) Business Responses to the Regulation of Health and Safety in England*. *Law & Policy* 15(3):219–233.
- Gioia DA (1992) Pinto Fires and Personal Ethics: A Script Analysis of Missed Opportunities. *Journal of Business Ethics* 11(5):379–389.
- Gmytrasiewicz PJ, Doshi P (2005) A Framework for Sequential Planning in Multi-Agent Settings. *Journal of Artificial Intelligence Research* 24:49–79.
- Greve HR, Palmer D, Pozner JE (2010) Organizations Gone Wild: The Causes, Processes, and Consequences of Organizational Misconduct. *Academy of Management Annals* 4(1):53–107.
- Guerette RT, Bowers KJ (2009) Chapter Assessing the Extent of Crime Displacement and Diffusion of Benefits: A Review of Situational Crime Prevention Evaluations. *Criminology* 47(4):1331–1368.
- Gupta A, Fung A, Murphy C (2021) Out of character: CEO political ideology, peer influence, and adoption of CSR executive position by Fortune 500 firms. *Strategic Management Journal* 42(3):529–557.
- Ha AY, Shang W, Wang Y (2023) Supplier Audit Information Sharing and Responsible Sourcing. *Management Science* 69(1):308–324.
- Hambrick DC, Cannella AA (2004) CEOs who have COOs: contingency analysis of an unexplored structural form. *Strategic Management Journal* 25(10):959–979.
- Hardcopf R, Shah R, Dhanorkar S (2021) The Impact of a Spill or Pollution Accident on Firm Environmental Activity: An Empirical Investigation. *Production and Operations Management* 30(8):2467–2491.
- Hartmann J, Forkmann S, Benoit S, Henneberg SC (2022) A consumer perspective on managing the consequences of chain liability. *Journal of Supply Chain Management* 58(4):58–89.

- Hawkins JT (2003) *A Lagrangian Decomposition Approach to Weakly Coupled Dynamic Optimization Problems and its Applications*. Doctor of philosophy in operations research, Massachusetts Institute of Technology.
- Heckman JJ (1976) The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models. *Annals of economic and social measurement* 5(4).
- Heitz A, Wang Y, Wang Z (2021) Corporate Political Connections and Favorable Environmental Regulatory Enforcement. *Management Science* .
- Hersel MC, Helmuth CA, Zorn ML, Shropshire C, Ridge JW (2019) The Corrective Actions Organizations Pursue Following Misconduct: A Review and Research Agenda. *Academy of Management Annals* 13(2):547–585.
- Hersher R (2018) After Decades Of Air Pollution, A Louisiana Town Rebels Against A Chemical Giant. <https://www.npr.org/sections/health-shots/2018/03/06/583973428/after-decades-of-air-pollution-a-louisiana-town-rebels-against-a-chemical-giant>.
- Ho DE (2012) Fudging the Nudge: Information Disclosure and Restaurant Grading. *The Yale Law Journal* 122(3).
- Ho DE, Ashwood ZC, Handan-Nader C (2019) New Evidence on Information Disclosure through Restaurant Hygiene Grading. *American Economic Journal: Economic Policy* 11(4):404–428.
- Hou J, Jiang DL, Ma X, Gopal A (Forthcoming) The Costs of Ambiguous Information Disclosure: On the Unintended Consequences of Providing Restaurant Hygiene Scores on Platforms in the United. *Information Systems Research* .
- Hu K, Chopra S, Chen Y (2021) The Effect of Tightening Standards on Automakers' Non-compliance. *Production and Operations Management* 30(9):3094–3115.
- Huang X, Tereagoglu N, Atasu A, Toktay LB (2022) Lemons, Trade-Ins and Certified Pre-Owned Programs. *Working Paper* .
- Huddart S, Ke B, Shi C (2007) Jeopardy, non-public information, and insider trading around SEC 10-K and 10-Q filings. *Journal of Accounting and Economics* 43(1):3–36.
- Hyndman RJ, Athanasopoulos G (2018) *Forecasting: Principles and practice* (Melbourne, Australia: OTexts), 2nd ed. edition.
- Ibanez MR, Toffel MW (2020) How Scheduling Can Bias Quality Assessment: Evidence from Food-Safety Inspections. *Management Science* 66(6):2396–2416.
- Jackson CH (2011) Multi-State Models for Panel Data: The msm Package for R. *Journal of Statistical Software* 38(8).
- Jacobs BW, Singhal VR (2017) The effect of the Rana Plaza disaster on shareholder wealth of retailers: Implications for sourcing strategies and supply chain governance*. *Journal of Operations Management* 49-51(1):52–66.
- Jagolinzer AD, Larcker DF, Ormazabal G, Taylor DJ (2020) Political Connections and the Informativeness of Insider Trades. *The Journal of Finance* 75(4):1833–1876.
- Jagolinzer AD, Larcker DF, Taylor DJ (2011) Corporate Governance and the Information Content of Insider Trades. *Journal of Accounting Research* 49(5):1249–1274.
- Jarrell G, Peltzman S (1985) The Impact of Product Recalls on the Wealth of Sellers. *Journal of Political Economy* 93(3):512–536.
- Jensen MC, Meckling WH (1976) Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of financial economics* 3(4):305–360.

- Jessop S (2021) Investors push world's top chemicals companies over hazardous substances. *Reuters* .
- Jiang B (2009) The effects of interorganizational governance on supplier's compliance with SCC: An empirical examination of compliant and non-compliant suppliers. *Journal of Operations Management* 27(4):267–280.
- Jiang X, Menon A, Wang S, Kim J, Ohno-Machado L (2012) Doubly Optimized Calibrated Support Vector Machine (DOC-SVM): an algorithm for joint optimization of discrimination and calibration. *PLoS one* 7(11):e48823.
- Jin GZ, Leslie P (2003) The Effect of Information on Product Quality: Evidence from Restaurant Hygiene Grade Cards. *The Quarterly Journal of Economics* 118(2):409–451.
- Johnson MS, Levine DI, Toffel MW (2023) Improving Regulatory Effectiveness through Better Targeting: Evidence from OSHA. *American Economic Journal: Applied Economics* 15(4):30–67.
- Joseph J, Wilson AJ (2018) The growth of the firm: An attention-based view. *Strategic Management Journal* 39(6):1779–1800.
- Kalkanci B, Rahmani M, Toktay LB (2019) The Role of Inclusive Innovation in Promoting Social Sustainability. *Production and Operations Management* 28(12):2960–2982.
- Kamalzadeh H, Ahuja V, Hahsler M, Bowen ME (2021) An Analytics-Driven Approach for Optimal Individualized Diabetes Screening. *Production and Operations Management* 30(9):3161–3191.
- Kanashiro P, Rivera J (2019) Do Chief Sustainability Officers Make Companies Greener? The Moderating Role of Regulatory Pressures. *Journal of Business Ethics* 155(3):687–701.
- Kang JS, Kuznetsova P, Luca M, Choi Y (2013) Where Not to Eat? Improving Public Policy by Predicting Hygiene Inspections Using Online Reviews. *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing* 1443–1448.
- Karpoff JM, Lee DS, Vondracik VP (1999) Defense Procurement Fraud, Penalties, and Contractor Influence. *Journal of Political Economy* 107(4):809–842.
- Kassinis G, Vafeas N (2002) Corporate boards and outside stakeholders as determinants of environmental litigation. *Strategic Management Journal* 23(5):399–415.
- Kelley HH, Michela JL (1980) Attribution theory and research. *Annual review of psychology* 31:457–501.
- Kim S, Oh S (2023) Outside directors' insider trading around board meetings. *Review of Accounting Studies* .
- Kim S, Wagner SM (2021) Examining the Stock Price Effect of Corruption Risk in the Supply Chain. *Decision Sciences* 52(4):833–865.
- Kim Y, Xu Y (2024) Operational Risk Management: Optimal Inspection Policy. *Management Science* 70(6):4087–4104.
- Kistler JT, Sharma L, Jayaram J, Eckerd S (2024) Does history really repeat itself? An empirical investigation of recurring misconduct violations in public procurement. *Journal of Purchasing and Supply Management* 30(1):100893.
- Kochenderfer MJ, Wheeler TA, Wray KH (2022) *Algorithms for decision making* (Cambridge Massachusetts: The MIT Press), ISBN 9780262047012.
- Kraft T, Valdés L, Zheng Y (2022) Consumer trust in social responsibility communications: The role of supply chain visibility. *Production and Operations Management* .
- Kuhn M (2019) The caret Package. <https://topepo.github.io/caret/index.html>.

- Kumar A, Paraskevas JP (2018) A Proactive Environmental Strategy: Analyzing the Effect of SCM Experience, Age, and Female Representation in TMT s. *Journal of Supply Chain Management* 54(4):20–41.
- Kurniawati H, Hsu D, Lee WS (2009) SARSOP: Efficient Point-Based POMDP Planning by Approximating Optimally Reachable Belief Spaces. Brock O, Trinkle JC, Ramos F, eds., *Robotics: Science and Systems IV*, 65–72 (Cambridge Mass.: MIT Press), ISBN 9780262513098, <http://dx.doi.org/10.15607/RSS.2008.IV.009>.
- Lam HKS, Ding L, Dong Z (2022) The impact of foreign competition on domestic firms' product quality: Evidence from a quasi-natural experiment in the United States. *Journal of Operations Management* 68(8):881–902.
- Lange D, Washburn NT (2012) Understanding Attributions of Corporate Social Irresponsibility. *Academy of Management Review* 37(2):300–326.
- Law AM (2013) *Simulation modeling and analysis*. McGraw-Hill series in industrial engineering and management science (Dubuque: McGraw-Hill Education), fifth edition edition, ISBN 9780073401324.
- Lehman DW, Kovács B, Carroll GR (2014) Conflicting Social Codes and Organizations: Hygiene and Authenticity in Consumer Evaluations of Restaurants. *Management Science* 60(10):2602–2617.
- Lemert EM (1951) *Social Pathology: A Systematic Approach To The Theory Of Sociopathic Behavior* (New York: McGraw-Hill).
- Levi R, Singhvi S, Zheng Y (2020) Economically Motivated Adulteration in Farming Supply Chains. *Management Science* 66(1):209–226.
- Li H, Lam HKS, Ho W, Yeung ACL (2022) The impact of chief risk officer appointments on firm risk and operational efficiency. *Journal of Operations Management* 68(3):241–269.
- Li W, Xu Q, Zhu Q (2025) CEO Hometown Preference in Corporate Environmental Policies. *Management Science* .
- Littman ML, Cassandra AR, Kaelbling LP (1995) Learning policies for partially observable environments: Scaling up. *Proceedings of the Twelfth International Conference on Machine Learning* 362–370.
- Liu AX, Liu Y, Luo T (2016) What Drives a Firm's Choice of Product Recall Remedy? The Impact of Remedy Cost, Product Hazard, and the CEO. *Journal of Marketing* 80(3):79–95.
- Liu X, Mishra A, Goldstein S, Sinha KK (2019) Toward Improving Factory Working Conditions in Developing Countries: An Empirical Analysis of Bangladesh Ready-Made Garment Factories. *Manufacturing & Service Operations Management* 21(2):379–397.
- Lo CKY, Tang CS, Zhou Y, Yeung ACL, Di Fan (2018) Environmental Incidents and the Market Value of Firms: An Empirical Investigation in the Chinese Context. *Manufacturing & Service Operations Management* 20(3):422–439.
- Luca M, Zervas G (2016) Fake It Till You Make It: Reputation, Competition, and Yelp Review Fraud. *Management Science* 62(12):3412–3427.
- Mackenzie C (2007) Boards, Incentives and Corporate Social Responsibility: the case for a change of emphasis. *Corporate Governance: An International Review* 15(5):935–943.
- Makofske MP (2020) The Effect of Information Salience on Product Quality: Louisville Restaurant Hygiene and Yelp.com*. *The Journal of Industrial Economics* 68(1):52–92.
- Mani V, Muthulingam S (2019) Does Learning from Inspections Affect Environmental Performance? Evidence from Unconventional Well Development in Pennsylvania. *Manufacturing & Service Operations Management* 21(1):177–197.

- Massdevice (2011) MassDevice Exclusive: Former Medtronic CEO Hawkins on the Sprint Fidelis recall. <https://www.massdevice.com/massdevice-exclusive-former-medtronic-ceo-hawkins-sprint-fidelis-recall/>.
- Mayo K, Ball G, Mills A (2022) CEO Tenure and Recall Risk Management in the Consumer Products Industry. *Production and Operations Management* 31(2):743–763.
- McKendall M, DeMarr B, Jones-Rikkens C (2002) Ethical Compliance Programs and Corporate Illegality: Testing the Assumptions of the Corporate Sentencing Guidelines. *Journal of Business Ethics* 37(4):367–383.
- McKendall MA, Wagner III JA (1997) Motive, Opportunity, Choice, and Corporate Illegality. *Organization Science* 8(6):624–647.
- McWilliams A, Siegel D (1997) EVENT STUDIES IN MANAGEMENT RESEARCH: THEORETICAL AND EMPIRICAL ISSUES. *Academy of Management Journal* 40(3):626–657.
- Mead PS, Slutsker L, Dietz V, McCaig LF, Bresee JS, Shapiro C, Griffin PM, Tauxe RV (1999) Food-related illness and death in the United States. *Emerging infectious diseases* 5(5):607–625.
- Mejia J, Mankad S, Gopal A (2019) A for Effort? Using the Crowd to Identify Moral Hazard in New York City Restaurant Hygiene Inspections. *Information Systems Research* 30(4):1363–1386.
- Merton RK (1938) Social Structure and Anomie. *American Sociological Review* 3(5):672.
- Meyer JW, Rowan B (1977) Institutionalized Organizations: Formal Structure as Myth and Ceremony. *American Journal of Sociology* 83(2):340–363.
- Miller K, Serafeim G (2014) Chief Sustainability Officers: Who Are They and What Do They Do? *Harvard Business School Working Paper* 15-011.
- Mohai P, Pellow D, Roberts JT (2009) Environmental Justice. *Annual Review of Environment and Resources* 34(1):405–430.
- Mohai P, Saha R (2007) Racial Inequality in the Distribution of Hazardous Waste: A National-Level Reassessment. *Social Problems* 54(3):343–370.
- Mohai P, Saha R (2015) Which came first, people or pollution? Assessing the disparate siting and post-siting demographic change hypotheses of environmental injustice. *Environmental Research Letters* 10(11):115008.
- Mohliver A (2019) How Misconduct Spreads: Auditors’ Role in the Diffusion of Stock-option Backdating. *Administrative Science Quarterly* 64(2):310–336.
- Mohliver A, Ody-Brasier A (2023) Religious Affiliation and Wrongdoing: Evidence from U.S. Nursing Homes. *Management Science* 69(1):533–554.
- Mollenkopf DA, Peinkofer ST, Chu Y (2022) Supply chain transparency: Consumer reactions to incongruent signals. *Journal of Operations Management* 68(4):306–327.
- Mukherjee UK, Ball GP, Wowak KD, Natarajan KV, Miller JW (2022) Hiding in the Herd: The Product Recall Clustering Phenomenon. *Manufacturing & Service Operations Management* 24(1):392–410.
- Mukherjee UK, Sinha KK (2018) Product Recall Decisions in Medical Device Supply Chains: A Big Data Analytic Approach to Evaluating Judgment Bias. *Production and Operations Management* 27(10):1816–1833.
- Mukherjee UK, Ye H, Chhajed D (2025) Encounter Decisions for Patients With Diverse Sociodemographic Characteristics: Predictive Analytics of EMR Data From a Large Chain of Clinics. *Journal of Operations Management* .

- Muthulingam S, Dhanorkar S, Corbett CJ (2022) Does Water Scarcity Affect Environmental Performance? Evidence from Manufacturing Facilities in Texas. *Management Science* 68(4):2785–2805.
- Ni JZ, Flynn BB, Jacobs FR (2014) Impact of product recall announcements on retailers' financial value. *International Journal of Production Economics* 153:309–322.
- NYT (2022) Frustrations Grow Over Company's Response to CPAP Recalls. <https://www.nytimes.com/2022/11/15/health/cpap-philips-breathing-device-recall.html>.
- Ocasio W (1997) Towards an attention-based view of the firm. *Strategic Management Journal* 18(S1):187–206.
- Ocasio W (2011) Attention to Attention. *Organization Science* 22(5):1286–1296.
- Oliver J (2022) Environmental Racism. <https://www.youtube.com/watch?v=-v0XiUQ1RLw>.
- O'Neil SG (2007) Superfund: evaluating the impact of executive order 12898. *Environmental health perspectives* 115(7):1087–1093.
- Owens Corning (2024) Sustainability Report. <https://www.owenscorning.com/en-us/corporate/sustainability/docs/2025/2024-Owens-Corning-Sustainability-Report.pdf>.
- Oxley JE, Rivkin JW, Ryall MD (2010) The Strategy Research Initiative: Recognizing and encouraging high-quality research in strategy. *Strategic Organization* 8(4):377–386.
- Pagell M, Parkinson M, Veltri A, Gray J, Wiengarten F, Louis M, Fynes B (2020) The Tension Between Worker Safety and Organization Survival. *Management Science* 66(10):4863–4878.
- Pagitsas (2022) *Chief Sustainability Officers At Work* ([S.l.]: Apress), 1st ed. edition, ISBN 978-1484278659.
- Palmer D (2012) *Normal Organizational Wrongdoing: A Critical Analysis of Theories of Misconduct in and by Organizations* (New York: Oxford University Press), ISBN 978-0-19-967742-9.
- Palmer D, Smith-Crowe K, Greenwood R (2016a) The imbalances and limitations of theory and research on organizational wrongdoing. Palmer D, Smith-Crowe K, Greenwood R, eds., *Organizational Wrongdoing*, 1–16 (Cambridge: Cambridge University Press), ISBN 9781316338827, <http://dx.doi.org/10.1017/CB09781316338827.002>.
- Palmer D, Smith-Crowe K, Greenwood R, eds. (2016b) *Organizational Wrongdoing* (Cambridge: Cambridge University Press), ISBN 9781316338827.
- Pastor M, Sadd J, Hipp J (2001) Which Came First? Toxic Facilities, Minority Move-In, and Environmental Justice. *Journal of Urban Affairs* 23(1):1–21.
- Patel V, Putnins TJ (2020) How Much Insider Trading Really Happens in Stock Markets? *SSRN Electronic Journal* .
- Pebesma E (2018) Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal* 10(1):439.
- Pebesma E, Bivand R (2023) *Spatial Data Science* (Boca Raton: Chapman and Hall/CRC), ISBN 9780429459016.
- Pendse SG (2012) Ethical Hazards: A Motive, Means, and Opportunity Approach to Curbing Corporate Unethical Behavior. *Journal of Business Ethics* 107(3):265–279.
- Perrow C (1984) *Normal Accidents: Living with High Risk Technologies* (Basic Books), 1 edition, ISBN 0-465-05142-1.
- Peters GF, Romi AM, Sanchez JM (2019) The Influence of Corporate Sustainability Officers on Performance. *Journal of Business Ethics* 159(4):1065–1087.

- Petersen F, Dhingra V, Shah R (2024) Are Chief Sustainability Officers Guardians of Environmental Justice? An Empirical Evaluation. Taneja S, ed., *Proceedings of the Eighty-fourth Annual Meeting of the Academy of Management* (Online), ISBN 2151-6561.
- Pfeffer J, Salancik GR (1978) *The external control of organizations: A resource dependence perspective* (New York: Harper & Row), ISBN 9780060451936.
- Philips (2022) Philips parts ways with CEO in midst of massive recall. <https://www.reuters.com/business/philips-says-ceo-van-houten-leave-october-2022-08-16/>.
- Pierce L, Snyder J (2008) Ethical Spillovers in Firms: Evidence from Vehicle Emissions Testing. *Management Science* 54(11):1891–1903.
- Pineau J, Gordon G, Thrun S (2003) Point-based value iteration: An anytime algorithm for POMDPs. *Proceedings of the 18th International Joint Conference on Artificial Intelligence (IJCAI)* 3:1025–1032.
- Plambeck EL, Taylor TA (2016) Supplier Evasion of a Buyer’s Audit: Implications for Motivating Supplier Social and Environmental Responsibility. *Manufacturing & Service Operations Management* 18(2):184–197.
- Platt JC (2000) Probabilities for SV Machines. Smola AJ, Bartlett P, Schölkopf B, Schuurmans D, eds., *Advances in Large-Margin Classifiers* (The MIT Press), ISBN 9780262283977, <http://dx.doi.org/10.7551/mitpress/1113.003.0008>.
- Porteous AH, Rammohan SV, Lee HL (2015) Carrots or Sticks? Improving Social and Environmental Compliance at Suppliers Through Incentives and Penalties. *Production and Operations Management* 24(9):1402–1413.
- Powell JS (1984) Political Difficulties Facing Waste-to-Energy Conversion Plant Siting.
- Puterman ML (2005) *Markov Decision Processes: Discrete Stochastic Dynamic Programming* (Wiley-Interscience), ISBN 978-0471727828.
- Ramonas A (2022) SEC Toughens Insider Trading Rules for Execs, Limits Protections. *Bloomberg Law* .
- Ravina E, Sapienza P (2010) What Do Independent Directors Know? Evidence from Their Trading. *Review of Financial Studies* 23(3):962–1003.
- Ren CR, Hu Y, Hu Y, Hausman J (2011) Managing Product Variety and Collocation in a Competitive Environment: An Empirical Investigation of Consumer Electronics Retailing. *Management Science* 57(6):1009–1024.
- Rerup C (2009) Attentional Triangulation: Learning from Unexpected Rare Crises. *Organization Science* 20(5):876–893.
- Roh J, Krause R, Swink M (2016) The appointment of chief supply chain officers to top management teams: A contingency model of firm-level antecedents and consequences. *Journal of Operations Management* 44(1):48–61.
- Romero D (2021) Ex-CEO of Chicago-area tech firm sentenced for role in Capitol riot. *NBC News* .
- Roth J, Sant’Anna PH, Bilinski A, Poe J (2023) What’s trending in difference-in-differences? A synthesis of the recent econometrics literature. *Journal of Econometrics* 235(2):2218–2244.
- Rousseau DM (1995) *Psychological contracts in organizations: Understanding written and unwritten agreements* (Thousand Oaks: SAGE Publications), ISBN 9780803971059.
- Schmenner RW, Swink ML (1998) On theory in operations management. *Journal of Operations Management* 17(1):97–113.
- Schnatterly K, Gangloff KA, Tuschke A (2018) CEO Wrongdoing: A Review of Pressure, Opportunity, and Rationalization. *Journal of Management* 44(6):2405–2432.

- Schomberg JP, Haimson OL, Hayes GR, Anton-Culver H (2016) Supplementing Public Health Inspection via Social Media. *PloS one* 11(3):e0152117.
- Schonlau M (2005) Boosted Regression (Boosting): An Introductory Tutorial and a Stata Plugin. *The Stata Journal: Promoting communications on statistics and Stata* 5(3):330–354.
- Scott A, Balthrop A, Miller JW (2021) Unintended responses to IT-enabled monitoring: The case of the electronic logging device mandate. *Journal of Operations Management* 67(2):152–181.
- Scott A, Nyaga GN (2019) The effect of firm size, asset ownership, and market prices on regulatory violations. *Journal of Operations Management* 65(7):685–709.
- Scurlock S (2021) Justin J. Pearson brings fight against oil pipeline from his Memphis neighborhood to the nation. *WREG.com* .
- Sengupta J, Storey K, Casey S, Trager L, Buescher M, Horning M, Gornick C, Abdelhadi R, Tang C, Brill S, Ashbach L, Hauser RG (2020) Outcomes Before and After the Recall of a Heart Failure Pacemaker. *JAMA internal medicine* 180(2):198–205.
- Shah R, Ball GP, Netessine S (2017) Plant Operations and Product Recalls in the Automotive Industry: An Empirical Investigation. *Management Science* 63(8):2439–2459.
- Shani G, Pineau J, Kaplow R (2013) A survey of point-based POMDP solvers. *Autonomous Agents and Multi-Agent Systems* 27(1):1–51.
- Shaver JM (1998) Accounting for Endogeneity When Assessing Strategy Performance: Does Entry Mode Choice Affect FDI Survival? *Management Science* 44(4):571–585.
- Shive SA, Forster MM (2020) Corporate governance and pollution externalities of public and private firms. *The Review of Financial Studies* 33(3):1296–1330.
- Short JL, Toffel MW, Hugill AR (2016) Monitoring global supply chains. *Strategic Management Journal* 37(9):1878–1897.
- Shover N (1978) Defining Organizational Crime. Ermann MD, Lundman RJ, eds., *Corporate and governmental deviance*, 37–40 (New York: Oxford University Press), ISBN 9780195023244.
- Shubham A, Subramanian R (2025) Playing Fair? Environmental Impacts and Practices of Facilities in Minority Communities.
- Siering M (2021) Leveraging online review platforms to support public policy: Predicting restaurant health violations based on online reviews. *Decision Support Systems* 143:1–11.
- Simon H (1955) A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics* 69(1).
- Skilton PF, Bernardes E (2022) Normal misconduct in the prescription opioid supply chain. *Journal of Supply Chain Management* 58(4):6–29.
- Sodnik EJ (1971) *The Optimal Control of Partially Observable Markov Decision Processes* Stanford, California. Phd, Stanford, California.
- Southern Nevada Health District (2010) Regulations Governing the Sanitation of Food Establishments.
- Southern Nevada Health District (2022) Restaurant Inspection Search. <https://www.southernnevadahealthdistrict.org/permits-and-regulations/restaurant-inspections/restaurant-inspection-search/>.
- Southern Nevada Health District (2023) Food Establishment Inspection Process. <http://www.southernnevadahealthdistrict.org/permits-and-regulations/food-establishment-operations/food-establishment-inspection-process/>.
- Spence M (1973) Job Market Signaling. *The Quarterly Journal of Economics* 87(3):355.

- Steinker S, Hoberg K, Thonemann UW (2017) The Value of Weather Information for E-Commerce Operations. *Production and Operations Management* 26(10):1854–1874.
- Steurer R, Langer ME, Konrad A, Martinuzzi A (2005) Corporations, Stakeholders and Sustainable Development I: A Theoretical Exploration of Business–Society Relations. *Journal of Business Ethics* 61(3):263–281.
- Steven AB, Dong Y, Corsi T (2014) Global sourcing and quality recalls: An empirical study of outsourcing–supplier concentration–product recalls linkages. *Journal of Operations Management* 32(5):241–253.
- Stock J, Yogo M (2002) Testing for Weak Instruments in Linear IV Regression. <http://dx.doi.org/10.3386/t0284>.
- Sun L, Abraham S (2021) Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics* 225(2):175–199.
- Sun Y, Wong AKC, Kamel MS (2009) Classification of Imbalanced Data: A Review. *International Journal of Pattern Recognition and Artificial Intelligence* 23(04):687–719.
- Taddy M (2013a) Measuring Political Sentiment on Twitter: Factor Optimal Design for Multinomial Inverse Regression. *Technometrics* 55(4):415–425.
- Taddy M (2013b) Multinomial Inverse Regression for Text Analysis. *Journal of the American Statistical Association* 108(503):755–770.
- Taddy M (2015) Distributed multinomial regression. *The Annals of Applied Statistics* 9(3).
- Taylor DE (1997) American Environmentalism: The Role of Race, Class and Gender in Shaping Activism 1820-1995. *Race, Gender & Class* 5(1):pp. 16–62.
- Tenbrunsel AE, Smith-Crowe K (2008) Ethical Decision Making: Where We’ve Been and Where We’re Going. *Academy of Management Annals* 2(1):545–607.
- The White House (1994) Federal Actions To Address Environmental Justice in Minority Populations and Low-Income Populations: Executive Order 12898 of February 11, 1994. <https://www.archives.gov/files/federal-register/executive-orders/pdf/12898.pdf>.
- Thirumalai S, Sinha KK (2011) Product Recalls in the Medical Device Industry: An Empirical Exploration of the Sources and Financial Consequences. *Management Science* 57(2):376–392.
- Thornton D, Kagan RA, Gunningham N (2009) When Social Norms and Pressures Are Not Enough: Environmental Performance in the Trucking Industry. *Law & Society Review* 43(2):405–435.
- Tibshirani R (1996) Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* 58(1):267–288.
- Tucker AL (2016) The Impact of Workaround Difficulty on Frontline Employees’ Response to Operational Failures: A Laboratory Experiment on Medication Administration. *Management Science* 62(4):1124–1144.
- Unruh G (2015) What Does a Corporate Responsibility Manager Do? <https://sloanreview.mit.edu/article/what-does-a-corporate-responsibility-manager-do/>.
- Vaughan D (1999) The Dark Side Of Organizations: Mistake, Misconduct, and Disaster. *Annual Review of Sociology* 25(1):271–305.
- Vroom VH (1964) *Work and Motivation* (New York: Wiley), ISBN 9780471912057.
- Walker K (2023) tigris: Load Census TIGER/Line Shapefiles. <https://github.com/walkerke/tigris>.
- Walker K, Herman M (2023) tidycensus: Load US Census Boundary and Attribute Data as ‘tidyverse’ and ‘sf’-Ready Data Frames. <https://walker-data.com/tidycensus/>.

- Wang Z, Lee A, Polonsky M (2018) Egregiousness and Boycott Intensity: Evidence from the BP Deepwater Horizon Oil Spill. *Management Science* 64(1):149–163.
- Wiengarten F, Di Fan, Lo CK, Pagell M (2017a) The differing impacts of operational and financial slack on occupational safety in varying market conditions. *Journal of Operations Management* 52(1):30–45.
- Wiengarten F, Di Fan, Pagell M, Lo CK (2019) Deviations from aspirational target levels and environmental and safety performance: Implications for operations managers acting irresponsibly. *Journal of Operations Management* 65(6):490–516.
- Wiengarten F, Lo CKY, Lam JYK (2017b) “How does Sustainability Leadership Affect Firm Performance? The Choices Associated with Appointing a Chief Officer of Corporate Social Responsibility”. *Journal of Business Ethics* 140(3):477–493.
- Wolf FG (2001) OPERATIONALIZING AND TESTING NORMAL ACCIDENT THEORY IN PETROCHEMICAL PLANTS AND REFINERIES. *Production and Operations Management* 10(3):292–305.
- Wooldridge JM (2010) *Econometric analysis of cross section and panel data* (Cambridge Mass.: MIT Press), 2nd ed. edition, ISBN 9780262232586.
- Wowak AJ, Mannor MJ, Wowak KD (2015) Throwing caution to the wind: The effect of CEO stock option pay on the incidence of product safety problems. *Strategic Management Journal* 36(7):1082–1092.
- Wowak KD, Ball GP, Post C, Ketchen DJ (2021) The Influence of Female Directors on Product Recall Decisions. *Manufacturing & Service Operations Management* 23(4):895–913.
- Wowak KD, Boone CA (2015) So Many Recalls, So Little Research: A Review of the Literature and Road map for Future Research. *Journal of Supply Chain Management* 51(4):54–72.
- Xiong X, Png IPL (2019) Location of U.S. Manufacturing, 1987-2014: A New Dataset. *SSRN Electronic Journal* .
- Ye S, Aydin G, Hu S (2015) Sponsored Search Marketing: Dynamic Pricing and Advertising for an Online Retailer. *Management Science* 61(6):1255–1274.
- Yelp Inc (2021) Yelp Dataset. <https://www.yelp.com/dataset>.
- Yi Z, Yu M, Cheung KL (2022) Impacts of Counterfeiting on a Global Supply Chain. *Manufacturing & Service Operations Management* 24(1):159–178.
- Zhang H, Aydin G, Parker RP (2022) Social Responsibility Auditing in Supply Chain Networks. *Management Science* 68(2):1058–1077.
- Zipp R (01/12/2022) Philips ups recall provisions by \$250M, impacted devices to more than 5M. *MedTech Dive* .

Appendix A

Appendix to Introduction

A.1 Tables

Table A.1: Types of Organizational Misconduct in the OSCM Literature

Type	Phenomenon	Definition	Stakeholder	Judge
	Spill and pollution accidents	Controversies regarding spills of chemicals, oils, and fuels. (Hardcopf et al. 2021)	Natural environment	Media
Sustainability violations	Environmental regulation violations	Non-compliance with nitrogen dioxide (NOx) emission standards based on vehicles' on-road NOx emissions. (Hu et al. 2021)	Natural environment	Researcher
		Violations of regulations regarding the construction and operation of unconventional wells. (Mani and Muthulingam 2019)	Natural environment	P-DEP
Employee mistreatment	Workplace safety regulation violations	Working condition risks (structural, fire, electrical) in factories in the Bangladesh RMG industry. (Liu et al. 2019)	Workers	Alliance; Accord
		Workplace safety violation. (Wiengarten et al. 2017a)	Workers	OSHA
	Performance opacity	Misrepresentation of product features and performance. (Belavina et al. 2020)	Backers	n.a.
Fraud	Funds misappropriation	Keeping backers' money despite not delivering the product. (Belavina et al. 2020)	Backers	n.a.
	Counterfeiting	Offering copies of name-brand goods to take advantage of the economic value associated with the brand. (Yi et al. 2022)	Name brand; Consumer	n.a.

Continued on next page

Table A.1: Types of Organizational Misconduct in the OSCM Literature (Continued)

Type	Phenomenon	Definition	Stakeholder	Judge
	Opioid oversupply	Manufacturing, distributing, or retailing Opioids in excess of the safe market needs that are determined by social control agents seeking to avoid harm to users or society. (Attari et al. 2024, Skilton and Bernardes 2022)	Society	Researcher
Product and service safety issues	Road safety regulation violations	Violations of regulations regarding truck driver driving time and other hours-of-service aspects. (Scott et al. 2021, Scott and Nyaga 2019)	Road users	FMCSA
		Moving violations like speeding, failing to obey traffic signals, following too closely, and improper lane changes. (Scott et al. 2021)	Road users	Police; FMCSA
	Hygiene regulation violations	Violation of restaurant hygiene regulations. (Ibanez and Toffel 2020, Mejia et al. 2019)	Consumers	Health departments

Table A.2: Selected Organizational Misconduct Definitions

Label	Definition	Source
Organizational crime	“(illegal). .. acts (of omission or commission) by individuals... during the course of... work... (in legitimate)... organizations, which they intend to contribute to the achievement of (operative) goals (of) the organization, some subunit within the organization, or their own job duties.”	Finney and Lesieur (1982, p. 266) (adapted from Shover (1978))
Illegal corporate behavior	“acts violating administrative and civil law, resolved through various procedures such as fines, settlements, consent decrees, judgments by government agencies against the firm, and so forth.”	Baucus (1994, p. 700)
Organizational deviance	“an event, activity, or circumstance, occurring in and/or produced by a formal organization, that deviates from both formal design goals and normative standards or expectations, either in the fact of its occurrence or in its consequences, and produces a suboptimal outcome.”	Vaughan (1999, p. 273)
Misconduct	“Acts of omission or commission committed by individuals or groups of individuals acting in their organizational roles who violate internal rules, laws, or administrative regulations on behalf of organizational goals. The extent of adverse consequences and harm to the public will vary with the act, so social cost may be contained or diffused.”	Vaughan (1999, p. 288)
Organizational misconduct	“behavior in or by an organization that a control agent judges to transgress a line separating right from wrong; where such a line can separate legal, ethical, and socially responsible behavior from their antitheses. We define a control agent, in turn, as an actor that represents a collectivity and that can impose sanctions on that collectivity’s behalf.”	Greve et al. (2010, p. 56)

Continued on next page

Table A.2: Selected Organizational Misconduct Definitions (Continued)

Label	Definition	Source
Organizational wrongdoing	“any behavior that organizational participants perpetrate in the course of fulfilling their organizational roles (e.g., as directors, managers, and/or employees) that the state judges to be wrongful. [...] I will consider and act to be wrongful if the state makes a concerned effort to designate the act a violation of the criminal or civil codes and/or imposes on the perpetrator some form of punishment.”	Palmer (2012, p. 35)
Organizational misconduct	“a wide range of behaviors – violations of criminal, civil, and administrative law; transgressions of explicit industry and professional codes; and contraventions of less codified organizational rules, social norms, and ethical principles.”	Palmer (2012, p. 1)
Organizational misconduct	“an illegal, unethical, or socially irresponsible behavior performed by an organization that directly harms its stakeholders (Greve et al., 2010). Typically, behaviors that qualify as misconduct include actions that cause harm to stakeholders and are explicitly illegal, punishable by law (Braithwaite, 1989), or are unethical and deemed unacceptable according to societal norms (Sharpe, 1993). These actions have the potential to harm stakeholders physically, emotionally, and/or economically.”	Hersel et al. (2019, p. 549)

Appendix B

Appendix to Chapter 3

B.1 Overview of the Quality and Operations Officer Roles

Table B.1: Detailed Overview of the Quality and Operations Officer Roles

Rolename	Rolename
Senior VP - Ops	Chief Operations Officer
Vice President - Ops	President - Quality
Executive VP - Ops	Regional President - Operations
Vice President - Quality	Senior VP/Chief Quality Officer
President - Operations	Vice President - Global Operations
Senior VP - Quality	Division Executive VP - Ops
Vice President - Regulatory Affairs/Quality	Executive VP - Quality
Senior VP - Quality & Regulatory Affairs	Senior VP - Worldwide Ops
Division Senior VP - Ops	Vice President - Quality Affairs
Executive VP - Technical Operations	Chief Medical Advisor
Executive VP - International Operations	Corporate VP - Operations
Division VP - Ops	Executive VP - Global Operations
Regional VP - Operations	Group VP - Ops
Regional Senior VP - Ops	Senior VP - Global Operations
Vice Chairman - Operations	Senior VP - HR/Ops
Vice President - Worldwide Ops	Chief Regulatory Officer/VP
Division President - Ops	Senior VP/Chief of Operations
Executive VP/Chief Compliance Quality & Risk Officer	Regional Executive VP - Ops
Senior VP - Sales/Ops	Senior VP - Ops/COO
Senior VP/Chief Operations Officer	Senior VP/Chief of Operations
Vice President - Global Operations/Supply Chain	Vice President - Ops/Development
Senior VP/Chief Technical Operations Officer	Vice President - Operation/Quality
Vice President - Technical Operations	Vice President - Ops/Sales

Notes: Following established practice to identify specific officers (e.g., Fu et al. 2020, Roh et al. 2016, Hambrick and Cannella 2004), we identify quality and operations officers if their titles include relevant keywords such as “Ops,” “Operations,” or “Quality.”

Appendix C

Appendix to Chapter 4

C.1 Variable Definitions

Table C.1: Description of Relevant Variables

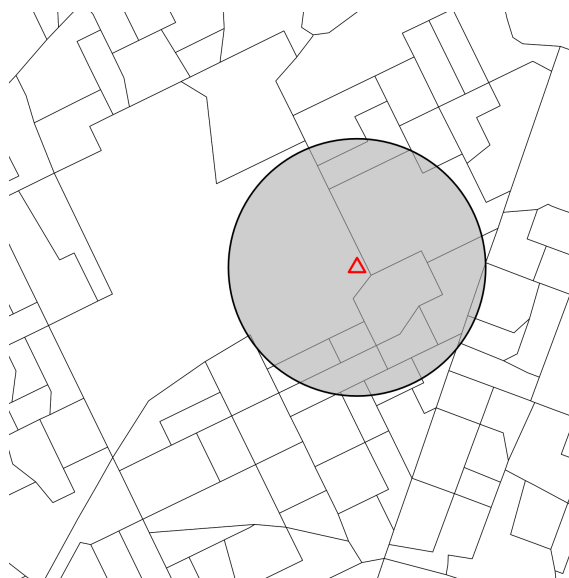
Variable	Unit	Definition	Source
Toxic Releases (in 1,000s lb)	1000 Pounds	The total sum of a specific chemical that a facility releases in a year.	TRI
After CSO Appointment	{0,1}	A binary variable that equals 1 if the observation is after the parent firm created a CSO role and is zero otherwise.	BoardEx
Underserved	{0,1}	A binary variable that equals 1 if the percentile of Black/African American in a facility's host neighborhood is higher than the median value for Black/African American in the facility's state.	ACS
Production Ratio	n.a.	The ratio of the current year's production to the prior year's production of the product(s) that contribute to the generation of the chemical as reported by the facility.	TRI
Toxicity Ratio	n.a.	The ratio of the facility's total toxicity-weighted chemical releases in the current year to the prior year. The variable is log-scaled to reduce the skew.	TRI, RSEI
Tenure of Certifying Official	Years	The number of subsequent years in which the same official signed the TRI submission form for the chemical.	TRI
Tenure of Technical Person	Years	The number of subsequent years in which the same technical person was listed on the TRI submission form for the chemical.	TRI
Environmental Management System (EMS)	{0,1}	A binary variable that equals 1 if the facility reports any source reduction activities attributable to an EMS in the last three years and is zero otherwise.	TRI
Regulatory Pressure	[0,100]	The extent to which the U.S. House of Representatives members from a state voted in favor of more stringent environmental regulations.	LCV
Republican Vote in County	[0,100]	The percentage of voters in the county that voted for the Republican nominee in the last Presidential election.	MEDSL
Total Population (in 100s)	Count	The total population (in 100s) residing in the 1-mile host neighborhood of a facility.	ACS
Black/African American	[0,100]	The percentile of Black or African American persons residing in the 1-mile host neighborhood of a facility.	ACS

Note. TRI: Toxic Release Inventory; ACS: American Community Survey; RSEI: Risk-Screening Environmental Indicators; LCV: League of Conservation Voters; MEDSL: MIT Election Data and Science Lab

C.2 Computing the Demographics Around a Facility

To compute the demographic characteristics of the population in a facility's host neighborhood we combine facility longitude and latitude information from the TRI data with block group level information from the U.S. Census American Community Survey using the R packages `tidycensus` (Walker and Herman 2023), `tigris` (Walker 2023), and `sf` (Pebesma and Bivand 2023, Pebesma 2018). Figure C.1 illustrates the computation of the demographics in a 1-mile-radius host neighborhood around a facility, using the example of an Ecolab facility in Philadelphia County, Pennsylvania. The figure displays part of the county, with the triangular marker indicating the facility's location. The borders within the figure represent census block groups for which we collect demographic information. The circle is a 1-mile-radius host neighborhood around the facility. We compute the demographic characteristics as the weighted average of the percentage of Black or African American persons in the host neighborhood, where the overlap between the host neighborhood and the census block group is the weight.

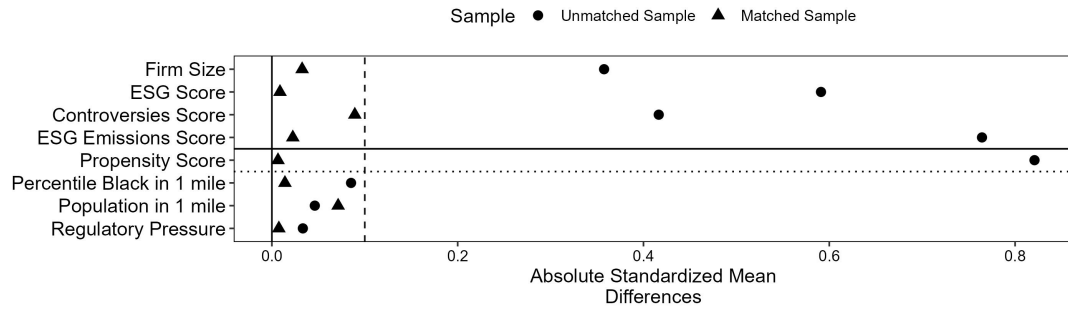
Figure C.1: Example of 1-mile-Radius Host Neighborhood



The figure presents the 1-mile-radius host neighborhood around an Ecolab facility in Philadelphia County, Pennsylvania.

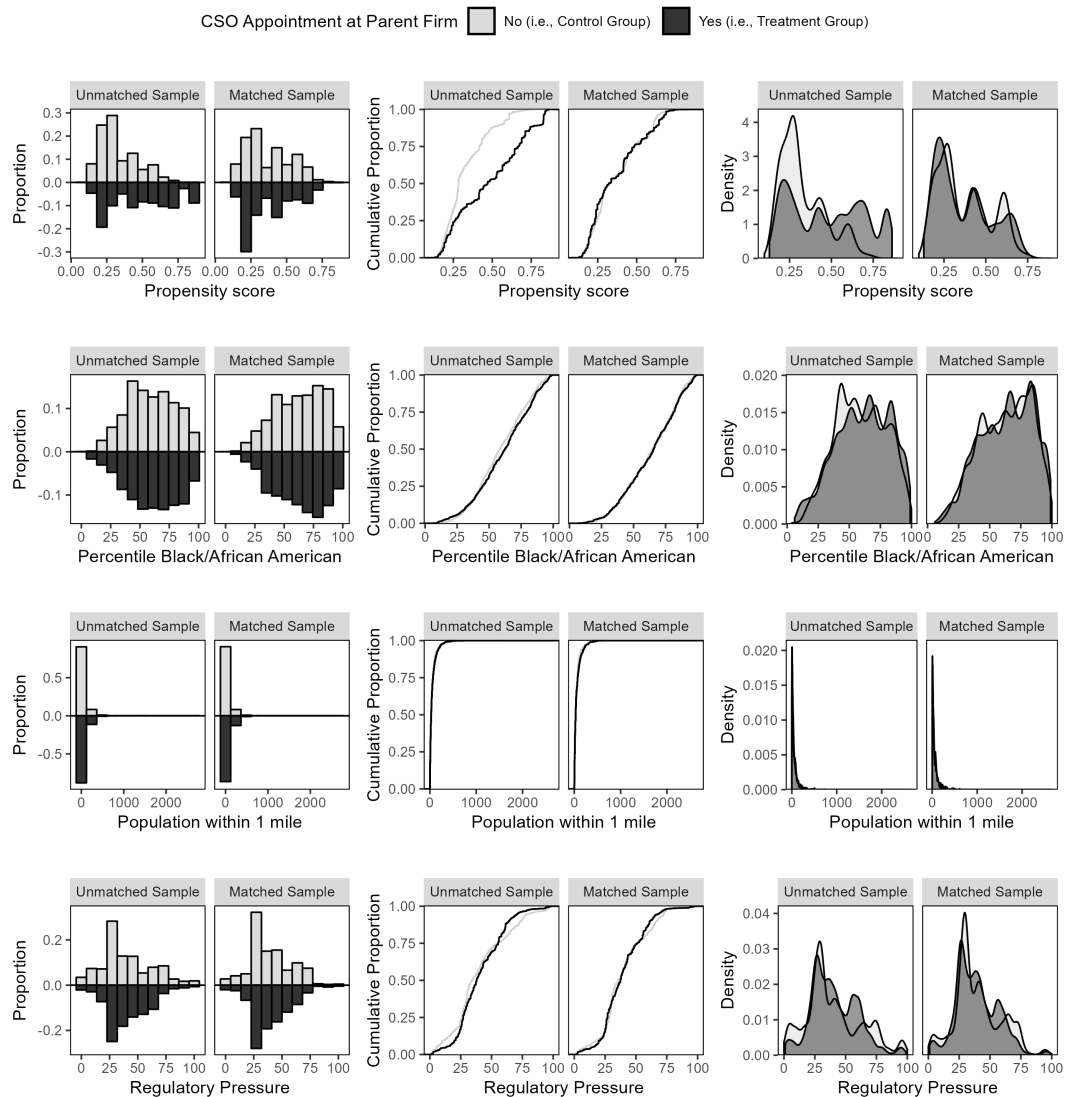
C.3 Variable Balance After Matching

Figure C.2: Variable Balance After Matching: Covariate Balance



The plot presents the matching variables on the y-axis and their standardized mean differences between the unmatched and matched samples on the x-axis. Circular points indicate the standardized mean difference in the unmatched sample; triangular points signify the corresponding values in the matched sample. Variables above the solid line are those used to compute the propensity score. Variables below the solid line are those used in the CEM step. The figure highlights that the matching procedure balances the sample's characteristics.

Figure C.3: Variable Balance After Matching: Distributional Balance



Akin to Figure 4.2, these plots present the distributional balance for the 'Unmatched' and 'Matched' samples but focus on the balance of the individual matching variables. The sub-figures in the left column display histograms of the distributional densities, the sub-figures in the middle column display the empirical cumulative distribution functions, and the sub-figures in the right column display overlapping density plots. The figure highlights that the matching procedure balances the sample's characteristics.

C.4 Linear Regression for Pre-Treatment Trends

Table C.2: Linear Regression Results of Pre-Treatment Trends

Two-Way Fixed Effects Coefficients	
Intercept	1.075 (1.268)
Time to CSO Appointment	0.194 (0.383)
Num.Obs.	64

Note. The table presents the results of a pre-treatment trend test. Each observation is a group–relative year coefficient from the dynamic treatment effects event study in Section 4.6.1. Here, a group refers to a CSO appointment year (e.g. 2005 or 2013), and a relative year refers to the time to CSO appointment (we include -5 to -2 in the analysis because -1 is the omitted reference category). The insignificant intercept indicates that the pre-treatment coefficients are not different from zero and the insignificant Time to CSO Appointment coefficient indicates that there is no trend in the pre-treatment coefficients. Heteroskedasticity robust standard errors are in parentheses. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

C.5 Alternative Matching Methods

C.6 Alternative Measures of Underserved Communities

In our main analysis, we define underserved communities based on the percentage of Black or African American individuals residing within a 1-mile radius of the host neighborhood. As a robustness check, we assess whether our findings hold under alternative definitions of underserved communities, varying our approach along three dimensions:

1. Expanding the definition to include both *Black/African American* and the so-called *Demographic Index*, which accounts for both race and income levels (EPA 2022).
2. Adjusting the geographic scope by considering both a 1-mile and 3-mile radius to define the host neighborhood.
3. Using different threshold values, specifically the 50th and 80th percentile, to classify communities as underserved.

We compute the *Demographic Index* as follows:

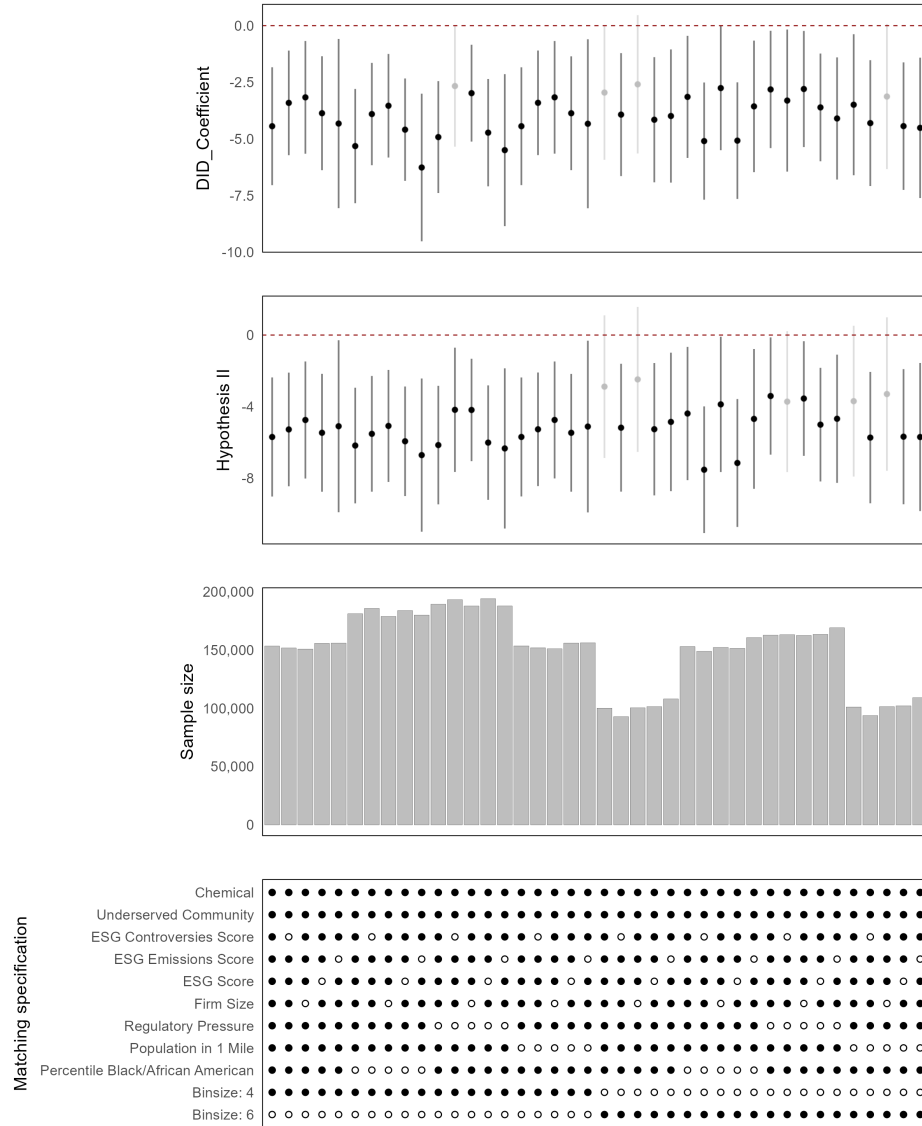
$$Demographic\ Index_{it} = 1 - \frac{Black/African\ American_{it} + Low\ Income_{it}}{2}, \quad (C.1)$$

where *Black/African American*_{it} is defined in Eq. (4.1) and *Low Income*_{it} is defined according to (EPA 2022):

$$Low\ Income_{it} = 1 - \frac{(People\ with\ Ratio\ of\ Income\ to\ Poverty\ from\ 2.00\ and\ Above)_{it}}{Total\ Population\ Whose\ Poverty\ Status\ is\ Known_{it}}. \quad (C.2)$$

Table C.3 presents the results of these alternative specifications. Consistent with our main findings (Panel A, Columns 1-3), we find that our results remain robust across different definitions of underserved communities. Specifically, appointing a CSO is associated with a reduction in toxic releases, with a stronger effect observed for facilities located in underserved communities.

Figure C.4: Specification Analysis of Alternative Matching Settings



This figure shows the robustness of the analysis to different matching settings. The first panel presents the coefficient and 95% confidence interval corresponding to our main estimate of ‘After CSO Appointment’ (Column 3, Table 4.3). The second panel presents the coefficient and 95% confidence interval corresponding to ‘After CSO Appointment’ in underserved communities (Column 5, Table 4.3). The third panel presents the sample size of the matched sample. The fourth panel presents the matching settings with full bullets indicating the setting that has been used. For instance, the leftmost specification uses exact matching on Chemical and community type combined with coarsened exact matching using four bins for Regulatory Pressure, Population in 1 Mile, Percentile of Black/African American, and a CSO appointment propensity score (calculated from ESG Controversies Score, ESG Emission Score, ESG Score, and Firm Size).

Table C.3: Robustness Results for the Moderation Effect Analysis with Alternative Measures for Underserved Community

	Toxic Releases (in 1,000s lb)					
	Cutoff: 50 th percentile			Cutoff: 80 th percentile		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Using <i>Black/African American</i> within 1 mile to identify underserved communities						
After CSO Appointment	-4.582** (1.427)	-1.046 (1.692)	-6.352*** (1.910)	-4.525** (1.374)	-3.021* (1.382)	-9.154** (3.449)
Num.Obs.	125,878	43,412	82,466	130,579	104,538	26,041
R^2	0.852	0.876	0.843	0.859	0.859	0.862
Panel B: Using <i>Demographic Index</i> [†] within 1 mile to identify underserved communities						
After CSO Appointment	-5.002*** (1.488)	-2.288 (1.940)	-6.753** (2.091)	-4.267** (1.316)	-3.222* (1.336)	-9.448* (3.762)
Num.Obs.	122,892	47,088	75,804	128,634	108,656	19,978
R^2	0.852	0.838	0.858	0.851	0.848	0.867
Panel C: Using <i>Black/African American</i> within 3 miles to identify underserved communities						
After CSO Appointment	-4.931*** (1.495)	-2.843 (2.391)	-5.638** (1.840)	-6.023*** (1.531)	-5.451*** (1.593)	-8.404* (3.778)
Num.Obs.	130,831	37,870	92,961	133,823	106,155	27,668
R^2	0.847	0.860	0.843	0.853	0.854	0.851
Panel D: Using <i>Demographic Index</i> [†] within 3 mile to identify underserved communities						
After CSO Appointment	-4.626*** (1.368)	-2.897 (1.586)	-5.608** (1.931)	-4.463** (1.413)	-3.289* (1.362)	-11.864* (4.609)
Num.Obs.	125,306	45,940	79,366	128,950	114,708	14,242
R^2	0.855	0.869	0.848	0.852	0.852	0.855
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Facility-Chemical Pair FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Sample	All	Better-Served Community	Underserved Community	All	Better-Served Community	Underserved Community

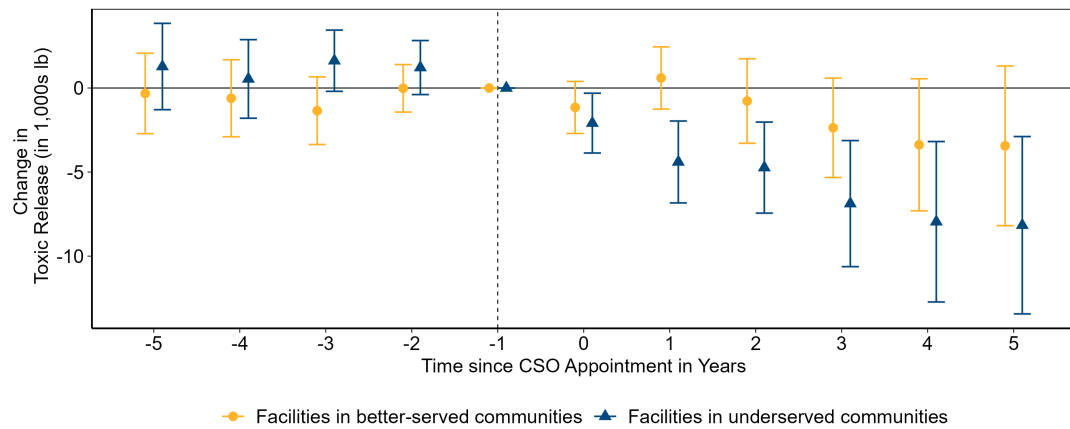
Note. The table presents the results of a two-way fixed effects regression with Toxic Release Ratio as the dependent variable. Each panel represents a different sample based on alternative definitions of underserved communities. The unit of analysis is facility-chemical-year. The controls include all variables from Section 4.4.2. Clustered standard errors in parentheses. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

†: We compute the *Demographic Index* for facility i as the average between the percentage of Black or African American persons and the percentage of persons living below twice the poverty line (i.e., low-income persons) in the facility's host neighborhood.

C.7 Sub-Sample Event Study

Section 4.6.1 presents an event study plot using the dynamic treatment effects estimator from Sun and Abraham (2021). Here, we present the same event study plot but distinguish between facilities in better-served and underserved communities. The regression coefficients corresponding to this plot are in Columns (2) and (3) of Table 4.4.

Figure C.5: Event Plot of the Effect of CSO Appointments on Toxic Releases by Community Type

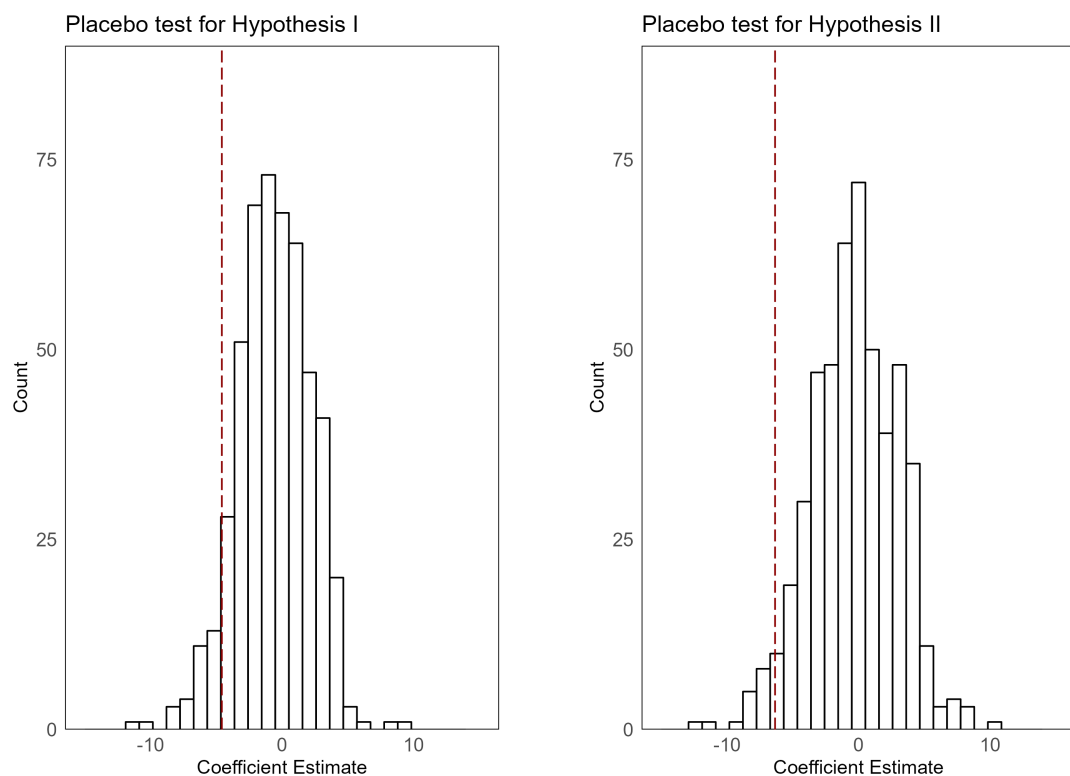


This figure shows the dynamic treatment effects for five years before to five years after the appointment of a Chief Sustainability Officer (CSO) by the type of community in which a facility is located. The y-axis reports the estimated coefficients and 95% confidence intervals for the dependent variable: toxic releases in 1,000 pounds. The x-axis shows the relative time from the CSO appointment, with negative values indicating years before the appointment.

C.8 Placebo Analysis

As discussed in Section 4.6.1, we implement a placebo analysis to further reduce concerns that our results might be driven by chance (Dhanorkar and Muthulingam 2020, Dhanorkar 2019). In this analysis, we assign observations to the treatment or control group by randomly allocating a CSO appointment year at the firm level. The below Figure shows the distribution of coefficients from 500 placebo samples.

Figure C.6: Placebo Analysis Results



This figure displays the distribution of coefficients for ‘After CSO Appointment’ obtained by re-estimating Equation 4.2 in 500 pseudo-samples generated by assigning firms to the treatment or control group by randomly allocating a CSO appointment year to them. Panel A represents the coefficients corresponding to our main analysis (Column 3, Table 4.3. The dashed vertical line denotes our original estimate. It is the 34th largest negative value, implying a p-value of 0.068. Panel B represents the coefficients corresponding to our analysis for underserved communities (Column 5, Table 4.3. Our original estimate is the 18th largest negative value, implying a p-value of 0.036.

C.9 Instrumental Variable Approach

Thus far, we have relied on a TWFE approach to identify the effect of appointing a CSO on toxic releases. In this section, we employ an instrumental variable approach as an alternative identification strategy to strengthen the robustness of our findings. This approach helps to alleviate concerns that our main findings are driven by a broader shift in a firm’s environmental priorities rather than the CSO appointment itself (Section 4.6.2).

Instrumental variables are a common identification strategy in governance-related operations management research (Ambulkar et al. 2023, Wowak et al. 2021). Using an instrument allows us to isolate the effects of appointing a CSO from other factors. This approach not only provides an additional means of identification, bolstering our research’s internal validity, but also allows us to examine the full sample without losing observations to matching.

One special feature of our treatment is that it is binary. We leverage this by using a method proposed by Wooldridge (2010) that incorporates a binary treatment into a two-stage least squares (2SLS) approach. This method uses a probit model to predict whether a firm will appoint a CSO based on certain factors, then uses those predictions as the instrument for the CSO appointment in a 2SLS model. Specifically, we first estimate the following zero-stage probit regression, in which we regress CSO presence on our established control variables (X_{ijct}) and a firm-level instrument (Z_{jt}):

$$Pr(CSO = 1|X_{ijct}, Z_{jt}) = \Phi(\beta X_{ijct}, \gamma Z_{jt}) \quad (\text{C.3})$$

We then use the predicted values from Equation C.3, \widehat{CSO} , as the instrument in a 2SLS approach. We use this approach for two reasons. First, it yields more precise estimates than omitting the zero-stage probit estimation (Wooldridge 2010). Second, it does not technically require an excluded instrument for identification because of the nonlinear relationship between different stages (Ren et al. 2011).

However, to make the source of the identification clearer, we include a well-established instrumental variable: the proportion of directors on a focal firm’s board who also sit on other boards where the other firm has a CSO (Ambulkar et al. 2023, Wowak et al. 2021, Adams and Ferreira 2009). We lag this variable by one year to ensure that it does not capture changes in the priorities of the firm, which influence both the board composition and the CSO appointment.

A valid instrumental variable must satisfy two key assumptions: relevance and exclusion restriction. The relevance assumption is met if the instrument is correlated with the decision to appoint a CSO. The exclusion restriction holds if the instrument affects toxic releases only through the appointment of a CSO and not through other channels. We argue that our instrument fulfills both criteria. First, the exclusion restriction is upheld by the nature of our instrument. The directors that we use to construct our instrument come from a diverse set of industries. Many of these industries, such as finance and services, do not directly engage in activities related to toxic emissions or manufacturing. As a result, board composition shaped by these industries is unlikely to directly affect toxic releases at the focal

firm, except through the appointment of a CSO. Additionally, previous studies (Ambulkar et al. 2023, Wowak et al. 2021, Adams and Ferreira 2009) have used similar instruments in the context of governance decisions, demonstrating that while governance decisions across firms may influence one another, they do not directly impact operational outcomes like toxic releases, further supporting the exclusion restriction.

The relevance assumption holds because, as shown by Gupta et al. (2021), firm governance practices, including the appointment of a CSO, are subject to mimetic isomorphism. When directors hold board positions at firms that have appointed CSOs, they are likely to encourage the same practice at the focal firm, increasing the probability of CSO appointment. This mechanism establishes a clear link between our instrument and the likelihood of appointing a CSO.

Table C.4: 2SLS Analysis of the Effect of CSO Presence on Toxic Releases

	1 st Stage			2 nd Stage		
	(1)	(2)	(3)	(4)	(5)	(6)
\widehat{CSO} from (C.3)	1.066*** (0.024)	1.054*** (0.041)	1.075*** (0.028)			
Presence of a CSO				-9.266*** (1.981)	-5.669* (2.713)	-11.446*** (2.711)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Facility-Chemical Pair FE	No	No	No	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Sample	All	Better-Served Community	Underserved Community	All	Better-Served Community	Underserved Community
Num.Obs.	384,151	154,550	229,601	384,151	154,550	229,601
R^2	0.786	0.789	0.786	0.851	0.857	0.848
Kleibergen-Paap rk Wald F statistic	2033.45	662.75	1465.66			

Note. The table presents the results of the instrumental variable analysis with toxic releases in 1000 pounds as the dependent variable. The instrument is the fraction of directors on the board who sit on other boards for firms that have a CSO. The controls include all variables from Section 4.4.2. Clustered standard errors in parentheses. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Table C.4 presents the results of our 2SLS analysis. The independent variable of interest is the presence of a CSO. Columns (1) through (3) present the first stage results. Column (4) presents results corresponding to Hypothesis 1, and columns (5) and (6) present results corresponding to Hypothesis 2. The first stage coefficients for \widehat{CSO} as the instrument are significant ($p < 0.001$), and the F statistics exceed the Stock and Yogo (2002) critical value of 16.38. This indicates the strong statistical relevance of our instrument. Consistent with our main findings, the presence of a CSO results in a significant reduction of toxic releases (-9.266 ; $p < 0.001$). Furthermore, the effect is stronger for facilities in underserved communities (-11.446 ; $p < 0.001$) than for those in better-served communities (-5.669 ; $p < 0.05$). Together, these results lend confidence to our findings.

C.10 Alternative Explanation: Any CSO Role Affects Toxic Releases

A potential alternative explanation for our main findings is that any intervention aimed at increasing awareness of sustainability issues would suffice to create the observed effect (see Section 4.6.2 for a detailed discussion). We test this using the creation of ESG committees as an example. The results in Table C.5 show that creating ESG committees is not associated with changes in toxic releases—regardless of community type. These findings underscore that not all governance interventions designed to heighten sustainability awareness produce the same effects and reinforce the distinct role of CSOs. Thereby, ruling out this alternative explanation.

Table C.5: Two-Way Fixed Effects Results of the Effect of ESG Committee Creation on Toxic Releases

	Toxic Releases (in 1,000s lb)				
	(1)	(2)	(3)	(4)	(5)
After ESG Committee	1.532 (2.217)	1.689 (2.153)	1.689 (2.156)	0.071 (2.498)	2.492 (2.690)
Production Ratio		2.430** (0.852)	2.430** (0.853)	2.531 (1.326)	2.477* (1.055)
Toxicity Ratio		0.732** (0.252)	0.732** (0.252)	1.151** (0.408)	0.291 (0.253)
Tenure of Certifying Official		0.556 (0.360)	0.556 (0.361)	0.146 (0.260)	0.800 (0.526)
Tenure of Technical Person		0.431* (0.204)	0.431* (0.205)	0.685** (0.263)	0.323 (0.292)
Environmental Management System		-0.240 (1.005)	-0.240 (1.007)	0.411 (1.982)	-0.563 (1.167)
Regulatory Pressure		-0.072 (0.086)	-0.072 (0.087)	-0.056 (0.105)	-0.045 (0.102)
Republican Vote in County		0.010 (0.114)	0.010 (0.114)	-0.031 (0.194)	0.067 (0.079)
Total Population (in 100s)		-0.005 (0.163)	-0.005 (0.164)	0.040 (0.348)	-0.068 (0.168)
Black/African American		0.130 (0.145)	0.130 (0.146)	0.144 (0.145)	-0.244 (0.396)
Facility-Chemical Pair FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Firm FE	No	No	Yes	Yes	Yes
Sample	All	All	All	Better-Served Community	Underserved Community
Num.Obs.	62,448	62,448	62,448	24,584	37,864
R^2	0.803	0.804	0.804	0.835	0.787

Note. The table presents the results of a two-way fixed effects regression with toxic releases in 1000 pounds as the dependent variable. The unit of analysis is facility-chemical-year. Akin to the identification of CSOs, we identify ESG committees from the BoardEx database based on committee titles containing “sustainability,” “sustainable,” “responsibility,” “ethics,” or “environment.” Clustered standard errors in parentheses. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

C.11 Alternative Explanation: Outsourcing of Production

A potential alternative explanation for our main findings is that facilities outsource or reduce their production following the appointment of a CSO, leading to a reduction of toxic releases (see Section 4.6.2 for a detailed discussion). We test this using the production ratio as the dependent variable. The results in Table C.5 show that the reduction in toxic releases following a CSO appointment is not driven by outsourcing or production displacement. Thereby, ruling out this alternative explanation.

Table C.6: Two-Way Fixed Effects Results of the Effect of CSO Appointments on Production Ratio

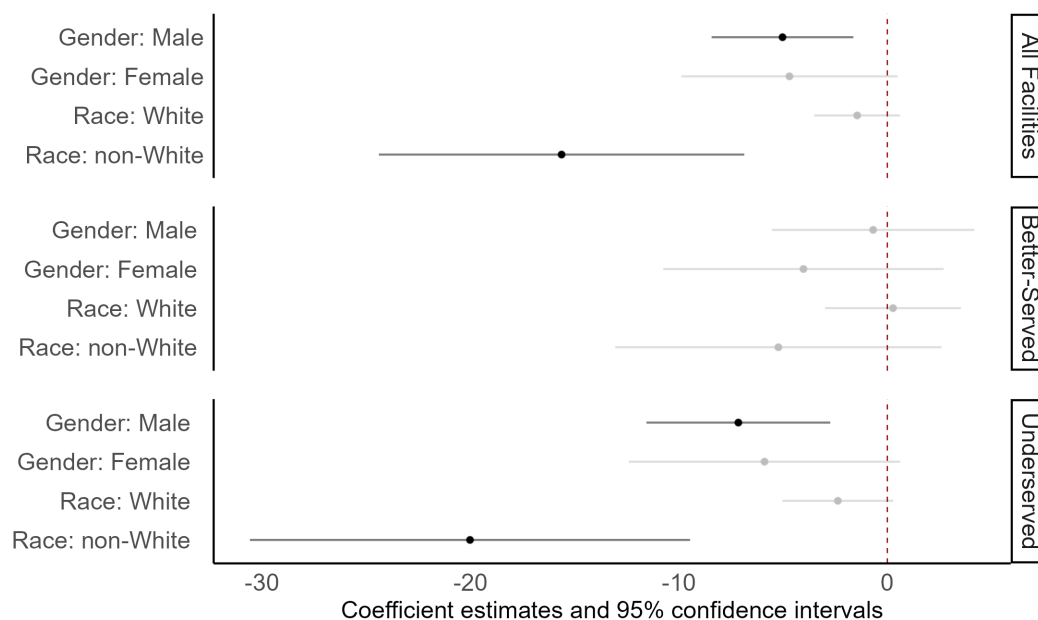
	Production Ratio				
	(1)	(2)	(3)	(4)	(5)
After CSO Appointment	0.010 (0.010)	0.008 (0.010)	0.008 (0.010)	0.003 (0.016)	0.009 (0.013)
Toxicity Ratio		0.038*** (0.004)	0.038*** (0.004)	0.042*** (0.008)	0.036*** (0.005)
Tenure of Certifying Official		0.002 (0.001)	0.002 (0.001)	0.003 (0.002)	0.001 (0.001)
Tenure of Technical Person		-0.002 (0.001)	-0.002 (0.001)	-0.006** (0.002)	0.000 (0.001)
Environmental Management System		-0.003 (0.010)	-0.003 (0.010)	-0.009 (0.015)	-0.001 (0.013)
Regulatory Pressure		0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Republican Vote in County		0.001 (0.001)	0.001 (0.001)	0.000 (0.001)	0.001 (0.001)
Total Population (in 100s)		0.001* (0.000)	0.001* (0.000)	-0.002 (0.001)	0.002** (0.001)
Black/African American		-0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)
Facility-Chemical Pair FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Firm FE	No	No	Yes	Yes	Yes
Sample	All	All	All	Better-Served Community	Underserved Community
Num.Obs.	125,878	125,878	125,878	43,412	82,466
R^2	0.142	0.149	0.149	0.151	0.150

Note. The table presents the results of a two-way fixed effects regression with Production Ratio as the dependent variable. The unit of analysis is facility-chemical-year. Clustered standard errors in parentheses. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

C.12 Heterogeneity of CSO Appointment Effect

The findings about the effect of CSO appointments prompt an interesting question: Does the effect of appointing a CSO differ depending on the characteristics of who fills the office? Factors like CSO gender and race could both have a significant impact on the outcomes. Figure C.7 examines these factors in the context of CSO appointments.

Figure C.7: Effect of CSO Appointments based on the CSO's Race and Gender



The coefficients are based on estimating Equation (4.2) with different CSO subsets as the treatment. We identify male and female CSOs based on the gender listed in BoardEx, and race/ethnicity using the R package `predictrace`, supplemented by a manual review of officers' online presence. Out of the 165 overall appointments, we classify 113 as male and 52 as female; 134 as White and 31 as non-White. In this context, "non-White" refers to individuals who likely self-identify with a race or ethnicity other than White, including but not limited to Hispanic, Indian, and Black). These classifications apply to each firm's first CSO appointment. They are used solely for explanatory analysis and should not be interpreted as definitive or exhaustive indicators of identity.

The results suggest that appointments of non-White CSOs are associated with greater reductions in toxic releases compared to their White counterparts, while no significant difference is observed between male and female CSOs. Although this analysis is exploratory and should not be interpreted as causal, it highlights a potentially important direction for future research on the intersection of leadership diversity and environmental outcomes.

Appendix D

Appendix to Chapter 5

D.2 Overview of Feature Set

Table D.1: Description of Predictors for Machine Learning Model

Feature Set	Predictor	Description	Source
Time-Only	WSL	The time (in weeks) since the last routine or re-inspection for the restaurant.	SNHD
	Lagged Violation	An indicator variable taking the value 1 if the last routine inspection resulted in a Violation and zero otherwise.	SNHD
	Lagged Grade “A”	An indicator variable taking the value 1 if the last routine inspection resulted in an “A” grade and zero otherwise.	SNHD
Inspection-Record	Lagged Grade “B”	An indicator variable taking the value 1 if the last routine inspection resulted in an “B” grade and zero otherwise.	SNHD
	Lagged Grade “C”	An indicator variable taking the value 1 if the last routine inspection resulted in an “C” grade and zero otherwise.	SNHD
	Lagged Demerits	The demerits count from the last routine inspection.	SNHD
Review-Information	Count of Reviews	The number of Yelp reviews posted since the previous inspection.	Yelp
	Star Rating	The average star rating for reviews posted since the last inspection.	Yelp
	Bad Reviews	The number of reviews with three or fewer stars since the last inspection.	Yelp
	Review Text	A review text score based on Multinomial Inverse Regression (see Section D.3 for details).	Yelp
Full-Model	Yelp Category [†]	Indicator variables taking the value 1 if the restaurant is classified in the respective category and zero otherwise.	Yelp

Notes. † These categories include: American (New), Asian Fusion, Barbeque, Bars, Breakfast & Brunch, Burgers, Cafes, Chicken Wings, Chinese, Coffee & Tea, Delis, Deserts, Event Planning & Services, Fast Food, Italian, Juice Bars & Smoothies, Mediterranean, Mexican, Nightlife, Pizza, Salad, Seafood, Specialty Food, Sports Bars, Sushi Bars, Thai.

D.3 Multinomial Inverse Regression for Review Text Variable Construction

To construct our review text measure, we used multinomial inverse regression, which was first suggested by Taddy (Taddy 2013b, 2015) and recently used by Campbell and Shang (2022) to predict misconduct from text measures. Specifically, to capture consumer sentiment about restaurant hygiene before an inspection, we start with all reviews that were posted on Yelp after the last inspection and before the focal inspection. We then clean up the review texts by removing all punctuation, stop words, and numbers using the Python Natural Language Toolkit (Loper and Bird 2002). We also replace each word with its word stem using the “Snowball Stemmer” which is an extension of the Porter stemming algorithm (Porter 1980).

We then extract all words from the preprocessed review text in the period between 2012 and 2015 - our training period. This step results in an inspection-word matrix with 137,098 words corresponding to 11,341 routine inspections. The matrix includes the number of times a given word appeared in reviews before the inspection for each of the 1,554,828,418 possible inspection-word combinations. Only 4,691,471 of these have non-zero values. We then reduce the sparseness (i.e. the proportion of entries with 0s to entries with other values) of our data to 99% by discarding the words that appear in the fewest reviews. The reason for this step is that these words are unlikely to be useful in predicting future inspection outcomes (Campbell and Shang 2022). This step leaves us with 44,582 words. We could use the count of these words as predictor variables in a forward regression to predict inspection outcomes. However, this would be computationally extensive, and it is advisable to collapse the number of variables (Taddy 2013b).

We follow Campbell and Shang (2022) and reduce the number of dimensions of the review words to one using multinomial inverse regression (Taddy 2013b, 2015). Our data consists of the $11,341 \times 44,582$ inspection-word matrix \mathbf{C} , where each element $c_{i,d}$ is the word count for word d and inspection i . The total word count for inspection i sums to $m_i = \sum_{d=1}^{44,582} c_{i,d}$ which we capture for all inspections in the vector \vec{m} . Furthermore, we have the inspection outcome v_i and a vector of controls \vec{x}_i .

Then, we can estimate a separate Poisson regression for every word d :

$$E[c_i | x_i, v_i] = e^{\alpha_d + \beta x_i + \varphi v_i + \epsilon_i}$$

Each of the 44,582 separate Poisson regressions has a negative log likelihood that is proportional to (Taddy 2015):

$$l(\alpha, \beta, \varphi) = \sum_{i=1}^{11,341} \left[m_i e^{\alpha + \beta x_i + \varphi v_i} - c_i (\alpha + \beta x_i + \varphi v_i) \right] \quad (\text{D.1})$$

We estimate (D.1) using a L1 regularization (Lasso) (Tibshirani 1996) minimization approach to avoid overfitting and consequently improve the predictive ability. We minimize

the following specification using the “distrom” package in R (Taddy 2015):

$$\hat{\alpha}, \hat{\beta}, \hat{\varphi} = \arg \min_{\alpha, \beta, \varphi} \{l(\alpha, \beta, \varphi) + n\lambda[\beta + \varphi]\}, \quad (\text{D.2})$$

where $l(\alpha, \beta, \varphi)$ is as defined in (D.1) and $\lambda \geq 0$.

Afterwards, we extract φ for every word d and compute one reduced score z_i per inspection i as out review text measure (Campbell and Shang 2022).

$$z_i = \sum_d \varphi_d \frac{c_{id}}{m_i}$$

Clearly, if φ is positive for a word—meaning that this word appears more often before a negative inspection outcome—then it increases z_i . Similarly, if φ is negative—meaning that this word appears less frequent before a negative inspection outcome—then it decreases z_i . Thus, z_i is a sufficient reduction of the underlying review text information (Taddy 2013a).

D.4 Comparison of Different Machine Learning Algorithms

Table D.2: Comparing the Predictive Performance of Machine Learning Models

	GBRT ^a	Random Forest ^b	Neural Network ^c	Support Vector Machine ^d
	(1)	(2)	(3)	(4)
Brier [†]	0.180 (0.012)	0.188 (0.010)	0.231 (0.005)	0.209 (0.009)
LogLoss [†]	0.539 (0.029)	0.560 (0.026)	0.653 (0.011)	0.609 (0.022)
Accuracy	0.736 (0.023)	0.721 (0.019)	0.600 (0.011)	0.688 (0.018)
Precision	0.396 (0.014)	0.385 (0.020)	0.315 (0.024)	0.349 (0.021)

Notes. †: Lower values indicate better performance. Standard errors computed based on 25 rolling time samples in parentheses.

a: Gradient Boosted Regression Tree (GBRT). Implemented in R using the caret library and function *xgbTree* from the xgboost library. Based on hyperparameter tuning using 3-times, 5-fold cross validation we use the following hyperparameters: nrounds = 1000, max_depth = 6, eta = 0.001, gamma = 0, colsample_bytree = 0.6, min_child_weight = 1 and subsample = 1.

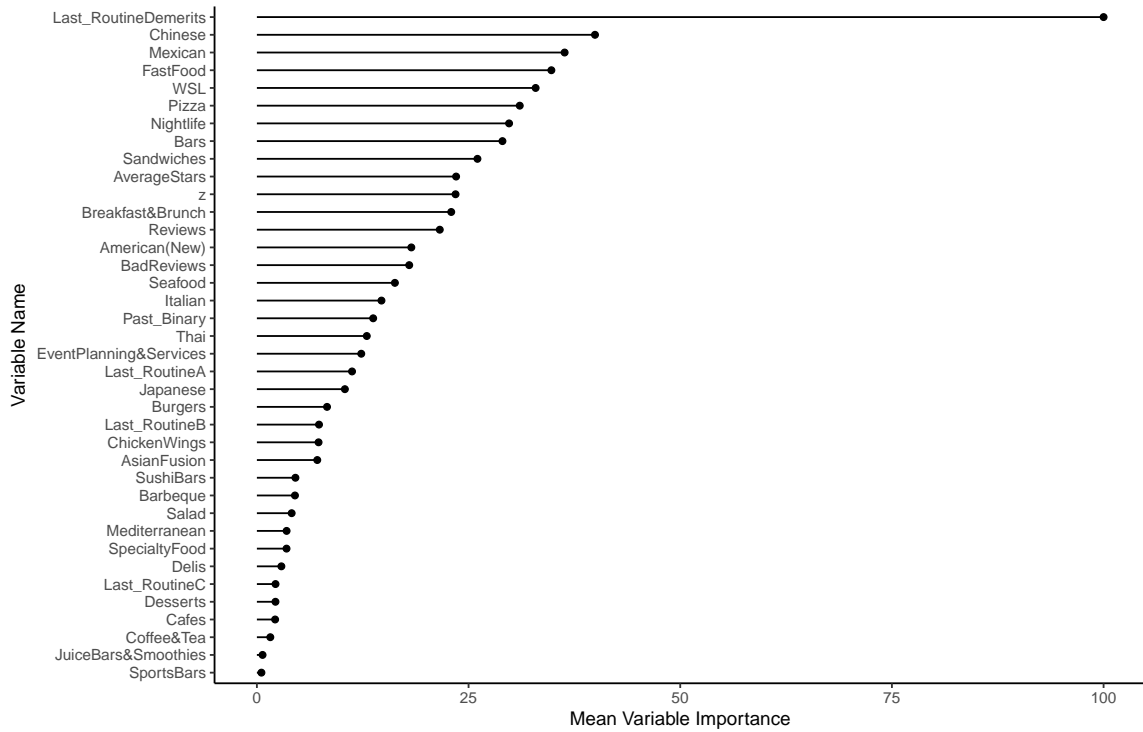
b: Implemented in R using the caret library and function *ranger* from the e1071 and ranger libraries. Based on hyperparameter tuning using 3-times, 5-fold cross validation we use the following hyperparameters: mtry = 2, splitrule = gini and min.node.size = 1.

c: Implemented in R using the caret library and function *ranger* from the e1071 and ranger libraries. Based on hyperparameter tuning using 3-times, 5-fold cross validation we use the following hyperparameters: size = 5 and decay = 1.

d: Implemented in R using the caret library and function *svmRadial* from the kernlab library. Based on hyperparameter tuning using 3-times, 5-fold cross validation we use the following hyperparameters: sigma = 0.01364814 and C = 4.

D.5 Variable Importance

Figure D.1: Variable Importance



Notes. The figure shows the scaled mean variable importance metrics for the full model across all 25 resamples.

D.6 Transition Matrix Construction

This appendix describes how the regulator’s transition probabilities $T_i(s, a_j, a_i, s')$ are derived, as shown in Table 5.3. Each transition probability reflects how the state of the system evolves from s to s' given a joint action (a_j, a_i) by the establishment and the regulator. The values are calculated using fixed probabilistic rules determined by the current state, the nature of the actions taken, and a set of underlying parameters. These rules can be applied manually to determine each entry in the transition matrix.

Transition Parameters. The transitions depend on the following:

- A probability matrix T^{Idle} describing transitions when the establishment makes no effort to improve (e.g., chooses action **Wait** or **Self-Inspect**).
- A probability matrix T^{Effort} describing transitions when the establishment exerts effort.
- A treatment matrix Q that governs how the system exits the inspected states, returning to uninspected ones over time.
- A detection probability μ , which indicates the chance that a violation is detected and corrected if the establishment self-inspects.

Construction of Transitions For each state-action combination (s, a_j, a_i) , the transition probabilities to each possible next state s' are determined as follows:

1. No Inspection ($a_{i,t}^j = \text{Wait}$):
 - If the current state is **Compliance** or **Violation**, then the system evolves according to T^{Idle} or T^{Effort} , depending on whether the establishment chooses **Wait/Self-Inspect** or **Effort**.
 - For example, starting in **Compliance** and with the establishment choosing **Effort**, the probability of remaining in **Compliance** is given by the corresponding row of T^{Effort} .
 - If the current state is inspected (either **Inspected_Compliance** or **Inspected_Violation**), the system transitions using the treatment matrix Q combined with the appropriate transition matrix (T^{Idle} or T^{Effort}) to determine the resulting uninspected state.
2. Inspection ($a_{i,t}^j = \text{Inspect}$):
 - If the current state is **Compliance** or **Violation**, inspection transitions the system into an *inspected* version of a state. The probability of entering **Inspected_Compliance** or **Inspected_Violation** depends on the base transition matrix (again, T^{Idle} or T^{Effort}), weighted by the detection probability.

- For example, starting in **Violation**, and assuming no self-inspection by the establishment, the probability of transitioning to **Inspected_Violation** comes directly from T^{Idle} .

Computing T^{Idle} and T^{Effort} To compute T^{Idle} and T^{Effort} , we first estimated a progression matrix T based on Las Vegas inspection data from 2012 to 2015—the same data that we used to construct our review text measure. If we had a common and constant inspection interval that equaled our cycle length of one-week then we could simply estimate the (s,s') th element of T as the ratio of restaurants that move from state s to state s' to the total number of restaurants that move out of state s . Similarly, if we had a common and constant inspection interval longer than our cycle length of one week, then we could estimate a multi-week progression matrix in the same way and derive the one-week transition probabilities from that matrix. (Craig and Sendi 2002) However, inspections in our data are not conducted at constant intervals but vary both within and between restaurant observations. We overcame this challenge by estimating the one-week progression matrix T based on maximum likelihood estimation given the observed states at each inspection and the time since the last inspection using the R package “msm” (Jackson 2011). We then altered the transition probabilities by slightly increasing the likelihood of staying in state compliance for T^{Effort} and by slightly decreasing the likelihood of staying in state compliance for T^{Idle} . The resulting matrices are as follows:

Table D.3: Transition Probability Matrices

T^{Idle}	T^{Effort}	Q^\dagger
$\begin{bmatrix} 0.965 & 0.035 & 0.000 & 0.000 \\ 0.000 & 1.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.975 & 0.025 & 0.000 & 0.000 \\ 0.000 & 1.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 1.000 & 0.000 \\ 0.000 & 0.000 & 0.990 & 0.010 \end{bmatrix}$

Notes. † To compute Q , we looked at the re-inspection following “B” or “C” grades during routine inspections. The SNHD conducts these within two weeks following the violation and they indicate how likely establishments are to return to a compliance state following a detected violation.

D.7 Simulation Results

Table D.4: Comparison Between Various Inspection Resource Allocation Policies

	Random (1)	Regular (2)	Predictions (3)	INDEX (4)	ML INDEX (5)	
Homogeneous	Reward	-17466 [-18545, -16387]	-18012 [-19250, -16775]	-14961 [-16066, -13856]	-13206 [-14082, -12330]	-12450 [-13450, -11449]
	Improvement	2117 [1675, 2558]	1570 [1354, 1787]	4622 [4030, 5214]	6377 [5835, 6919]	7133 [6524, 7743]
	Reward	-13462 [-14362, -12562]	-12368 [-13310, -11427]	-11381 [-12518, -10245]	-9936 [-10875, -8996]	-9755 [-10612, -8898]
	Improvement	6121 [5474, 6767]	7214 [6494, 7935]	8201 [7455, 8948]	9647 [8906, 10388]	9827 [9134, 10521]
Heterogeneous	Reward	-16371 [-17637, -15105]	-16707 [-18266, -15147]	-15031 [-16370, -13691]	-12706 [-13920, -11492]	-12266 [-13578, -10955]
	Improvement	2480 [1787, 3174]	2145 [1851, 2439]	3821 [3287, 4355]	6145 [5212, 7079]	6585 [5849, 7321]
	Reward	-13099 [-14059, -12139]	-12028 [-13286, -10770]	-11072 [-12242, -9902]	-9251 [-10061, -8441]	-9057 [-9890, -8223]
	Improvement	5753 [4765, 6740]	6824 [5671, 7976]	7780 [7113, 8447]	9601 [8392, 10809]	9795 [8585, 11005]

Notes. The table present the average discounted reward and average improvement over baseline across all settings in Table 5.5. It compares the results of five different policies: *Random* assigns inspection resources randomly, *Regular* assigns inspection resources to the establishments with the longest time since the last inspection, *ML* assigns inspection to the establishments with the highest predicted likelihood of violation, *INDEX* assigns inspection resources according to our algorithm outlined in Section 5.5, and *ML INDEX* assigns inspection resources according to our algorithm outlined in Section 5.5 while incorporating Machine Learning predictions as detailed in Section 5.5.4. 95% confidence intervals in square brackets.