Title
Essays on Food Safety Economics

Author
Wang, Yijing

Publication Date
2023

Essays on Food Safety Economics

By

YIJING WANG
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Agricultural and Resource Economics

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Timothy Beatty, Chair

Dalia Ghanem

Pierre Mérel

Daniel A. Sumner

Committee in Charge

2023

Abstract

This dissertation studies the economics of food regulation, with applications to food safety, third-party certification, and litigation. It contains 2 main parts. The first part of the dissertation contains two essays: the first essay studies the value of certification in signaling food safety; the second highlights the effectiveness of the common auditing system on firms' behavior change. In the second part, I study the impacts of food safety lawsuits on governmental regulatory behaviors and firms' food safety practices.

Third-party certification has been widely adopted in the food industry, but its credibility and effectiveness are controversial. In Part I, I assemble a comprehensive data set using plant inspection records from the U.S. Food Safety and Inspection Service (FSIS) and food safety certification results from both the Safe Quality Food (SQF) and the British Retail Consortium (BRC) global standards program. The first essay examines the role of third-party certification on certified plants' food safety practices and outcomes in the meat and poultry industry. I take the compliance rate of FSIS sanitary tasks as a measure of the level of food safety practices and the FSIS pathogen sampling test results as a measure of food safety outcomes. I find that certification ratings are informative of plants' food safety practices, but not their food safety outcomes. Leveraging the timing of initial and re-certification audits, the second essay finds SQF certified plants have a gradual increase in their average food safety practice level after initial audits, but not BRC certified plants. I do not find a significant impact of re-certification audits on food safety practices.

In Part II, I study the net impact of food safety lawsuits brought by those who claim harm on public food safety regulation intensity and defendant plants' food safety practices by combining a food safety lawsuit dataset that I developed with FSIS inspection records and exploiting information on the timing of lawsuit filing dates. I find that food safety lawsuits have a crowd-out effect on the regulatory intensity in inspection tasks that are directly relevant to the food safety disputed issues in the specific lawsuit. There is no evidence that food safety lawsuits have a statistically significant impact on defendant plants' food safety practices. The empirical result expands the knowledge of the interactions of plaintiffs' claims in litigation, public regulation, and plant behaviors in the context of food safety issues in the US meat and poultry industry.

Keywords: food safety, food regulation, third-party certification, food safety litigation, event study, multiple hypothesis testing

JEL Codes: D22, I18, K32, L66, Q18

Acknowledgments

I am incredibly grateful for my PhD journey in the Department of Agricultural and Resource Economics at UC Davis. I want to express my heartfelt appreciation to my committee chair, Professor Timothy Beatty, for his invaluable guidance and support throughout my PhD and the writing of this dissertation. His expertise and insight played a vital role in shaping the direction of this research and helping me overcome numerous challenges along the way. I am also thankful to my committee members, Professors Dalia Gahanem, Pierre Mérel, and Daniel A. Sumner, for their patience, thoughtful feedback, and constructive criticism, which helped me refine my ideas and produce a stronger final product. I truly appreciate the encouragement they all gave me and the confidence they had in me.

I am also grateful to my classmates and cohort members, particularly Shanchao Wang, Jason Xiao, Bret Stevens, Charlotte Ambrozek, and Marieke Fenton, for their emotional and moral support during my PhD, especially in moments of crisis. Their compassion and love helped me persevere through some of the toughest times.

Finally, I want to extend my deepest gratitude to my family, including my husband Alexander, parents, grandparents, in-laws, and my dog Noodle, for their unwavering support and encouragement throughout this process. Without my husband's support, I would not have been able to stay strong during the most challenging moments. I am grateful for the unconditional love and wisdom my parents and parents-in-law provided me. Lastly, I want to thank Abby for her exceptional editorial efforts in reviewing my work.

Contents

List of Figures

List of Tables

Part I

Third-party certification and food safety

Chapter 1

Introduction

The use of third-party private certification in the food industry to disclose product information and regulate food quality is ubiquitous. The global food certification market is estimated to be worth \$4.7 billion in 2020 and is expected to grow at a compound annual growth rate of 5.4% (MarketsandMarkets, 2020). There were at least 425 international certification standards in 2016 directly relevant to the agri-food industry (Caswell et al., 2017). Food safety certification is one of the most popular types of certifications for agri-food firms. Many large retailers such as Walmart, Kroger, Costco, Target, and Safeway have made third-party food safety certifications mandatory for all of their suppliers.

Despite the widespread adoption of third-party food safety certification schemes, food safety incidents for firms with good audit ratings are not uncommon. For example, in 2019, an outbreak of ground beef Salmonella infections caused thirteen infections, including nine hospitalizations and one death in the United States. Some of the ground beef that caused people ill was traced back to a California meat producer, Central Valley Meat Co. Afterward, the company recalled 34,222 pounds of ground beef products due to Salmonella contamination. This is not the first time that Central Valley Meat Co. produced products with severe health hazard concerns. In 2013, Central Valley Meat Co. recalled 48,760 pounds of ground beef products that may have contained foreign materials. However, the company has a long-standing third-party certification record with excellent scores A or AA from September 2013 onward. These incidents naturally lead to the three following questions:

First, does the information from third-party food safety certification reflect underlying food safety practices and outcomes of a certified plant? Typically, the grade of a third-party certification summarizes how well a plant's food safety system complies with a specific food safety standard. If complying with private standards helps companies improve their food safety practice behaviors and produce safer food, and the certification grade contains valuable information on the compliance level, I expect to observe a positive relationship between certification scores and food safety conditions. Large retailers work with thousands of suppliers. They often use audit scores to select their suppliers. Therefore, it is important to understand the information value of the certification grades.

Second, is there any empirical evidence supporting the claim that third-party food safety certification helps firms improve food safety practices? Studies find that credible information disclosure can induce companies to improve product quality in various settings (Mérel, Ortiz-Bobea and Paroissien, 2021; Jin and Leslie, 2003). Crandall et al. (2017) found 74% of low-risk food producers and 64% of high-risk food producers agreed that the GFSI certification process helped them improve regulatory compliance for their business. However, anecdotal evidence suggests that certification is not effective in improving the quality system itself, but is simply a costly signal.

Lastly, do firms with third-party food safety certification maintain their effort in food safety practices once certified? Third-party certification bodies (CBs) use the annually scheduled re-audit to ensure that certified firms maintain food safety standards. Typically, CBs only capture a snapshot of the production site by conducting site visits lasting several days on a pre-specified date or time window. Auditors do not directly evaluate product safety through product testing. Instead, they use checklists, review company documentation, and perform on-site observations. Therefore, companies can temporarily comply with standards, pass an audit, and obtain good scores without maintaining private food safety standards once the auditors have left. Although increased food safety practices can benefit firms by reducing the reputation and financial liability costs resulting from food safety incidents, the compliance costs of maintaining standards, such as training costs, changes in the production process, and capital investments, may be high. It is an empirical question to test firms' "gaming" behavior in response to third-party audits.

Food safety regulation in the meat and poultry industry of the United States provides a unique setting to investigate these questions. In the US, meat and poultry plants are actively monitored by both public and private sectors, which provide two sets of food safety measurements for a plant at the same time. Public inspectors from the Food Safety and Inspection Service (FSIS) perform inspection tasks on site when the plant is operating, providing a continuous and frequent measure of food safety practices. FSIS inspectors also conduct periodic pathogen sampling tests to directly measure food safety results. Compared to FSIS, third-party certification, such as the SQF Institute (SQF) and British Retail Consortium (BRC), provides a less frequent plant behavior-based measure for food safety. Third-party auditors audit certified plants annually for several days and rate the food safety level based on how well plants comply with the certification standards when third-party auditors review plant documents and observe their food safety conditions on site. I combine these public and private third-party records to construct a plant-month panel to make answering the three questions above possible.

I address the first research question, the information value of third-party certification ratings, empirically with a simple and transparent comparison of means. The analysis first focuses on whether better performers, as recognized by third-party certification, also have better food safety practices. It compares the average food safety practice level, measured by FSIS sanitary task compliance rate, among plants with different certification ratings. To extend our understanding of the relationship between certification rating information and food safety outcomes directly, I further compare plants' FSIS product-pathogen specific sampling results among plants with different certification ratings. This generates two main results.

First, overall, SQF and BRC certification ratings are informative of an establishment's food safety practice level. Plants with higher ratings are generally associated with higher FSIS sanitary compliance rates. Second, there is no evidence that higher certification ratings are associated with statistically significantly better pathogen test results. This is in keeping with Ollinger and Bovay (2018): good performance on certain FSIS sanitary tasks does not necessarily mean good test outcomes. Therefore, third-party certification does convey some valuable information on plants' food safety behavior, but an observation-based third-party audit is not a replacement for direct pathogen sampling tests.

I study the second and third research questions by exploiting the variations of FSIS sanitary task compliance rate around certification audit dates. I adapt and apply the two-way fixed-effect event study method to estimate the certified plants' behavior response to initial certification audits and annually repeating recertification audits. Results on whether certification can help plants improve their food safety practices are mixed. I find that there is a slow gradual increase in their average food safety practice level for SQF certified plants after initial certification, but not for BRC certified plants. The reason for the difference is not conclusive and open to further investigation. Plants' responses to recertification audits are similar for SQF and BRC certification. Compared to initial certification, the impact of re-certification on plant food safety level is limited. We do observe a micro pattern of plants' sanitary task compliance rate ramping up and down before and after re-certification audits, which implies plants could temporarily "perform" better for the re-certification audits without fundamentally improving their food safety practice level in the long run. However, the estimated effect is imprecise. Therefore, results are only suggestive and should be interpreted with caution.

This study makes three contributions.

First, this study is the first to empirically research the role of third-party certification on plants' food safety behavior and outcomes. Most previous work on third-party certification and food safety regulation uses qualitative analysis or surveys (Tanner, 2000; Duflo et al., 2013; Castka et al., 2015; Crandall et al., 2017). For the limited amount of empirical studies using observational data, most of them focus on the credibility of third-party certifiers and the incentives of certified firms (Albersmeier et al., 2009; Anders, Souza-Monteiro and Rouviere, 2010; Zheng and Bar, 2019).

Second, this study extends the literature on food safety regulation. The effects of public policies on improving food safety performance in the meat and poultry industry are well studied by Ollinger and Bovay (2018, 2019). However, the effect of private food safety regulation, third-party certification, is understudied. I find that certification is a tool for firms to signal good food safety practices, but the improvement effect of certification in plants' food safety behavior is limited.

Third, this research extends the literature on the welfare effects of quality disclosure and certification. Third-party certifications can improve welfare in three ways: by disclosing product information to allow consumers to choose the products that best match their preferences (sorting effect, e.g. see Akerlof (1970)), by incentivizing firms to invest and improve quality (incentivization effect, e.g. see Spence (1973)), and by providing cost-effective tools to help firms improve product quality (education effect, e.g. see Crandall et al. (2017)). The first two effects are based on the credibility of the certification grades, which I test empirically. As regards the educational effect, this research draws mixed conclusions on whether an initial certification can help plant food safety practices. Thus, my research complements the theoretical literature of certification by providing empirical evidence on the effects of certification.

The remainder of Part I of the dissertation is organized as follows. Chapter 2 provides additional background and institutional details. Chapter 3 is a review of the literature. Chapter 4 presents the main data set and descriptive statistics. Chapter 5 analyzes the relationship between certification ratings and certified plants' food safety conditions. Chapter 6 presents estimates of the average effects of initial and re-certification audits on plants' food safety practices.

Chapter 2

Background

2.1 What is private third-party food safety certification?

Food safety certification is the verification of products, processes, or systems in the food supply chain that meet specific accepted food safety standards. Unlike other food-related certifications focusing on quality or other food attributes such as organic, non-GMO, and sustainability, food safety is a common requirement for all participants in the food industry.

The auditing process against specific standards can be done by multiple parties. The first, second, and third parties can be defined differently according to context (Tanner, 2000; Rosenthal and Kunreuther, 2010). In this paper, first-party audits refer to internal audits by suppliers themselves according to standards specific to the company. Second-party audits mean that purchasers or users inspect and certify the suppliers against the requirements defined in the contract. Third-party audits are performed by accredited independent bodies with expertise against recognized sets of standards. The relative advantages of third-party auditing primarily relate to its objectivity, transparency, cost-effectiveness, and professionalism of the auditors.

The three main components of the third-party food safety certification industry are standard owners that set food safety standards, certification bodies (CBs) that conduct third-party audits, and accreditation bodies (ABs) that are approved by standard owners to accredit CBs. In general, standard owners, ABs, and CBs are governments, non-government organizations (NGOs), or private firms (Boys, Caswell and Hoffmann, 2015). The word "private" here differentiates licensed private companies from public agencies that conduct government regulatory and certification activities according to public food safety standards. The "USDA inspected" logo commonly seen on meat packages in the US is an example of a public food safety certification, indicating that the firm conforms to public food safety standards. Unless otherwise specified, third-party food safety certification refers to third-party private certification.

2.2 Development of private third-party food safety certification

2.2.1 Private food safety standards

Private food safety standards are the main driver of the development of third-party food safety certification. In the food industry, large retailers and food processors often impose private food safety standards on suppliers, and third-party auditors have become the primary tool for enforcing these private standards (Hatanaka, Bain and Busch, 2005). Governments typically set and regulate minimum food safety standards. Private companies establish more stringent and comprehensive food safety standards in response to consumers' growing food safety concerns and advances in technology and understanding of production practices. The private food safety standard has become a substitute for inadequate public standards and a tool for companies to cover liability risks and protect reputation (Henson and Reardon, 2005).

Before 2000, each retailer applied their own standards to their suppliers. Firms were burdened with all kinds of complex and redundant private standards, which increased trade barriers and costs in the global agri-food system. Meanwhile, consumers were losing trust in food safety management in the food industry after several food safety incidents in the 1990s (Weinroth, Belk and Belk, 2018). Against this background, the Global Food Safety

Initiative (GFSI) was created in 2000 as a collaboration from major retailers, manufacturers, and foodservice operators to improve the food safety management system, build consumer trust, and harmonize different private food safety standards. GFSI Board members created guideline documents in 2001 that set the benchmarking requirements for different food safety schemes with the goal of “once certified, accepted everywhere”. The GFSI guidelines are frequently updated to keep up with food safety issues and best practices. Currently, there are 12 GFSI-recognized certification program owners (CPOs), and the scope of certification programs covers the entire supply chain from farm to fork (see Table 2.1 for details).

GFSI-recognized food safety standards are internationally accepted by important players in the global food industry. In 2008, Wal-Mart became the first nationwide US grocery chain to request that suppliers get certified against one of the GFSI-benchmarked standards. Later, more retailers started to accept or require food suppliers to be certified by GFSI-recognized standards. For example, Safeway requires all human and animal food and food-contact packaging material to be certified against various food safety schemes with minimum certification score requirements. Costco does not require GFSI certification audits but accepts it along with minimum score requirements similar to those used by Safeway.

2.2.2 Market structure of the private third-party certification industry

The value of certification is highly dependent on how well standard owners monitor the compliance of certified firms. The third-party certification industry developed a multi-layer market organization to achieve this goal, as shown in Figure 2.1. Standard owners partner with third-party private CBs worldwide to conduct audits and issue certifications. Typically, CBs must be licensed by the standard owner and accredited and regularly assessed by licensed ABs such as ANSI National Accreditation Board (ANAB), the largest AB in North America. Each third-party CB hires certified individual auditors to conduct inspections either as full-time CB employees or as subcontracted auditors who bid for audits. Firms seeking to be certified against certain standards hire one of the licensed CBs to schedule their annual

Table 2.1: GFSI-recognized certification program owners (as of April 2020)

Certification program owner (CPO)	Industry scopes covered in GFSI-benchmarked standards[1]	# of certified sites		# of CBs		Score
		World	Us	World	Us	
Food Safety System Certification 22000 (FSSC 22000)	C, D, EI, EII, EIII, EIV, F, L; J; M	23366	1445	128	20	Certified or not, not graded
SQF Institute (SQF)[2]	AI, BI; C, D, EI, EII, EIII, EIV, F, L; M; J	10054	7161	38	28	Excellent, good, complies, fails
British Retail Consortium (BRC)	C, D, EI, EII, EIII, EIV, F, L; J; M; N	29871	2715	97	28	AA, A, B, C, D, not certified; if unannounced, add + after letter grades
International Featured Standards (IFS)[3]	C, D, EI, EII, EIII, EIV, L; J; M; N	16800*	35*	132	11	Level of compliance in percentage, not certified (final score <75%)
Global Good Agricultural Practices (Global GAP)[4]	AII, BI	18000+*	Na*	145	Na*	Certified or not, not graded
Primus GFS Standard (Primus GFS)	BI, BII, D, EII, EIII, EIV, J	Na	Na	14	9	Percentage scores: audit percentage score > 90%; module percentage scores > 85%
Global Aquaculture Alliance (GAA)	EI	83	2346	7	7	Certified or not; 1 to 4 start designation logo means the seafood was bap-certified all the way from feed, farm, hatchery and processor.
Global Red Meat Standard (GRMS)	C, EI	46	0	4	0	Level i, ii, iii
CANADAGAP	BI, D	43	2193	2	2	Certified (>85% & fail in implementing corrective actions) or not, not graded
Japan Food Safety Management Association (JFSM)	EI, EIII, EIV	897	0	10	0	Certified or not, not graded
ASIAGAP	BI, BII, D	400	0	6	0	Certified or not, not graded
Freshcare	BI, D	Na	Na	7	Na	Certified or not, not graded

Source: GFSI, CPO websites and personal contacts

Note: * represents the number of certified sites and CBs are not available directly from publicly available records on current CPO websites. The notes below address how to i get the numbers.

---

[1] Industry Scope Code: AI Farming of Animals, AII Farming of Fish, BI Farming of Plants, BII Farming of Grains and Pulses, C Animal Conversion, D Pre Processing Handling of Plant Products, EI Processing of Animal Perishable Products, EII Processing of Plant Perishable Products, EIII Processing of Animal and Plant Perishable Products, EIV Processing of Ambient Stable Products, F Production of Feed, J Provision of Storage and Distribution Services, L Production of (Bio) Chemicals, M Production of Food Packaging, N Agents and Brokers

[2] SQF # of certified sites are calculated by adding the current certification number of SQF Food Safety Audit 8.1 and 8.0 on SQF websites; SQF # of CBs are from personal contact with SQF

[3] IFS does not have public available certification data. IFS # of US certified sites are the number of IFS Food 6 suppliers from https://www.ifs-certification.com/index.php/en/ifs; # of world certified sites are from the estimated IFS certificates per year data from https://www.ifs-certification.com/index.php/en/ifs

[4] # of Global G.A.P. certified sites is 18000+*, North America comprises 1.4%. Available at https://www.globalgap.org/export/sites/default/.content/.galleries/Documents_for_Mailings/170712_GG_IntroPPT_EN_Session_KM.pdf

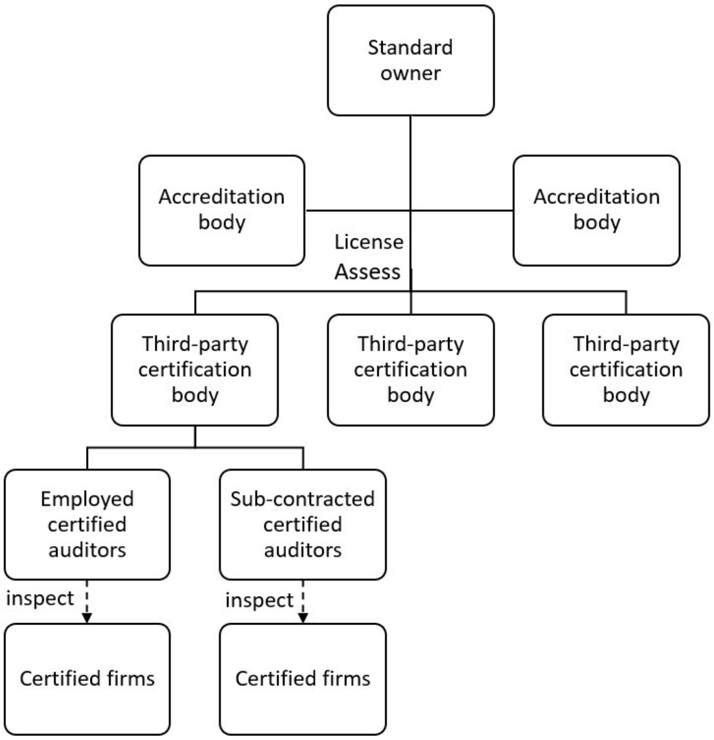


Figure 2.1: Simplified framework of key players in the private third-party certification industry

audits. With the increasing popularity of private standards, the number of CBs grew rapidly in the 1980s and 1990s, with 581 CBs in the world and 358 CBs covering at least one food safety standard (Caswell et al., 2017) . However, the current market structure can negatively affect the objectiveness of private third-party CBs and individual auditors (Jahn, Schramm and Spiller, 2005; Albersmeier et al., 2009; Anders, Souza-Monteiro and Rouviere, 2010; Zheng and Bar, 2019).

2.3 Meat & poultry food safety regulation in the US

In the US, the safety of meat and poultry products is regulated by both public agencies and the private sector, providing a unique setting to assess the value of third-party food safety certification. In the public sector, the Food Safety and Inspection Service (FSIS) is responsible for inspecting all meat, poultry, and egg products sold in interstate and foreign commerce for safety, wholesomeness, and proper labeling. FSIS employs approximately 8,000 in-plant and other front-line personnel in more than 6,000 federally inspected slaughter and processing establishments, in laboratories, and in commerce nationwide (USDA, 2019). FSIS inspection program personnel (IPP) inspect slaughter establishments during production and operations; they also inspect all the meat and poultry processing plants for a portion of production days to ensure products meet minimum legal requirements. In the private sector, suppliers and buyers address food safety problems through industry-wide initiatives and vertical contracts with specific food safety requirements (Pouliot and Wang, 2018). Out of all the GFSI-benchmarked food safety standards, SQF, BRC, and FSSC22000 are the most commonly used international food safety standards in the US meat and poultry industry.

2.3.1 Public regulation: FSIS inspection & sampling

FSIS inspection

All meat and poultry plants that sell products across state lines must be inspected by FSIS inspectors while operating in the US. Inspection tasks carried out routinely, continuously, or on a planned basis under normal conditions are called routine tasks. Most routine inspection tasks are based on Sanitation Performance Standards (SPS), Sanitation Standard Operating Procedures (SSOP), and Hazard Analysis and Critical Control Points (HACCP) standards. SPS primarily addresses specific sanitary issues within and around the establishment, such as grounds and facilities of the establishment, equipment and utensils, sanitary operations, employee hygiene, and tagging equipment, rooms, or compartments, to prevent the creation of unsanitary environments. SSOP consists of all procedures a plant must conduct daily, including pre-operational procedures (procedures before production operation) and operational procedures (procedures during production). Facilities must develop, implement, and maintain written SSOPs to prevent direct contamination or adulteration of products. HACCP is an analysis of food safety hazards during production and identifies preventive measures that could be used to mitigate potential harms.

FSIS inspectors perform SPS, SSOP (pre-operational and operational), HACCP, and other inspection tasks by conducting recordkeeping, review, and observation activities to ensure implementation and maintenance of the regulatory rules. Noncompliance records (NRs) are generated if establishments fail to comply with the inspection tasks. The NR serves as a notification and documentation of firms' non-compliance with regulatory standards. After receiving NRs, firms need to take corrective actions to reach regulatory requirements. Inspectors then verify their corrective actions and close the NRs. When there is a danger of adulterated, contaminated, misbranded, or hazardous products leaving the establishment, inspectors will follow the FSIS Rules of Practice (ROP) to take enforcement actions, such as regulatory control, withholding, and suspension actions. Figure 2.2 shows the regulatory process of an FSIS inspector.

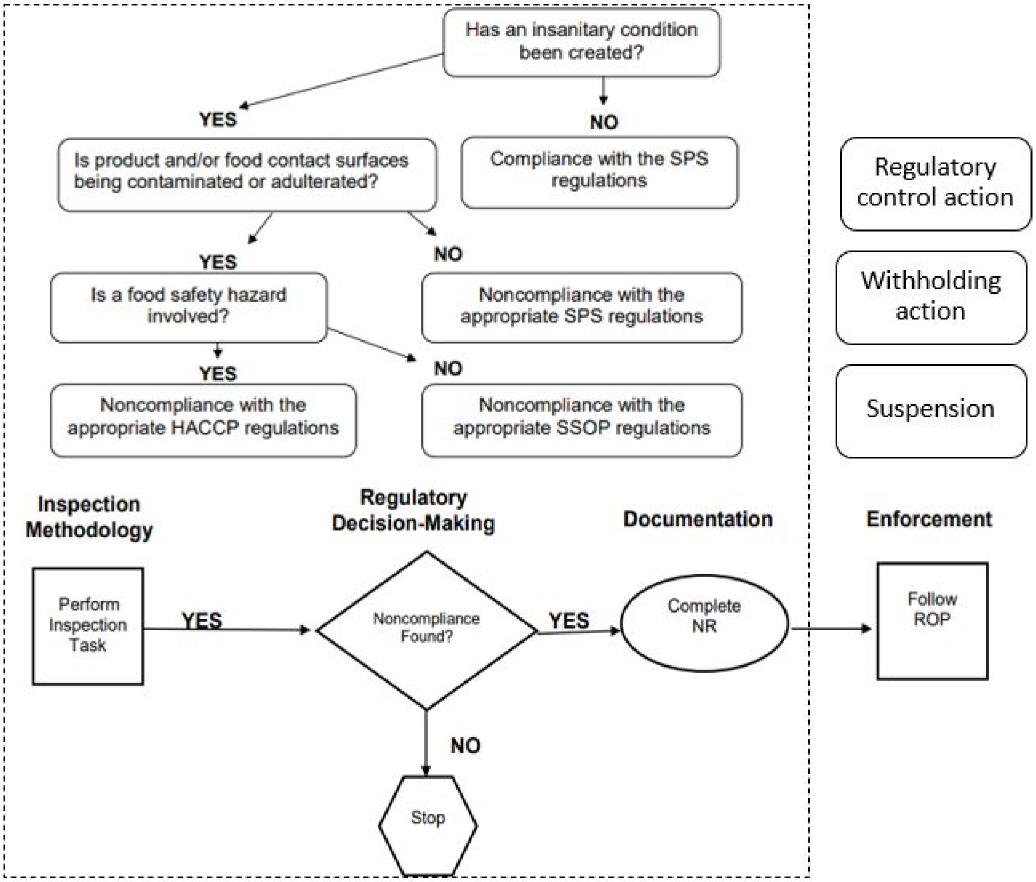
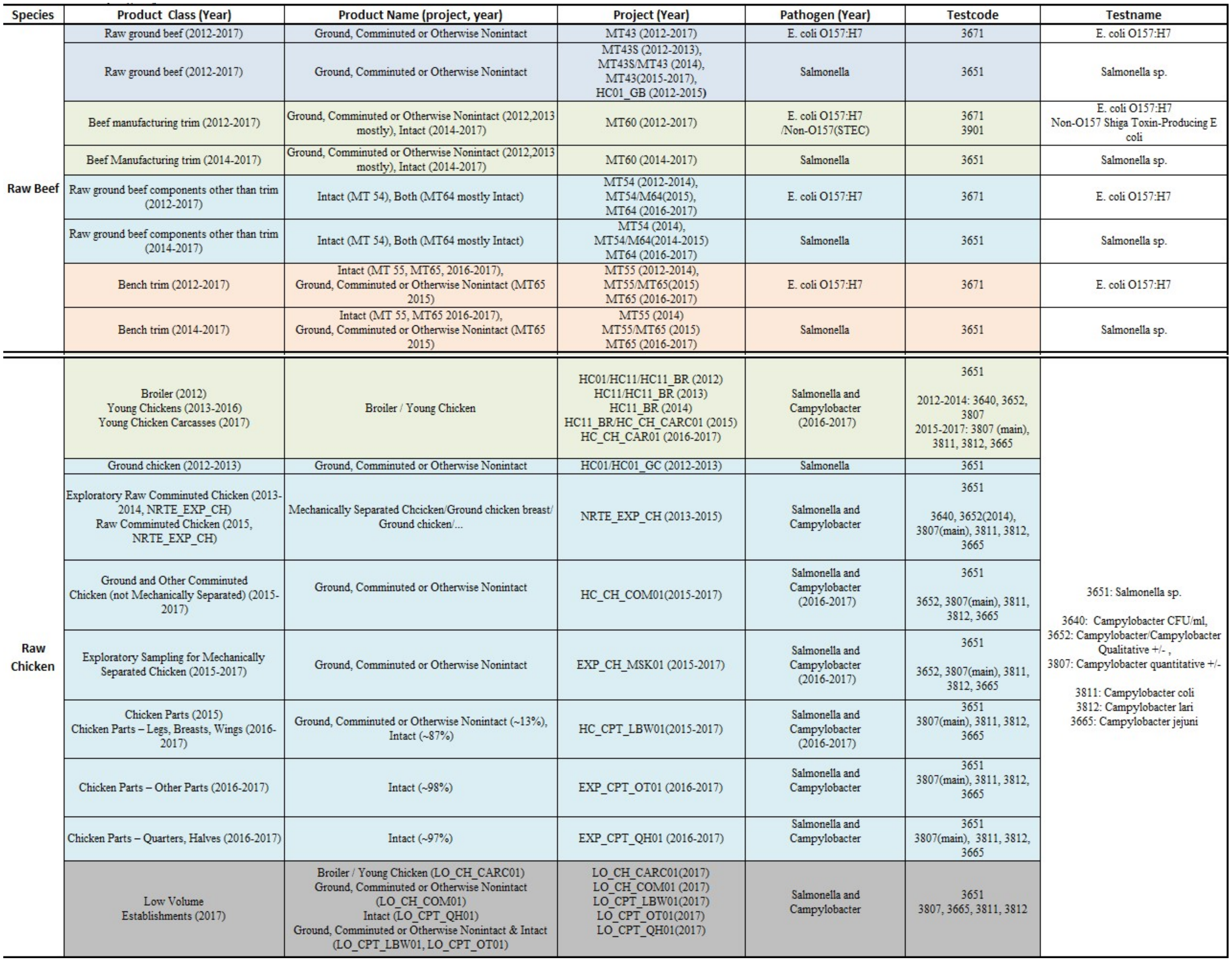


Figure 2.2: FSIS inspector regulatory process

FSIS sampling programs

To ensure that meat and poultry products are free of pathogen contamination and chemical residues, FSIS implements many regular sampling programs in domestic establishments, imports, and in-commerce facilities, and updates/introduces new sampling programs over the years. These sampling programs are critical components of HACCP and FSIS's verification responsibilities for the effectiveness of process control. FSIS's microbiological testing is pathogen-specific. Each product class is subject to different pathogen sampling programs and tests identified by sampling project codes and test codes. Through the years, FSIS has created new sampling tests and updated existing sampling tests for each species, which makes it harder to compare sampling results across time. However, I used the FSIS sampling program plan published annually on the FSIS website and the sampling program data set to created Table 2.2, summarizing the main FSIS sampling programs and the major tests from 2012-2017 for the main species. This can be used to identify establishments with similar production activities and help calculate the comparable failure rate of sampling tests across establishments and over time.

Table 2.2: FSIS domestic sampling projects



Species	Product Class (Year)	Product Name (project, year)	Project (Year)	Pathogen (Year)	Testcode	Testname
Raw Beef	Raw ground beef (2012-2017)	Ground, Comminuted or Otherwise Nonintact	MT43 (2012-2017)	E. coli O157:H7	3671	E. coli O157:H7
	Raw ground beef (2012-2017)	Ground, Comminuted or Otherwise Nonintact	MT43S (2012-2013), MT43S/MT43 (2014), MT43(2015-2017), HC01_GB (2012-2015)	Salmonella	3651	Salmonella sp.
	Beef manufacturing trim (2012-2017)	Ground, Comminuted or Otherwise Nonintact (2012,2013 mostly), Intact (2014-2017)	MT60 (2012-2017)	E. coli O157:H7 /Non-O157(STEC)	3671 3901	E. coli O157:H7 Non-O157 Shiga Toxin-Producing E coli
	Beef Manufacturing trim (2014-2017)	Ground, Comminuted or Otherwise Nonintact (2012,2013 mostly), Intact (2014-2017)	MT60 (2014-2017)	Salmonella	3651	Salmonella sp.
	Raw ground beef components other than trim (2012-2017)	Intact (MT 54), Both (MT64 mostly Intact)	MT54 (2012-2014), MT54/M64(2015), MT64 (2016-2017)	E. coli O157:H7	3671	E. coli O157:H7
	Raw ground beef components other than trim (2014-2017)	Intact (MT 54), Both (MT64 mostly Intact)	MT54 (2014), MT54/M64(2014-2015) MT64 (2016-2017)	Salmonella	3651	Salmonella sp.
	Bench trim (2012-2017)	Intact (MT 55, MT65, 2016-2017), Ground, Comminuted or Otherwise Nonintact (MT65 2015)	MT55 (2012-2014), MT55/MT65(2015) MT65 (2016-2017)	E. coli O157:H7	3671	E. coli O157:H7
	Bench trim (2014-2017)	Intact (MT 55, MT65 2016-2017), Ground, Comminuted or Otherwise Nonintact (MT65 2015)	MT55 (2014) MT55/MT65 (2015) MT65 (2016-2017)	Salmonella	3651	Salmonella sp.
Raw Chicken	Broiler (2012) Young Chickens (2013-2016) Young Chicken Carcasses (2017)	Broiler / Young Chicken	HC01/HC11/HC11_BR (2012) HC11/HC11_BR (2013) HC11_BR (2014) HC11_BR/HC_CH_CARC01 (2015) HC_CH_CAR01 (2016-2017)	Salmonella and Campylobacter (2016-2017)	3651 2012-2014: 3640, 3652, 3807 2015-2017: 3807 (main), 3811, 3812, 3665	3651: Salmonella sp. 3640: Campylobacter CFU/ml, 3652: Campylobacter/Campylobacter Qualitative +/- , 3807: Campylobacter quantitative +/- 3811: Campylobacter coli 3812: Campylobacter lari 3665: Campylobacter jejuni
	Ground chicken (2012-2013)	Ground, Comminuted or Otherwise Nonintact	HC01/HC01_GC (2012-2013)	Salmonella	3651	
	Exploratory Raw Comminuted Chicken (2013-2014, NRTE_EXP_CH) Raw Comminuted Chicken (2015, NRTE_EXP_CH)	Mechanically Separated Chcicken/Ground chicken breast/ Ground chicken/...	NRTE_EXP_CH (2013-2015)	Salmonella and Campylobacter	3651 3640, 3652(2014), 3807(main), 3811, 3812, 3665	
	Ground and Other Comminuted Chicken (not Mechanically Separated) (2015-2017)	Ground, Comminuted or Otherwise Nonintact	HC_CH_COM01(2015-2017)	Salmonella and Campylobacter (2016-2017)	3651 3652, 3807(main), 3811, 3812, 3665	
	Exploratory Sampling for Mechanically Separated Chicken (2015-2017)	Ground, Comminuted or Otherwise Nonintact	EXP_CH_MSK01 (2015-2017)	Salmonella and Campylobacter (2016-2017)	3651 3652, 3807(main), 3811, 3812, 3665	
	Chicken Parts (2015) Chicken Parts – Legs, Breasts, Wings (2016-2017)	Ground, Comminuted or Otherwise Nonintact (~13%), Intact (~87%)	HC_CPT_LBW01(2015-2017)	Salmonella and Campylobacter (2016-2017)	3651 3807(main), 3811, 3812, 3665	
	Chicken Parts – Other Parts (2016-2017)	Intact (~98%)	EXP_CPT_OT01 (2016-2017)	Salmonella and Campylobacter	3651 3807(main), 3811, 3812, 3665	
	Chicken Parts – Quarters, Halves (2016-2017)	Intact (~97%)	EXP_CPT_QH01 (2016-2017)	Salmonella and Campylobacter	3651 3807(main), 3811, 3812, 3665	
	Low Volume Establishments (2017)	Broiler / Young Chicken (LO_CH_CARC01) Ground, Comminuted or Otherwise Nonintact (LO_CH_COM01) Intact (LO_CPT_QH01) Ground, Comminuted or Otherwise Nonintact & Intact (LO_CPT_LBW01, LO_CPT_OT01)	LO_CH_CARC01(2017) LO_CH_COM01 (2017) LO_CPT_LBW01(2017) LO_CPT_OT01(2017) LO_CPT_QH01(2017)	Salmonella and Campylobacter	3651 3807, 3665, 3811, 3812	

Table continued

Raw Turkey	Turkeys (2012)	Turkey	HC01 (2012)	Salmonella	3651	3651: Salmonella sp. 3640: Campylobacter CFU/ml, 3804: Campylobacter CFU / sq cm 3652: Campylobacter/Campylobacter Qualitative +/- , 3807: Campylobacter quantitative +/- 3811: Campylobacter coli 3812: Campylobacter lari
	Turkeys (2012-2016) Turkey Carcasses (2017)	Turkey	HC11/HC11_TU (2012) HC11/HC11_TU (2013) HC11_TU (2014) HC11_TU/HC_TU_CARC01 (2015) HC_TU_CARC01 (2016-2017)	Salmonella and Campylobacter (2016-2017)	3651 3652, 3804, 3807	
	Ground turkey (2012-2013)	Ground, Comminuted or Otherwise Nonintact	HC01_GT (2012-2013)	Salmonella	3651	
	Exploratory Raw Comminuted Turkey (2013-2015)	ground boneless wings/ground turkey/Turkey sausage/turkey leg/turkey thigh meat etc	NRTE_EXP_TU (2013-2015)	Salmonella and Campylobacter	3651 3807, 3640 (2014 only), 3811(2014-2015) , 3812 (2014-2015)	
	Ground and Other Comminuted Turkey (not Mechanically Separated) (2015-2017)	Sampling for Ground and Other Comminuted turkey (not Mechanically Separated)	HC_TU_COM01(2015-2017)	Salmonella & Campylobacter (2016-2017)	3651 3652, 3807, 3811, 3812	
	Exploratory Sampling for Mechanically Separated Turkey (2015-2017)	Exploratory Sampling for Mechanically Separated Turkey	EXP_TU_MSK01 (2015-2017)	Salmonella and Campylobacter (2016-2017)	3651 3652, 3807, 3811, 3812	
	Low Volume Establishments (2017)	Turkey (LO_TU_CARC01) Ground, Comminuted or Otherwise Nonintact (LO_TU_COM01)	LO_TU_CARC01 LO_TU_COM01	Salmonella, Campylobacter and	3807	
Raw Pork	Exploratory Sampling for Pork - Comminuted (2016-2017)	Ground, Comminuted or Otherwise Nonintact	EXP_PK_COM01 (2016-2017)	Salmonella	3651	Salmonella sp.
	Comminuted Pork Exploratory Sampling (2017)	Ground, Comminuted or Otherwise Nonintact	EXP_PK_COM02 (2017)	Salmonella E. coli O157:H7 /Non-O157(STEC)	3651, 3671, 3901	Salmonella sp. E. coli O157:H7 Non-O157 Shiga Toxin-Producing E coli
	Exploratory Sampling for Pork - Intact Cuts (2016-2017)	Intact	EXP_PK_ICT01 (2016-2017)	Salmonella Indicator Organisms	3651	Salmonella sp.
	Intact Pork Cuts Exploratory Sampling (2017)	Intact	EXP_PK_ICT02 (2017)	Salmonella E. coli O157:H7 /Non-O157(STEC)	3651, 3671, 3901	Salmonella sp. E. coli O157:H7 Non-O157 Shiga Toxin-Producing E coli
	Exploratory Sampling for Pork - Intact Other (2016-2017)	Intact	EXP_PK_IOT01 (2016-2017)	Salmonella	3651	Salmonella sp.
	Exploratory Sampling for Pork - Non Intact Cuts (2016-2017)	Ground, Comminuted or Otherwise Nonintact	EXP_PK_NCT01 (2016-2017)	Salmonella	3651	Salmonella sp.
	Non- Intact Pork Cuts Exploratory Sampling (2017)	Ground, Comminuted or Otherwise Nonintact	EXP_PK_NCT02 (2017)	Salmonella E. coli O157:H7 /Non-O157(STEC)	3651, 3671, 3901	Salmonella sp. E. coli O157:H7 Non-O157 Shiga Toxin-Producing E coli
	Non-Intact Other Pork Exploratory Sampling (2016-2017)	Ground, Comminuted or Otherwise Nonintact	EXP_PK_NOT01 (2016-2017)	Salmonella	3651	Salmonella sp.
Ready-to-Eat	Both post-lethality exposed and non post-lethality exposed RTE products (2012-2017)	Acidified/Fermented/RTE/Fully cooked, …/Other Fully cooked/Dried/Salt Cured etc.	ALLRTE (2012) ALLRTE/RTEPROD_RAND (2013) RTEPROD_RAND(2014-2017)	Salmonella, Lm	3651 3662	Salmonella sp. Listeria monocytogenes
	Post-lethality exposed RTE products (2012-2017)	Acidified/Fermented/RTE/Fully cooked, …/Other Fully cooked/Dried/Salt Cured etc.	RTE001 (2012) RTE001/RTEPROD_RISK (2013) RTEPROD_RISK(2014-2017)	Salmonella, Lm	3651 3662	Salmonella sp. Listeria monocytogenes
	RLm product samples (Composited 5-sample Units) (2012-2017)	RTE/Fully cooked, …/Other Fully cooked/Dried etc.	RLMPROD (2012) RLMPRODC(2013-2017)	Lm	3662	Listeria monocytogenes
	RLm food contact surface samples (2012-2017)	Product Contact Surface Sponge, Water/Brine (2015-2017)	RLMCONT (2012-2017)	Lm	3662	Listeria monocytogenes
	RLm non-food contact environ. samples (Comp. 5-sample Units; Lm) (2012-2017)	Non/Product Contact Surface Sponge	RLMENVC (2012 - 2017)	Lm	3662	Listeria monocytogenes

2.3.2 Private regulation: GFSI-benchmarked food safety certification SQF, BRC

Of the GFSI-benchmarked food safety certifications listed in Table 2.1, SQF, BRC, and FSSC 22000 are the most commonly used private food safety standards in the US meat and poultry industry. This paper focuses on SQF and BRC only, since FSSC 22000 covers less than 10% of the market share (Hu and Zheng, 2019) and does not provide historical audit records.

SQF

SQF is a process and product certification standard based on HACCP to control food safety and quality hazards. It was first developed in Australia in 1997, and the SQF level 2 program became GFSI-benchmarked in 2004. To keep up with the best practices, SQF upgrades its code requirements on an ongoing basis. SQF Code 7.2 (July 3, 2014 - January 2, 2018) is the edition of the standard that is relevant to this study.

To prepare for an SQF certification, firms must first decide on the relevant SQF Food Sector Categories (FSC) and the level at which they would like to be certified, and then they must document and implement the requirements in the corresponding SQF Code Modules. Module 2: SQF System Elements applies to all industries. The requirements in other modules are based on different FSCs. Suppliers can choose one of the three certification levels in SQF Code 7.2 based on how well their food safety and quality management system has developed and the requirements of their buyers. Level 1 (Food Safety Fundamentals) is an entry-level for new business and only covers Module 2 - level 1 and other basic requirements. Level 2 (Certified HACCP Based Food Safety Plans) is GFSI-benchmarked, incorporates Level 1 requirements, and adds the HACCP food safety plan and Module 2 - level 2. Level 3 (Comprehensive Food Safety and Quality Management System) incorporates both levels and adds the HACCP food quality plan and Module 2 - level 3.

The typical process of initial certification involves the following steps. Facilities first learn about the SQF code, select relevant modules, and register on the SQF assessment dataset. Then they designate SQF Practitioners to oversee the development and implementation of the SQF system. Facilities that want to obtain SQF certification need to have a minimum of two months of records after implementing the SQF system. Afterward, companies can select an SQF-licensed CB, ask for price quotes, and schedule an initial audit with their chosen CB on an agreed day. An initial audit includes a desk and a facility audit. Auditors conduct the desk audit to check whether the documentation meets the requirements of the SQF code. The facility audit focuses on whether plants effectively implement what they document. The CB makes a certification decision based on the number and severity of non-compliances with the SQF codes. An audit report and a rating of Excellent - A, Good - B, Complies - C, or Fails to Comply - No Grade is provided to suppliers, offering an overall evaluation of how well a site complies with the SQF standards. Certification is issued if a facility has a rating of A, B, or C and corrective actions of all non-compliances are verified by auditors within a required time frame. The grading details are shown in Table 2.3.

To maintain certification, firms must be recertified annually on an agreed day within 30 calendar days on either side of the anniversary of the last day of the initial audit. If a firm receives a Complies (C) rating, a surveillance audit shall be conducted within 30 calendar days on either side of the 6 months of the last day of the previous audit. If a plant fails to comply, the plant must re-apply for another facility audit. Figure 2.3 summarizes the timeline of the initial and maintenance audits.

Table 2.3: SQF audit score and rating

Panel A. Audit score

Types of non-conformity	Severity level	Assigned points for each non-conformity type (N)
Minor- non-conformity	May lead to a risk to food safety and quality but not likely to cause a system element breakdown.	1
Major non-conformity	Carry a food safety or quality risk and likely to result in a system element breakdown.	10
Critical non-conformity	Breakdown of control(s) at a critical control point, a pre-requisite program, or other process step and judged likely to cause a significant public health risk and/or where a product is contaminated.	50

Panel B. Audit rating

Score (100 - N)	Rating[1]	Critical	Major	Minor	Audit Frequency
96 - 100	A (Excellent)			4 or fewer	12 monthly recertification audit
86 - 95	B (Good)			5-14	12 monthly recertification audit
			1	4 or fewer	
70 – 85	C (Complies)			15-30	6 monthly surveillance audit
			1	5-20	
			2	10 or fewer	
			3		
0 - 69	No Grade (Fails to comply			31 or more	Reapplication within 6 months of the last audit date: if with the same CB, only desk audit is required. Reapplication after 6 months of the last audit date or with a new CB, both desk and facility audits are required.
			1	21 or more	
			2	11 or more	
			3 or more		
		1 or more			

Source: SQF Code Edition 7.2 https://www.sqfi.com/wp-content/uploads/2018/08/SQF-Code_Ed-7.2-July.pdf

[1] To be consistent with BRC rating, recode Excellent, Good, Complies, Fails to comply, to A, B, C, No Grade. Certification requires facility to have rating A, B, C and that all major non-conformities are closed out within fourteen (14) calendar days and minor non-conformities within thirty (30) calendar days of the completion of the facility audit.

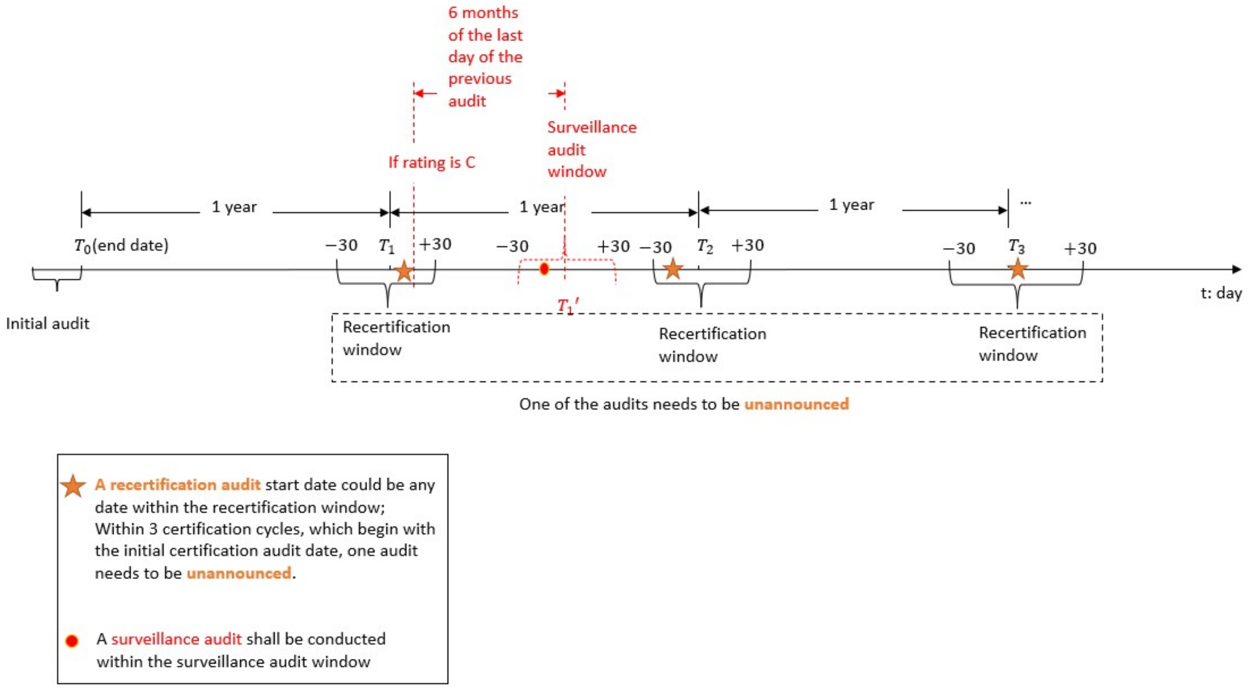


Figure 2.3: SQF certification initial and maintenance audit timeline

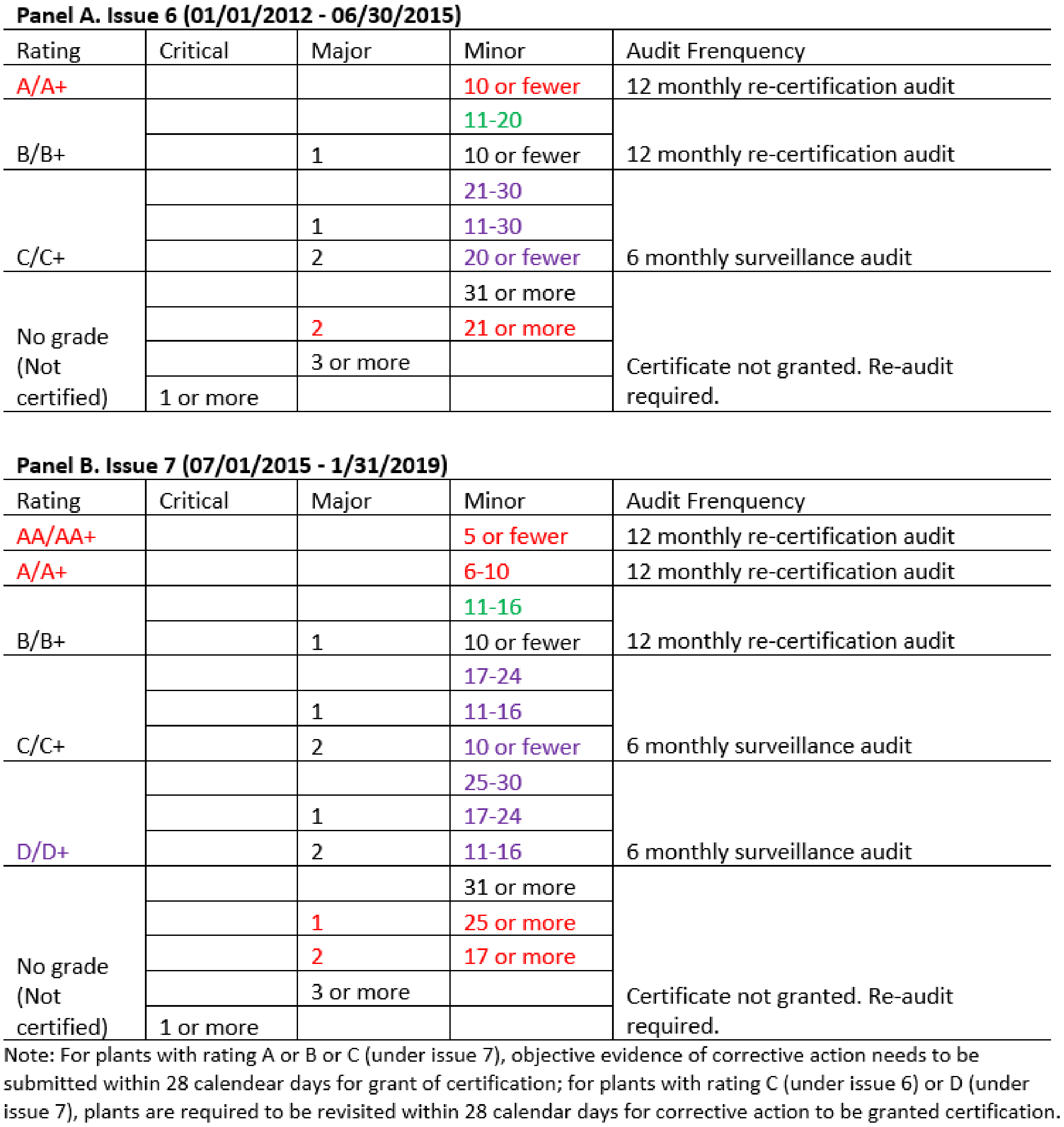
BRC

The BRC Food Safety Standard was developed in 1998 and has been updated regularly. It was the first standard to be GFSI benchmarked. It is broadly similar to SQF, requiring an HACCP-based food safety (quality) management system, senior management commitment, and prerequisite programs necessary for the production of safe food. BRC standard code Issue 6 (January 1, 2012 - June 30, 2015) and Issue 7 (January 1, 2015 - January 31, 2019) are the editions of BRC standards applicable during the time range of this study.

One major difference between Issue 6 and Issue 7 is BRC certification grading system. For Issue 6, auditing grades include A(A+), B(B+), C(C+), and No Grade (Not certified), which are assigned by third-party auditors based on the number of critical, major or minor non-conformities found in the plant during audits (see Table 2.4 Panel A). '+' indicates that the plant is enrolled in a voluntary unannounced audit instead of an announced audit and that the audit date is agreed upon with the CB in advance of the audit. Very few plants (less than 1.5%) are enrolled in this voluntary unannounced audit program. Plants are certified if they receive an audit grade and submit their corrective actions within 28 calendar days. In Issue 7, BRC makes its auditing grades finer including AA(AA+), A(A+), B(B+), C(C+), D(D+) and not certified (see Table 2.4 panel B). If what counts as critical, major, or minor nonconformities is not changed, Issue 7 has slightly more stringent criteria for getting certified and finer grades to distinguish certified plants compared to the Issue 6 audit grade rule.

The general audit protocols of BRC and SQF are very similar. For BRC certification preparation, plants select an announced or unannounced audit option, decide the scope of the audit, self-assess the compliance with the BRC standard (normally taking more than 3 months), and select a CB approved by BRC for the formal initial audit. Before initial audits, sites and auditors agree on a date and audit duration to make sure plants are prepared with documentation for auditors to assess. Based on the identification of non-conformities and grades given, sites must undertake corrective actions within 28 days and pass the review

Table 2.4: BRC audit rating for issue 6 and 7



Panel A. Issue 6 (01/01/2012 - 06/30/2015)

Rating	Critical	Major	Minor	Audit Frenquency
A/A+			10 or fewer	12 monthly re-certification audit
B/B+			11-20	12 monthly re-certification audit
		1	10 or fewer	
C/C+			21-30	6 monthly surveillance audit
		1	11-30	
		2	20 or fewer	
No grade (Not certified)			31 or more	Certificate not granted. Re-audit required.
		2	21 or more	
		3 or more		
	1 or more			

Panel B. Issue 7 (07/01/2015 - 1/31/2019)

Rating	Critical	Major	Minor	Audit Frenquency
AA/AA+			5 or fewer	12 monthly re-certification audit
A/A+			6-10	12 monthly re-certification audit
B/B+			11-16	12 monthly re-certification audit
		1	10 or fewer	
C/C+			17-24	6 monthly surveillance audit
		1	11-16	
		2	10 or fewer	
D/D+			25-30	6 monthly surveillance audit
		1	17-24	
		2	11-16	
No grade (Not certified)			31 or more	Certificate not granted. Re-audit required.
		1	25 or more	
		2	17 or more	
		3 or more		
	1 or more			

Note: For plants with rating A or B or C (under issue 7), objective evidence of corrective action needs to be submitted within 28 calendear days for grant of certification; for plants with rating C (under issue 6) or D (under issue 7), plants are required to be revisited within 28 calendar days for corrective action to be granted certification.

before being issued a certification. Similar to SQF, to maintain the certification, BRC certified plants with grade AA/A/B(+) are required to be re-audited every 12 months; plants with grade C(C+)/D(D+) need to be re-audited every 6 months. For an announced audit, re-audit will occur within a 28-day time period up to the 1-year or 6-month anniversary of the initial audit date; for an unannounced audit, re-audit will occur during a certain time-frame after the last audit date based on what unannounced audit options a plant chooses.

To summarize, there are 4 main differences between the SQF and BRC audit protocols. (1) BRC food safety program does not have different program levels like SQF (level 2: food safety, level 3: food safety and quality). (2) BRC allows firms to choose between announced or unannounced audit options. (3) BRC standard updated to a finer grading system from Issue 6 to Issue 7. (4) BRC's announced recertification is within 28 days up to the anniversaries of the initial audit. SQF's recertification is within 30-day either side of the initial audit anniversary.

Chapter 3

Literature review

Third-party food safety certification is relatively understudied. Of direct relevance to my research are the following empirical papers on the reliability of the food safety signal and the effect of the certification on company performance. My empirical work extends these topics.

The first strand of the literature suggests that certification grade may not be a good signal of food safety level because the objectiveness of certifiers may be jeopardized. Zheng and Bar (2019) empirically tested the link between market competition and audit grades using British Retail Consortium (BRC) global food safety standard data and found that a higher level of competition among certification bodies is correlated with higher certification grades. Albersmeier et al. (2009) focus on audit grade patterns alone. They conducted statistical analyses on the German certification system Quality and Safety database and found audit results differed hugely among different certification bodies and auditors. Although they were unable to establish the reasons for the differences, they raised doubts about the competence and economic pressure of the certifiers to perform objective audits. Without another set of independent and frequent food safety measurements, it is difficult to know how well audit scores reflect the actual food safety levels. My paper builds a unique dataset compiling both frequent measurements from public inspections and less frequent food safety measurements from private third parties to evaluate the reliability of third-party certification.

The second strand of literature analyzes the reliability concerns of third-party food safety certifications from the angle of certified firms' attitudes and behaviors toward third-party certifications. Bar and Zheng (2018) study what influences firms' choices of certification body using the BRC audit data of US firms. They found that firms tend to choose certification bodies that are geographically close, considered more lenient, or have been used in previous years by the firm before. Castka et al. (2015) also found that the audit orientation (improvement or mere compliance with the standard) of a firm could influence the choice of certifiers and the overall satisfaction with certifications by surveying companies certified to a quality management standard ISO 9000 in Australia and New Zealand. These two studies together suggest that it matters what motivates firms to seek certification; if their aim is not to improve food safety but simply to pass the audit, they will choose auditors that will allow them to easily obtain good grades. This would be detrimental to the role of certifications as a credible signal of a firm's true food safety level. My paper adds to the literature by testing whether initial certification is helpful in improving food safety practices and whether certified maintain their efforts once certified.

The third strand of the literature examines the impact of food safety certifications on food safety outcomes and other aspects of company performance. Zheng and Bar (2019) observed that sites audited for the first time have lower grades than more experienced ones, suggesting that firms learn from certification and improve their food safety practices. However, it is unclear what firms learn over time. Do they learn to game the system to get high grades, or do they actually learn how to improve food safety levels? Hu and Zheng (2019) found that firms with food safety certifications are associated with better pathogen test results based on private food safety certifications and FSIS laboratory test data. However, it is difficult to draw causal conclusions from this due to sample selection bias. My study contributes to the food safety certification literature by providing a better identification of the impact of food safety certification on food safety practice behaviors.

My study also relates to the food safety regulation literature. The most relevant papers to this work are those by Ollinger and Bovay (2018, 2019), who also conducted their analysis in the US meat and poultry industry. They use FSIS pathogen test data to evaluate the effects of two specific public policies on improving firms' food safety performances: public disclosure of FSIS Salmonella test results for chicken carcasses with poor performance and the zero-tolerance standard for Salmonella in beef for the National School Lunch Program (NSLP). They found that a credible threat of public disclosure of poor performance firms improved chicken Salmonella test results, and firms with a better performance show a higher rate of improvement (Ollinger and Bovay, 2019). My paper instead focuses on the reliability of information disclosure from private third-party certifications, and the research scope is beyond just the poultry and beef industry. Compared to public disclosure, food safety certifications are perceived as not only a signaling tool, but also a cost-effective tool for buyers to ensure that suppliers meet higher private standards and provide expertise to help suppliers improve food safety practices. This paper contributes by empirically testing the expected roles of private regulation on food safety.

This research is also related to the broad literature on quality disclosure, certification, and reputation. When there is asymmetric information, there will be loss of welfare due to quality under-provision (Akerlof, 1970) and allocation inefficiency (Spence, 1973). When buyers face quality uncertainty, they are only willing to pay an average price for both types of sellers (high quality and low quality), which leads to an equilibrium where low quality goods drive high quality goods out of the market. Reliable information disclosed by certification can reduce information asymmetry, improving social welfare by encouraging firms to provide high quality products (incentivization effect), and better matching buyers and sellers (sorting effect). Although in the long run buyers may be able to learn about product quality through channels such as reputation, rational sellers may still lack incentives to provide quality because reputation can take a long time to build (Bai, 2018). Therefore, reliable quality disclosure provided by third-party certification is especially important for goods with credence characteristics (such as food safety) and in developing countries, where reputation is hard to build (Bai, Gazze and Wang, 2022; Bai, 2018). However, multiple parties found third-party certification unreliable in different settings. Dranove and Jin (2010) pointed out in their extensive review of the literature on certification and quality disclosure that the potential conflict of interest of certifiers could be detrimental to the value of information from third-party disclosures, and theoretical models found that competition between certifiers, reputation concerns, and external monitoring all have a mixed effect on solving incentive problems of auditors. Empirically, Duflo et al. (2013) found that, in the environmental regulation market where third-party auditors are chosen and paid by the firms they audited, auditors systematically underreport plant emissions just below the standards. In addition to incentive issues, third-party certifiers may not be able to measure quality accurately. This article contributes to the literature by providing empirical support of how well certification grades could reflect the quality level of a product in the food safety setting.

Last but not least, the literature on dynamic certification and monitoring is also relevant to this study. Marinovic, Skrzypacz and Varas (2018) showed that in a dynamic setting, when firms could decide whether and when to be certified, if the industry could coordinate a good certification rule, voluntary certification could encourage firms to maintain investment rather than temporary quality provision. Since a dynamic monitoring policy is essential, their subsequent theoretical paper studied the trade-offs of deterministic policies versus random inspection policies in designing optimal dynamic monitoring systems (Varas, Marinovic and Skrzypacz, 2020). Deterministic inspections may give strong incentives for firms to pass the inspection, but they give weak incentives for maintaining standards after an audit, and random inspection policies may not be ideal for acquiring valuable quality information. In their model, they assumed that a firm could affect the persistent quality of its products by exerting unobservable effort and that firms have reputation concerns. Monitoring policy must balance two roles: incentivizing firms to reduce moral hazard and acquiring information about the firm's quality. They concluded that the optimal policy is a mixture of the two types of inspections; when the moral hazard problem is weak, the optimal policy is close to deterministic inspection. My paper contributes to the literature by providing empirical evidence on firms' maintenance/shirking behaviors in the food safety certification market, where there is a combination of deterministic and random inspection policies.

Chapter 4

Data & descriptive statistics

I use 3 main data sources: The FSIS data set is obtained through a data sharing agreement with FSIS; the BRC data set is obtained from the BRC standards program with the help of Antony Harrison, Directory Services Manager at BRC; SQF data are obtained from the SQF standards program IT system manager, Daniel Akinmolayan, by extracting the publicly available US certification information from the SQF website. Figure 4.1 shows the data sources and the data keys by which the datasets are joined. In the following, Sections 4.1,4.3, and 4.2 summarize the content in each data set in detail, and Section 4.4 describes the sample construction.

4.1 FSIS dataset

4.1.1 FSIS establishment demographics data

FSIS establishment cross-sectional demographic data set provided by FSIS captures a snapshot of plants' characteristics. It includes plant name, address, circuit, district, [1] inspection activities, and HACCP processing size (large/small/very small/missing).

The 6,394 distinct establishments are assigned to 191 circuits of 10 districts. Based on the HACCP processing size, about 6% of the plants are large, 38% small, 54% very small,

---

[1] Circuit is an organizational structure of plants and positions designed to deliver program services and provide supervision in an efficient and effective manner to in-plant personnel. The goal is to maintain a balanced workload among circuits. Each circuit belongs to one of the 10 Office of Field Operations (OFO) Districts.

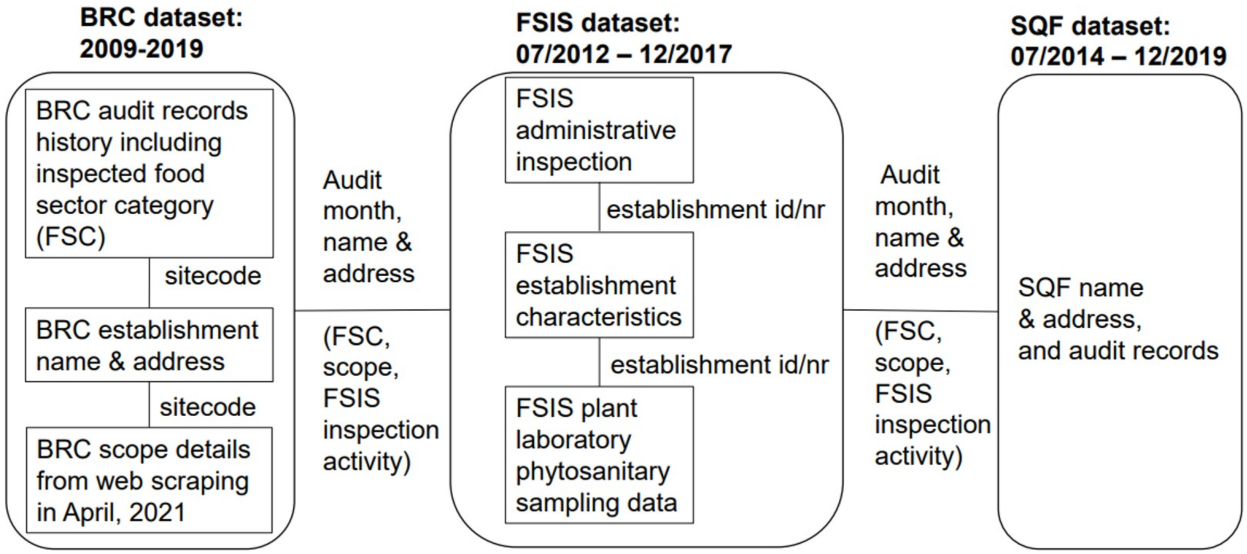


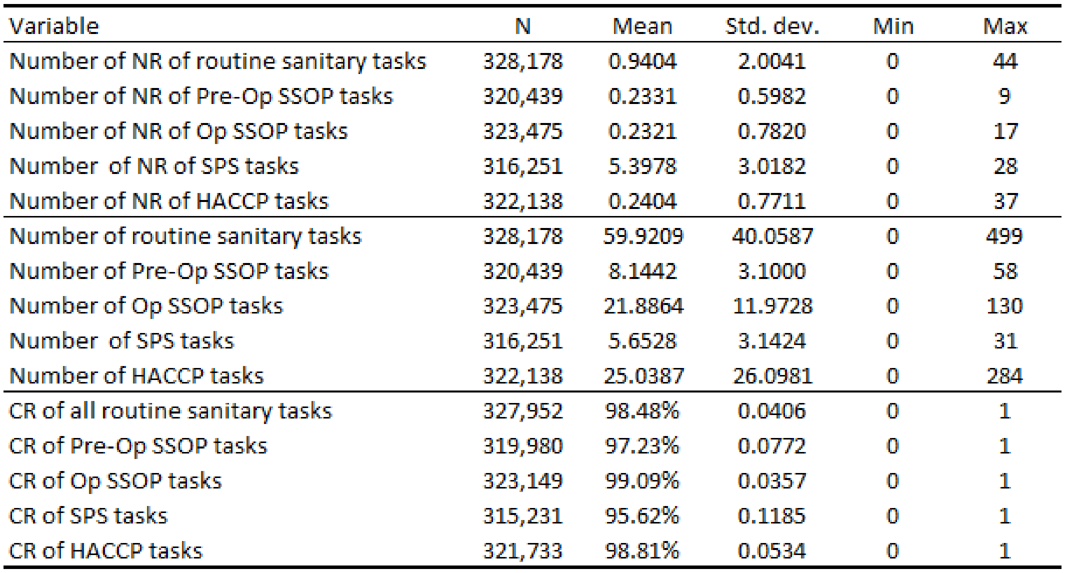
Figure 4.1: Data schema

and less than 2% are missing. Inspection activities can be used to identify establishments' general production activity, such as meat processing (MP), poultry processing (PP), meat slaughter (MS), and poultry slaughter (PS): 16% plant activities include MP, 8% include PS, 92% include MP and 72 % include PP. In general, almost all establishments process something, and plants have various production activities and products. Furthermore, by dividing all establishments mutually exclusive and collectively exhaustive, I observed that about half of plants process both meat and poultry: around 58% establishments process both meat and poultry (MP+PP), 17% only process meat (MP), 3% only process poultry (PP), 10% slaughter and process meat (MS+MP), 4% slaughter and process poultry (PS+PP), 2% slaughter and process both meat and poultry (MS+PS+MP+PP), and around 5% are others. Poultry-only plants tend to have a large HACCP processing size, and meat processing plants tend to have a small or very small HACCP processing size.

4.1.2 FSIS administrative inspection data

FSIS detailed administrative inspection data are provided by FSIS. It contains all daily inspection activities at over 6000 federal inspected facilities from July 2012 to Dec 2017. The data set is collapsed into the establishment and monthly level to construct 5 key outcomes

Table 4.1: Summary statistics on the panel of FSIS inspection tasks at the establishment/-month level

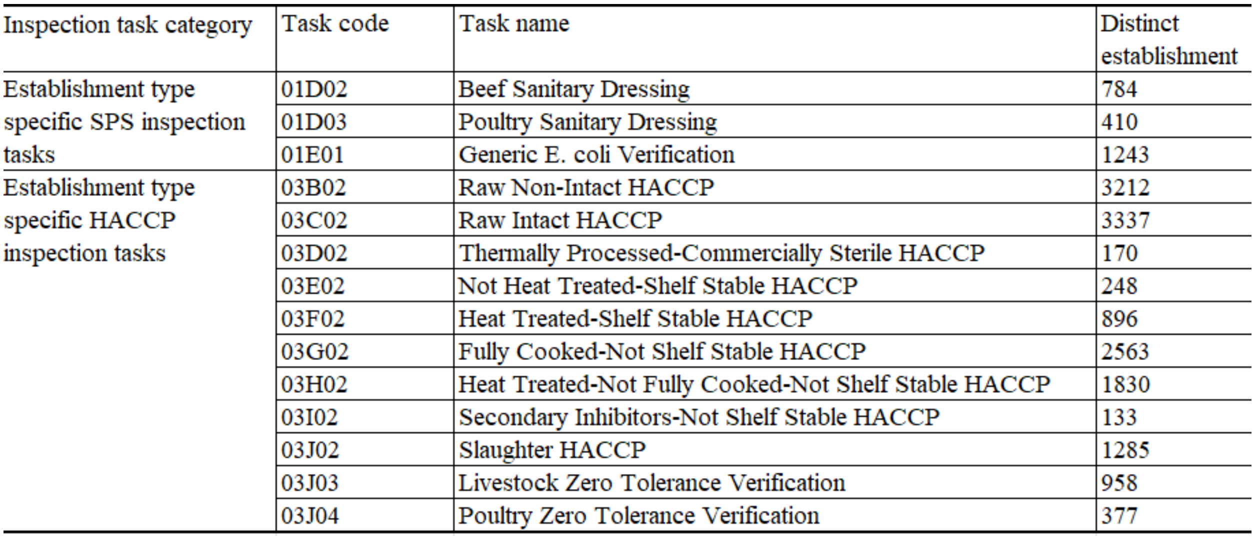


Variable	N	Mean	Std. dev.	Min	Max
Number of NR of routine sanitary tasks	328,178	0.9404	2.0041	0	44
Number of NR of Pre-Op SSOP tasks	320,439	0.2331	0.5982	0	9
Number of NR of Op SSOP tasks	323,475	0.2321	0.7820	0	17
Number of NR of SPS tasks	316,251	5.3978	3.0182	0	28
Number of NR of HACCP tasks	322,138	0.2404	0.7711	0	37
Number of routine sanitary tasks	328,178	59.9209	40.0587	0	499
Number of Pre-Op SSOP tasks	320,439	8.1442	3.1000	0	58
Number of Op SSOP tasks	323,475	21.8864	11.9728	0	130
Number of SPS tasks	316,251	5.6528	3.1424	0	31
Number of HACCP tasks	322,138	25.0387	26.0981	0	284
CR of all routine sanitary tasks	327,952	98.48%	0.0406	0	1
CR of Pre-Op SSOP tasks	319,980	97.23%	0.0772	0	1
CR of Op SSOP tasks	323,149	99.09%	0.0357	0	1
CR of SPS tasks	315,231	95.62%	0.1185	0	1
CR of HACCP tasks	321,733	98.81%	0.0534	0	1

of interest: compliance rate (CR) of all routine sanitary tasks (including SPS, SSOP, and HACCP tasks), routine SPS tasks, routine pre-operational SSOP (Pre-Op SSOP), routine operational SSOP (Op SSOP), and routine HACCP tasks as measurements for plants' food safety practice levels. Table 4.1 presents some summary statistics on the panel of FSIS administrative inspection results. The compliance rate of routine sanitary tasks (Sanitary CR) is equal to 1 minus the total number of non-compliance records (NRs) divided by the total number of sanitary inspection tasks in that month. The compliance rate of each specific task denoted as Pre-Op SSOP CR, Op SSOP CR, HACCP CR is calculated the same way but using the number of NRs and inspection tasks in each standard category. The average compliance rate of all sanitary tasks and tasks in each category is 98.48%, 97.23%, 99.09%, 95.62%, and 98.81%. CR of SPS tasks has the lowest compliance rate among all task categories. On average, poultry/large plants have lower CR.

It should be noted that different establishments are subject to different specific SPS tasks and HACCP tasks based on the process categories listed in Table 4.2. 14 Dummy variables of these establishment specific tasks are created. They are set equal to 1 as long as the plant was inspected under the specific task during the observed period. For example, the dummy

Table 4.2: FSIS inspection tasks



Inspection task category	Task code	Task name	Distinct establishment
Establishment type specific SPS inspection tasks	01D02	Beef Sanitary Dressing	784
	01D03	Poultry Sanitary Dressing	410
	01E01	Generic E. coli Verification	1243
Establishment type specific HACCP inspection tasks	03B02	Raw Non-Intact HACCP	3212
	03C02	Raw Intact HACCP	3337
	03D02	Thermally Processed-Commercially Sterile HACCP	170
	03E02	Not Heat Treated-Shelf Stable HACCP	248
	03F02	Heat Treated-Shelf Stable HACCP	896
	03G02	Fully Cooked-Not Shelf Stable HACCP	2563
	03H02	Heat Treated-Not Fully Cooked-Not Shelf Stable HACCP	1830
	03I02	Secondary Inhibitors-Not Shelf Stable HACCP	133
	03J02	Slaughter HACCP	1285
	03J03	Livestock Zero Tolerance Verification	958
	03J04	Poultry Zero Tolerance Verification	377

of the Raw Non-Intact HACCP inspection task is set equal to 1 as long as the plant was inspected under this task category at some point during the observation period. They are part of plant characteristics used to make apple-to-apple comparisons among a diversified group of establishments.

4.1.3 FSIS laboratory sampling data

FSIS sampling data for microbiological contaminants are provided by FSIS. It contains all the results of the sampling tests for different products and tests for each establishment from June 2012 to December 2017. To evaluate a plant's food safety outcome, This study uses the passing rates of product-pathogen specific sampling tests as the measure for food safety outcomes. I construct the main outcome variables, product-pathogen specific sampling passing rate, for each plant-month based on Table 2.2 and the summary statistics are shown in Table 4.3 below.

The passing rate of product-pathogen samples for a given establishment in a certain month is constructed in the following way:

- For a specific product-pathogen category (i.e. ground beef - salmonella), calculate the total samples taken under the corresponding project code and pathogen test codes

Table 4.3: Summary statistics of product - pathogen specific sampling test passing rate

	Product	Pathogen	N	Mean	Std	Min	Max
Beef	Ground beef	Salmonella	35,191	98.57%	0.1032	0	1
	Ground beef	Ecoli	39,734	99.90%	0.0292	0	1
	Beef trim	Salmonella	6,052	97.54%	0.1303	0	1
	Beef trim	Ecoli	8,593	99.70%	0.0475	0	1
	Beef trim	Non-O157	8,224	99.04%	0.0883	0	1
	Raw ground beef components other than trim	Salmonella	1,352	93.18%	0.2386	0	1
	Raw ground beef components other than trim	Ecoli	2,089	99.69%	0.0546	0	1
	Bench trim	Salmonella	3,616	98.88%	0.1013	0	1
	Bench trim	Ecoli	5,664	99.93%	0.0266	0	1
Chicken	Young chicken	Salmonella	8,359	95.08%	0.1476	0	1
	Young chicken	Campylobacter	8,322	93.65%	0.1902	0	1
	Ground chicken etc	Salmonella	12,965	74.63%	0.3746	0	1
	Ground chicken etc	Campylobacter	12,450	90.21%	0.2553	0	1
Turkey	Turkey carcass	Salmonella	1,875	98.24%	0.0923	0	1
	Turkey carcass	Campylobacter	1,857	98.11%	0.0859	0	1
	Ground turkey etc	Salmonella	2,232	82.50%	0.2859	0	1
	Ground turkey etc	Campylobacter	2,129	97.48%	0.1174	0	1
Pork	Ground pork	Salmonella	2,471	77.88%	0.4029	0	1
	Ground pork	Ecoli	206	99.51%	0.0697	0	1
	Ground pork	Non-O157	206	100.00%	0.0000	0	0
	Intact pork	Salmonella	2,240	84.83%	0.3462	0	1
	Intact pork	Ecoli	146	100.00%	0.0000	0	0
	Intact pork	Non-O157	146	100.00%	0.0000	0	0
	Non intact Prok	Salmonella	923	89.30%	0.2913	0	1
	Non intact Prok	Ecoli	64	100.00%	0.0000	0	0
	Non intact Prok	Non-O157	64	100.00%	0.0000	0	0
Ready-to-Eat	Both post-lethality exposed and non post-lethality exposed RTE	Salmonella	21,853	99.95%	0.0214	0	1
	Both post-lethality exposed and non post-lethality exposed RTE	Lm	21,850	99.74%	0.0510	0	1
	Post-lethality exposed RTE product	Salmonella	45,926	99.95%	0.0233	0	1
	Post-lethality exposed RTE product	Lm	45,925	99.73%	0.0513	0	1
	RLm product samples (Composited 5-sample Units)	Lm	1,215	98.98%	0.0909	0	1
	RLm food contact surface samples	Lm	1,220	99.52%	0.0373	0.3333	1
	RLm non-food contact environ. samples (Comp. 5-sample Units; Lm)	Lm	1,220	99.52%	0.0373	0.3333	1

shown in Table 2.2 for each plant-month, and calculate the number of each corresponding sample tested positive.

- A product-pathogen specific passing rate for each plant-month is equal to 1 minus the ratio of the number of product-pathogen samples tested positive and the total number of these product-pathogen samples.

Table 4.3 shows that the mean of product-pathogen sampling tests varies a considerable amount, ranging from 78% to 100%. As expected, for extremely detrimental pathogens like E. coli, the passing rate is high and close to 100%. For pathogens with higher tolerance by FSIS standards such as Salmonella, the passing rate is relatively low. Overall, ground poultry/meat has the lowest Salmonella passing rate. Summary statistics indicate that to consistently compare the food safety results of certified plants with different certification ratings or levels, we should include a fixed effect for each product-pathogen-specific sampling program and time dummies. Also, as noted, compared to the FSIS sanitary inspection data (see Table 4.1), the phytosanitary test data has fewer observations. The reason is that sampling tests are not conducted for each plant every month. FSIS conducts testing based on volumes of production and past performances. This means to use passing rate of phytosanitary sampling tests as outcome variables, we need to take care of the data sparsity problem, which is discussed in detail in section 5.2.1.

4.2 SQF dataset

The SQF audit record data is obtained from the SQF standard program by request. SQF data is on an establishment-date level from July 2014 to December 2019, which contains over 7500 unique establishments and over 33,000 auditing records. The key variables from the dataset are demographic information of the certified plants (name, address, food sector category, scope), and variables related to audits such as SQF level (level 1, 2 or 3), audit rating (A,B,C or No Grade due to failure of certification), audit start and end date, certification body, and audit type (initial audit, announced re-audit announced re-audit, surveillance audit). In general, approximately 50% of the audit records have A ratings, 45% of the records have B ratings, and less than 5% of the records have a rating of C or No Grade. Less than 1% of the SQF records are level 1, about 73% are level 2 records, and around 26% are SQF level 3 audit records.

4.3 BRC dataset

The BRC audit record data are obtained from the BRC standards program on request. BRC data covers the audit history of all BRC certified US food plants since 2009. In total, there are over 2500 locations and more than 1600 auditing records. Key variables are similar to the SQF dataset, which includes demographic information about certified plants (name, address, food sector category, scope), audit rating, audit start and end date, and certification body. The audit type variable (initial audit, re-audit, or surveillance audit) is inferred using the audit start date and previous audit date variable. If the previous audit date variable is missing, then I infer that the establishment had an initial audit on this audit start date. It should be noted that the audit rating rule changed on July 1, 2015 (see Table 2.4) from 4 to 6 ratings. Roughly speaking, rating A/A + before July 1, 2015, is split into two levels AA/AA+ and A/A+ after July 1, 2015; rating C/C+ before July 1, 2015 is split into two levels C/C+ and D/D+ after July 1, 2015. BRC certification tries to improve rating granularity to further distinguish plants between good-performing and bad-performing plants. To make the BRC audit rating more comparable between issue 6 and issue 7, I recorded the ratings into 4 levels: A, B, C/D, No Grade (not certified). AA/AA+ / A/A+ is coded as A, and C/C+/D/D+ in issue 7 is coded as C/D. Compared to the rating distribution of SQF, BRC has more firms in the top tier rating: over 85% audit records have A or better than A ratings, followed by 12% with a B/B+. Only about 2% of the audit records show a result worse than C. One of the reasons is that SQF/BRC rates plants with the same number of critical, major, and minor non-conformity slightly differently (see Tables 2.3 and 2.4). Assuming that what qualifies critical, major, and minor non-conformity is similar under SQF and BRC certification, we can see that SQF A rating has a way higher cut-off for non-conformity than BRC A rating under issue 6 and a slightly higher cut-off for non-conformity than AA rating under issue 7. This makes getting an A in SQF harder than AA/A in BRC. Thus, BRC has a higher portion of A and above ratings.

4.4 Sample construction

I identified all BRC/SQF certified establishments in the FSIS data set through fuzzy matching using the name and address of the establishment. Fuzzy matching is done in multiple rounds using the stata package reclink2. I found that 1086 FSIS establishments had SQF certification during the SQF audit observation window from July 2014 to December 2019, 616 had BRC certification during the BRC audit observation window from January 2009 to December 2019, 41 FSIS establishments match both BRC and SQF records, and the rest 4651 FSIS establishments did not have any SQF/BRC certification record observed.

To ensure the quality of the matching, I manually checked whether the food sector category (FSC) and scope are related to meat, poultry, and eggs for the matched establishments, to the best of my knowledge. Then, I extracted the month of the audit start date of SQF/BRC dataset to covert SQF/BRC dataset to be a month-site unbalanced panel. Finally, I merged the SQF/BRC inspection records with the FSIS administrative dataset to construct a site-month panel (referred to as FSIS - SQF/BRC sample) containing both FSIS inspection and sampling test outcomes, site demographics, and SQF/BRC auditing outcomes for the ever certified plants observed in the certification dataset. During this process, the following 282 establishments are excluded from the FSIS - SQF/BRC analysis sample: (1) FSIS establishments matched to both SQF and BRC records because it is hard to tell whether there is a mismatch of sites based on the information in all the data sets or the establishments are audited by both at the same time; (2) FSIS establishments that only have one plant but two or more plants in certification dataset with the same name and address, FSIS establishments that have two or more plants at the same location with the same company name but only one plant in the certification dataset, and FSIS establishments with multiple plants of the same company name and location but unable to be one-to-one linked to the multiple establishments in the certification dataset; (3) FSIS establishments with obvious and unfixable invalid certification record dates, i.e. negative audit duration; (4) SQF/BRC establishments that are fuzzy matched with FSIS establishments according to address and name, but with questionable food safety category and audit scopes. (5) FSIS establishments with SQF/BRC certification but with all their records outside the FSIS inspection window from July 2012 to December 2017.

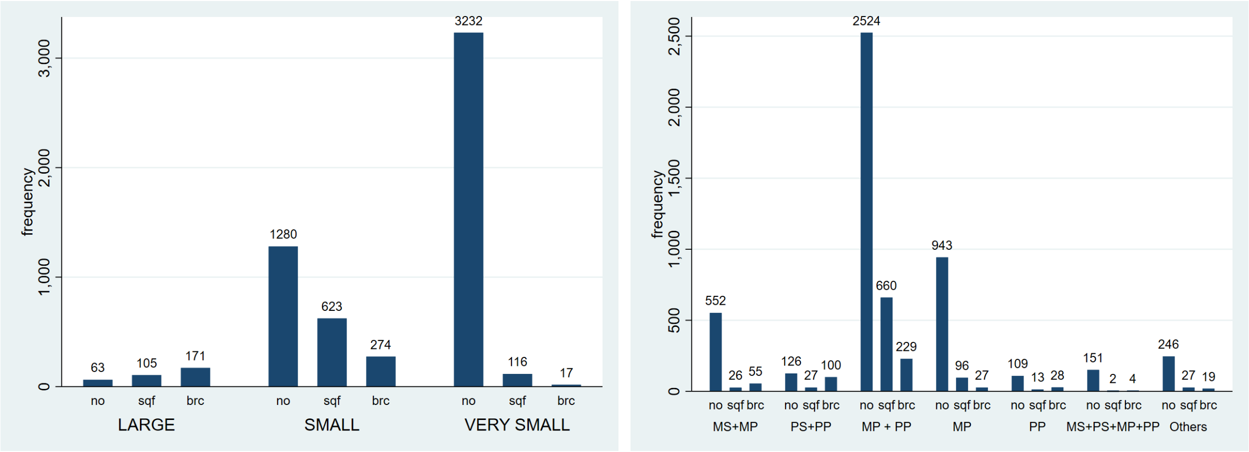
Therefore, the final FSIS - SQF/BRC sample for main analysis is an establishment month panel of 5964 FSIS establishments and 66 months (07/2012 - 12/2017). 851 FSIS establishments have 2543 distinct SQF audit records available from July 2014 to December 2017. 462 establishments have 2133 BRC audits during the FSIS inspection observation window from July 2012 to December 2017. The rest 4651 FSIS establishments serve as control groups that have not observed any certification record in the SQF/BRC dataset. The main dependent variables of interest are food safety practice level measured by compliance rate of FSIS routine sanitary tasks and food safety outcome measured by passing rate of FSIS laboratory sampling tests. The key independent variables of interest for question 1 are certification rating and level, and time dummies for initial and re-certification event time for questions 2 and 3. Relevant control variables include plants' characteristics: HACCP processing size, plant production activity category, FSIS district to which a plant belongs, inspection task category, and product-pathogen sampling test category.

4.5 Summary statistics

Figure 4.2 shows the number of not certified, SQF certified, and BRC certified establishments within each HACCP processing size, processing activity, and district.[2] For large FSIS establishments, most of them are certified by SQF/BRC: about 50% large FSIS plants are BRC certified, 30% are SQF certified, and less than 20% are not certified by either. For small FSIS plants, around 60% are not certified by BRC/SQF, about 28% are certified by SQF,

---

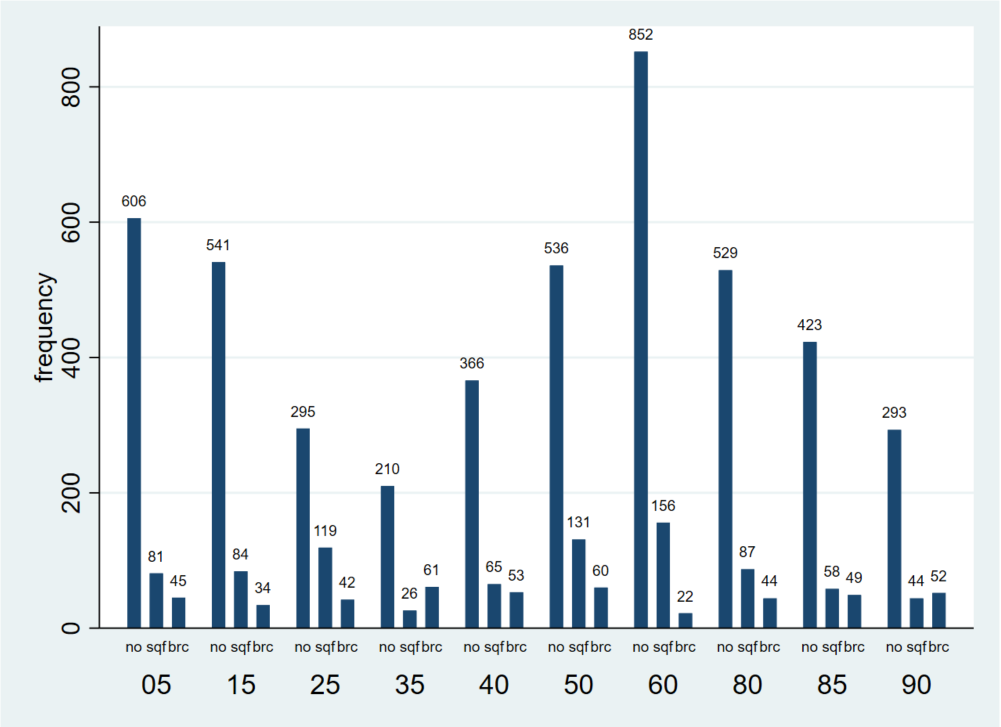
[2] In this dissertation, the term "not certified" specifically refers to plants that were not observed to be certified by SQF or BRC.



(a) HACCP processing sizes

(b) Processing activities

Note: MS = Meat Slaughter, MP = Meat Process, PS = Poultry Slaughter, PP= Poultry Process



(c) FSIS districts

Figure 4.2: Number of establishments by certification type within each HACCP processing size/processing activity/FSIS district

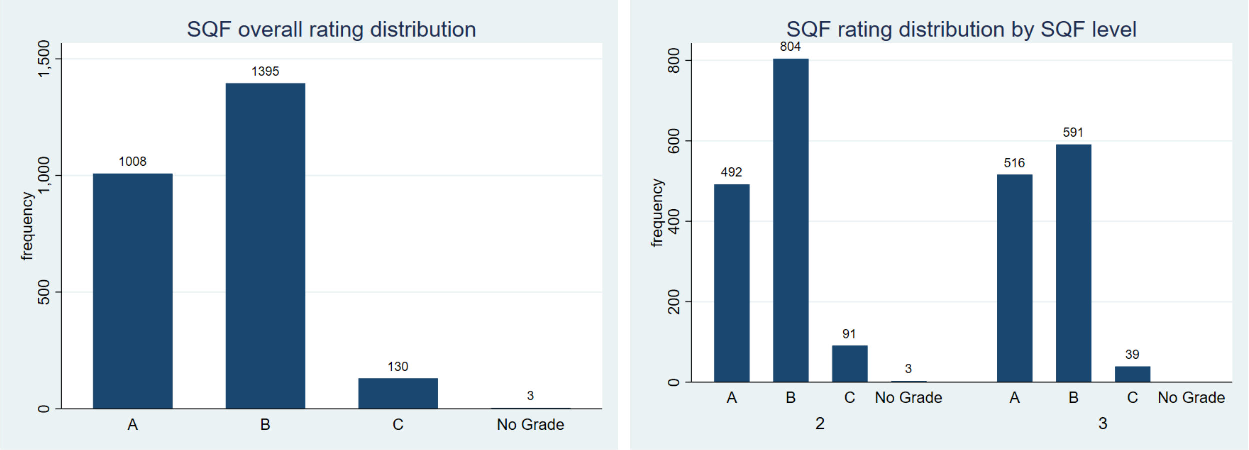
and 12% are certified by BRC. Most of the very small FSIS establishments are not certified by BRC/SQF standards. Since poultry-only plants tend to have large HACCP sizes, it is not surprising to see in Figure 4.2b that poultry-only plants (PS+PP or PP) tend to have a larger portion of SQF/BRC certified establishments than FSIS establishments of all other processing activities. When only looking at SQF/BRC certified FSIS establishments, I find that the top 4 types (86%) of SQF certified plants are small, very small, and large meat and poultry processing plants (MP+PP, 61%, 11% and 6%) and meat processing plants (MP, 8%); the top 4 types (71%) of BRC certified plants are small and large meat and poultry processing plants (MP+PP, 40%, and 7%), large poultry slaughtering and processing plants (PS+PP, 17%), and large meat slaughtering and processing plants (MS+MP, 7%). MS+MP, PS+PP, MP+PP plants tend to choose BRC over SQF for certification, and MP+PP and MP plants tend to choose SQF over BRC for certification. Most certified large plants are BRC certified, and most certified small and very small plants are SQF certified.

When focusing on SQF/BRC certified plants, I find that SQF and BRC have different rating distributions for good plants (above the C rating) as shown in Figure 4.3: the most common grade in the SQF audit record is B, but the most common grade in BRC is the best grade (A/AA). [3] This observation is unsurprising, because as shown in Tables 2.3 and 2.4, SQF requires a lower number of non-conformities for plants to get the highest rating than BRC, which indicates that it is harder for SQF certified plants to get the highest rating than BRC certified plants. In terms of certification level, 1391 (53.76%) SQF certification records are level 2, 1147 (46.15%) are SQF level 3, and SQF level 1 records are dropped. As shown in 4.3b, the certification records of level 3 SQF have a higher proportion of A ratings compared to level 2 SQF certification. BRC certification does not contain different certification levels, but BRC issue 7 tries to provide finer audit grades for certified firms.

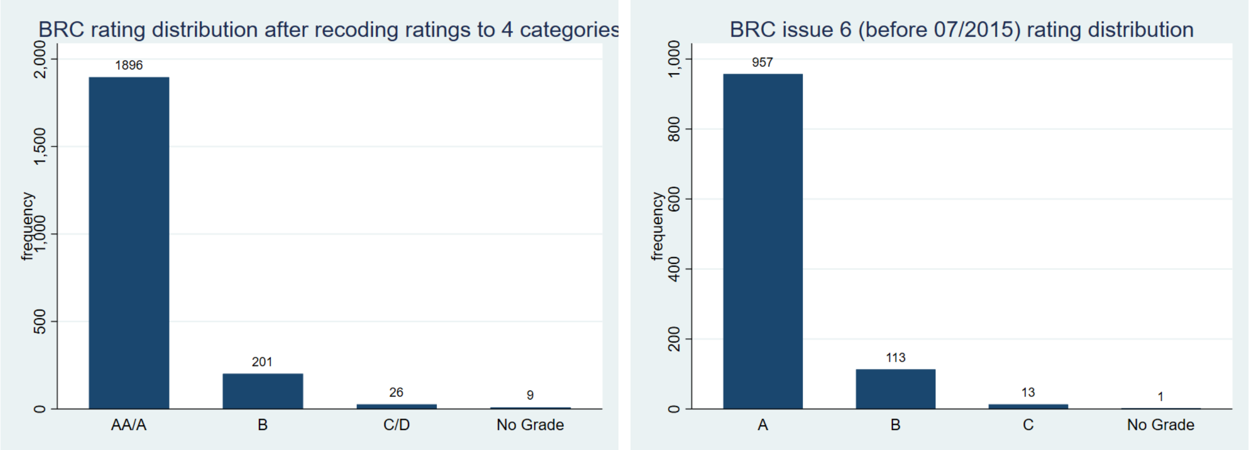
SQF and BRC audit rating distributions also vary by audit type and have different

---

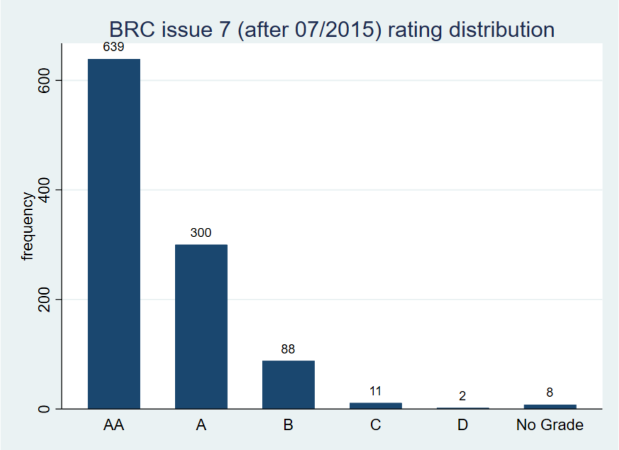
[3] Compared to the rating distribution of all SQF records in the SQF dataset (see 4.2, the SQF certified FSIS establishments have slightly smaller portions (40% vs 50%) of A and slightly bigger portion (55% vs 45%) of B ratings.



(a) SQF overall rating distribution (b) SQF rating distribution by SQF level



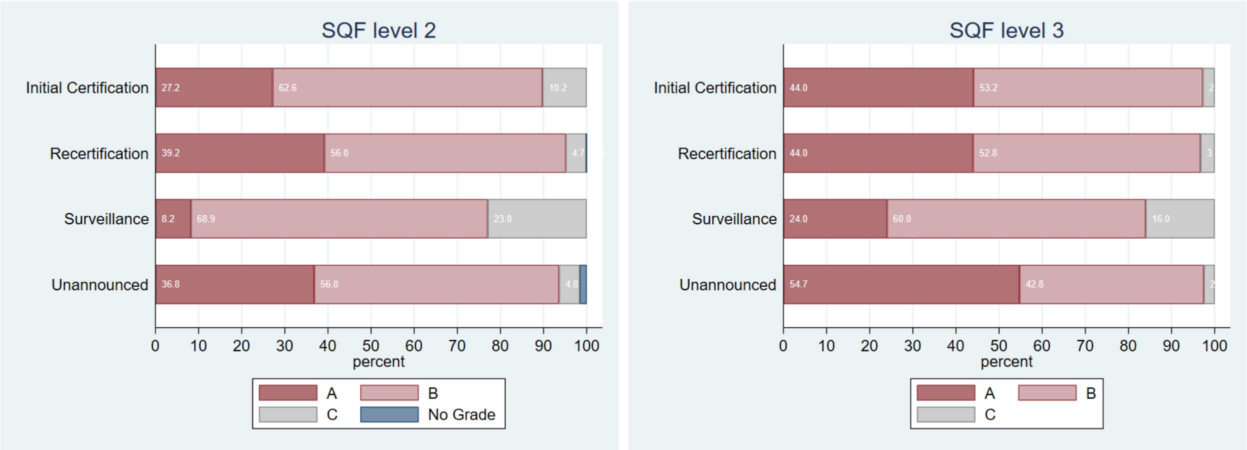
(c) BRC overall rating distribution (d) BRC issue 6 rating distribution



(e) BRC issue 7 rating distribution

Figure 4.3: SQF & BRC certification rating distribution

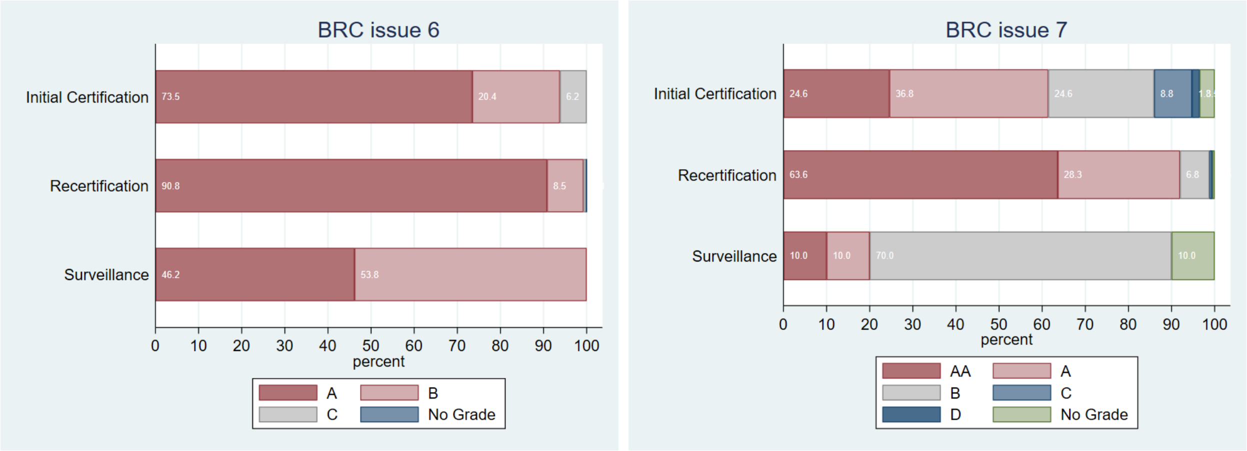
(a) SQF



SQF Level 2	Rating				
Audit type	A	B	C	No Grade	Total
Initial Certification	69	159	26	0	254
Recertification	372	532	45	1	950
Surveillance	5	42	14	0	61
Unannounced	46	71	6	2	125

SQF Level 3	Rating				
Audit type	A	B	C	No Grade	Total
Initial Certification	48	58	3	0	109
Recertification	375	450	28	0	853
Surveillance	6	15	4	0	25
Unannounced	87	68	4	0	159

(b) BRC



BRC issue 6	Rating				
Audit type	A	B	C	No Grade	Total
Initial Certification	83	23	7	0	113
Recertification	868	81	6	1	956
Surveillance	6	7	0	0	13

BRC issue 7	Rating						
Audit type	AA	A	B	C	D	No Grade	Total
Initial Certification	14	21	14	5	1	2	57
Recertification	624	278	67	6	1	5	981
Surveillance	1	1	7	0	0	1	10

Figure 4.4: SQF & BRC certification ratings by audit types (percentage & frequency)

patterns. SQF and BRC audits are categorized into two main types: initial audits and re-audits. The initial audits are all announced audits. SQF re-audits include announced, unannounced, and surveillance audits. BRC re-audits are all announced, which are either recertification or surveillance audit. As listed in Figure and Table 4.4a, in the FSIS - SQF sample (level 2 + level 3 SQF), there are a total of 374 (254 + 109) initial audits, 1803 (950+853) announced and 284 (125+159) unannounced recertifications, and 86 (61+25) surveillance audits. Initial certification has a slightly lower share of A ratings than announced and unannounced certification for SQF level 2, but a similar percentage of A ratings with announced recertification for SQF level 3. Unlike SQF, BRC recertifications have a much higher share of A (or AA) than initial certifications (see Figure 4.4b). For surveillance audits, both the SQF and BRC certification have a relatively smaller portion of the top grades, as expected, because those who experience surveillance audits are plants with relatively worse performance who got a rating below C 6 months ago.

Table 4.5 shows the transition matrix of SQF levels and SQF/BRC certification ratings. Panel 4.4b indicates that plants do not change their SQF certification levels much during the sample period. If a plant chooses to be certified against level 2, the probability of changing to level 3 in the next period is 4.95%. If a plant chooses to be certified in SQF level 3, it only has a 0.91% probability of moving to level 2. In contrast, as shown in panel 4.4a, there are more time variations in SQF ratings, although there is some persistence for a firm to remain in its current grade if it received an A or a B: a plant with the SQF A rating has a 67% probability to get an A in its next audit, and 31% to get a B; a plant with the SQF B rating has a 67% chance to maintain a B and 28% to get an A during its next audit. However, plants with BRC certification tend to have a greater chance of getting better grades during their next audit than the current one: as seen in 4.4c and 4.4e, plants with the highest ratings under BRC issue 6 or issue 7 have an over 82% chance to maintain this rating in the next audit, which is way higher than SQF certification; plants with middle ratings (B/C under issue 6, A/B/C under issue 7) have around 40% - 90% chance to get a higher rating

Table 4.4: SQF & BRC transition matrix of certification ratings and levels

(a) SQF ratings

	Current inspection				
Previous inspection	A	B	C	No Grade	Total
A	67.34% (435)	31.42% (203)	1.24% (8)	0% (0)	100% (646)
B	27.49% (259)	67.41 % (635)	4.78% (45)	0.32% (3)	100% (942)
C	15.53% (16)	57.28 % (59)	27.18% (28)	0% (0)	100% (103)
No Grade	0% (0)	50% (1)	50% (1)	0% (0)	100% (2)

(b) SQF levels

	Current inspection		
Previous inspection	2	3	Total
2	94.05% (869)	5.95% (55)	100% (924)
3	0.91% (7)	99.09% (763)	100% (770)

(c) BRC ratings (issue 6)

	Current inspection (issue 6)				
Previous inspection (issue 6)	A	B	C	No Grade	Total
A	95.76% (565)	4.07% (24)	0.17% (1)	0% (0)	100% (590)
B	73.08% (57)	24.36% (19)	2.56% (2)	0% (0)	100% (78)
C	45.45% (5)	54.55% (6)	0% (0)	0% (0)	100% (11)
No Grade	100% (1)	0% (0)	0% (0)	0% (0)	100% (1)

(d) BRC ratings transition from issue 6 to issue 7

	Current inspection (issue 7)						
Previous inspection (issue 6)	AA	A	B	C	D	No Grade	Total
A	65.99% (227)	28.49% (98)	4.94% (17)	0.58% (2)	0% (0)	0% (0)	100% (344)
B	23.33% (7)	36.67% (11)	40% (12)	0% (0)	0% (0)	0% (0)	100% (30)
C	0% (0)	0% (0)	100% (2)	0% (0)	0% (0)	0% (0)	100% (2)
No Grade	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)

(e) BRC ratings (issue 7)

	Current inspection (issue 7)						
Previous inspection (issue 7)	AA	A	B	C	D	No Grade	Total
AA	82.80% (313)	14.55% (55)	1.85% (7)	0% (0)	0% (0)	0.79% (3)	100% (378)
A	40.96% (68)	50% (83)	7.83% (13)	0.6% (1)	0% (0)	0.6% (1)	100% (166)
B	9.26% (5)	51.85% (28)	33.33% (18)	3.70% (2)	1.85% (1)	0% (0)	100% (54)
C	22.22% (2)	22.22% (2)	44.44% (4)	11.11 % (1)	0% (0)	0% (0)	100% (9)
D	0% (0)	0% (0)	50% (1)	0%(0)	0% (0)	50% (1)	100% (2)
No Grade	60% (3)	20% (1)	0% (0)	0% (0)	0% (0)	20% (1)	0% (5)

Note: Numbers of observations are in parentheses.

during the next audit, which is also higher than SQF certification. Two of the reasons could be that (1) BRC certified plants improve/maintain their food safety practice better than SQF certified plants, or (2) auditors try to give relatively subsequently better grades for the less-performing BRC plants to maintain customers. Hypothesis (1) is further explored in Chapter 6.

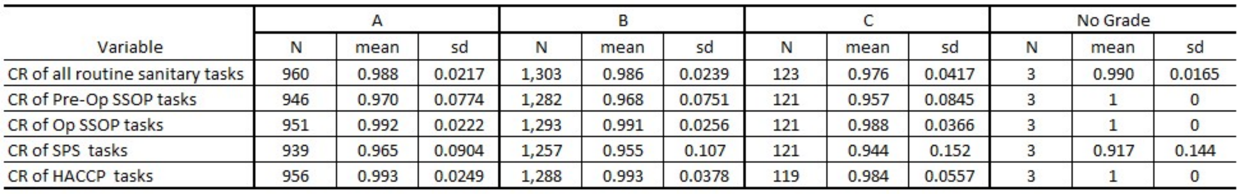
To examine the pattern of certification ratings/levels and their corresponding FSIS routine inspection compliance rate, I calculated the mean CR of different routine tasks in the audit month by SQF/BRC ratings and SQF levels, as shown in Table 4.5. In general, the CRs of different routine tasks shown in Tables 4.5a and 4.5c vary monotonically across certification ratings A, B, and C and within each SQF level. [4] It indicates that SQF ratings are positively associated with FSIS inspection results. However, CRs of routine tasks appear to be similar between SQF level 2 and SQF level 3. For BRC certification, audit ratings and FSIS inspection results depict a slightly different picture. As Table 4.5d and 4.5e show, for issue 6, CR of all routine sanitary tasks is positively correlated with BRC rating A, B and C, but the monotonical relationship is only driven by CR of HACCP tasks; for BRC issue 7, same CR patterns are observed among ratings AA, A, and B. This raw evidence leads to a more formal analysis in Chapter 5 to investigate the relationship between certification ratings and food safety practice.

---

[4] Only 3 SQF inspection records out of over 2000 audit records between July 2014 to December 2017 does not have grades (fails to comply, not certified). The summary statistics of SQF "No Grade" reported in Tables 4.5a and 4.5c are only for completeness. The sample size is too small for any meaningful analysis. BRC "No Grade" group has the same issue.

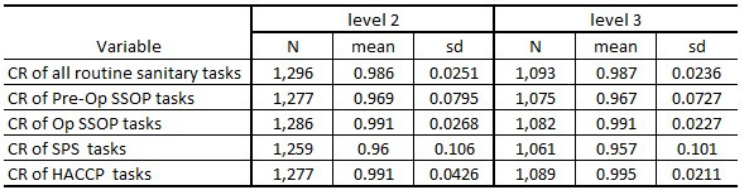
Table 4.5: Mean compliance rate of FSIS inspection tasks at the audit month

(a) By SQF rating



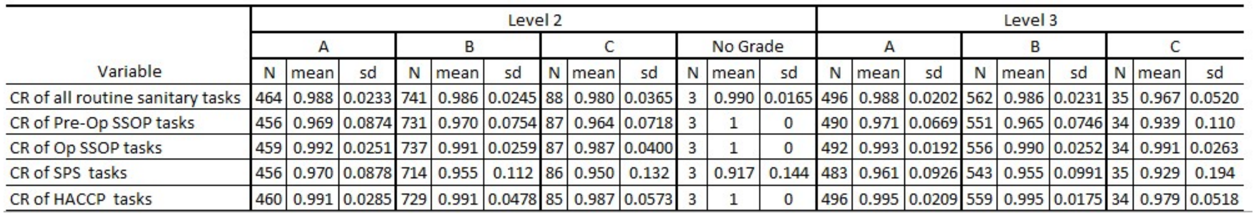
	A			B			C			No Grade		
Variable	N	mean	sd	N	mean	sd	N	mean	sd	N	mean	sd
CR of all routine sanitary tasks	960	0.988	0.0217	1,303	0.986	0.0239	123	0.976	0.0417	3	0.990	0.0165
CR of Pre-Op SSOP tasks	946	0.970	0.0774	1,282	0.968	0.0751	121	0.957	0.0845	3	1	0
CR of Op SSOP tasks	951	0.992	0.0222	1,293	0.991	0.0256	121	0.988	0.0366	3	1	0
CR of SPS tasks	939	0.965	0.0904	1,257	0.955	0.107	121	0.944	0.152	3	0.917	0.144
CR of HACCP tasks	956	0.993	0.0249	1,288	0.993	0.0378	119	0.984	0.0557	3	1	0

(b) By SQF level



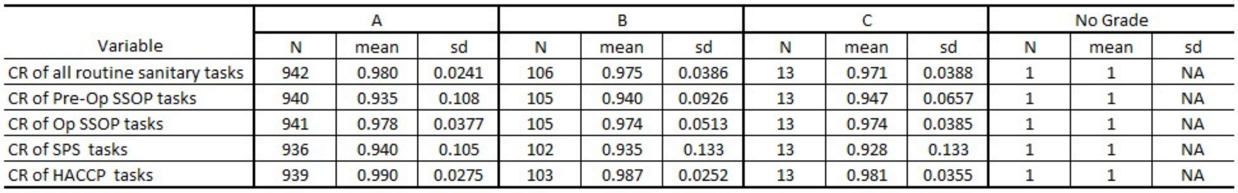
	level 2			level 3		
Variable	N	mean	sd	N	mean	sd
CR of all routine sanitary tasks	1,296	0.986	0.0251	1,093	0.987	0.0236
CR of Pre-Op SSOP tasks	1,277	0.969	0.0795	1,075	0.967	0.0727
CR of Op SSOP tasks	1,286	0.991	0.0268	1,082	0.991	0.0227
CR of SPS tasks	1,259	0.96	0.106	1,061	0.957	0.101
CR of HACCP tasks	1,277	0.991	0.0426	1,089	0.995	0.0211

(c) By SQF level and rating



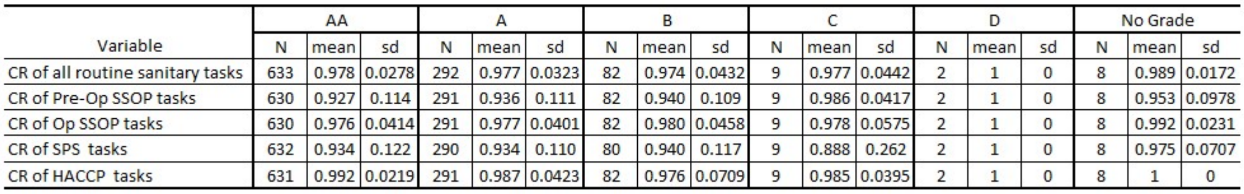
	Level 2												Level 3								
	A			B			C			No Grade			A			B			C		
Variable	N	mean	sd	N	mean	sd	N	mean	sd	N	mean	sd	N	mean	sd	N	mean	sd	N	mean	sd
CR of all routine sanitary tasks	464	0.988	0.0233	741	0.986	0.0245	88	0.980	0.0365	3	0.990	0.0165	496	0.988	0.0202	562	0.986	0.0231	35	0.967	0.0520
CR of Pre-Op SSOP tasks	456	0.969	0.0874	731	0.970	0.0754	87	0.964	0.0718	3	1	0	490	0.971	0.0669	551	0.965	0.0746	34	0.939	0.110
CR of Op SSOP tasks	459	0.992	0.0251	737	0.991	0.0259	87	0.987	0.0400	3	1	0	492	0.993	0.0192	556	0.990	0.0252	34	0.991	0.0263
CR of SPS tasks	456	0.970	0.0878	714	0.955	0.112	86	0.950	0.132	3	0.917	0.144	483	0.961	0.0926	543	0.955	0.0991	35	0.929	0.194
CR of HACCP tasks	460	0.991	0.0285	729	0.991	0.0478	85	0.987	0.0573	3	1	0	496	0.995	0.0209	559	0.995	0.0175	34	0.979	0.0518

(d) By BRC rating (issue 6)



	A			B			C			No Grade		
Variable	N	mean	sd	N	mean	sd	N	mean	sd	N	mean	sd
CR of all routine sanitary tasks	942	0.980	0.0241	106	0.975	0.0386	13	0.971	0.0388	1	1	NA
CR of Pre-Op SSOP tasks	940	0.935	0.108	105	0.940	0.0926	13	0.947	0.0657	1	1	NA
CR of Op SSOP tasks	941	0.978	0.0377	105	0.974	0.0513	13	0.974	0.0385	1	1	NA
CR of SPS tasks	936	0.940	0.105	102	0.935	0.133	13	0.928	0.133	1	1	NA
CR of HACCP tasks	939	0.990	0.0275	103	0.987	0.0252	13	0.981	0.0355	1	1	NA

(e) By BRC rating (issue 7)



	AA			A			B			C			D			No Grade		
Variable	N	mean	sd	N	mean	sd	N	mean	sd	N	mean	sd	N	mean	sd	N	mean	sd
CR of all routine sanitary tasks	633	0.978	0.0278	292	0.977	0.0323	82	0.974	0.0432	9	0.977	0.0442	2	1	0	8	0.989	0.0172
CR of Pre-Op SSOP tasks	630	0.927	0.114	291	0.936	0.111	82	0.940	0.109	9	0.986	0.0417	2	1	0	8	0.953	0.0978
CR of Op SSOP tasks	630	0.976	0.0414	291	0.977	0.0401	82	0.980	0.0458	9	0.978	0.0575	2	1	0	8	0.992	0.0231
CR of SPS tasks	632	0.934	0.122	290	0.934	0.110	80	0.940	0.117	9	0.888	0.262	2	1	0	8	0.975	0.0707
CR of HACCP tasks	631	0.992	0.0219	291	0.987	0.0423	82	0.976	0.0709	9	0.985	0.0395	2	1	0	8	1	0

Chapter 5

Are certification rating and level informative of an establishment's food safety condition?

One of the key roles of certification rating and level is to help certified plants to signal their underlying food safety condition, assist downstream manufacturers or retailers to regulate plants' food safety practices, and select plants with better food safety levels. However, it is unclear whether the third-party certification rating and level information are indeed useful to reveal plants' food safety conditions. This section empirically tests the question by analyzing the relationship between certification information and food safety practice level measured by CR of FSIS routine inspection tasks in section 5.1, the association between certification and food safety outcomes measured by passing rate of FSIS pathogen sampling test in section 5.2.

5.1 Certification information and food safety practice level

To formally answer the question of whether certification rating/level signals a plant's food safety practice level, I compared the average FSIS routine inspection results in the certification (SQF/BRC) audit month (or around the audit months) among different certification ratings and levels. Then I test whether a better grade of SQF/BRC certification or SQF level is associated with a better food safety practice level after controlling for various variables of plant characteristics and time trends.

The key assumptions are as follows. First, I assume that the FSIS inspection results and the SQF and BRC certification ratings are decided independently. This seems plausible, as certification bodies do not require sites to notify them when USDA issues NRs and FSIS inspectors conduct inspections based on the task lists. Second, I assume that FSIS routine inspection outcomes, measured by the compliance rate of inspection tasks, reflect a firm's level of food safety practice. A higher compliance rate indicates a higher level of food safety practice if the food safety and quality management system helps a firm effectively manage food safety risks. The measurements are bounded by 0 % and 100%, which may hinder differentiating firms that have 0% or 100% compliance. However, a robustness check is carried out using the tobit model to resolve the complication of censoring below 0 and above 100% instead of using the main specification, linear regression. The main conclusion does not change. Lastly, a better rating or higher level of SQF / BRC certification means that the firm complies better with the food safety standards. This is plausible as shown in Table 2.3 and Table 2.4 that both SQF and BRC ratings are according to the number of non-conformity subject to the certification standards. The fewer and less severe the non-conformity a certification auditor finds according to the certification standard, the higher the rating a plant has. Furthermore, SQF Level 3 incorporates all elements of Level 2 and goes beyond addressing food quality risks. The food safety practices of firms with SQF level 3 are expected to be at least as good as those with level 2. Under these assumptions, we could conclude that if the SQF certification is a credible signal of a plant's level of food safety practice, a higher rating and SQF level will be associated with a higher compliance rate of routine sanitary inspections.

5.1.1 Methodology

SQF

I use the following specification (regression (5.1)) with CR of all routine sanitation tasks as the main summary outcome variable and apply multiple sets of one-sided t-tests explained in detail below to examine the hypotheses for SQF certification:

1. Better SQF ratings for SQF certification level 2 and 3 indicate a higher level of plant food safety practices (hypothesis family 1: hypotheses (5.2), (5.3)).
2. With the same SQF ratings, plants with the SQF certification level 3 indicate a better food safety practice than those with the SQF level 2 (hypothesis family 2: hypotheses (5.4)).

In addition, I explore the relationship between SQF ratings/levels and CR of specific sanitary tasks (Pre-Op SSOp, OP SSOP, SPS, and HACCP inspection tasks).

$$y_{it} = \sum_{j=2,3} \sum_{r=A,B,CNoGrade} \beta_{rj} 1\{rating_{it} = r, level_{it} = j\} + \lambda_t + X_i + \epsilon_{it} \tag{5.1}$$

where i denotes an establishment, t denotes a year-month, y_{it} is the FSIS routine inspection outcomes of establishment i at year-month t, $1\{rating_{it} = r, level_{it} = j\}$ is a dummy variable that equals to 1 when establishment i is audited against level j and receives rating r during the year-month t, [1] λ_t is time-fixed effect, and X_i are firm characteristic controls including dummies for HACCP plant size, dummies for plant production activities, dummies for FSIS districts (see Figure 4.2) and dummies for FSIS inspection task categories (see Table 4.2). Equation (5.1) is estimated using OLS regression and the errors are clustered to the establishment level.

The main outcome of interest is FSIS CR of all routine sanitary tasks (y_{0it}) for establishment i at year-month t as a summary measurement of FSIS food safety practice level. It

---

[1] Rating C and "No Grade" are combined to be one single category because very few observations have "No Grade" as shown in Table 4.5.

incorporates the information of individual measurement from different angles: CR of Pre-Op SOOP tasks (y_{1it}), CR of Op SSOP tasks (y_{2it}), CR of SPS tasks (y_{3it}) and CR of HACCP tasks (y_{4it}). Using the summary measurement will inform us of the relationship between the overall food safety practice and SQF certification level and ratings. However, the use of individual measurements ($y_{1it}, y_{2it}, y_{3it}, y_{4it}$) can inform us about the specific aspects of establishments' food safety practice levels.

When conducting the above OLS regression (5.1), it is important to consider whether using only the certification auditing month data for regression estimation is appropriate. This approach can suffer from three potential issues: small sample size, the randomness of FSIS inspection outcomes in a particular month, and the lack of FSIS inspection information in months adjacent to the certification audit months. To address these issues, I include the FSIS CR outcomes from one month before and after the certification audit month in the regression sample, and forward and back fill the rating and level information for those months. Most of the certification audits happen at least 4 months apart (4-8 months apart for C/No Grade establishments, and around 12 months apart for A/B establishments). Therefore I chose a balanced total 3 months window including the certification month to prevent FSIS inspection outcomes for the same establishment and month from being double-counted into different level-rating certification groups. One may be concerned that plants might only perform during the audit month but not before and after the audit (I provide a detailed analysis in Chapter 6), which will lead to a lower CR in general than using only the FSIS CR during the audit month. However, both of the measurements are interesting and the results should be interpreted in a slightly different way. The regression and testing results using only data from the audit month indicate whether certification ratings/levels provide information on the plant's food safety practice level at that month. Instead, the results using data around the audit month can provide evidence of whether a certification rating/level is informative of a plant's food safety level after taking into consideration that they may game the system by only performing well during the audit month.

For the different outcome variables mentioned above, we are interested in 2 groups of null hypotheses: The first family of hypotheses includes 6 null hypotheses as shown in (5.2) and (5.3). They compare the level of food safety practice among all pairwise rating groups for level 2 and level 3 SQF certification. The second family of hypotheses includes 3 null hypotheses as shown in (5.4). They compare the outcome variables between SQF certification level 2 and level 3 when the ratings are held constant at A, B, or C/No Grade.

$$H_{0,1}^{sqf}: \beta_{A,2} \leq \beta_{B,2}, \quad H_{0,2}^{sqf}: \beta_{A,2} \leq \beta_{C/NoGrade,2}, \quad H_{0,3}^{sqf}: \beta_{B,2} \leq \beta_{C/NoGrade,2} \tag{5.2}$$

$$H_{0,4}^{sqf}: \beta_{A,3} \leq \beta_{B,3}, \quad H_{0,5}^{sqf}: \beta_{A,3} \leq \beta_{C/NoGrade,3}, \quad H_{0,6}^{sqf}: \beta_{B,3} \leq \beta_{C/NoGrade,3} \tag{5.3}$$

$$H_{0,7}^{sqf}: \beta_{A,3} \leq \beta_{A,2}, \quad H_{0,8}^{sqf}: \beta_{B,3} \leq \beta_{B,2}, \quad H_{0,9}^{sqf}: \beta_{C/NoGrade,3} \leq \beta_{C/NoGrade,2} \tag{5.4}$$

BRC

For BRC certification, I adopt the same method as described above and adapted them to the BRC certification case. BRC rating regime changed from "A, B, C, No Grade" in BRC issue 6 to rating "AA, A, B, C, D, No Grade" in BRC issue 7 as shown in Table 2.4, so I analyze the research question, whether certification is informative of a plant's food safety practice level, separately for BRC issue 6 and issue 7. I use regressions (5.5) and the hypothesis family (5.6) to analyze whether the ratings in BRC issue 6 are informative, and I use regression (5.7) and the hypothesis family (5.8) for BRC issue 7.

$$\text{BRC issue 6:} \quad y_{it} = \sum_{r=A,B,CNoGrade} \beta_{rj} 1\{rating_{it} = r\} + \lambda_t + X_i + \epsilon_{it} \tag{5.5}$$

$$H_{0,1}^{brc}: \beta_A \leq \beta_B, \quad H_{0,2}^{brc}: \beta_A \leq \beta_{C/NoGrade}, \quad H_{0,3}^{brc}: \beta_B \leq \beta_{C/NoGrade} \tag{5.6}$$

$$\text{BRC issue 7:} \quad y_{it} = \sum_{r=AA,A,B,CDNoGrade} \beta_{rj} 1\{rating_{it} = r\} + \lambda_t + X_i + \epsilon_{it} \tag{5.7}$$

$$\begin{gathered} H_{0,4}^{brc} : \beta_{AA} \leq \beta_A, \quad H_{0,5}^{brc} : \beta_{AA} \leq \beta_B, \quad H_{0,6}^{brc} : \beta_A \leq \beta_B, \\ H_{0,7}^{brc} : \beta_{AA} \leq \beta_{C/D/NoGrade}, \ H_{0,8}^{brc} : \beta_A \leq \beta_{C/D/NoGrade}, \ H_{0,9}^{brc} : \beta_B \leq \beta_{C/D/NoGrade} \end{gathered} \tag{5.8}$$

where i denotes an establishment, t denotes a year-month, y_{it} is the FSIS routine inspection outcomes of establishment i at year-month t, $1\{rating_{it} = r\}$ is a dummy variable that equals to 1 when establishment i receives rating r during the year-month t, [2] λ_t is time-fixed effect, and X_i are firm characteristic controls such as dummies for HACCP plant size, dummies for FSIS inspection activities, dummies for FSIS districts (see Figure 4.2) and dummies for FSIS inspection task categories (see Table 4.2). Equations (5.5) and (5.7) are both estimated using OLS regression and the errors are clustered to the establishment level.

Below is a list of hypothesis families I will test for BRC certification:

1. Hypothesis family 1 (3 hypotheses): hypotheses shown in (5.6) for regression (5.5) with the main outcome variable (CR of routine sanitary tasks $y_{0,it}$);

2. Hypothesis family 2 (12 hypotheses): hypotheses shown in (5.6) for regression (5.5) with outcome variables CR of routine Pre-Op SSOP tasks ($y_{1,it}$), CR of routine Op SSOP tasks ($y_{2,it}$), CR of routine SPS tasks ($y_{3,it}$) and CR of routine HACCP tasks ($y_{4,it}$);

3. Hypothesis family 3 (6 hypotheses): hypotheses shown in (5.8) for regression (5.7) with main outcome variable $y_{0,it}$;

4. Hypothesis family 4 (24 hypotheses): hypotheses shown in (5.8) for regression (5.7) with outcome variables $y_{1,it}$, $y_{2,it}$, $y_{3,it}$, and $y_{4,it}$.

---

[2] Rating C and "No Grade" in BRC issue 6 are combined to be one single category because very few observations have "No Grade" as shown in Table 4.5. For the same reason, Ratings C, D and "No Grade" in BRC issue 7 are combined into one single category.

Multiple hypothesis testing

In this section, I will use the case of SQF as an example to discuss the multiple hypothesis testing issues in my research context.

The naive way to test the SQL hypotheses listed above (5.2, 5.3, 5.4) is to conduct t-tests for the two families of total 9 hypotheses for each of the 5 outcome variables y_{it} using only the audit month data or 3-month window data and set the Type I error rate at 5% (or 10%). This means that, for each individual test, the probability of rejecting the null hypothesis when it is true is 5% (or 10%) regardless of whether the remaining hypotheses are rejected. The p-values of each hypothesis that is conducted this way are normally called the marginal probability of falsely rejecting the hypothesis (Veazie, 2006). However, to carefully conclude whether better ratings are associated with higher sanitary CRs of the FSIS inspection, it makes more sense to look at the tests in groups rather than individual hypotheses. Therefore, certain hypotheses are grouped as a family, as explained above, to control the probability of false rejection among these particular hypotheses.

The individual reported p-values are correct for tests conducted in isolation but fail to address the multiple hypothesis situation. When a large number of tests are conducted due to situations such as multiple outcome variables and hypothesis testing among multiple groups, significant results are more likely to simply emerge by chance even if the null hypotheses are true (Anderson, 2008). For example, when the marginal probability of falsely rejecting each hypothesis in (5.2) & (5.3) with y_{0it} as the outcome variable is set to 5%, the probability of falsely rejecting one or more hypotheses in a family of hypotheses (Familywise Error Rate, FWER) is equal to $1 - 95\%^6 = 26.49\%$. In other words, without adjusting the p-values of the individual hypothesis to control for the FWER, the probability of falsely rejecting at least one hypothesis in the hypothesis family 1 is more than 26%. Furthermore, when using outcome variables $y_{1it}, y_{2it}, y_{3it}, y_{4it}$ to explore the relationship of certification rating and FSIS inspection CR of different process aspects, the hypothesis family will have $M = 24$ hypotheses, where M = number of outcome variables × total number of tests per outcome variable $= 4 \times 6 = 24$. If we set the significance level for each hypothesis in the hypothesis family at 5%, the FWER is much higher, equal to $= 1 - 95\%^{24} = 70.80\%$. Cherry picking individual p-values $< 5\%$ and interpreting the results will lead to the wrong conclusions.

To solve the multiple hypothesis testing problems, I could either choose to control the familywise error rate (FWER), the probability of rejecting at least one true null hypothesis, or control the false discover rate (FDR), the proportion of rejections that are falsely rejected, by adjusting the p-values of each test within a hypothesis family. As defined, FWER controls the probability of making any false rejection mistakes in a hypothesis family, so it is well suited for cases where the cost of having one single rejection is high (Anderson, 2008). However, when testing a large number of hypotheses, the FWER adjustments will have very low power. Therefore, we face a type I error (falsely reject) and power (correctly reject) trade-off, which could be addressed by controlling FDR instead. According to the definition of FWER and FDR, when all null hypotheses are true, FWER is equal to FDR; when some null hypotheses are correctly rejected, FDR is less than FWER. Thus, when controlling FWER and FDR at the same level, FDR has better power. As stated in Anderson (2008), “FDR control is well suited to exploratory analysis because it allows a small number of type I errors in exchange for greater power than FWER control”. Therefore, in our case, it is reasonable to control FWER for hypothesis family 1 of 6 hypotheses: (5.2), (5.3) and family 2 of 3 hypotheses: (5.4), respectively, when the outcome variable is the most comprehensive measurement for the overall food safety practice, CR of all sanitary tasks (y_{0it}). When exploring which aspect of food safety practices (Pre-Op SSOP, Op SSOP, SPS, HACCP) is associated with SQF ratings/levels, we can also control for FWER, but controlling FDR may better balance type I error and power trade-off, since there will be 24 (= 4 outcome variables $\times$ 6 hypotheses) hypotheses for hypothesis family 1, and 12 (= 4 outcome variables $\times$ 3 hypotheses) for hypothesis family 2. For the sake of transparency, I will report both FWER and FDR for these two situations with multiple approaches as described below.

There are multiple approaches to asymptotically control FWER; one of the most popular approaches is the Bonferroni correction (Bonferroni, 1935) and the Holm correction (Holm, 1979). Bonferroni correction controls FWER by multiplying each p-value by the number of tests in a hypothesis family and maxing it at one. Though the method is very simple, it does not incorporate information of the joint dependence structure of the tests, which gives poor power. Holm correction applies a "step-down" procedure; it rejects all hypotheses rejected by Bonferroni, but will reject some more. Both generate very conservative adjusted p-values by assuming that the p-values are independent of each other. More powerful approaches include the Westfall-Young stepdown adjusted p-values (Westfall and Young, 1993) and the Romano-Wolf stepdown adjusted p-values (Romano and Wolf, 2005*a*,*b*, 2016) because they both take into account the dependence structure of the test statistics by resampling from the original data. Compared to Westfall and Young p-values, Romano-Wolf stepdown adjusted p-values eliminate a key assumption of resampling-based procedures, the "subset pivotality assumption", the distribution of any subset of the family of test statistics depends only on the validity of the hypotheses in that subset (see more details in Westfall and Young (1993)). In our setting for hypothesis family 1 (5.2) and hypothesis family 2 (5.3), the subset pivotality assumption may not be satisfied because it is very likely that if B-rated plants have better food safety practices than plants with C/No Grade, then A-rated plants may also have better CR than C/No Grade plants. In other words, the subset of the hypotheses in the family of test statistics depends on the hypotheses outside the subset. Practically, I calculated the Holm correction and Westfall-Young p-values by adjusting the STATA program 'wyoung' to accommodate one-sided tests and used the STATA program 'rwolf2' to calculate the Romano-Wolf stepdown adjusted p-values. For detailed procedures of the algorithms for Holm, Westfall-Young and Romano-Wolf p-values, please refer to Holm (1979), Romano and Wolf (2016), and Westfall and Young (1993). In conclusion, though for completeness and transparency, I report all the methods mentioned to adjust p-values, the most relevant method (no special assumption and better power) is the Romano-Wolf stepdown adjusted p-value estimators.

To control FDR, Anderson's sharpened q-values are used. The specific procedure is discussed in Anderson (2008) and Michael Anderson's code is easy to use to compute the q-values. This method has greater power than the Benjamini and Hochberg (1995) method for controlling FDR. However, it does not incorporate information on the joint dependence structure of different tests, therefore, not allowing p-values to be correlated. Though Benjamini, Krieger and Yekutieli (2006) demonstrated that this procedure works well for positively dependent p-values, for negatively dependent p-values, one needs to adopt a more conservative method.

Finally, it is worth mentioning that the Westfall-Young, Romano-Wolf stepdown p-values, and Anderson's sharpened q-values can be smaller than unadjusted p-values. Also, though Anderson's sharpened q-values are used to control for FDR, they could be bigger than Westfall-Young and Romano-Wolf stepdown p-values, which incorporate the correlations of test statistics to control for FWER. However, Anderson's sharpened q-values will always be no bigger than the Holm adjusted p-values. The reason is that q-values control for FDR, but Holm adjusted p-values are adjusted to control the FWER; both are based on step-down procedures and do not incorporate the correlation among test statistics.

In the Results and Discussion section below, Romano-Wolf,Westfall-Young, Holm, Bonferroni adjusted p-values, and Andersonâs sharpened q-values are all reported to address the multiple hypothesis testing issues.

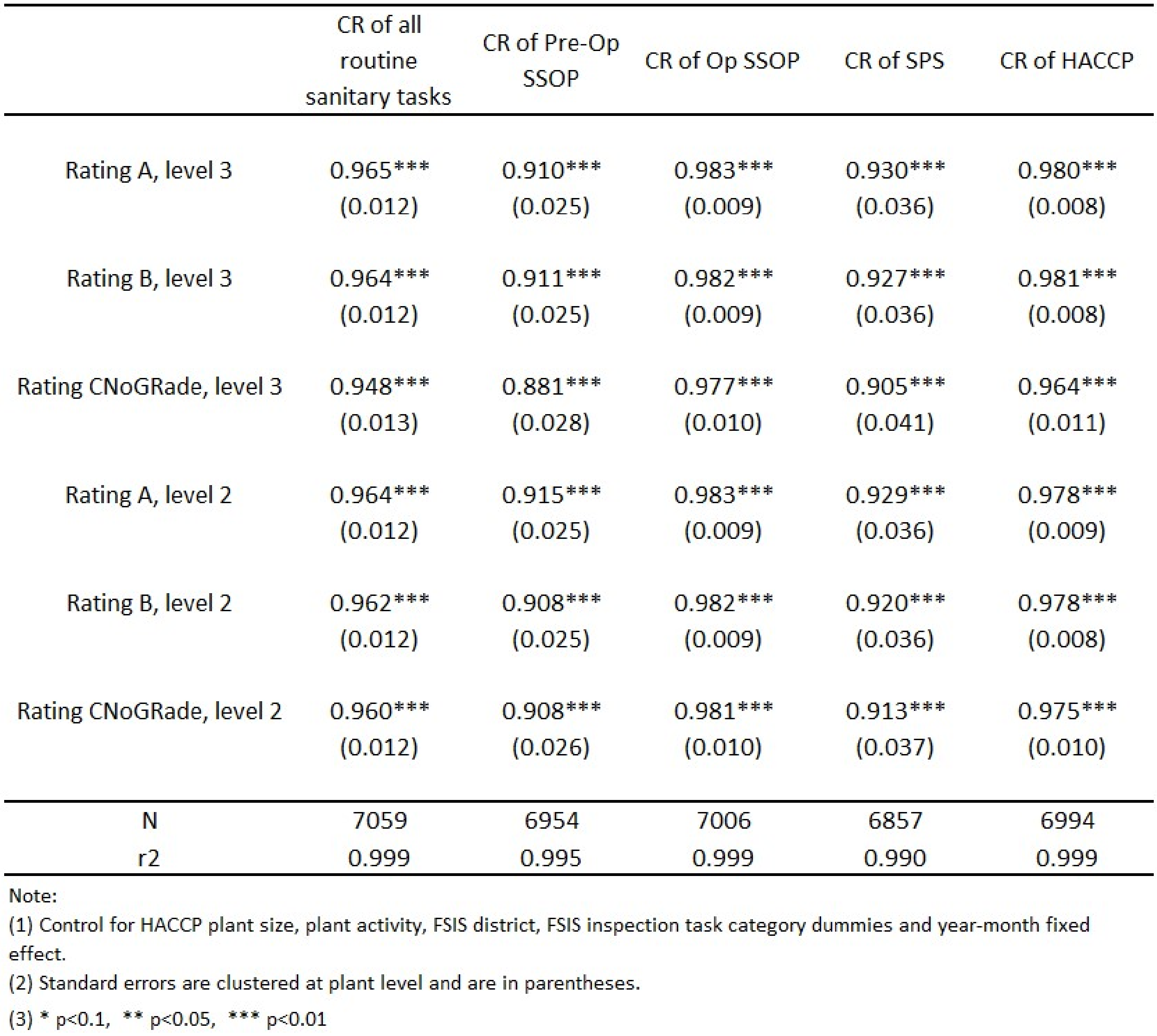
5.1.2 Results and discussion

SQF

At a high level, I find that, in general, SQF ratings are informative of plants' food safety practice behavior, but SQF levels are not.

Table 5.1 shows the results of the OLS regression (5.1) for different outcome variables ($y_{0,it} - y_{4,it}$) using data within a three-month window around audit months. As a robustness check, the results using outcome variables only in the certification audit month are reported in Table 5.A.1. These results are consistent with each other: regardless of whether plants

Table 5.1: Regression result of food safety practice level on SQF rating and level



	CR of all routine sanitary tasks	CR of Pre-Op SSOP	CR of Op SSOP	CR of SPS	CR of HACCP
Rating A, level 3	0.965***	0.910***	0.983***	0.930***	0.980***
	(0.012)	(0.025)	(0.009)	(0.036)	(0.008)
Rating B, level 3	0.964***	0.911***	0.982***	0.927***	0.981***
	(0.012)	(0.025)	(0.009)	(0.036)	(0.008)
Rating CNoGRade, level 3	0.948***	0.881***	0.977***	0.905***	0.964***
	(0.013)	(0.028)	(0.010)	(0.041)	(0.011)
Rating A, level 2	0.964***	0.915***	0.983***	0.929***	0.978***
	(0.012)	(0.025)	(0.009)	(0.036)	(0.009)
Rating B, level 2	0.962***	0.908***	0.982***	0.920***	0.978***
	(0.012)	(0.025)	(0.009)	(0.036)	(0.008)
Rating CNoGRade, level 2	0.960***	0.908***	0.981***	0.913***	0.975***
	(0.012)	(0.026)	(0.010)	(0.037)	(0.010)
N	7059	6954	7006	6857	6994
r2	0.999	0.995	0.999	0.990	0.999

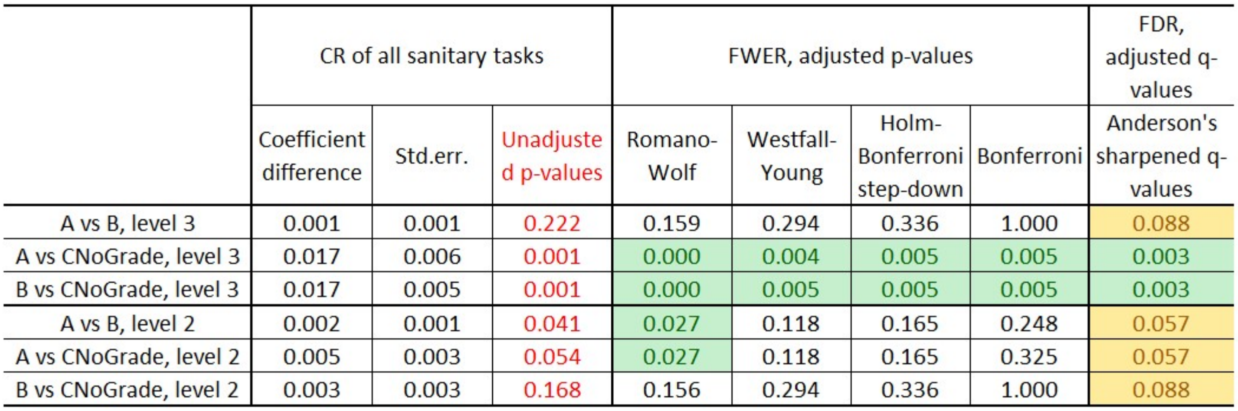
Note:

(1) Control for HACCP plant size, plant activity, FSIS district, FSIS inspection task category dummies and year-month fixed effect.

(2) Standard errors are clustered at plant level and are in parentheses.

(3) * p<0.1, ** p<0.05, *** p<0.01

Table 5.2: One-sided hypothesis tests:
Are better SQF ratings correlated with higher CR of all sanitary tasks?



	CR of all sanitary tasks			FWER, adjusted p-values				FDR, adjusted q-values
	Coefficient difference	Std.err.	Unadjusted p-values	Romano-Wolf	Westfall-Young	Holm-Bonferroni step-down	Bonferroni	Anderson's sharpened q-values
A vs B, level 3	0.001	0.001	0.222	0.159	0.294	0.336	1.000	0.088
A vs CNoGrade, level 3	0.017	0.006	0.001	0.000	0.004	0.005	0.005	0.003
B vs CNoGrade, level 3	0.017	0.005	0.001	0.000	0.005	0.005	0.005	0.003
A vs B, level 2	0.002	0.001	0.041	0.027	0.118	0.165	0.248	0.057
A vs CNoGrade, level 2	0.005	0.003	0.054	0.027	0.118	0.165	0.325	0.057
B vs CNoGrade, level 2	0.003	0.003	0.168	0.156	0.294	0.336	1.000	0.088

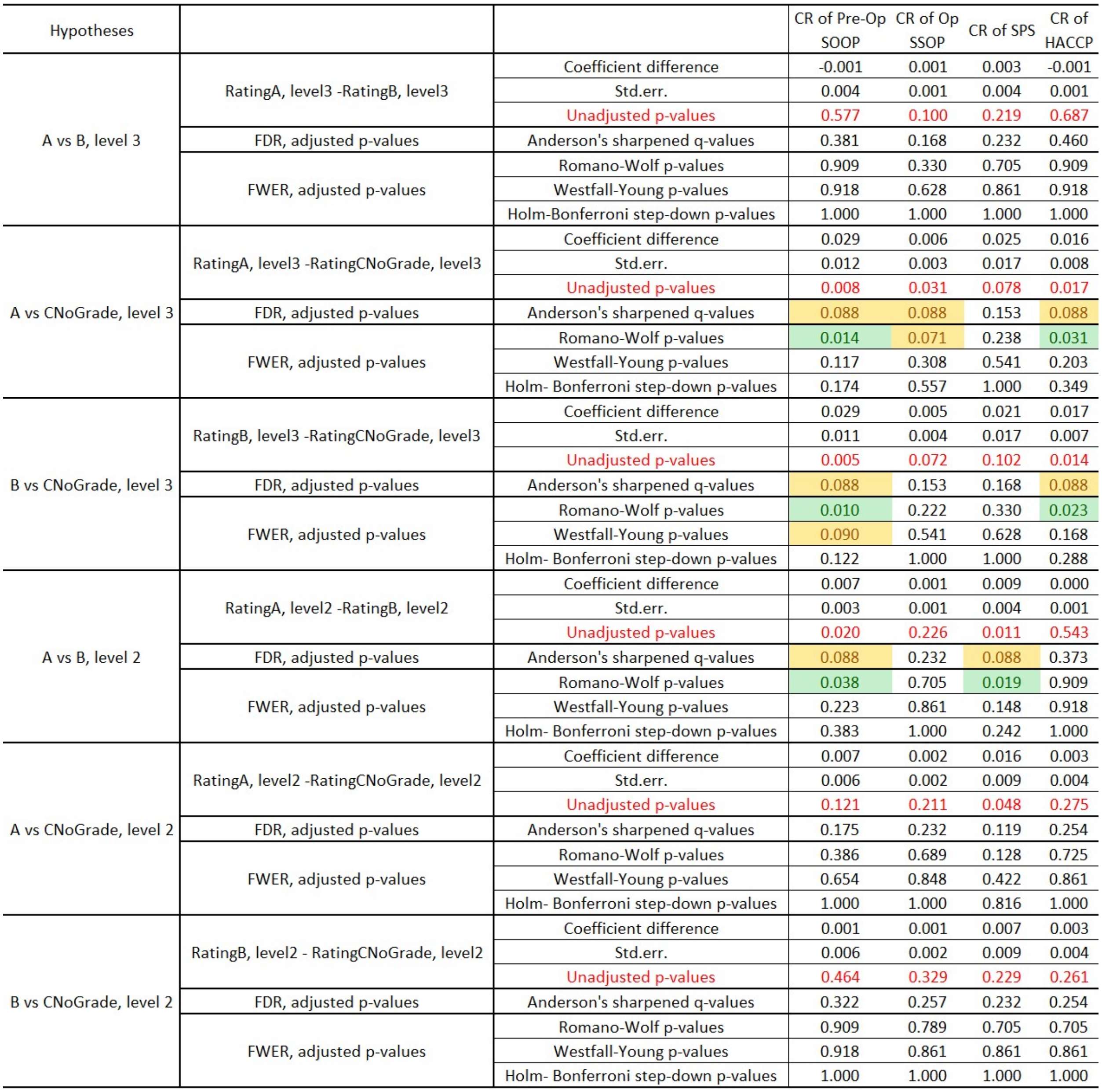
are certified at SQF level 3 or level 2, plants with rating A are associated with the highest CR of all routine sanitary tasks, followed by rating B, and plants with rating C/No Grade are associated with the lowest CR. Similar patterns are seen for CR of Op SSOP and CR of SPS tasks.

From Table 5.2, we can see that most of the unadjusted p-values of the hypotheses in family 1 (hypotheses in (5.2) & (5.3)) are significant at 5%. After adjusting the p-values for multiple hypothesis testing to control FWER, the conclusion does not change when using the Romano-Wolf method: for SQF level 3, we find that plants with A or B rating have a significantly higher CR of overall routine sanitary tasks than plants with C rating, but CR of sanitary tasks for plants with A and B rating are not significantly different from each other with 95% confidence. For SQF level 2, plants with rating A have a significantly higher CR of all sanitary tasks than plants with rating B or C, but CR of sanitary tasks for B and C rating plants are not found significantly different from each other. Figure 5.A.4 shows the null distribution of the t statistics used to calculate the adjusted p-values for each hypothesis using the Romano-Wolf procedure, which allows us to empirically observe how demanding the Romano-Wolf correction is compared to the uncorrected test. When we look at the Westfall-Young and more conservative methods for FWER in Table 5.3, the hypothesis testing results do not change for SQF level 3, but become insignificant for SQF level 2. If we allow for a higher type 1 error in exchange for a higher power, Anderson's sharpened q-values show that with FDR less than 10%, all the hypotheses in hypothesis family 1 are rejected. In other words, we find evidence that SQF plants with better ratings are associated with higher food safety practice levels measured by FSIS inspection CR of all sanitary tasks while holding the SQF certification level constant. Therefore, the overall SQF ratings could help distinguish plants with relatively better food safety practices.

Furthermore, Table 5.3 presents the results on which aspects of food safety practice the higher-rated plants perform better. For plants with SQF level 3 certification, plants with A ratings perform significantly better, under 10% significant level, in Pre-Op SOOP, Op SOOP, and HACCP tasks, as shown by the Romano-Wolf p-values and Anderson's sharpened q-values; B rating plants perform significantly better in Pre-Op SSOP and HACCP inspection tasks. For plants with SQF level 2 certification, rating A plants seem to perform significantly better than rating B plants in terms of Pre-Op SSOP and SPS tasks.

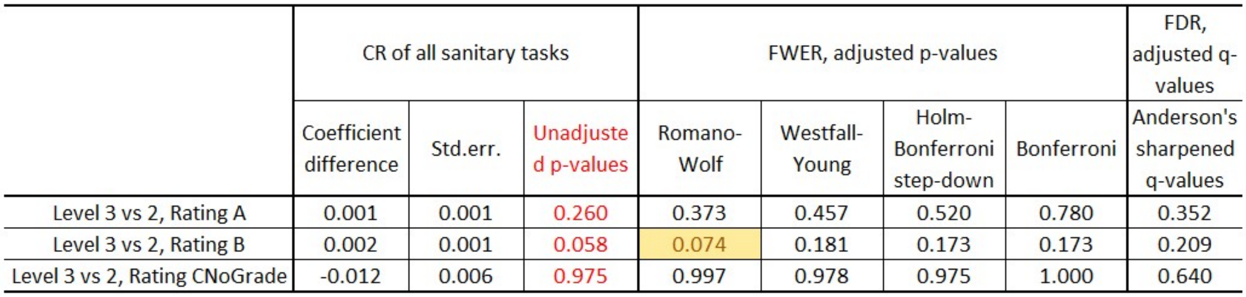
When comparing the coefficients $\beta_{A,3}$ & $\beta_{A,2}$, $\beta_{B,3}$ & $\beta_{B,2}$, and $\beta_{C/NoGrade,3}$ & $\beta_{C/NoGrade,2}$ of Table 5.1, we observe a slightly higher CR of all routine sanitary tasks for SQF level 3 plants whose ratings are A and B. However, the CRs of SQF level 3 and 2 plants are not significantly different from each other at 5% significant level as seen in Table 5.4. Only the Romano-Wolf adjusted p-value is under 10% for the one-sided test $H_{0,8}^{sqf} : \beta_{B,3} \leq \beta_{B,2}$ in the hypothesis family. This indicates that for plants with B ratings, SQF level 3 plants have a slightly better overall CR than those with SQF level 2 certification at 10% significant level and the overall difference in sanitary CR is mainly due to slightly better performance on HACCP tasks, as shown in Table 5.5. However, the average routine HACCP compliance rates for both level 2 and level 3 sites are very high, above 99% as seen in Table 4.5c. Therefore, in general, levels 2 and 3 have similar FSIS routine compliance rates. It makes intuitive sense because SQF level 3 builds on level 2 and emphasizes more on the food quality system, so we expect plants of SQF level 2 and 3 certifications to have very similar levels of food safety practice.

Table 5.3: One-sided hypothesis tests:
What food safety practices do plants with higher SQF ratings do better?



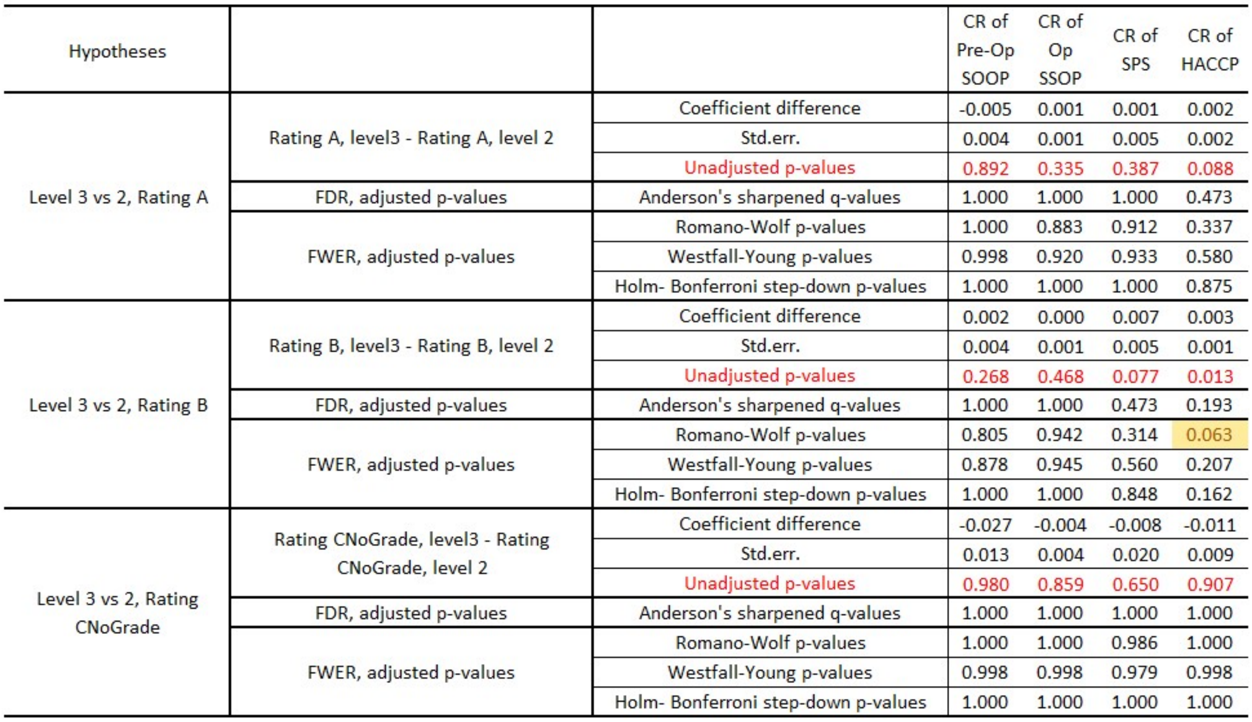
Hypotheses			CR of Pre-Op SOOP	CR of Op SSOP	CR of SPS	CR of HACCP
A vs B, level 3	RatingA, level3 -RatingB, level3	Coefficient difference	-0.001	0.001	0.003	-0.001
		Std.err.	0.004	0.001	0.004	0.001
		Unadjusted p-values	0.577	0.100	0.219	0.687
	FDR, adjusted p-values	Anderson's sharpened q-values	0.381	0.168	0.232	0.460
	FWER, adjusted p-values	Romano-Wolf p-values	0.909	0.330	0.705	0.909
		Westfall-Young p-values	0.918	0.628	0.861	0.918
		Holm-Bonferroni step-down p-values	1.000	1.000	1.000	1.000
A vs CNoGrade, level 3	RatingA, level3 -RatingCNoGrade, level3	Coefficient difference	0.029	0.006	0.025	0.016
		Std.err.	0.012	0.003	0.017	0.008
		Unadjusted p-values	0.008	0.031	0.078	0.017
	FDR, adjusted p-values	Anderson's sharpened q-values	0.088	0.088	0.153	0.088
	FWER, adjusted p-values	Romano-Wolf p-values	0.014	0.071	0.238	0.031
		Westfall-Young p-values	0.117	0.308	0.541	0.203
		Holm- Bonferroni step-down p-values	0.174	0.557	1.000	0.349
B vs CNoGrade, level 3	RatingB, level3 -RatingCNoGrade, level3	Coefficient difference	0.029	0.005	0.021	0.017
		Std.err.	0.011	0.004	0.017	0.007
		Unadjusted p-values	0.005	0.072	0.102	0.014
	FDR, adjusted p-values	Anderson's sharpened q-values	0.088	0.153	0.168	0.088
	FWER, adjusted p-values	Romano-Wolf p-values	0.010	0.222	0.330	0.023
		Westfall-Young p-values	0.090	0.541	0.628	0.168
		Holm- Bonferroni step-down p-values	0.122	1.000	1.000	0.288
A vs B, level 2	RatingA, level2 -RatingB, level2	Coefficient difference	0.007	0.001	0.009	0.000
		Std.err.	0.003	0.001	0.004	0.001
		Unadjusted p-values	0.020	0.226	0.011	0.543
	FDR, adjusted p-values	Anderson's sharpened q-values	0.088	0.232	0.088	0.373
	FWER, adjusted p-values	Romano-Wolf p-values	0.038	0.705	0.019	0.909
		Westfall-Young p-values	0.223	0.861	0.148	0.918
		Holm- Bonferroni step-down p-values	0.383	1.000	0.242	1.000
A vs CNoGrade, level 2	RatingA, level2 -RatingCNoGrade, level2	Coefficient difference	0.007	0.002	0.016	0.003
		Std.err.	0.006	0.002	0.009	0.004
		Unadjusted p-values	0.121	0.211	0.048	0.275
	FDR, adjusted p-values	Anderson's sharpened q-values	0.175	0.232	0.119	0.254
	FWER, adjusted p-values	Romano-Wolf p-values	0.386	0.689	0.128	0.725
		Westfall-Young p-values	0.654	0.848	0.422	0.861
		Holm- Bonferroni step-down p-values	1.000	1.000	0.816	1.000
B vs CNoGrade, level 2	RatingB, level2 - RatingCNoGrade, level2	Coefficient difference	0.001	0.001	0.007	0.003
		Std.err.	0.006	0.002	0.009	0.004
		Unadjusted p-values	0.464	0.329	0.229	0.261
	FDR, adjusted p-values	Anderson's sharpened q-values	0.322	0.257	0.232	0.254
	FWER, adjusted p-values	Romano-Wolf p-values	0.909	0.789	0.705	0.705
		Westfall-Young p-values	0.918	0.861	0.861	0.861
		Holm- Bonferroni step-down p-values	1.000	1.000	1.000	1.000

Table 5.4: One-sided hypothesis tests:
Are higher SQF levels correlated with higher CR of all sanitary tasks?



	CR of all sanitary tasks			FWER, adjusted p-values				FDR, adjusted q-values
	Coefficient difference	Std.err.	Unadjusted p-values	Romano-Wolf	Westfall-Young	Holm-Bonferroni step-down	Bonferroni	Anderson's sharpened q-values
Level 3 vs 2, Rating A	0.001	0.001	0.260	0.373	0.457	0.520	0.780	0.352
Level 3 vs 2, Rating B	0.002	0.001	0.058	0.074	0.181	0.173	0.173	0.209
Level 3 vs 2, Rating CNoGrade	-0.012	0.006	0.975	0.997	0.978	0.975	1.000	0.640

Table 5.5: One-sided hypothesis tests:
What food safety practices do plants with higher SQF levels do better?

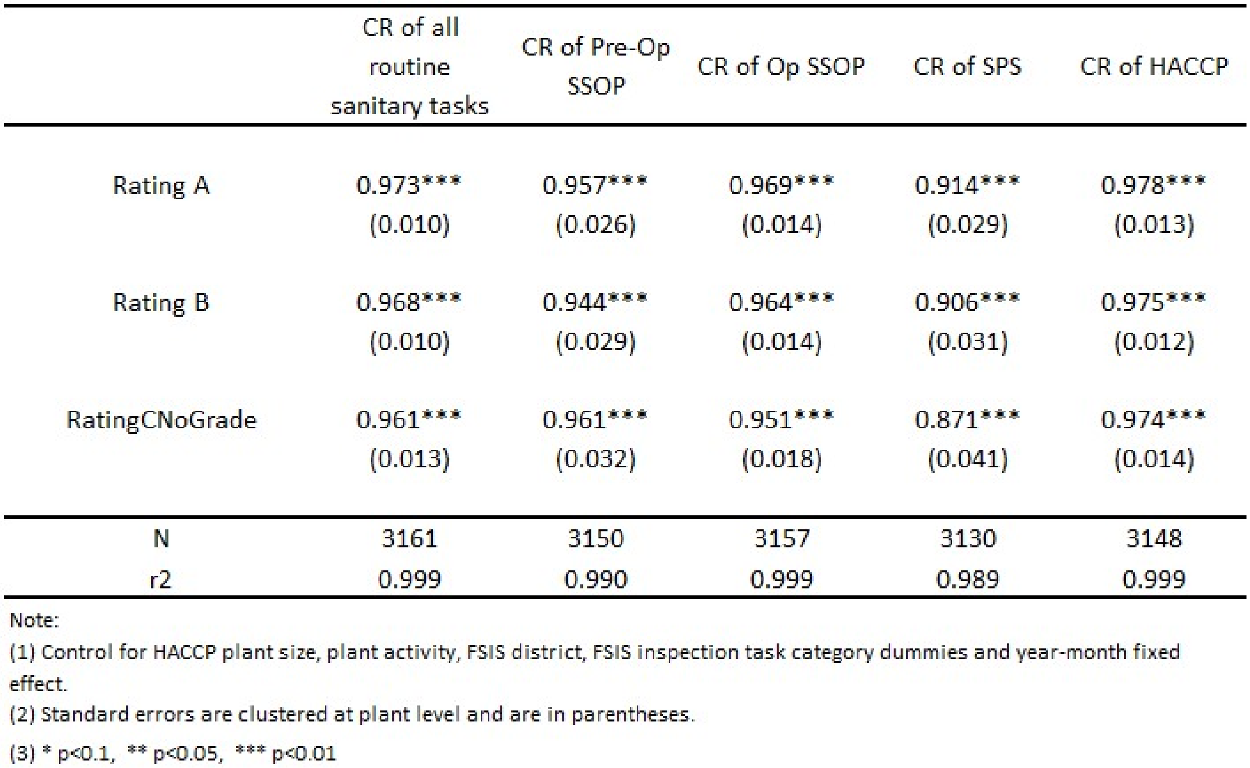


Hypotheses			CR of Pre-Op SOOP	CR of Op SSOP	CR of SPS	CR of HACCP
Level 3 vs 2, Rating A	Rating A, level3 - Rating A, level 2	Coefficient difference	-0.005	0.001	0.001	0.002
		Std.err.	0.004	0.001	0.005	0.002
		Unadjusted p-values	0.892	0.335	0.387	0.088
	FDR, adjusted p-values	Anderson's sharpened q-values	1.000	1.000	1.000	0.473
	FWER, adjusted p-values	Romano-Wolf p-values	1.000	0.883	0.912	0.337
		Westfall-Young p-values	0.998	0.920	0.933	0.580
		Holm- Bonferroni step-down p-values	1.000	1.000	1.000	0.875
Level 3 vs 2, Rating B	Rating B, level3 - Rating B, level 2	Coefficient difference	0.002	0.000	0.007	0.003
		Std.err.	0.004	0.001	0.005	0.001
		Unadjusted p-values	0.268	0.468	0.077	0.013
	FDR, adjusted p-values	Anderson's sharpened q-values	1.000	1.000	0.473	0.193
	FWER, adjusted p-values	Romano-Wolf p-values	0.805	0.942	0.314	0.063
		Westfall-Young p-values	0.878	0.945	0.560	0.207
		Holm- Bonferroni step-down p-values	1.000	1.000	0.848	0.162
Level 3 vs 2, Rating CNoGrade	Rating CNoGrade, level3 - Rating CNoGrade, level 2	Coefficient difference	-0.027	-0.004	-0.008	-0.011
		Std.err.	0.013	0.004	0.020	0.009
		Unadjusted p-values	0.980	0.859	0.650	0.907
	FDR, adjusted p-values	Anderson's sharpened q-values	1.000	1.000	1.000	1.000
	FWER, adjusted p-values	Romano-Wolf p-values	1.000	1.000	0.986	1.000
		Westfall-Young p-values	0.998	0.998	0.979	0.998
		Holm- Bonferroni step-down p-values	1.000	1.000	1.000	1.000

BRC

At a high level, I find evidence that BRC ratings are helpful in distinguishing plants with the highest ratings from those with lower ratings (i.e. A vs B/C in issue 6, AA vs A/B in issue 7) in terms of their level of food safety practice. Increasing the rating granularity from issue 6 to issue 7 (see Table 2.4) helps differentiate food safety practice top performers from the good performers.

Table 5.6: Regression result of food safety practice level on BRC rating (issue 6)



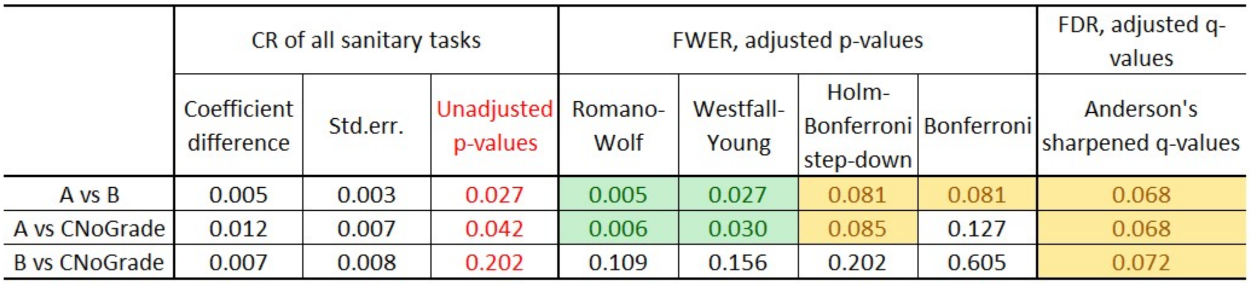
	CR of all routine sanitary tasks	CR of Pre-Op SSOP	CR of Op SSOP	CR of SPS	CR of HACCP
Rating A	0.973*** (0.010)	0.957*** (0.026)	0.969*** (0.014)	0.914*** (0.029)	0.978*** (0.013)
Rating B	0.968*** (0.010)	0.944*** (0.029)	0.964*** (0.014)	0.906*** (0.031)	0.975*** (0.012)
RatingCNoGrade	0.961*** (0.013)	0.961*** (0.032)	0.951*** (0.018)	0.871*** (0.041)	0.974*** (0.014)
N	3161	3150	3157	3130	3148
r2	0.999	0.990	0.999	0.989	0.999

Note:
(1) Control for HACCP plant size, plant activity, FSIS district, FSIS inspection task category dummies and year-month fixed effect.
(2) Standard errors are clustered at plant level and are in parentheses.
(3) * p<0.1, ** p<0.05, *** p<0.01

Table 5.6 shows the results of OLS regressions (5.5) for the BRC issue 6 regime with different outcome variables ($y_{0,it} - y_{4,it}$) within a three-month window around audit months. As a robustness check, results using outcome variables only in the certification audit month are reported in Table 5.A.2. It shows that plants with an A rating are associated with the highest CR of overall routine sanitary tasks, followed by the B and C ratings. Similar patterns are seen for CR of SSOP, SPS, and HACCP tasks.

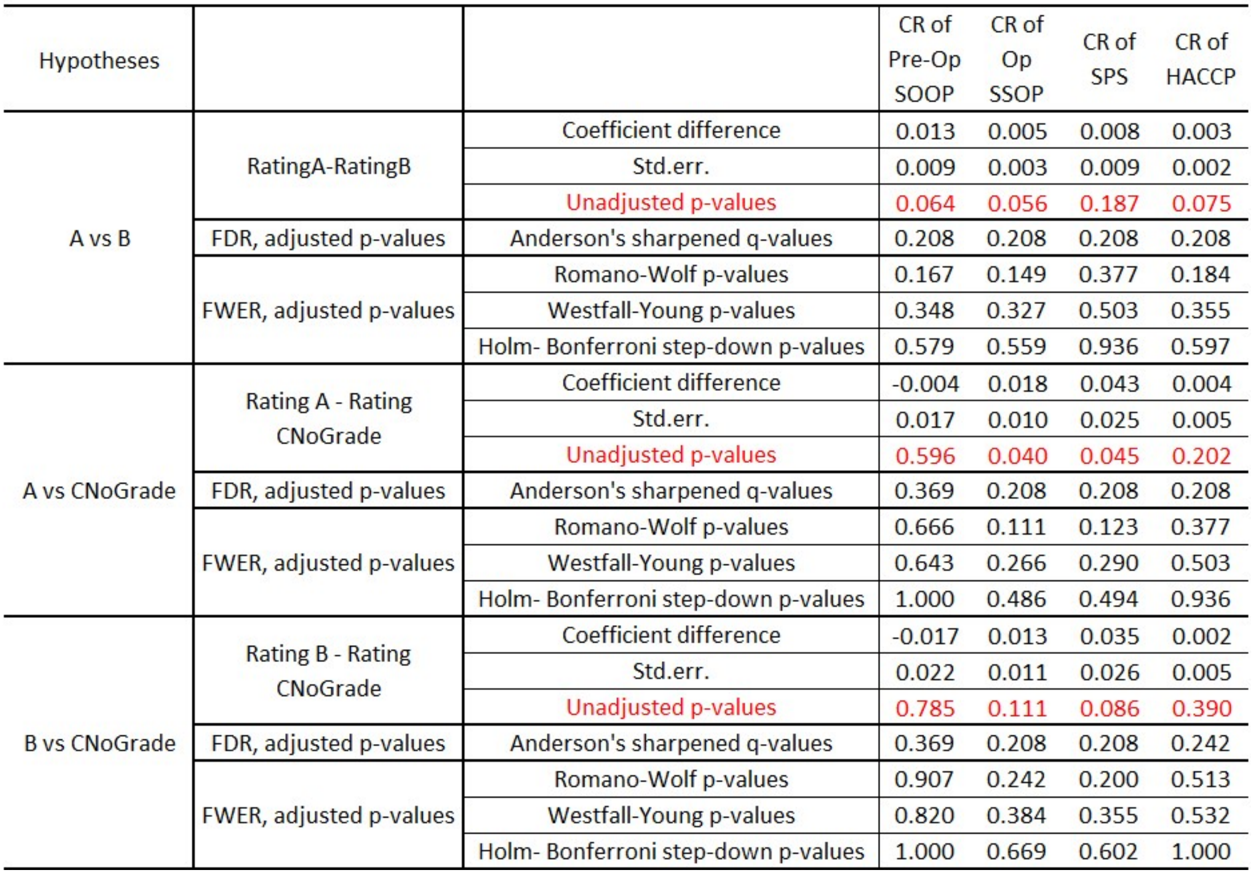
Furthermore, from Table 5.7, we can see that two of the three unadjusted p-values of the hypotheses family (5.6) are significant at 5%. After adjusting the p-values for multiple hypothesis testing to control FWER, the conclusion does not change when using the Romano-Wolf and Westfall-Young method: plants with BRC issue 6 rating A have a significantly higher CR of overall sanitary tasks than plants with rating B or C/No Grade; no significant differences in CR of sanitary tasks for B and C/No Grade plants are found with 95% confidence, but the p-values are borderline 10%. Figure 5.A.6 in Appendix 5.A shows the null distribution of the t statistics used to calculate the adjusted p-values for each hypothesis using the Romano-Wolf procedure, which allows us to empirically observe how demanding the Romano-Wolf correction is compared to the uncorrected test. When we look at the more conservative methods, Holm-Bonferroni and Bonferroni, for FWER in Table 5.3, the hypothesis testing results become insignificant at 5% significance level but significant at 10% significance level. This is due to the lower power of these more conservative tests. If we allow a higher type 1 error in exchange for a higher power, Anderson's sharpened q-values show that with FDR less than 10%, all the hypotheses in the hypothesis family (5.6) are rejected. In other words, we find evidence that for BRC issue 6 certified plants, these with better ratings are associated with higher food safety practice levels measured by CR of FSIS routine sanitary inspection tasks. Therefore, BRC ratings under the issue 6 regime are able to help distinguish plants with different levels of food safety practice. Furthermore, though the unadjusted p-values in Table 5.8 show plants with rating A have better CR in a lot of different food safety practice aspects than plants with B and C/No Grade at 10% significance level, the p-values adjusted for multiple hypothesis tests, the CR differences are not significant anymore.

Table 5.7: One-sided hypothesis tests:
Are better BRC issue 6 ratings correlated with higher CR of all sanitary tasks?



	CR of all sanitary tasks			FWER, adjusted p-values				FDR, adjusted q-values
	Coefficient difference	Std.err.	Unadjusted p-values	Romano-Wolf	Westfall-Young	Holm-Bonferroni step-down	Bonferroni	Anderson's sharpened q-values
A vs B	0.005	0.003	0.027	0.005	0.027	0.081	0.081	0.068
A vs CNoGrade	0.012	0.007	0.042	0.006	0.030	0.085	0.127	0.068
B vs CNoGrade	0.007	0.008	0.202	0.109	0.156	0.202	0.605	0.072

Table 5.8: One-sided hypothesis tests:
What food safety practices do plants with higher BRC issue 6 ratings do better?

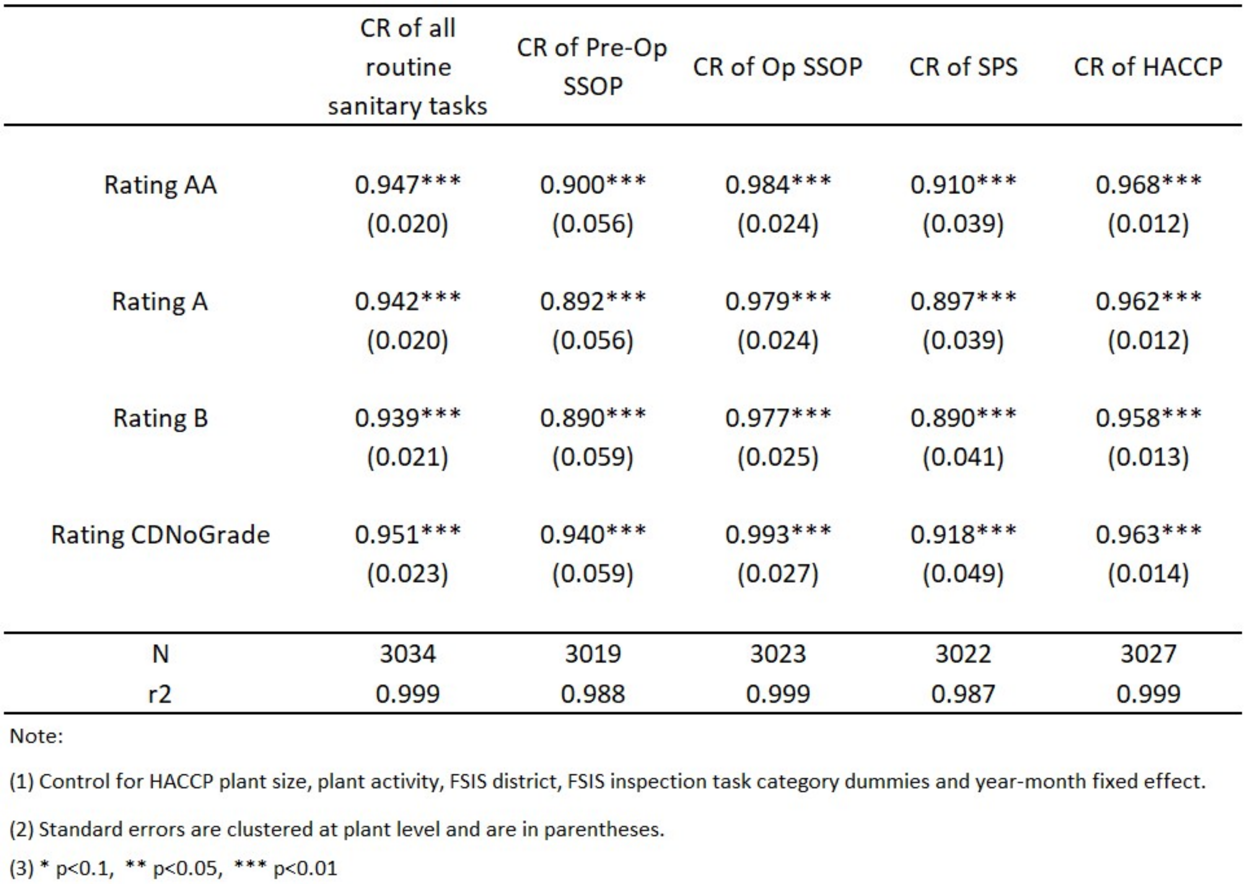


Hypotheses			CR of Pre-Op SOOP	CR of Op SSOP	CR of SPS	CR of HACCP
A vs B	RatingA-RatingB	Coefficient difference	0.013	0.005	0.008	0.003
		Std.err.	0.009	0.003	0.009	0.002
		Unadjusted p-values	0.064	0.056	0.187	0.075
	FDR, adjusted p-values	Anderson's sharpened q-values	0.208	0.208	0.208	0.208
	FWER, adjusted p-values	Romano-Wolf p-values	0.167	0.149	0.377	0.184
		Westfall-Young p-values	0.348	0.327	0.503	0.355
		Holm- Bonferroni step-down p-values	0.579	0.559	0.936	0.597
A vs CNoGrade	Rating A - Rating CNoGrade	Coefficient difference	-0.004	0.018	0.043	0.004
		Std.err.	0.017	0.010	0.025	0.005
		Unadjusted p-values	0.596	0.040	0.045	0.202
	FDR, adjusted p-values	Anderson's sharpened q-values	0.369	0.208	0.208	0.208
	FWER, adjusted p-values	Romano-Wolf p-values	0.666	0.111	0.123	0.377
		Westfall-Young p-values	0.643	0.266	0.290	0.503
		Holm- Bonferroni step-down p-values	1.000	0.486	0.494	0.936
B vs CNoGrade	Rating B - Rating CNoGrade	Coefficient difference	-0.017	0.013	0.035	0.002
		Std.err.	0.022	0.011	0.026	0.005
		Unadjusted p-values	0.785	0.111	0.086	0.390
	FDR, adjusted p-values	Anderson's sharpened q-values	0.369	0.208	0.208	0.242
	FWER, adjusted p-values	Romano-Wolf p-values	0.907	0.242	0.200	0.513
		Westfall-Young p-values	0.820	0.384	0.355	0.532
		Holm- Bonferroni step-down p-values	1.000	0.669	0.602	1.000

Similarly, Table 5.9 shows the results of OLS regression (5.7) for the BRC issue 7 regime with different outcome variables ($y_{0,it} - y_{4,it}$) within a three-month window around audit months. We can see that plants with rating AA are associated with the highest CRs, followed by plants with rating A and rating B. Table 5.10 reports the results for the hypotheses in (5.8). We can see that certified BRC plants with AA ratings have significantly higher CR of overall sanitary tasks than plants with rating A and B at 5% significance level; however, the difference between plants with rating A and B is not statistically significant. The conclusion does not change no matter which method is used to adjust for multiple hypothesis testing. The evidence that rating AA plants are associated with significantly higher CR of overall routine sanitary tasks than A rating plants indicates that with a finer rating scheme, BRC issue 7 certification is able to further distinguish among the good performers, i.e. these A rating plants under BRC issue 6.

In terms of what specific sanitary tasks that higher rating plants perform better, we can see from Table 5.11 that AA rating plants perform significantly better than A plants, under 10% significant level, in Op SOOP, SPS and HACCP tasks (see Romano-Wolf p-values and Anderson's sharpened q-values). Furthermore, plants with AA ratings have significantly higher CR of HACCP tasks than plants with a B rating, which is supported by the hypothesis testing result that all adjusted p-values for this one-sided hypothesis test are less than 5% in Table 5.11.

Table 5.9: Regression result of food safety practice level on BRC rating (issue 7)



	CR of all routine sanitary tasks	CR of Pre-Op SSOP	CR of Op SSOP	CR of SPS	CR of HACCP
Rating AA	0.947***	0.900***	0.984***	0.910***	0.968***
	(0.020)	(0.056)	(0.024)	(0.039)	(0.012)
Rating A	0.942***	0.892***	0.979***	0.897***	0.962***
	(0.020)	(0.056)	(0.024)	(0.039)	(0.012)
Rating B	0.939***	0.890***	0.977***	0.890***	0.958***
	(0.021)	(0.059)	(0.025)	(0.041)	(0.013)
Rating CDNoGrade	0.951***	0.940***	0.993***	0.918***	0.963***
	(0.023)	(0.059)	(0.027)	(0.049)	(0.014)
N	3034	3019	3023	3022	3027
r2	0.999	0.988	0.999	0.987	0.999

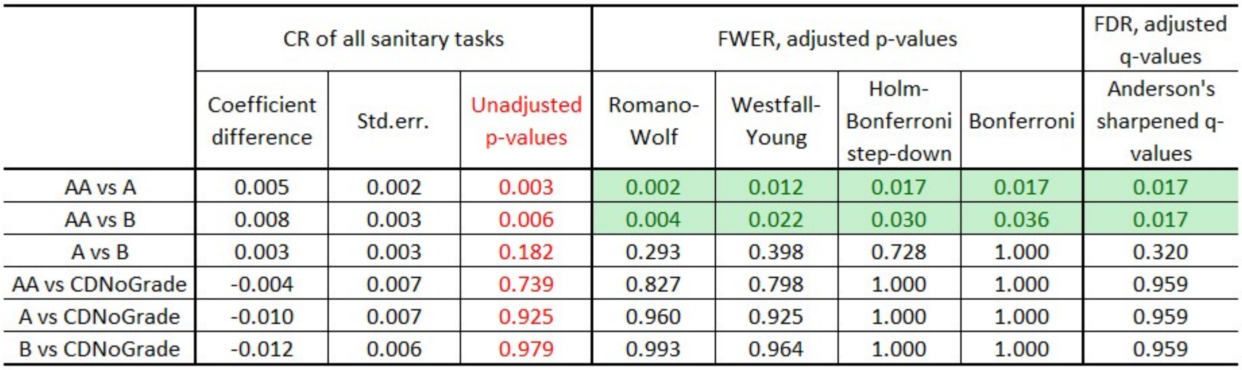
Note:

(1) Control for HACCP plant size, plant activity, FSIS district, FSIS inspection task category dummies and year-month fixed effect.

(2) Standard errors are clustered at plant level and are in parentheses.

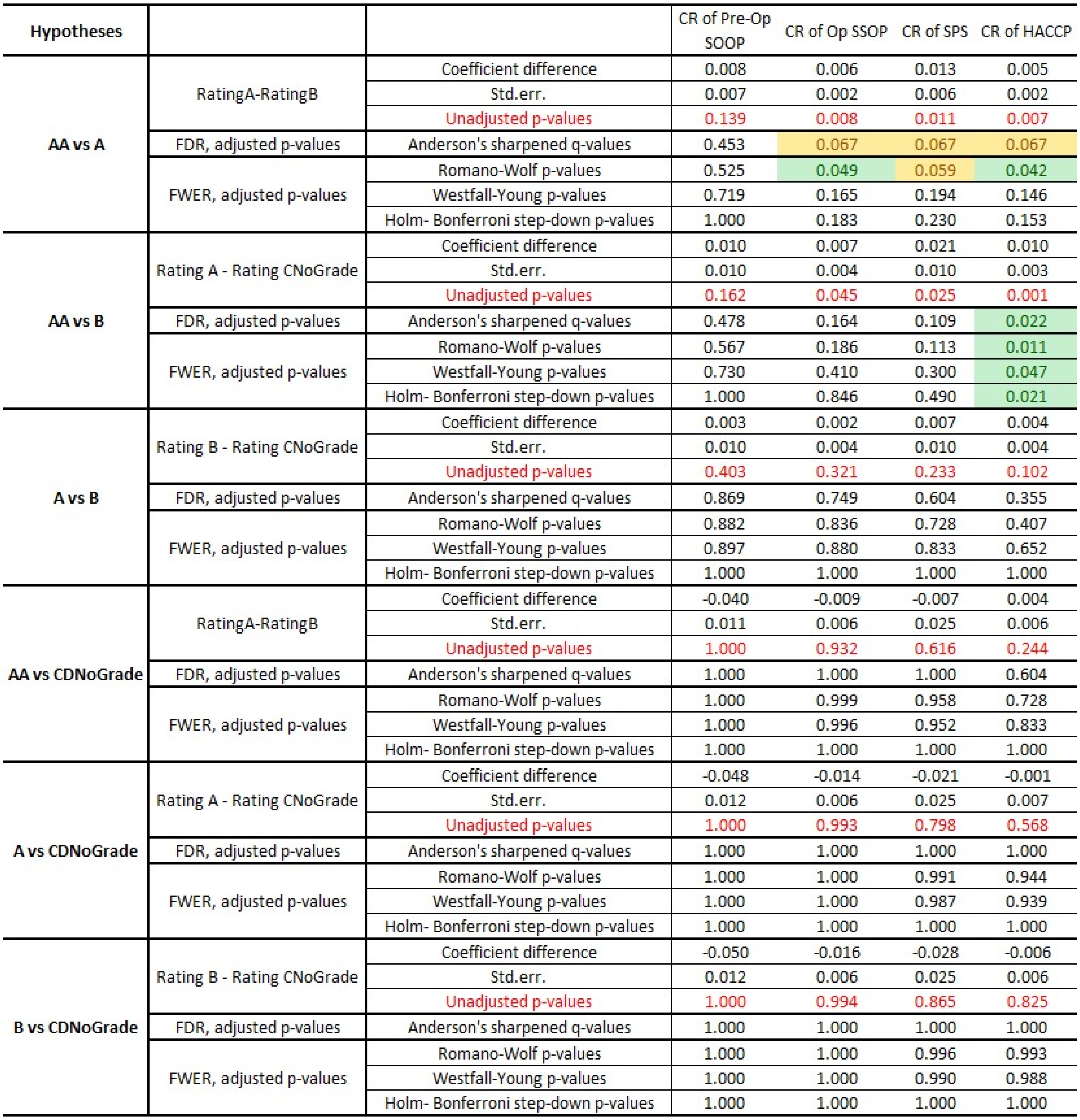
(3) * p<0.1, ** p<0.05, *** p<0.01

Table 5.10: One-sided hypothesis tests:
Are better BRC issue 7 ratings correlated with higher CR of all sanitary tasks?



	CR of all sanitary tasks			FWER, adjusted p-values				FDR, adjusted q-values
	Coefficient difference	Std.err.	Unadjusted p-values	Romano-Wolf	Westfall-Young	Holm-Bonferroni step-down	Bonferroni	Anderson's sharpened q-values
AA vs A	0.005	0.002	0.003	0.002	0.012	0.017	0.017	0.017
AA vs B	0.008	0.003	0.006	0.004	0.022	0.030	0.036	0.017
A vs B	0.003	0.003	0.182	0.293	0.398	0.728	1.000	0.320
AA vs CDNoGrade	-0.004	0.007	0.739	0.827	0.798	1.000	1.000	0.959
A vs CDNoGrade	-0.010	0.007	0.925	0.960	0.925	1.000	1.000	0.959
B vs CDNoGrade	-0.012	0.006	0.979	0.993	0.964	1.000	1.000	0.959

Table 5.11: One-sided hypothesis tests:
What food safety practices do plants with higher BRC issue 7 ratings do better?



Hypotheses			CR of Pre-Op SOOP	CR of Op SSOP	CR of SPS	CR of HACCP
AA vs A	RatingA-RatingB	Coefficient difference	0.008	0.006	0.013	0.005
		Std.err.	0.007	0.002	0.006	0.002
		Unadjusted p-values	0.139	0.008	0.011	0.007
	FDR, adjusted p-values	Anderson's sharpened q-values	0.453	0.067	0.067	0.067
	FWER, adjusted p-values	Romano-Wolf p-values	0.525	0.049	0.059	0.042
		Westfall-Young p-values	0.719	0.165	0.194	0.146
		Holm- Bonferroni step-down p-values	1.000	0.183	0.230	0.153
AA vs B	Rating A - Rating CNoGrade	Coefficient difference	0.010	0.007	0.021	0.010
		Std.err.	0.010	0.004	0.010	0.003
		Unadjusted p-values	0.162	0.045	0.025	0.001
	FDR, adjusted p-values	Anderson's sharpened q-values	0.478	0.164	0.109	0.022
	FWER, adjusted p-values	Romano-Wolf p-values	0.567	0.186	0.113	0.011
		Westfall-Young p-values	0.730	0.410	0.300	0.047
		Holm- Bonferroni step-down p-values	1.000	0.846	0.490	0.021
A vs B	Rating B - Rating CNoGrade	Coefficient difference	0.003	0.002	0.007	0.004
		Std.err.	0.010	0.004	0.010	0.004
		Unadjusted p-values	0.403	0.321	0.233	0.102
	FDR, adjusted p-values	Anderson's sharpened q-values	0.869	0.749	0.604	0.355
	FWER, adjusted p-values	Romano-Wolf p-values	0.882	0.836	0.728	0.407
		Westfall-Young p-values	0.897	0.880	0.833	0.652
		Holm- Bonferroni step-down p-values	1.000	1.000	1.000	1.000
AA vs CDNoGrade	RatingA-RatingB	Coefficient difference	-0.040	-0.009	-0.007	0.004
		Std.err.	0.011	0.006	0.025	0.006
		Unadjusted p-values	1.000	0.932	0.616	0.244
	FDR, adjusted p-values	Anderson's sharpened q-values	1.000	1.000	1.000	0.604
	FWER, adjusted p-values	Romano-Wolf p-values	1.000	0.999	0.958	0.728
		Westfall-Young p-values	1.000	0.996	0.952	0.833
		Holm- Bonferroni step-down p-values	1.000	1.000	1.000	1.000
A vs CDNoGrade	Rating A - Rating CNoGrade	Coefficient difference	-0.048	-0.014	-0.021	-0.001
		Std.err.	0.012	0.006	0.025	0.007
		Unadjusted p-values	1.000	0.993	0.798	0.568
	FDR, adjusted p-values	Anderson's sharpened q-values	1.000	1.000	1.000	1.000
	FWER, adjusted p-values	Romano-Wolf p-values	1.000	1.000	0.991	0.944
		Westfall-Young p-values	1.000	1.000	0.987	0.939
		Holm- Bonferroni step-down p-values	1.000	1.000	1.000	1.000
B vs CDNoGrade	Rating B - Rating CNoGrade	Coefficient difference	-0.050	-0.016	-0.028	-0.006
		Std.err.	0.012	0.006	0.025	0.006
		Unadjusted p-values	1.000	0.994	0.865	0.825
	FDR, adjusted p-values	Anderson's sharpened q-values	1.000	1.000	1.000	1.000
	FWER, adjusted p-values	Romano-Wolf p-values	1.000	1.000	0.996	0.993
		Westfall-Young p-values	1.000	1.000	0.990	0.988
		Holm- Bonferroni step-down p-values	1.000	1.000	1.000	1.000

5.2 Certification information and food safety outcomes

This section explores the relationship between food safety certification ratings/levels and food safety outcomes. Since the routine FSIS inspection tasks do not include product testing, the CR of routine sanitary inspection tasks is more like a measure of plants' food safety practice behavior than plants' food safety outcomes. It is possible that a plant has a relatively high CR of routine sanitary inspection tasks, but its food safety outcome (passing rate of pathogen sampling tests) is not satisfactory. Ollinger and Bovay (2018) found limited effects of performing HACCP and SSOP tasks on Salmonella test outcomes. Furthermore, third-party auditors for SQF and BRC certification do not perform product tests. Their site audits are observation-based to determine the effective implementation of the site's documented food safety practices. Therefore, it is of great interest to understand whether certification ratings and levels are informative of a plant's food safety outcome on top of their food safety practice behaviors.

5.2.1 Data preparation

Overall, sanitary test data is sparse compared to the FSIS sanitary inspection data. Each plant is subjected to a variety of different testing regimes as shown in Table 2.2 based on what products they produce. The product-pathogen specific test results are not available for each plant-year-month because FSIS conducts tests based on the volume of production and past performances. If we only use the passing rate of each product-pathogen sampling test in the certification audit month as a measurement of the plant food safety outcome, many observations with certification audit ratings will be dropped. Thus, to address the data sparsity problem, I calculated the total number of sampling tests for each product-pathogen sampling program 5 months around each plant's certification audit time (including the audit month, plus 2 months before and after the audit month) and compute the corresponding passing rate using data inside the 5-month window.[3] Additionally, I reshaped the data to

---

[3] The main results reported use the pathogen sampling data 2 months around the certification audit. Estimation results are robust to the window duration, i.e. 3 months and 7 months.

the establishment-sampling program-month level for the regression analysis in subsection 5.2.2, and created a key outcome variable, passing rate, and a new set of controls, product-pathogen sampling test dummies.

The key reason for the data transformation is as follows. As we saw in Table 4.3, different product-pathogen tests do have non-negligible differences in the mean and standard deviation of the passing rates. It is inappropriate to aggregate across different tests and calculate an overall sampling test passing rate. To analyze the test results of different sampling programs, we can conduct regression analysis on samples of plants with the same product-pathogen program separately. However under this method the statistical power of each sample will be low because of the small sample size. Instead, an establishment-sampling program-month level dataset not only allows us to control for the product-pathogen fixed effect to compare test results of certified plants in the same product-pathogen program but also ensures a bigger sample size in consideration of the statistical power.

5.2.2 Methodology

To test whether higher SQF ratings and levels are associated with better food safety outcomes measured by passing rates of phytosanitary sampling tests, I use a similar specification as shown in equation (5.1) with the passing rate as the outcome variable and slightly different controls X_i, including dummies for HACCP plant size, FSIS district to which a plant belongs, and product-pathogen sampling program fixed effect. The standard error is clustered at the establishment level. Similarly, we are interested in testing 2 family hypotheses:

1. holding SQF levels constant, test whether higher ratings are associated with better passing rate: hypotheses (5.2) and (5.3);
2. holding SQF rating constant, test whether higher certification levels are associated with better food safety outcomes: hypotheses (5.4).

To address multiple hypothesis testing problem in each hypothesis family, in addition to the unadjusted p-values of each one-sided test, I calculate the Romano-Wolf, Westfall-Young, Holm-Bonferroni step-down, and Bonferroni correction to control for FWER and Anderson's sharpened q-values to control for FDR as explained in details in section 5.1.1.

For BRC issue 6 and issue 7, I adjust the regression (5.5) and (5.7) by using the new outcome variable and controls as SQF regression mentioned above. Due to the data sparsity of phytosanitary sampling programs, for BRC issue 6, no certified plants with pathogen test passing rates get a rating C or below. Therefore, we only need to test whether the food safety outcome of A rating plants is significantly better than the B rating plants. No multiple hypothesis testing adjustments are needed. For issue 7, we conduct the 6 hypotheses shown in (5.8) and adjust the p-values as described above for SQF certification.

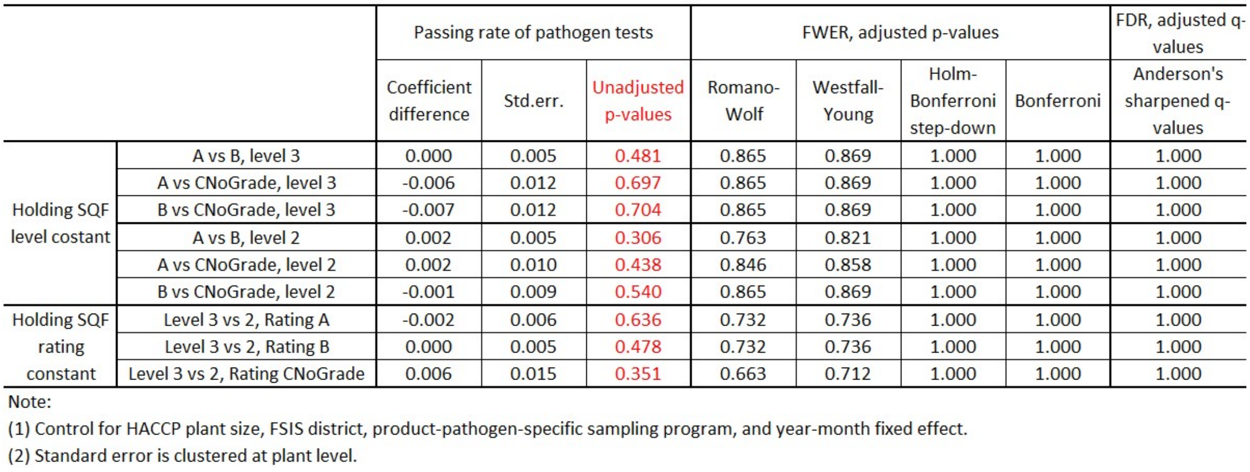
5.2.3 Results and discussion

Overall, there is no evidence that higher certification ratings and levels are associated with a statistically significant higher passing rates of pathogen tests after controlling for plant size, the FSIS district plants belong to, product-pathogen sampling program fixed effect, and year-month fixed effect. Though the point estimators for BRC certification issue 7 suggest that higher ratings do have a higher passing rate of pathogen tests, they are not estimated precisely.

SQF

The upper panel of Table 5.12 reports the pairwise coefficient differences of all rating groups for SQF certification level 2 and 3. The point estimates show that on average, among SQF certification level 2 plants, those with A ratings are associated with around 0.2% higher passing rate than those with worse ratings. All other point estimate differences between higher ratings and lower rates are non-positive. However, none of these coefficient differences are significantly different from zero at 10% significance level. Thus, our pathogen test data cannot reject the hypotheses in (5.2) and (5.3). There is no statistically significant evidence to support the claim that SQF ratings reflect plants' underlying food safety outcomes. Similarly, the lower panel of Table 5.12 displays the hypothesis testing results for hypotheses in (5.4).

Table 5.12: One-sided hypothesis tests:
Are better SQF ratings and levels correlated with higher pathogen test passing rates?



		Passing rate of pathogen tests			FWER, adjusted p-values				FDR, adjusted q-values
		Coefficient difference	Std.err.	Unadjusted p-values	Romano-Wolf	Westfall-Young	Holm-Bonferroni step-down	Bonferroni	Anderson's sharpened q-values
Holding SQF level costant	A vs B, level 3	0.000	0.005	0.481	0.865	0.869	1.000	1.000	1.000
	A vs CNoGrade, level 3	-0.006	0.012	0.697	0.865	0.869	1.000	1.000	1.000
	B vs CNoGrade, level 3	-0.007	0.012	0.704	0.865	0.869	1.000	1.000	1.000
	A vs B, level 2	0.002	0.005	0.306	0.763	0.821	1.000	1.000	1.000
	A vs CNoGrade, level 2	0.002	0.010	0.438	0.846	0.858	1.000	1.000	1.000
	B vs CNoGrade, level 2	-0.001	0.009	0.540	0.865	0.869	1.000	1.000	1.000
Holding SQF rating constant	Level 3 vs 2, Rating A	-0.002	0.006	0.636	0.732	0.736	1.000	1.000	1.000
	Level 3 vs 2, Rating B	0.000	0.005	0.478	0.732	0.736	1.000	1.000	1.000
	Level 3 vs 2, Rating CNoGrade	0.006	0.015	0.351	0.663	0.712	1.000	1.000	1.000

Note:
(1) Control for HACCP plant size, FSIS district, product-pathogen-specific sampling program, and year-month fixed effect.
(2) Standard error is clustered at plant level.

As we can see, plants with higher SQF levels do not have significantly better pathogen test results.

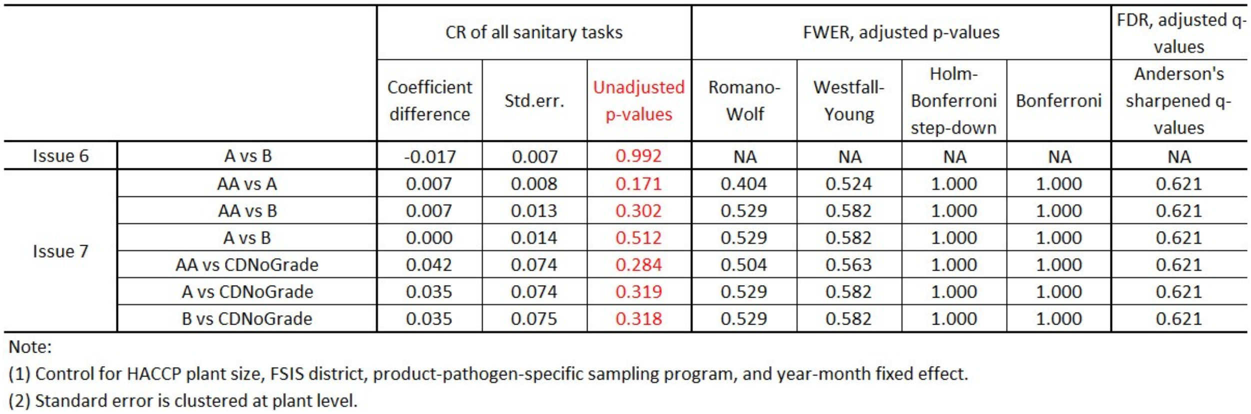
BRC

Table 5.13 summarizes the hypothesis testing results on whether plants with higher ratings of BRC certification are associated with better food safety outcomes measured by pathogen test passing rates. The results indicate that there is no evidence that rating A plants have significantly better pathogen test results than rating B plants under the BRC issue 6 regime. A finer rating scale is introduced in BRC certification issue 7. From the point estimates, we observe that higher grades are associated with higher passing rates of pathogen tests. However, all the unadjusted and adjusted p-values of the hypothesis tests 5.8 are over 10%, which indicates the passing rate differences among AA, A, B, and CDNoGrade rating groups are not significantly different at 10% level.

5.3 Conclusion

This chapter studies whether the third-party food safety certification rating and level are informative of a plant's food safety practice behavior and its underlying food safety outcomes. Using the two independent measures of plant food safety conditions from FSIS inspectors

Table 5.13: One-sided hypothesis tests:
Are better BRC (issue 6/7) ratings correlated with higher pathogen test passing rates?



		CR of all sanitary tasks			FWER, adjusted p-values				FDR, adjusted q-values
		Coefficient difference	Std.err.	Unadjusted p-values	Romano-Wolf	Westfall-Young	Holm-Bonferroni step-down	Bonferroni	Anderson's sharpened q-values
Issue 6	A vs B	-0.017	0.007	0.992	NA	NA	NA	NA	NA
Issue 7	AA vs A	0.007	0.008	0.171	0.404	0.524	1.000	1.000	0.621
	AA vs B	0.007	0.013	0.302	0.529	0.582	1.000	1.000	0.621
	A vs B	0.000	0.014	0.512	0.529	0.582	1.000	1.000	0.621
	AA vs CDNoGrade	0.042	0.074	0.284	0.504	0.563	1.000	1.000	0.621
	A vs CDNoGrade	0.035	0.074	0.319	0.529	0.582	1.000	1.000	0.621
	B vs CDNoGrade	0.035	0.075	0.318	0.529	0.582	1.000	1.000	0.621

Note:
(1) Control for HACCP plant size, FSIS district, product-pathogen-specific sampling program, and year-month fixed effect.
(2) Standard error is clustered at plant level.

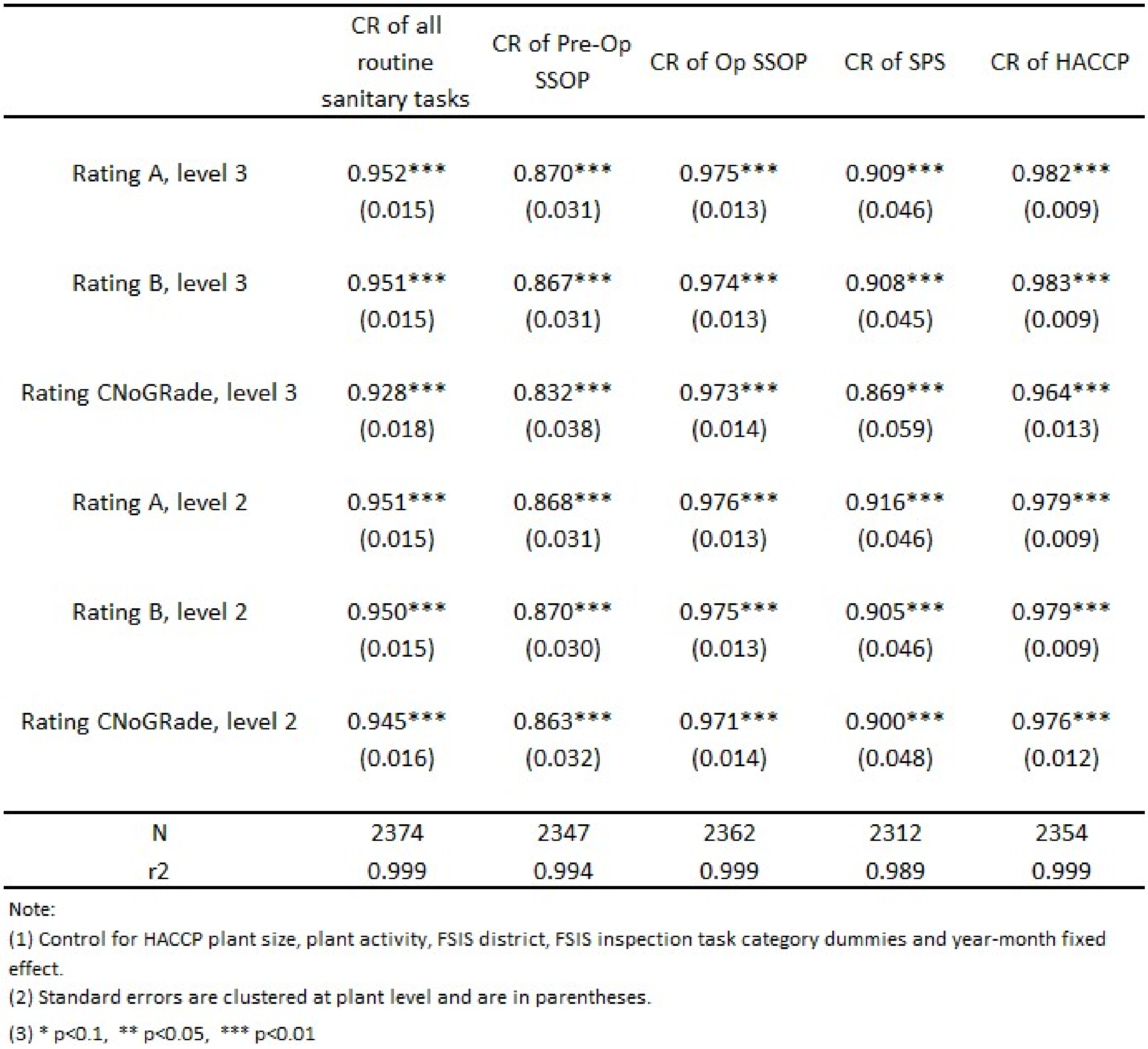
and private third-party auditors, I find statistically significant evidence that SQF and BRC ratings are informative of an establishment's level of food safety practice. I also find that by introducing a finer grading scale, BRC certification can better distinguish the top performers in food safety practices from the bad ones.

As found in a previous study by Ollinger and Bovay (2018), good performance on certain FSIS sanitary tasks may not necessarily translate into good pathogen test outcomes. Consistent with this, my analysis did not find that plants with higher ratings and certification levels are associated with significantly better pathogen test results. Although the point estimates of the average passing rate for pathogen tests among BRC issue 7 ratings follow the positive direction, they are not precisely estimated. The findings imply the important and irreplaceable role of conducting sampling tests to evaluate plants' food safety conditions.

5.A Appendix

5.A.1 Robustness check

Table 5.A.1: Regression result of food safety practice level on SQF rating and level only using the certification audit month data



	CR of all routine sanitary tasks	CR of Pre-Op SSOP	CR of Op SSOP	CR of SPS	CR of HACCP
Rating A, level 3	0.952***	0.870***	0.975***	0.909***	0.982***
	(0.015)	(0.031)	(0.013)	(0.046)	(0.009)
Rating B, level 3	0.951***	0.867***	0.974***	0.908***	0.983***
	(0.015)	(0.031)	(0.013)	(0.045)	(0.009)
Rating CNoGRade, level 3	0.928***	0.832***	0.973***	0.869***	0.964***
	(0.018)	(0.038)	(0.014)	(0.059)	(0.013)
Rating A, level 2	0.951***	0.868***	0.976***	0.916***	0.979***
	(0.015)	(0.031)	(0.013)	(0.046)	(0.009)
Rating B, level 2	0.950***	0.870***	0.975***	0.905***	0.979***
	(0.015)	(0.030)	(0.013)	(0.046)	(0.009)
Rating CNoGRade, level 2	0.945***	0.863***	0.971***	0.900***	0.976***
	(0.016)	(0.032)	(0.014)	(0.048)	(0.012)
N	2374	2347	2362	2312	2354
r2	0.999	0.994	0.999	0.989	0.999

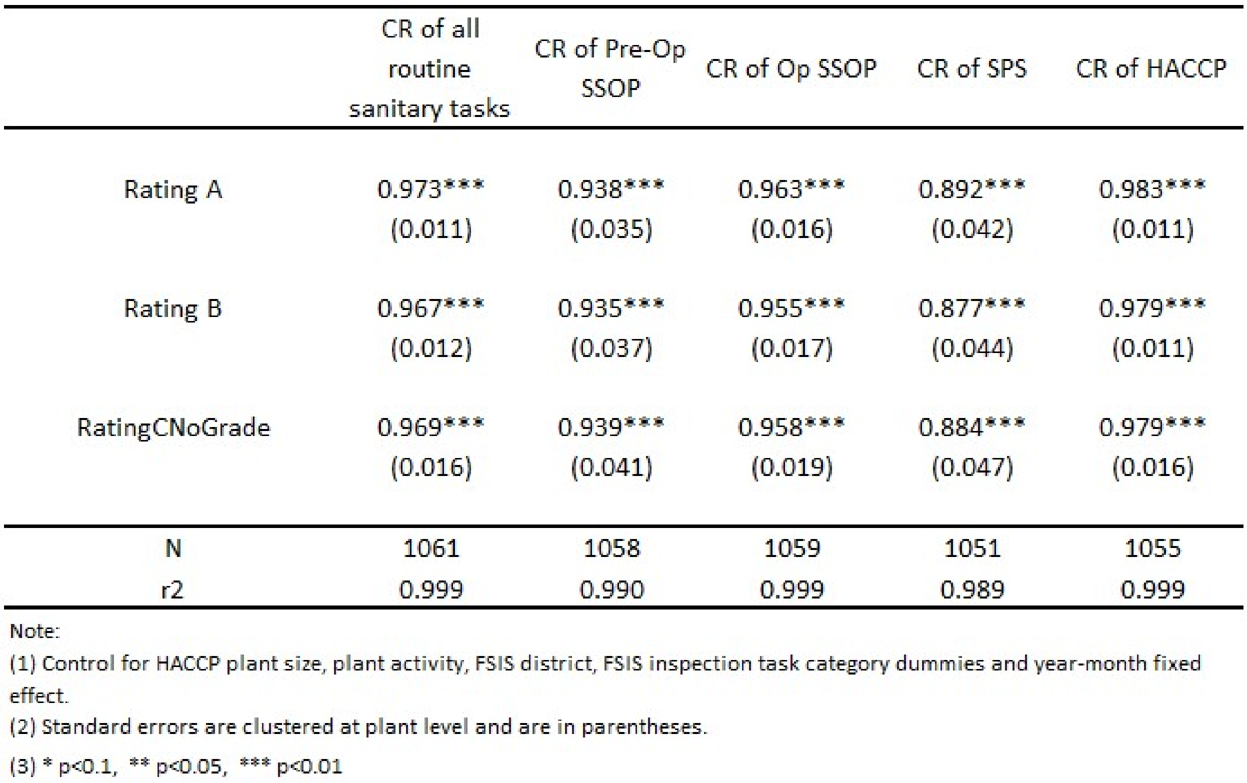
Note:

(1) Control for HACCP plant size, plant activity, FSIS district, FSIS inspection task category dummies and year-month fixed effect.

(2) Standard errors are clustered at plant level and are in parentheses.

(3) * p<0.1, ** p<0.05, *** p<0.01

Table 5.A.2: Regression result of food safety practice level on BRC rating (issue 6) only using the certification audit month data



	CR of all routine sanitary tasks	CR of Pre-Op SSOP	CR of Op SSOP	CR of SPS	CR of HACCP
Rating A	0.973***	0.938***	0.963***	0.892***	0.983***
	(0.011)	(0.035)	(0.016)	(0.042)	(0.011)
Rating B	0.967***	0.935***	0.955***	0.877***	0.979***
	(0.012)	(0.037)	(0.017)	(0.044)	(0.011)
RatingCNoGrade	0.969***	0.939***	0.958***	0.884***	0.979***
	(0.016)	(0.041)	(0.019)	(0.047)	(0.016)
N	1061	1058	1059	1051	1055
r2	0.999	0.990	0.999	0.989	0.999

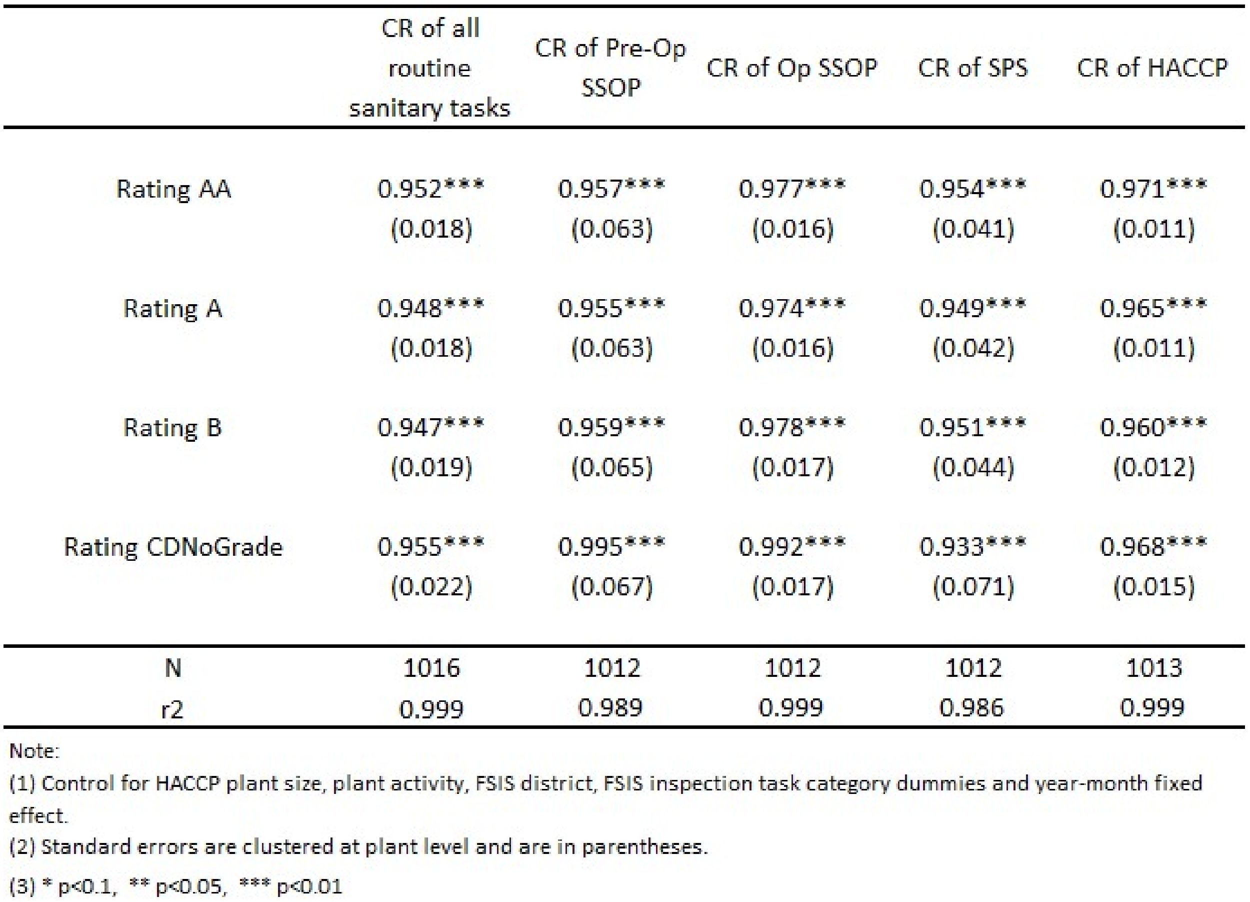
Note:

(1) Control for HACCP plant size, plant activity, FSIS district, FSIS inspection task category dummies and year-month fixed effect.

(2) Standard errors are clustered at plant level and are in parentheses.

(3) * p<0.1, ** p<0.05, *** p<0.01

Table 5.A.3: Regression result of food safety practice level on BRC rating (issue 7) only using the certification audit month data



	CR of all routine sanitary tasks	CR of Pre-Op SSOP	CR of Op SSOP	CR of SPS	CR of HACCP
Rating AA	0.952***	0.957***	0.977***	0.954***	0.971***
	(0.018)	(0.063)	(0.016)	(0.041)	(0.011)
Rating A	0.948***	0.955***	0.974***	0.949***	0.965***
	(0.018)	(0.063)	(0.016)	(0.042)	(0.011)
Rating B	0.947***	0.959***	0.978***	0.951***	0.960***
	(0.019)	(0.065)	(0.017)	(0.044)	(0.012)
Rating CDNoGrade	0.955***	0.995***	0.992***	0.933***	0.968***
	(0.022)	(0.067)	(0.017)	(0.071)	(0.015)
N	1016	1012	1012	1012	1013
r2	0.999	0.989	0.999	0.986	0.999

Note:

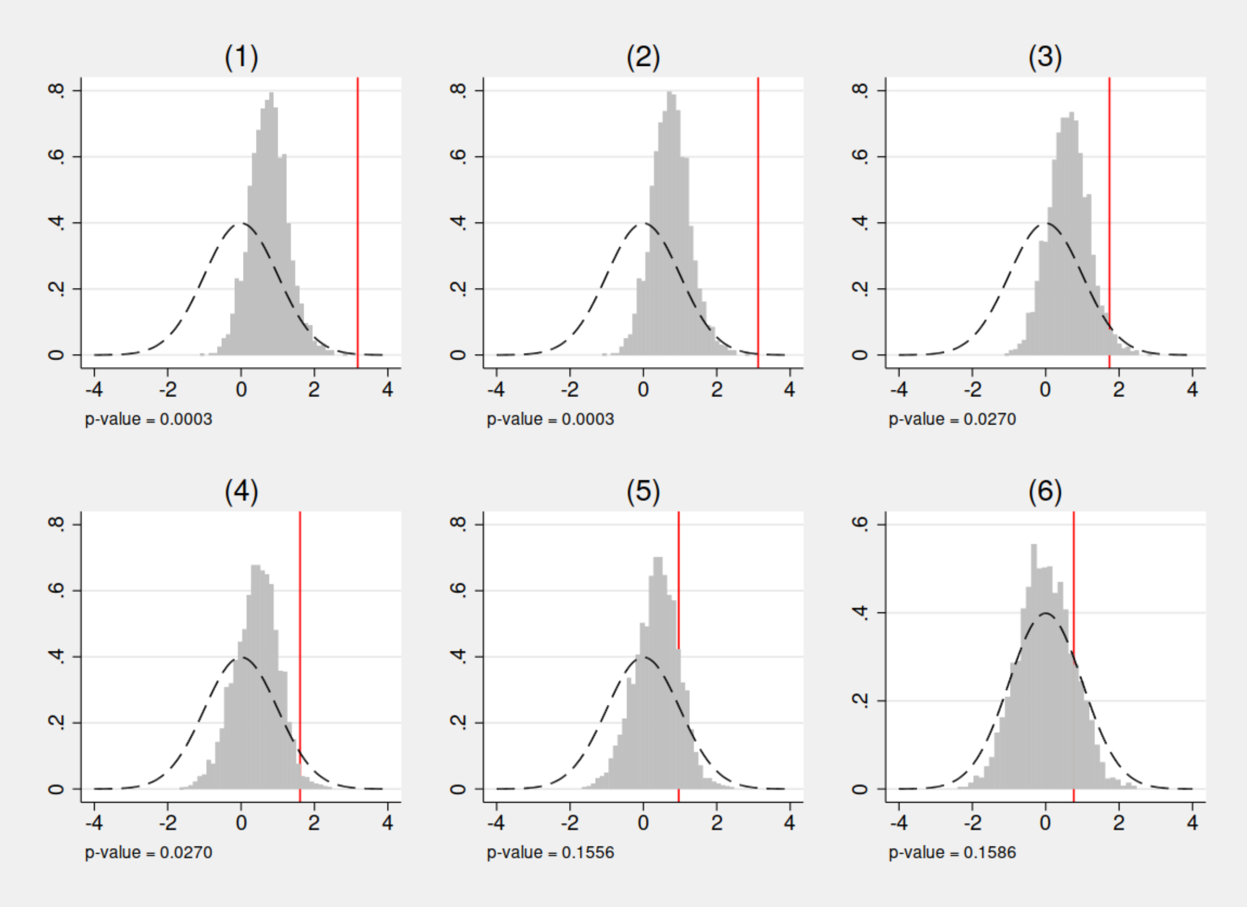
(1) Control for HACCP plant size, plant activity, FSIS district, FSIS inspection task category dummies and year-month fixed effect.

(2) Standard errors are clustered at plant level and are in parentheses.

(3) * p<0.1, ** p<0.05, *** p<0.01

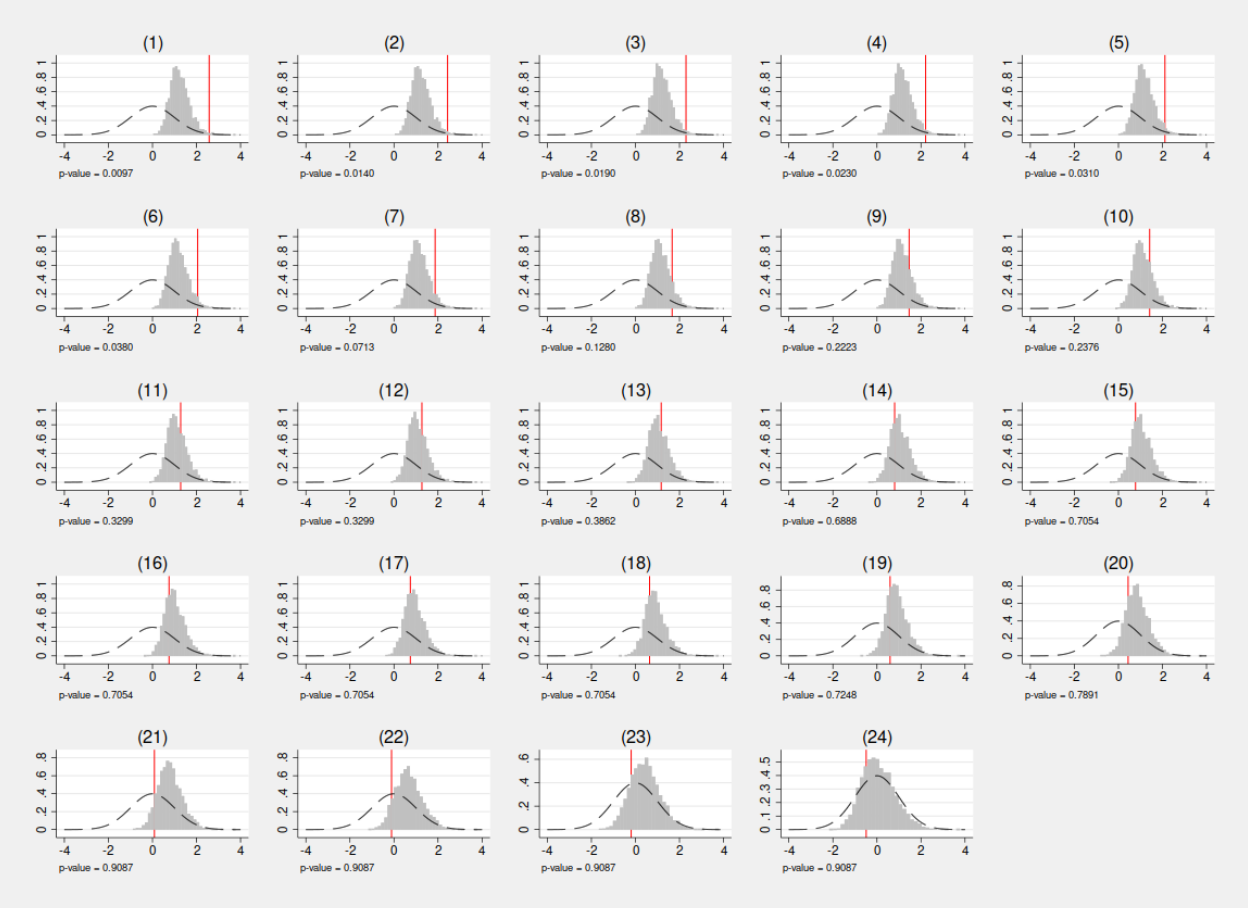
5.A.2 Details on the Romano-Wolf procedure

The following figures show the null distributions used to calculate the p-values in each one-sided hypothesis for the Romano-Wolf procedure. The black dotted line presents an exact normal distribution. The solid vertical line presents the t-statistics for each hypothesis in the hypothesis family. The empirical distribution used to calculate the corrected p-values looks more demanding than the dotted distribution for most of the figures except for Figures 5.A.6, 5.A.3, and 5.A.6, where the tail is less heavy than the corresponding standard normal distribution. This is why most of the Romano-Wolf corrected p-values are bigger than the unadjusted p-values except for the ones in the tables corresponding to the figures mentioned.



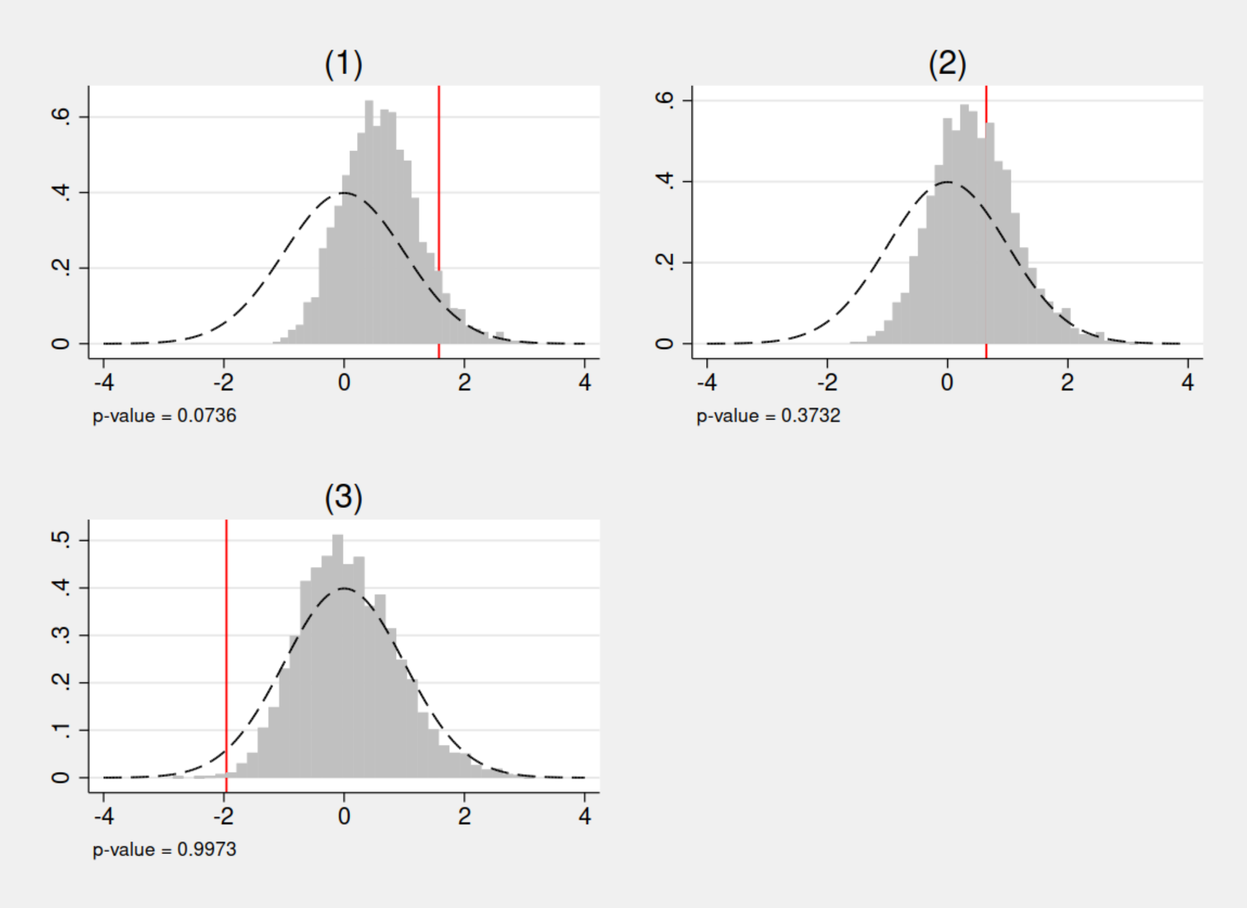
Note: SQF ratings, y_{0i}, hypothesis family 5.2, 5.3

Figure 5.A.1: Null distribution using Romano-Wolf procedure and original t-statistics for one-sided tests in Table 5.2



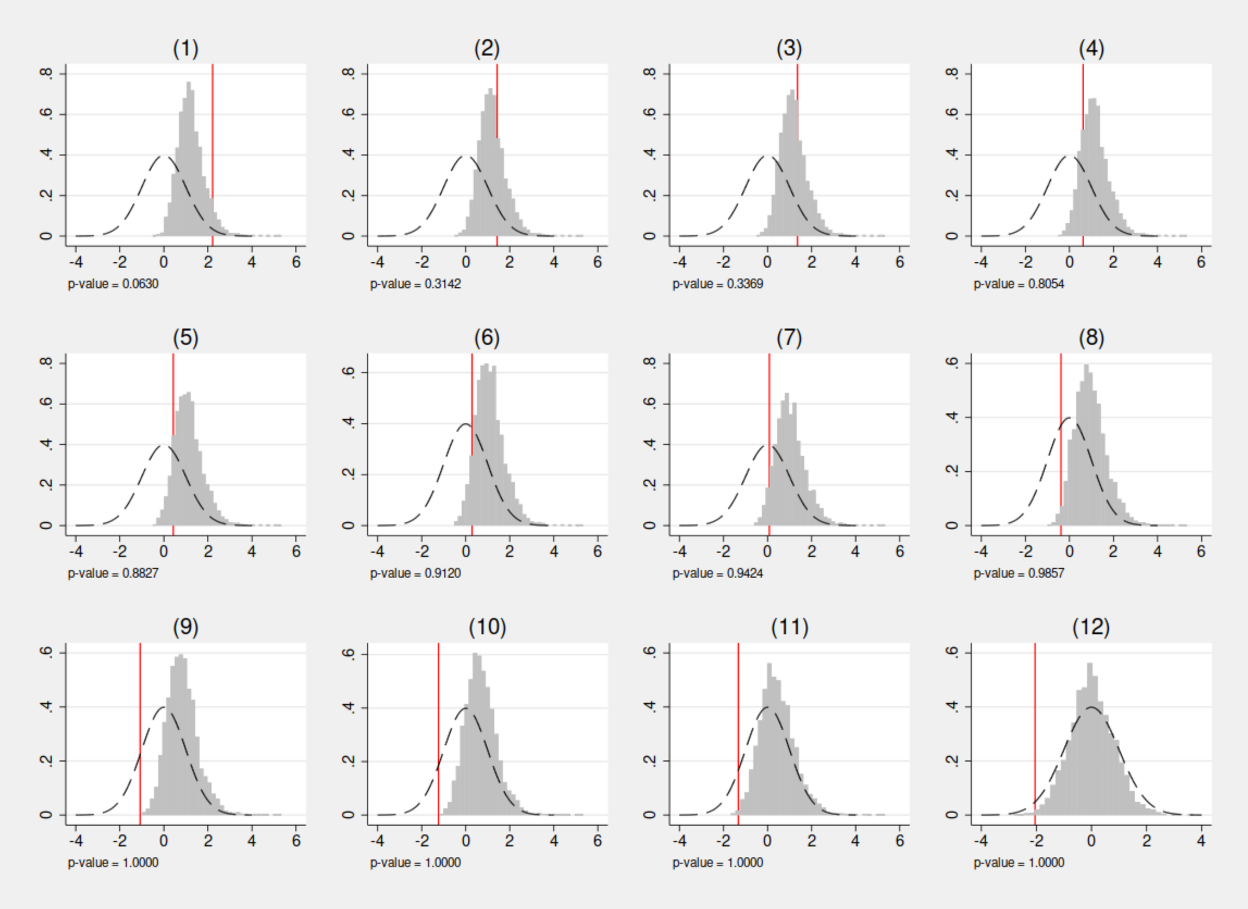
Note: SQF ratings, $y_{1i}, y_{2i}, y_{3i}, y_{4i}$, hypothesis family 5.2, 5.3

Figure 5.A.2: Null distribution using Romano-Wolf procedures and original t-statistics for one-sided tests in Table 5.3



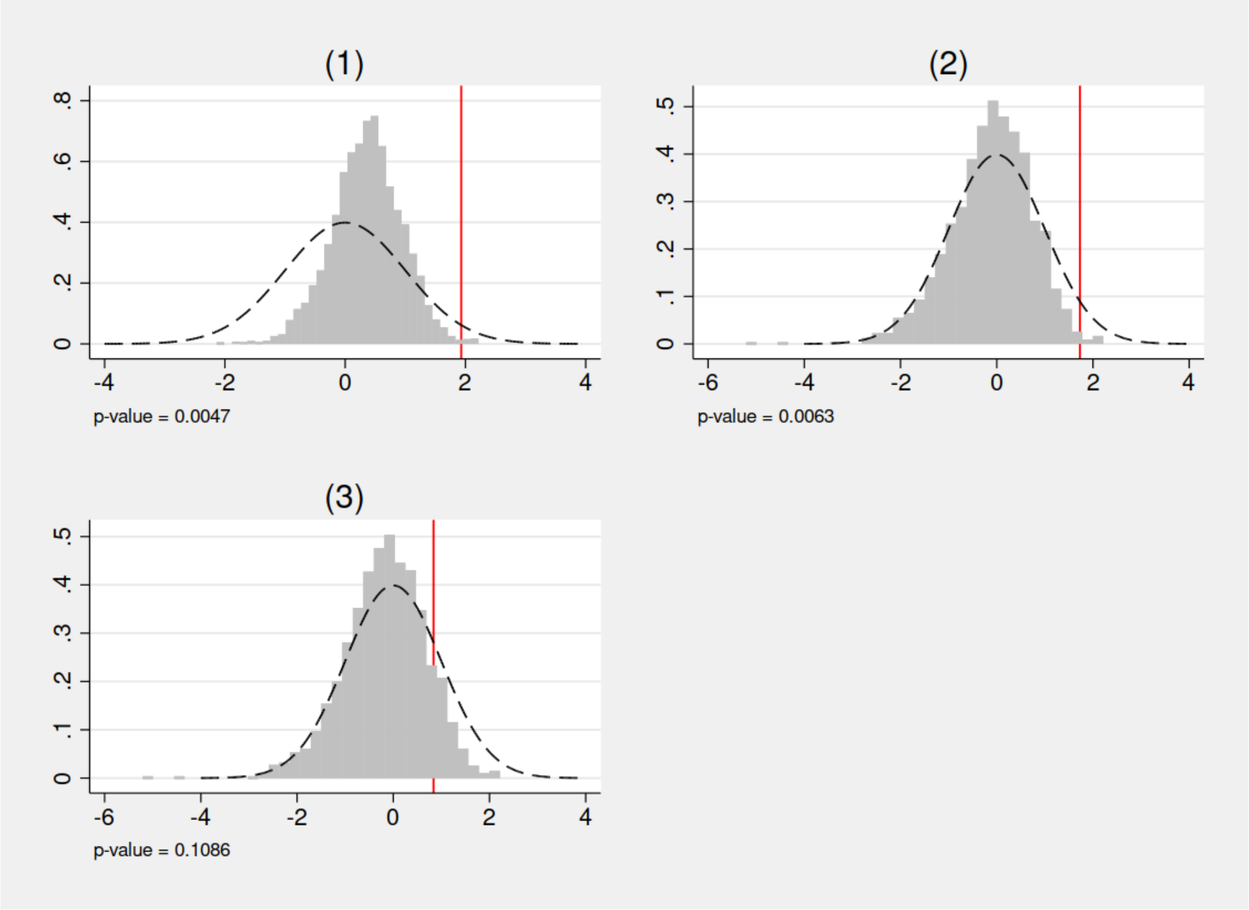
Note: SQF levels, y_{0i}, hypothesis family 5.4

Figure 5.A.3: Null distribution using Romano-Wolf procedure and original t-statistics for one-sided tests in Table 5.4



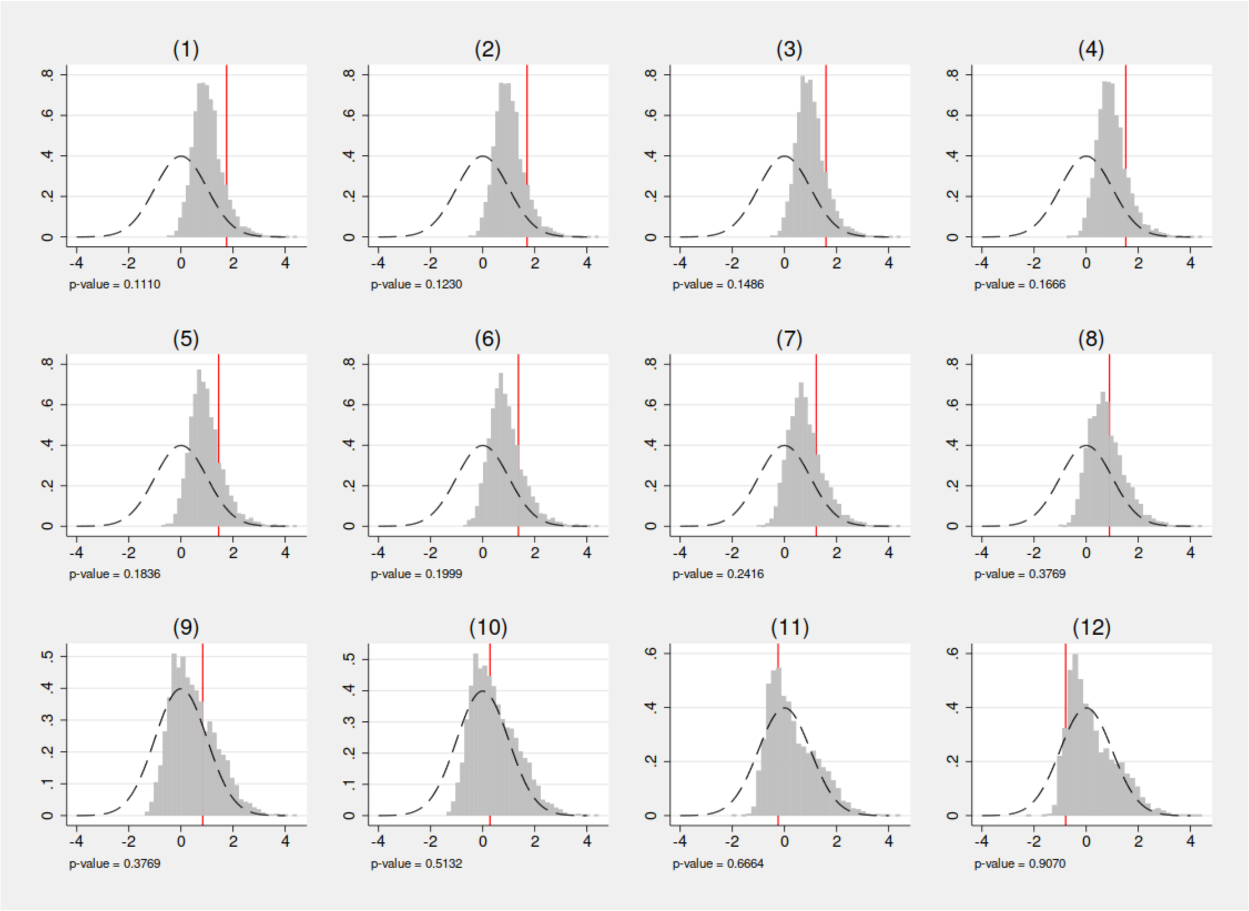
Note: SQF levels, $y_{1i}, y_{2i}, y_{3i}, y_{4i}$, hypothesis family 5.4

Figure 5.A.4: Null distribution using Romano-Wolf procedure and original t-statistics for one-sided tests in Table 5.5



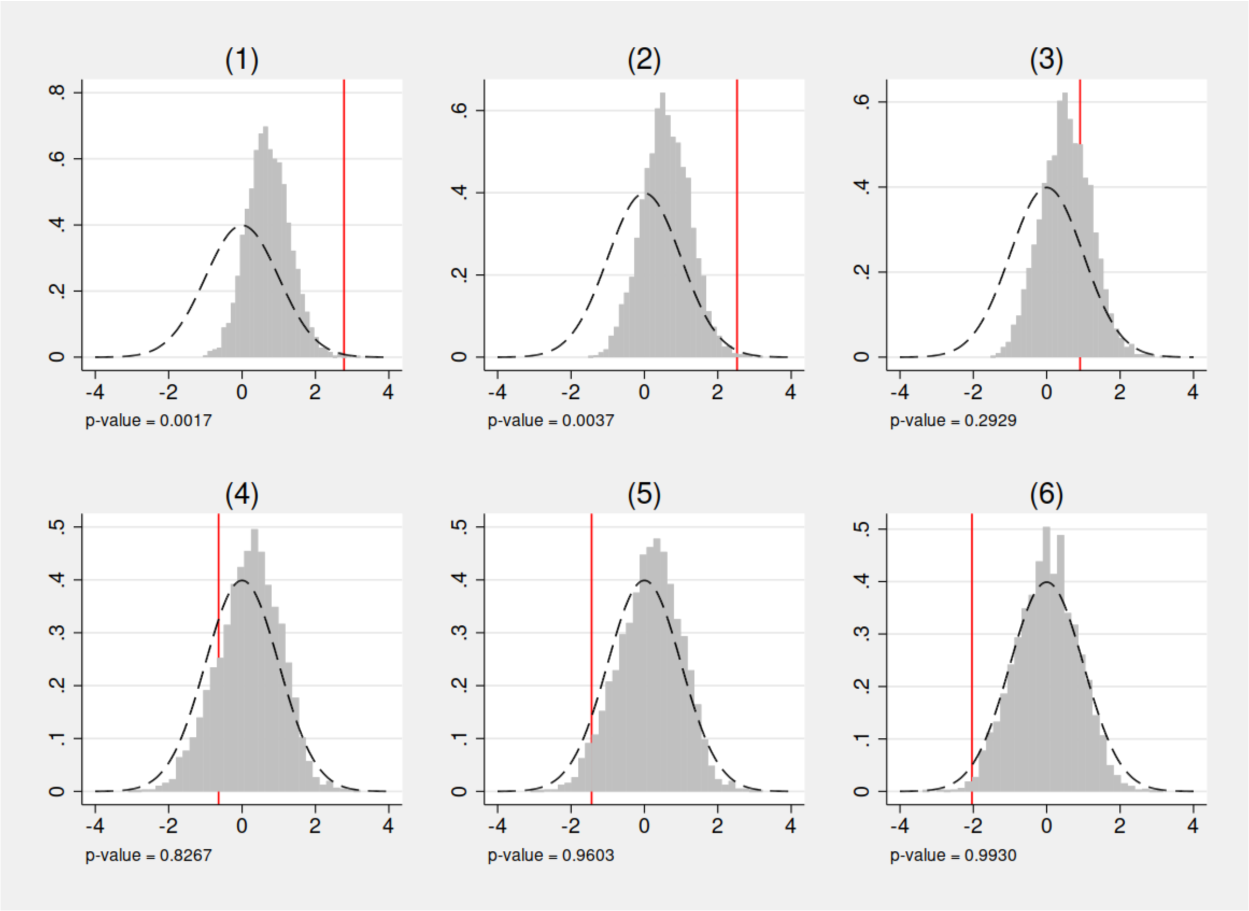
Note: BRC issue 6 ratings, y_{0i}, hypothesis family 5.6

Figure 5.A.5: Null distribution using Romano-Wolf procedure and original t-statistics for one-sided tests in Table 5.7



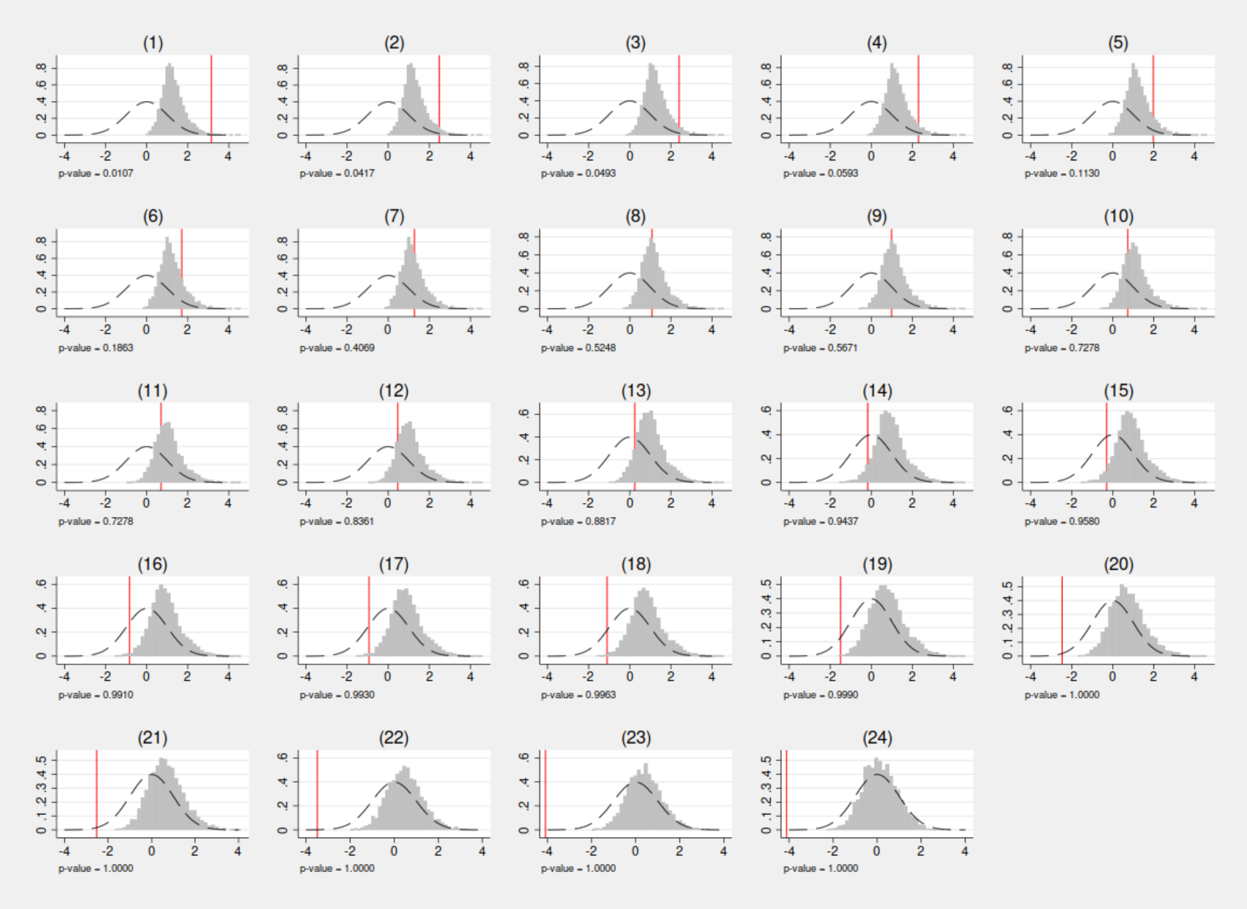
Note: BRC issue 6 ratings, $y_{1i}, y_{2i}, y_{3i}, y_{4i}$, hypothesis family 5.6

Figure 5.A.6: Null distribution using Romano-Wolf procedure and original t-statistics for one-sided tests in Table 5.8



Note: BRC issue 7 ratings, y_{0i}, hypothesis family 5.8

Figure 5.A.7: Null distribution using Romano-Wolf procedure and original t-statistics for one-sided tests in Table 5.10



Note: BRC issue 7 ratings, $y_{1i}, y_{2i}, y_{3i}, y_{4i}$, hypothesis family 5.8

Figure 5.A.8: Null distribution using Romano-Wolf procedure and original t-statistics for one-sided tests in Table 5.11

Chapter 6

What is the effect of initial and re-certification audits on plants' food safety practices?

The primary objective of this chapter is to evaluate whether certification has a positive and statistically significant impact on plants' food safety practices. In other words, how do plants' food safety practices evolve due to third-party certification? I will answer the research question from two aspects:

(1) Do certified plants' food safety practices improve due to the initial certification audit?

(2) Do certified plants maintain food safety practice level when BRC/SQF third-party auditors are not present between re-certification audits?

The main research questions are analyzed using a two-way fixed effect event study approach and a balanced panel around each event (initial certification audit and re-certification audit events). Using variations in audit dates for SQF/BRC certified plants, I study whether plants improve their food safety practices around the initial audit time and maintain them afterward. Each audit that an SQF/BRC certified plant experiences is treated as a separate event/treatment. I observe that every SQF certified plant experiences one or multiple events from July 2014 to December 2019. The audit types are given in the data set as either initial audits, announced/unannounced re-certification audits, or surveillance audits (see details in 4.2). I also observe the auditing histories of BRC certified plants from January 2009 to December 2019. The type of each audit in the BRC data set is deduced from the previous audit date variable and the current audit date variable. The series of audits usually include an initial audit, where the previous audit date is empty, and multiple re-certification audits, where the previous audit date is around 12 months after the current audit date. If the previous and current audit dates are approximately 6 months apart and the previous audit grade is C or below, the audit is a surveillance audit (see the details in 4.3).

Our identification strategy is not perfect in this research setting where plants select into initial audit and experience anticipated repeated re-certification events. This research design may overrule the classic fixed effect event study assumption, no anticipation effect and parallel trend before each initial/re-certification audit event.[1] However, currently, there is no established econometrics literature on the causal impact of a repeated occurring event for different types of events. Thus, to the best I can do, I control for potential confounders and carefully construct the samples for the event study analysis, test for the parallel trend assumptions, and conduct multiple robustness checks for our results.

6.1 Sample construction for event studies

6.1.1 Initial certification audit

The samples for the initial audit analysis of SQF and BRC are constructed following the steps described below.

1. Set the pre-treatment period to be 12 months before the initial audit month and the post-treatment period to 7 months, because normally plants take around a year to

---

[1] One way to deal with the anticipation effect is to set the treatment date earlier than when the plant gets an audit. Our goal is to create a "pre-effect" period when the plant has not responded to the audit. However, to avoid the overlap of the pre-effect period of the current audit and the post-treatment period of the previous audit, I can only recode the "treatment date" at most a few months before the audit month, because the re-certification events repeat around every 12 months and the post-audit window is set to 6 months. Furthermore, this can potentially make the "pre-effect" window too short.

prepare for it and I want to test whether there is any pre-trend. I am also interested in looking at a plant's food safety practice behavior within 7 months after an initial audit event to avoid the initial audit effects being contaminated by the next re-certification audit which typically happens 11 to 13 months after the initial certification. [2] Therefore, I identify the audit months of all the initial certifications and only keep the certified plants' data where calendar dates are within the -12, +6 event window.

2. (Balancing step) Restrain the samples by keeping the initial events that have no missing values of the main variables of interest, y_{0it} routine CR of sanitary tasks, throughout the 12-month pre-audit window and 7-month after-audit window to ensure a balanced panel of treated plants in event time.

3. Drop the initial event whose next re-audit is within 10 months of the initial audit. This step excludes the initial audit followed by a surveillance audit or with an abnormal re-certification schedule.

4. Drop the initial events that are not the first observed events in the SQF sample. It could either be a data entry mistake or a special case for some plant.

5. (Adding all control units step) Combine the treated plant sample with all control plants. Control plants refer to those that are never observed being certified within our SQF & BRC certification data set.

6. (Adding only "balanced" control units step) Identify all the calendar times of the treated

---

[2] Typically, if a plant gets an above C rating at its initial audit and decides to maintain its certification, it will be re-certified between 11 to 13 months after the initial audit. A plant might respond to the subsequent re-certification event earlier than the audit month. To exclude the influence of the re-certification event, I am interested in looking at up to 7 months after the initial event, which is 3 - 5 months before the next re-certification event, instead of a longer period. I can also set pre-treatment to be longer such as 18 months or shorter such as 6 months. A longer pre-treatment period can potentially help check whether there is any anticipation effect, but the long pre-treatment will make a balanced panel around the event date lose many treated plants leading to a smaller sample size for estimation. By setting pre-treatment to 18 months and bringing forward the initial audit date, I did not find any pre-trend or anticipation effect. For robustness check, I also shorten the pre-treatment period to 6 months for comparison.

plant sample and only keep the control plants whose outcome variables exist on all of these calendar dates.

Samples for initial certification audit analyses generated from steps 1-4 will be referred to as the initial balanced sample. It will be used to create our baseline result. For robustness checks, I also prepare the following samples: (1) samples created from only step 1, 3 & 4 will be referred to as the initial unbalanced, (2) from step 1, 2 & 4 will be referred to as the initial balanced influenced sample, (3) from step 1-6 will be referred as the initial balanced + control balanced, and (4) from step 1 - 5 will be referred to as initial balanced + control unbalanced.

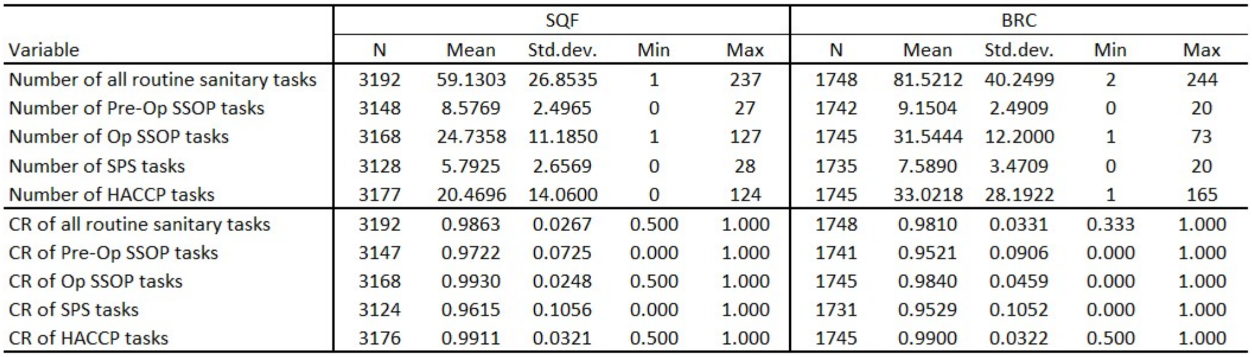
Our main sample is the initial balanced sample. It is a balanced panel in terms of the main outcome of interest CR of all sanitary tasks at the establishment and event time level from event time -12 to 6 with 0 representing when a plant gets initially audited. For SQF initial balanced sample, 168 plants/initial events were left after the sampling construction steps. For BRC, 92 plants/initial events are used for the initial certification analysis. As we can see from Table 6.1b, the BRC initial balanced sample has a higher percentage of large plants or poultry (PP, PS+PP) plants than the SQF initial balanced sample. Table 6.1a shows that, in general, plants in the BRC initial certification balanced sample on average have a higher number of inspection tasks per month and lower CRs than plants in the SQF initial certification balanced sample.

6.1.2 Re-certification audit

To keep their certification status, all certified plants need to be re-audited according to the rules shown in Table 2.3 & 2.4. One way to deal with this situation is to regard each re-audit as a separate single event and assume plants react to each re-audit similarly and independently, and the effect stays constant after a short period window. Then, we can turn the situation where certified plants experience multiple subsequent re-audits into a classic single-event study setting. However, it is worth noting that we need to carefully construct samples to avoid using the same observation multiple times to estimate the subsequent

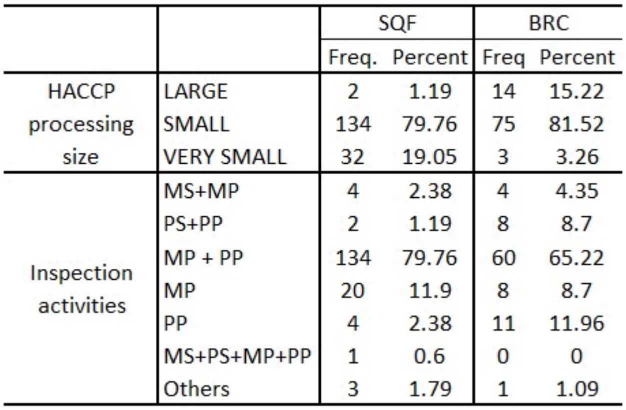
Table 6.1: Summary statistics of the SQF & BRC initial certification balanced sample

(a) FSIS inspection tasks



	SQF					BRC				
Variable	N	Mean	Std.dev.	Min	Max	N	Mean	Std.dev.	Min	Max
Number of all routine sanitary tasks	3192	59.1303	26.8535	1	237	1748	81.5212	40.2499	2	244
Number of Pre-Op SSOP tasks	3148	8.5769	2.4965	0	27	1742	9.1504	2.4909	0	20
Number of Op SSOP tasks	3168	24.7358	11.1850	1	127	1745	31.5444	12.2000	1	73
Number of SPS tasks	3128	5.7925	2.6569	0	28	1735	7.5890	3.4709	0	20
Number of HACCP tasks	3177	20.4696	14.0600	0	124	1745	33.0218	28.1922	1	165
CR of all routine sanitary tasks	3192	0.9863	0.0267	0.500	1.000	1748	0.9810	0.0331	0.333	1.000
CR of Pre-Op SSOP tasks	3147	0.9722	0.0725	0.000	1.000	1741	0.9521	0.0906	0.000	1.000
CR of Op SSOP tasks	3168	0.9930	0.0248	0.500	1.000	1745	0.9840	0.0459	0.000	1.000
CR of SPS tasks	3124	0.9615	0.1056	0.000	1.000	1731	0.9529	0.1052	0.000	1.000
CR of HACCP tasks	3176	0.9911	0.0321	0.500	1.000	1745	0.9900	0.0322	0.500	1.000

(b) Plant characteristics



		SQF		BRC	
		Freq.	Percent	Freq	Percent
HACCP processing size	LARGE	2	1.19	14	15.22
	SMALL	134	79.76	75	81.52
	VERY SMALL	32	19.05	3	3.26
Inspection activities	MS+MP	4	2.38	4	4.35
	PS+PP	2	1.19	8	8.7
	MP + PP	134	79.76	60	65.22
	MP	20	11.9	8	8.7
	PP	4	2.38	11	11.96
	MS+PS+MP+PP	1	0.6	0	0
	Others	3	1.79	1	1.09

Note: MS = Meat Slaughter, MP = Meat Process, PS = Poultry Slaughter, PP= Poultry Process multiple re-audit events of a certified plant. Specifically, a post-event observation of the first re-audit can enter as a pre-event observation of the second re-audit if the pre- and post-windows are too long. Therefore, the effect of surveillance audits on plants' food safety practice behavior will be hard to identify using this method because surveillance audits typically occur around 5 to 7 months away from the previous audits. To identify a clean effect of a surveillance audit alone, we need a pre-event window that does not overlap with the post-event window of the previous audit (initial/re-certification). As a result, we have to have a very short post-event window of the initial/re-certification audit and a very short pre-event window of the surveillance audit itself, such as 2 months respectively. Together with the small number of surveillance audits (less than 4% of re-audits for SQF certification and less than 1.5% for BRC), it is hard to estimate the effects of surveillance audits with precision and without assumptions that the re-audit effect is instant and without anticipation. Thus, given our context, it is preferred to focus on the plants that have never experienced surveillance audits, but only have a sequence of re-certification audits in our data set and interpret our results with a smaller scope as the re-certification effect of certified plants with above C ratings (A/B for SQF, AA/A/B for BRC).

I construct the re-certification audit sample for analysis based on the key rules discussed above by following the steps below:

1. Set the pre- and post-event window to be 5 months. [3] For SQF certified plants, drop the observations before the year-month November 2014 (= July 2014 + post-event window - 1). The reason is that we can only observe the SQF audit information since July 2014, so it is possible that the data before November 2014 is within a post-event window of an audit that we do not observe.

2. Drop the plants whose audit schedules have two subsequent audits within 10 months of each other because these plants may behave differently from plants that have a typical

---

[3] To avoid window overlap of two adjacent re-certification audits, which are typically 11-13 months apart, a 5-month window is the longest we can select. For robustness checks, we can shorten the window.

annual re-certification schedule, and it will cause the overlap of the post-event window of the current event and the pre-event window of the next event with event window set as -5. +5.

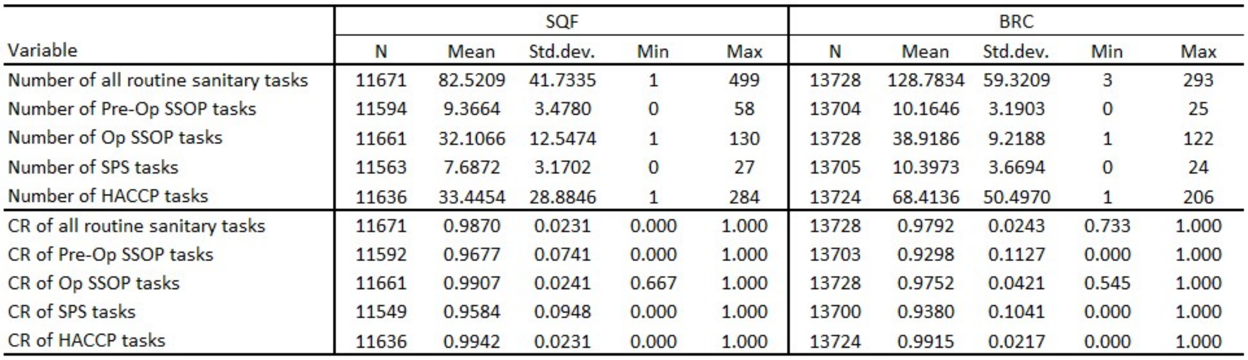
3. Identify the audit months of all the re-certification audits and only keep the certified plants' data where calendar dates are within the -5, +5 event window.
4. (Balancing step) Restrain samples by keeping re-certification events that have no missing values of the main variables of interest, y_{0it} routine CR of sanitary tasks, during the entire 5-month pre-audit window and 6-month after-audit window to ensure a balanced panel of treated plants in event time.
5. (Adding all control units step) Combine the treated plant sample with all control plants. Control plants refer to those that are never observed certified within our SQF & BRC certification data set.
6. (Step of adding only "balanced" control units) Identify all the calendar times of the treated plant sample and only keep the control plants whose outcome variables exist on all of these calendar dates.

We name the baseline sample created from steps 1-4 re-certification balanced, from steps 1-3 re-certification unbalanced, from steps 1-6 re-certification balanced + control balanced, and from steps 1-5 re-certification balanced + control.

Our main sample for analysis is the re-certification balanced sample. It is a balanced panel in terms of the main outcome of interest CR of all sanitary tasks at the establishment and re-certification event time level from event time -5 to 5 with 0 representing when a plant gets a re-certification audit. For SQF re-certification balanced sample, 547 unique plants and 1061 re-certification audit events are used for the re-certification analysis. For BRC, 332 unique plants and 1248 re-certification audit events are used for the analysis. The reason why BRC has a larger number of re-certification audit events is that we only observe SQF audit records since July 2014, but observe a longer period of audit records for BRC.

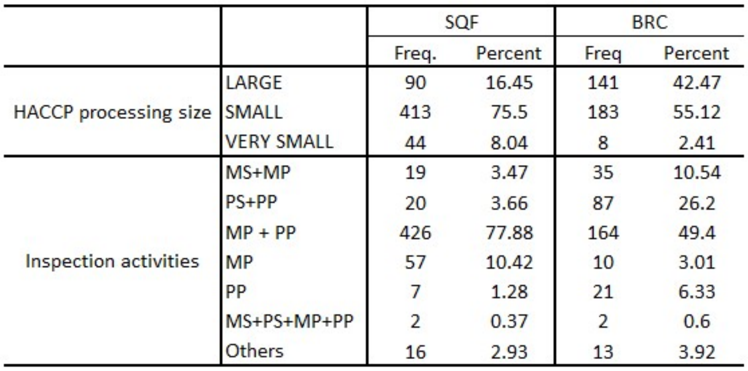
Table 6.2: Summary statistics of the SQF & BRC re-certification balanced sample

(a) FSIS inspection tasks



	SQF					BRC				
Variable	N	Mean	Std.dev.	Min	Max	N	Mean	Std.dev.	Min	Max
Number of all routine sanitary tasks	11671	82.5209	41.7335	1	499	13728	128.7834	59.3209	3	293
Number of Pre-Op SSOP tasks	11594	9.3664	3.4780	0	58	13704	10.1646	3.1903	0	25
Number of Op SSOP tasks	11661	32.1066	12.5474	1	130	13728	38.9186	9.2188	1	122
Number of SPS tasks	11563	7.6872	3.1702	0	27	13705	10.3973	3.6694	0	24
Number of HACCP tasks	11636	33.4454	28.8846	1	284	13724	68.4136	50.4970	1	206
CR of all routine sanitary tasks	11671	0.9870	0.0231	0.000	1.000	13728	0.9792	0.0243	0.733	1.000
CR of Pre-Op SSOP tasks	11592	0.9677	0.0741	0.000	1.000	13703	0.9298	0.1127	0.000	1.000
CR of Op SSOP tasks	11661	0.9907	0.0241	0.667	1.000	13728	0.9752	0.0421	0.545	1.000
CR of SPS tasks	11549	0.9584	0.0948	0.000	1.000	13700	0.9380	0.1041	0.000	1.000
CR of HACCP tasks	11636	0.9942	0.0231	0.000	1.000	13724	0.9915	0.0217	0.000	1.000

(b) Plant characteristics



		SQF		BRC	
		Freq.	Percent	Freq	Percent
HACCP processing size	LARGE	90	16.45	141	42.47
	SMALL	413	75.5	183	55.12
	VERY SMALL	44	8.04	8	2.41
Inspection activities	MS+MP	19	3.47	35	10.54
	PS+PP	20	3.66	87	26.2
	MP + PP	426	77.88	164	49.4
	MP	57	10.42	10	3.01
	PP	7	1.28	21	6.33
	MS+PS+MP+PP	2	0.37	2	0.6
	Others	16	2.93	13	3.92

Note: MS = Meat Slaughter, MP = Meat Process, PS = Poultry Slaughter, PP= Poultry Process

Comparing the sample compositions of the SQF and BRC re-certification balanced sample, I found similar patterns as the comparison between SQF and BRC initial balanced sample. From Table 6.2, we can see that the BRC sample has a larger percentage of large-size plants or poultry (PP, PS+PP) plants. In general, plants in the BRC sample on average have a higher number of inspection tasks per month and lower CRs than plants from SQF re-certification balanced sample.

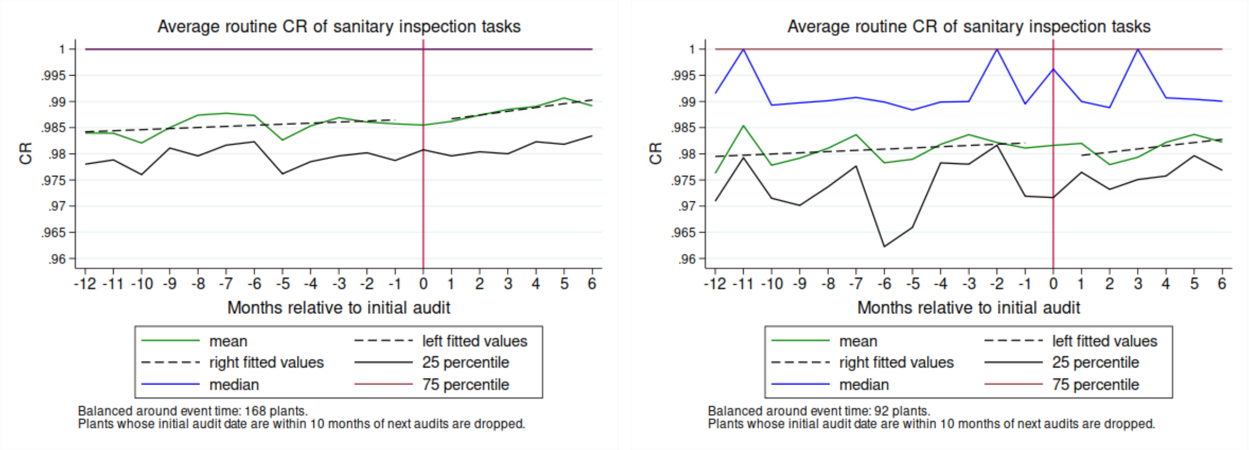
6.2 Raw evidence on the effect of initial certification & re-certification on plants' food safety practice

Before I turn to the formal analysis, it is helpful to look at the patterns of plants' routine sanitary task CR around initial certification and re-certification audits respectively in the raw data. The calculation is based on the initial and re-certification balanced samples for the subsequent event study analyses constructed in section 6.1. Similar figures of the initial and re-certification unbalanced samples are displayed in Appendix 6.A.1 and 6.A.2, which show similar pattern as described below.

Figure 6.1 depicts the mean, 25, 50, and 75 percentiles of FSIS routine sanitary task CRs of certified plants in each month relative to their initial audit months for both SQF and BRC initial certification. Subfigures 6.1a and 6.1b are graphed with the same scale for easy comparison of the trend and magnitude of CR evolution among SQF and BRC initial and re-certification audits. Sub-figures 6.1c and 6.1d are graphed at different scales, which are the zoomed-in version of subfigures 6.1a and 6.1b. They provide a closer look at the micro patterns of CR evolution around event time.

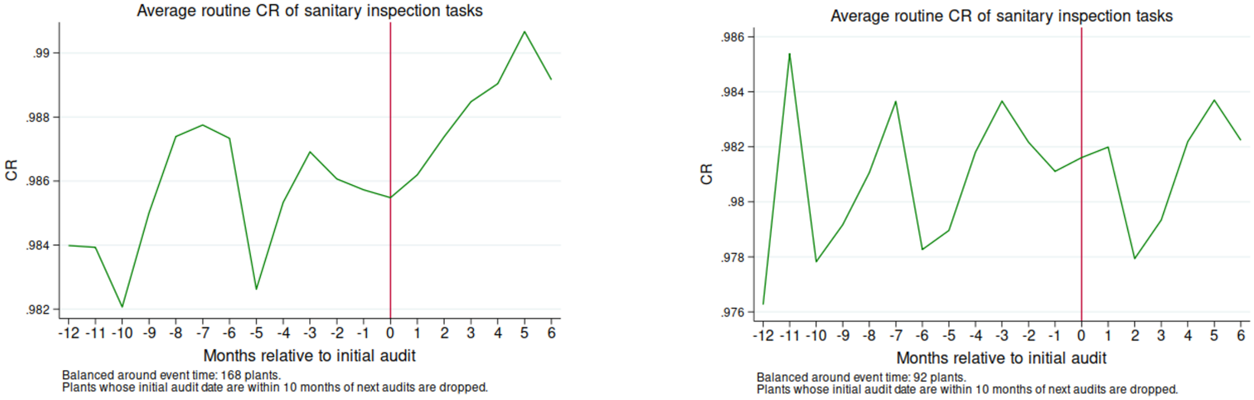
As we can see in Figure 6.1, there is basically no pre-trend before the initial audit for both SQF and BRC. For SQF, there is a slight increase in the average and 25 percentile of routine sanitary inspection CR after the initial audits; however, for BRC, there is no obvious trend or change after the initial audits. Also, from the summary statistics in 6.1a, we observe that, on average, BRC certified plants have a lower CR of sanitary inspection tasks.

Similarly, Figure 6.2 shows the mean and percentiles of sanitary task CRs of certified plants in each month relative to their re-certification audit months. Compared to the initial certification figures, the variations in the average routine CR of sanitary inspection tasks across re-certification audit event time are smaller. It possibly indicates that the impact of re-certification audits on plants' food safety practices measured by CR is very limited if there is any. When looking at the zoomed-in Figures (sub-figures 6.2c and 6.2d) of the mean evolution of FSIS routine sanitary CR around re-certification events, interestingly, we observe that there is a micro pattern of CR generally increasing before the re-certification audits and decreasing after the re-certification audits for both SQF and BRC certification. However, the magnitudes of the CR increase from event time -5 to 0 and decrease from event time 0 to 5 are very small (+0.13% and -0.0004% for SQF; +0.04% and -0.11% for BRC).



(a) SQF, initial audits (same scale)

(b) BRC, initial audits (same scale)



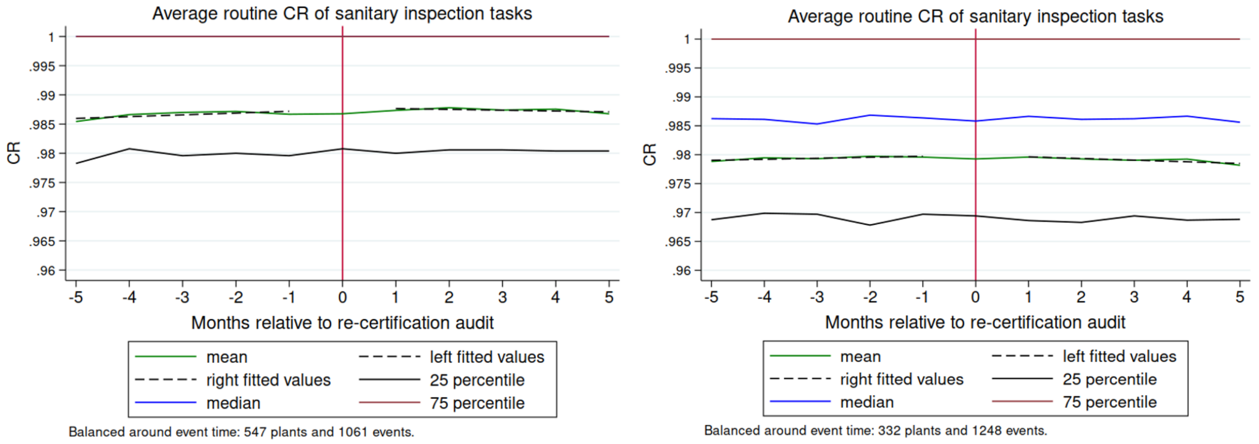
(c) SQF, initial audits (zoom in)

(d) BRC, initial (zoom in)

Figure 6.1: Average and percentiles of routine sanitary inspection CR before and after initial certification audits

Therefore, the raw evidence indicates a possibility that plants might temporarily improve food safety practices around the months of re-certification but the overall impact on CR is very small.

Overall, the raw evidence is only suggestive. In the next section, we will use econometrics to estimate and test the effect of initial certification and re-certification audits on plants' food safety practices statistically.



(a) SQF, re-certification audits (same scale) (b) BRC, re-certification audits (same scale)

(c) SQF, re-certification audits (zoom in) (d) BRC, re-certification audits (zoom in)

Figure 6.2: Average and percentiles of routine sanitary inspection CR before and after re-certification audits

6.3 Two-way fixed effect event study estimators

6.3.1 Initial certification audit

In this section, I focus on examining the changes in food safety practices that occur around initial certification audits by third-party certifiers. It is important to note that the status and timing of the initial certification may not be randomly assigned to establishments. Lacking a clean random research design, the best we can do in this context is to exploit the variation in the timing of initial audits among certified plants after controlling for possible confounding factors including time-invariant plant characteristics, seasonality, and common policy shocks.

From the raw evidence, we do not observe an obvious pre-trend before the initial audit for either SQF or BRC initial audit, which is somewhat comforting for the no anticipation effect. Therefore, formally, I use the baseline event study regression (6.1) and the SQF/BRC initial certification balanced sample constructed in Appendix 6.1 to study the effects of the SQF and BRC initial certification audit on food safety practices.

$$y_{it} = \sum_{j=-12, j\neq -1}^{6} \alpha_j int_{it}^{j} + \gamma_i + \lambda_t + \epsilon_{it} \tag{6.1}$$

where i denotes an establishment, t denotes a year-month, y_{it} is the FSIS routine inspection outcomes of establishment i at year-month t (CRs), int_{it}^{j} is a dummy variable that equals to 1 when establishment i is j month away from its initial certification month int_i at the year-month t ($int_{it}^{j} = 1\{t = int_i + j\}$, $j = -12, -11, ...6$), λ_t is time-fixed effect, γ_i is plant-fixed effect. I compute robust standard errors allowing for the correlation of errors over time but not between individual plants. Considering that the noise might be correlated among plants within the same FSIS circuit, I also clustered the standard error to the circuit level as a robustness check.

The SQF/BRC initial certification balanced sample contains all the certified plants that have a balanced CR of sanitary routine tasks around the initial audit event time. Because there is no consensus on samples in the event study analysis, I also re-estimated the same regression using different versions of the baseline sample, such as the initial certification unbalanced sample, and a sample that includes never certified plants as controls. Details on the construction of different versions of samples are described in Appendix 6.1. The main coefficients of interests are α_j ($j = -12, -11, ...6$). α_{-1} is normalized to zero to avoid multicollinearity issues.

However, it is important to consider the subtle differences in assumptions and interpretations when using only the treated units (plants with SQL/BRC initial certification audits) versus using both treated and control units (plants without SQF/BRC certification) in the estimation sample. When only treated units are included, the "not-yet-treated" plants serve as comparison groups, with the key identification assumption being the parallel trend of not-yet-treated and treated plants and no anticipation effect. In this case, α_j (with $j \neq -1$) is interpreted as the CR difference between the initial certification event time j and one month before the initial certification event, after controlling for plant and time fixed effects. On the other hand, when both treated and control plants are included in the estimation sample, both "never treated" and "not-yet treated" are used as comparison groups for the treated plants. This approach relies on the extra identification assumption of parallel trend between the control and treated units. Therefore, α_j (with $j \neq -1$) captures the dynamic effect of the initial certification audit events relative to an average of treated units in period -1 before treatment and non-treated units in all periods.

To test the parallel trend assumption and no anticipation effect, we can plot the estimates of the pre-event dummies and test whether each of them is significantly different from zero. we can also conduct the joint hypothesis 6.2. Failure to reject the null hypothesis gives some support to our identification assumptions.

$$H_0 : \alpha_{-12} = \alpha_{-11} = ... = \alpha_{-2} = 0 \tag{6.2}$$

$\alpha_0, \alpha_1, ...\alpha_6$ show how the average FSIS routine CR of certified plants evolves compared to the month before the initial certification audit. A priori, the effect is uncertain. It is possible that we observe an increase in CR after initial certification because initial certification could help plants correct their non-conformities, which improves the CR. It is also plausible that we will observe no effects because certification might only serve as a signal for a plant's food safety practices with no effect on plants' behaviors, or the positive changes in the plant's food safety condition are not reflected in the FSIS CR measurements. The patterns of the post-event coefficients will help identify the dynamic effect of the initial certification audits. Also, we can test whether the effect after the event is overall different from zero by conducting the joint hypothesis 6.3.

$$H_0 : \alpha_0 = \alpha_1 = ... = \alpha_6 = 0 \tag{6.3}$$

6.3.2 Re-certification audit

I estimate how certified plants behave around re-certification audits following a similar method explained in the above section. Since each certified plant's approximate re-certification date is set based on the initial certified date (see the timeline of re-certifications in Figure 2.3), plants have less control over when they get re-certification audits once the initial certification audit is done. This lends more credibility to the exogeneity of the re-certification audit dates. However, since the re-certification audit is expected to happen annually for these certified plants, plants can change their behaviors before the actual auditing months. Therefore, relying on an assumption of parallel trends in CR had the plants not been re-certified at that month, formally, I used the event study regression (6.4) and the baseline sample, re-certification balanced sample constructed in Appendix 6.1.2, to study the effects of re-certification audit on food safety practice behaviors. The re-certification balanced sample contains all the re-certification audit events of plants with a balanced sanitary CR around the re-certification audit event. The parameters of interest βs capture the effect of the re-certification audit event relative to the treated units in event time -1. For robustness checks, I also estimated the results using various versions of the baseline sample, such as samples including the never-treated establishments. The details of sample construction are documented in Appendix 6.1.2. When control units (plants without SQF/BRC certification) are included in the estimation sample, the parameters of interest βs reflect the effect relative to an average of treated units in event time -1 before treatment and non-treated units in all periods.

The re-certification audit event study regression is at the establishment (i) - reaudit event

(k) - year-month (t) level.

$$y_{ikt} = \sum_{j=-5, j \neq -1}^{5} \beta_j rec_{ikt}^{j} + \gamma_i + \lambda_t + \epsilon_{ikt} \tag{6.4}$$

where i denotes an establishment, t denotes a year-month, y_{it} is the FSIS routine inspection outcomes of establishment i at year-month t (CRs), rec_{ikt}^{j} is a dummy variable that equals to 1 when establishment i's kth re-certification audit is j month away from the establishment i's kth re-certification audit month rec_{ik} at the year-month t ($rec_{ikt}^{j} = 1\{t = rec_{ik} + j\}$, $j = -5, -4, ...5$), λ_t is time-fixed effect, γ_i is plant-fixed effect. The robust standard error is calculated which allows for the correlation of errors over time but not between individual plants. For robustness check, standard errors are also clustered at the circuit level to allow correlation among plants located in the same FSIS circuit.

The main parameters of interest are βs. The patterns of the coefficients indicate how plants react to the re-certification audit. There are multiple reasonable scenarios. First, it is possible that plants "perform" on certain tasks to prepare for the re-certification audit such as thorough cleaning on the premises that they would not have done normally. This could lead to the "pre-event parameters" β_j, $j < 0$ to be on the negative side and slowly increase, and the β_j, $j >= 0$ slowly drops if FSIS inspection task CR can successfully reflect the food safety practice changes from the plants. Second, it is also plausible that each re-certification audit can help plants improve their food safety practice fundamentally and increase their average CR. In this case, we might observe the β_j, $j >= 0$ is positive and slowly increases. Last but not least, we might not see a significant change of βs at all. If the FSIS inspection CR fails to capture the minute improvements made in plants' food safety practices, then even if plants "perform" better during the re-certification audit, our data may not reflect this improvement. If there is not much wiggle room for plants to "perform" temporarily to improve food safety level after the food safety system is set up, then we also do not expect the βs to fluctuate much.

To formally examine the most plausible cases for the re-certification audit effect, we can estimate βs, observe the pattern of point estimates and conduct pairwise hypothesis test against $\beta_j = 0$, $j = -5, -4, ... + 5$. Additionally, we can also conduct the joint hypotheses (6.5) and (6.6) to help understand whether there is any overall before and after effect of the re-certification audit.

$$H_0 : \beta_{-5} = \beta_{-4} = ... = \beta_{-2} = 0 \tag{6.5}$$

$$H_0 : \beta_0 = \beta_1 = ... = \beta_5 = 0 \tag{6.6}$$

6.4 Results and discussion

The main results for initial and re-certification audits are presented graphically in Figures 6.3 and 6.6: only SQF certified plants have a gradual improvement in sanitary CR after initial certification; there is no significant impact of re-certification audits and BRC initial certification audit on plants' food safety practices. The main results are robust to various specifications and sample constructions.

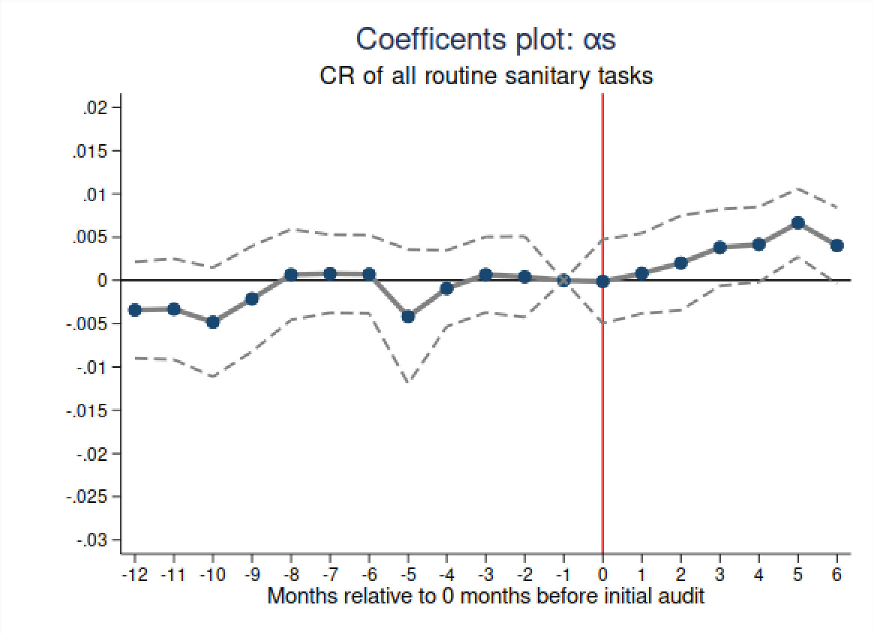
6.4.1 Initial certification audit

The baseline results of the initial certification audit effects (αs) on plants' food safety practices measured by CR of all routine sanitary tasks are shown in Table 6.5 and Table 6.6 column (1) respectively. Both of the tables are divided into 3 sections: the first top section reports the point estimate and standard error of $\alpha_j, j = 0, 1, ...6$, the middle section reports the number of observations, R^2, and the p-values of 2 joint hypotheses 6.2 and 6.3, and the bottom section reports the sample and regression version used. To intuitively look at the results, Figure 6.3a & Figure 6.3b graphically plot the αs, dynamic effect of SQF and BRC initial certification audit on the average food safety practice levels of SQF & BRC certified plants respectively.

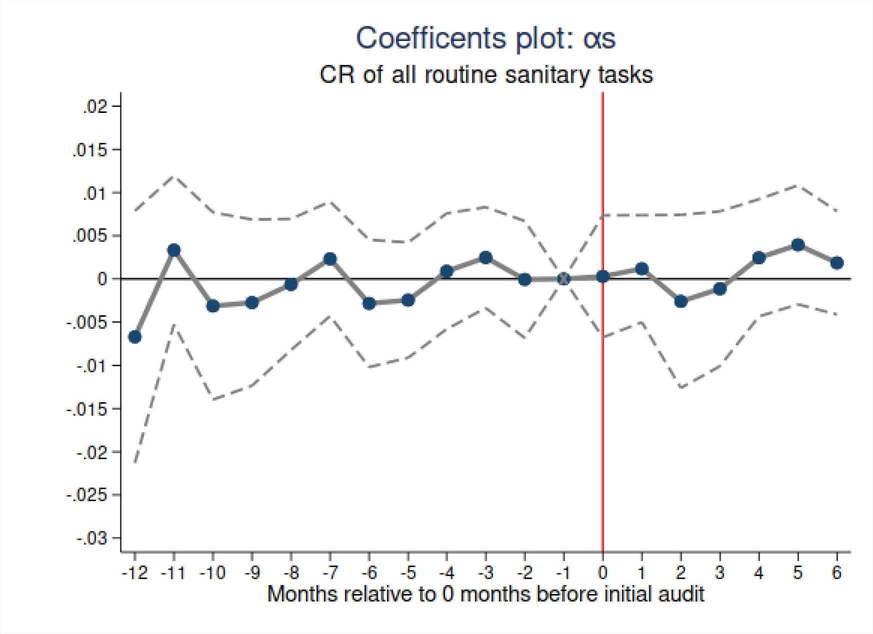
We observe no effect of initial audits on a plant's CR overall routine sanitary tasks before the initial event. I show in Figure 6.3a that each of the coefficients before audit events is not significantly different from zero, and so does the joint hypothesis (see the p-values of "pre-event coefficients all zero" in Table 6.5). The after-event point estimates of SQF initial certification effect are gradually increasing up to 5 months post the initial certification event and dropped slightly at month 6 after the initial audit. All the post-event point estimates are larger than those before the initial audits. It indicates that the effect of treatment could accumulate over time, so that the point estimate of α_j, $j = 0, 1, ...5$ increases in j. From Table 6.5 column (1), we can observe that the point estimates of $\alpha_3, \alpha_4,$ and α_6 are significantly different from zero at 10% significance level, and the only point estimate of α_5 is significantly different from zero at 5% level. It implies that certified SQF plants can slowly improve their food safety practice over time after initial certification. After 5 months of initial certification, on average, routine sanitary CR increases by 0.66% compared to the month before an initial audit. To understand the magnitude of this increase in CR (decrease in non-compliance rate), I calculated the average CR of routine sanitary tasks one month before the initial audit, which is 98.57%. Thus, the non-compliance rate (NCR) is about 1.43% (= 1- 98.57%). A 0.66% decrease in NCR is 46% of the base NCR during the month before the initial audit.

To further investigate what specific routine FSIS inspection tasks might drive improvement in overall routine sanitary task CR, I rerun the analysis with the CR of 4 specific inspection tasks, CR of Pre-Op SSOP, Op SSOP, SPS, and HACCP. The results are reported in Table 6.3 from column (2) to column (5), and coefficients αs are plotted in Figure 6.7. None of the pre-event estimates of the dummies are significantly different from zero with 95% confidence, though the joint hypothesis (6.2) with outcome variable Pre-Op SSOP CR is rejected. Thus, in general, I find no evidence violating the the parallel trend assumption and found no anticipation effect. For after-event coefficients, the post-event point estimates are in general on an increasing trend up to event month 5 for Pre-Op SSOP CR, OP SSOP CR and SPS CR though the coefficients are noisy. Specifically, α_5 is significantly different from zero with CR of Pre-OP SSOP, and SSOP as outcome variables at 5% significance level,

(a) SQF, initial certification

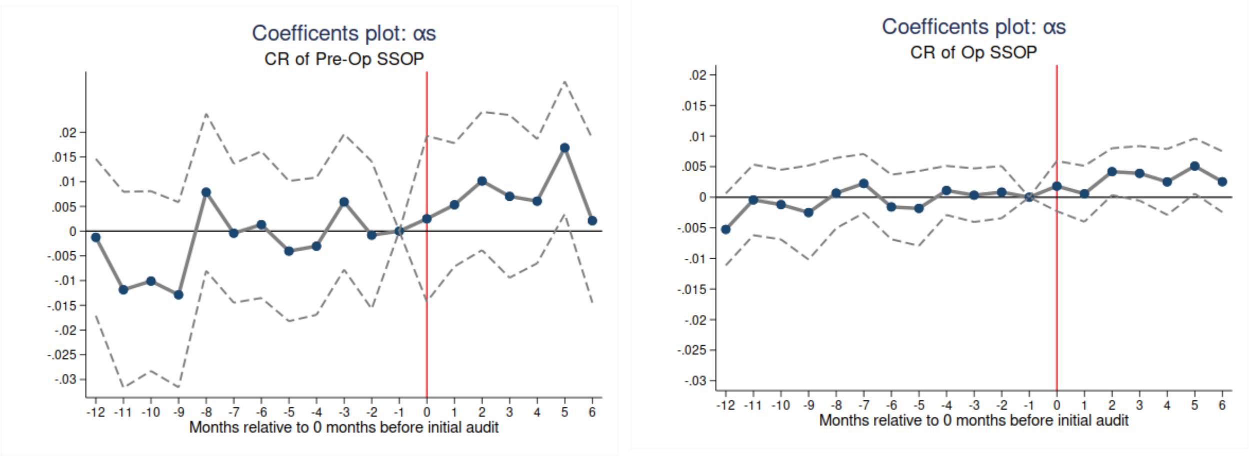


(b) BRC, initial certification



Note: the result is based on fixed effect estimator from regression 6.1 and SQF/BRC initial certification balanced sample. Blue dots represent the point estimates of pre- and post-event dummies. The dotted lines represents the 95% confidence interval of each point estimate.

Figure 6.3: Effect of SQF & BRC initial certification on plant's overall food safety practice level

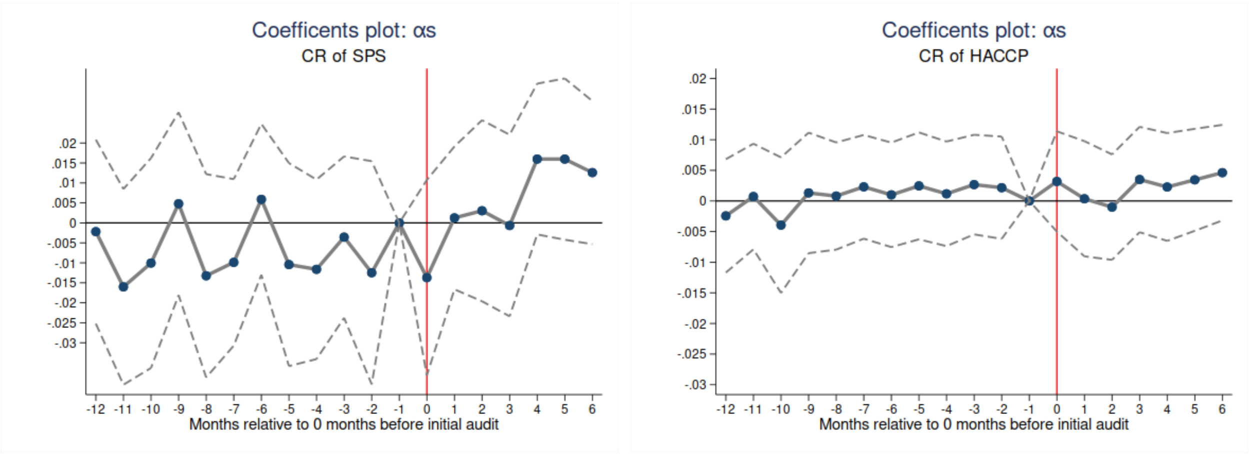


(a) Pre-Op SSOP

(b) Op SSOP

(c) SPS

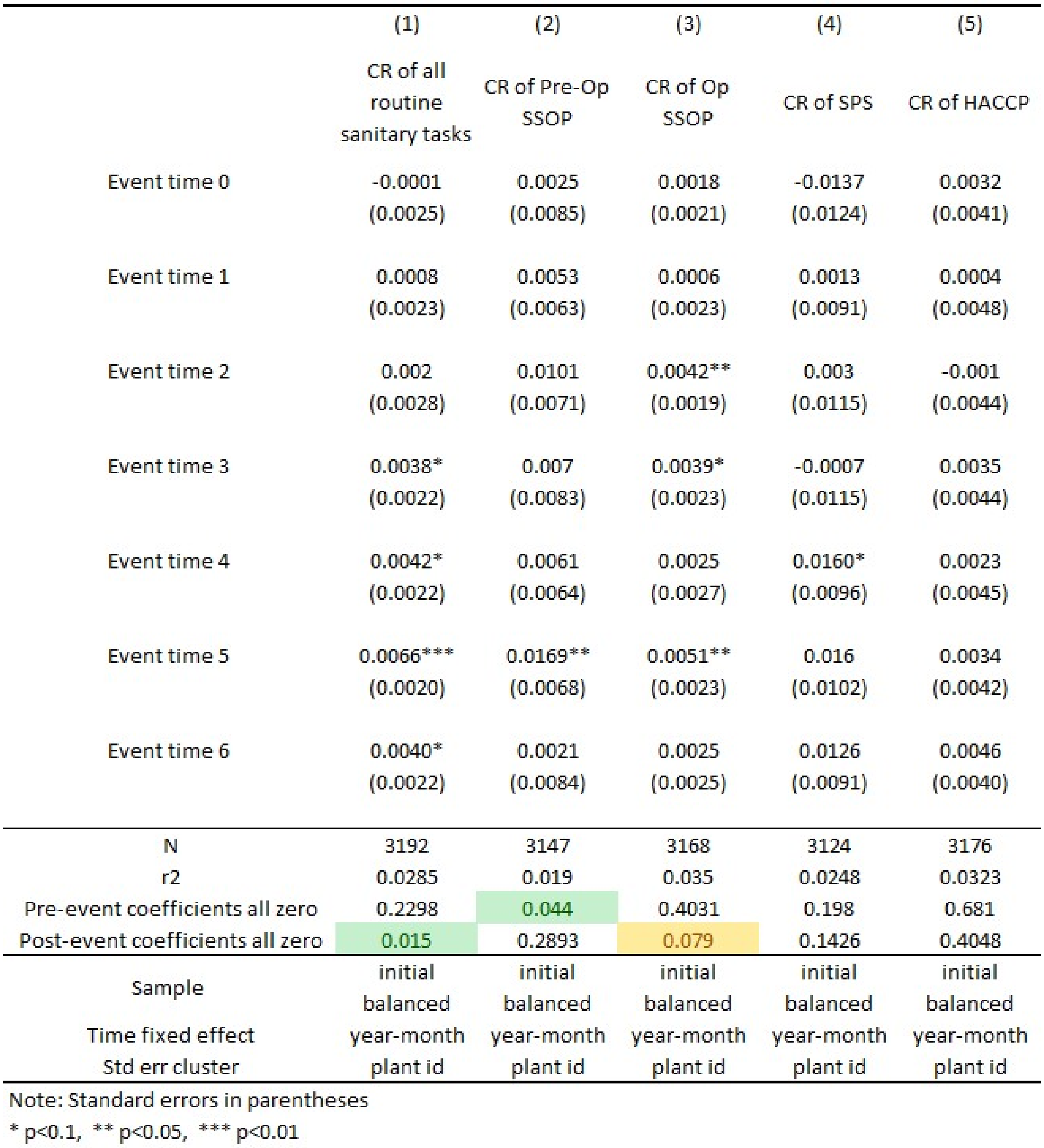
(d) HACCP



Note: the results are based on fixed effect estimator from regression 6.1 using different outcome variables as indicated in the caption of the sub-figures and SQF initial certification balanced sample. Blue dots represent the point estimates of pre- and post-event dummies. The dotted lines represents the 95% confidence interval of each point estimate.

Figure 6.4: Effect of SQF initial certification on plant's food safety practice level from 4 different aspects

Table 6.3: Regression results of SQF initial audit effect on CR of different routine inspection tasks



	(1) CR of all routine sanitary tasks	(2) CR of Pre-Op SSOP	(3) CR of Op SSOP	(4) CR of SPS	(5) CR of HACCP
Event time 0	-0.0001 (0.0025)	0.0025 (0.0085)	0.0018 (0.0021)	-0.0137 (0.0124)	0.0032 (0.0041)
Event time 1	0.0008 (0.0023)	0.0053 (0.0063)	0.0006 (0.0023)	0.0013 (0.0091)	0.0004 (0.0048)
Event time 2	0.002 (0.0028)	0.0101 (0.0071)	0.0042** (0.0019)	0.003 (0.0115)	-0.001 (0.0044)
Event time 3	0.0038* (0.0022)	0.007 (0.0083)	0.0039* (0.0023)	-0.0007 (0.0115)	0.0035 (0.0044)
Event time 4	0.0042* (0.0022)	0.0061 (0.0064)	0.0025 (0.0027)	0.0160* (0.0096)	0.0023 (0.0045)
Event time 5	0.0066*** (0.0020)	0.0169** (0.0068)	0.0051** (0.0023)	0.016 (0.0102)	0.0034 (0.0042)
Event time 6	0.0040* (0.0022)	0.0021 (0.0084)	0.0025 (0.0025)	0.0126 (0.0091)	0.0046 (0.0040)
N	3192	3147	3168	3124	3176
r2	0.0285	0.019	0.035	0.0248	0.0323
Pre-event coefficients all zero	0.2298	0.044	0.4031	0.198	0.681
Post-event coefficients all zero	0.015	0.2893	0.079	0.1426	0.4048
Sample	initial balanced	initial balanced	initial balanced	initial balanced	initial balanced
Time fixed effect	year-month	year-month	year-month	year-month	year-month
Std err cluster	plant id	plant id	plant id	plant id	plant id

Note: Standard errors in parentheses
* p<0.1, ** p<0.05, *** p<0.01

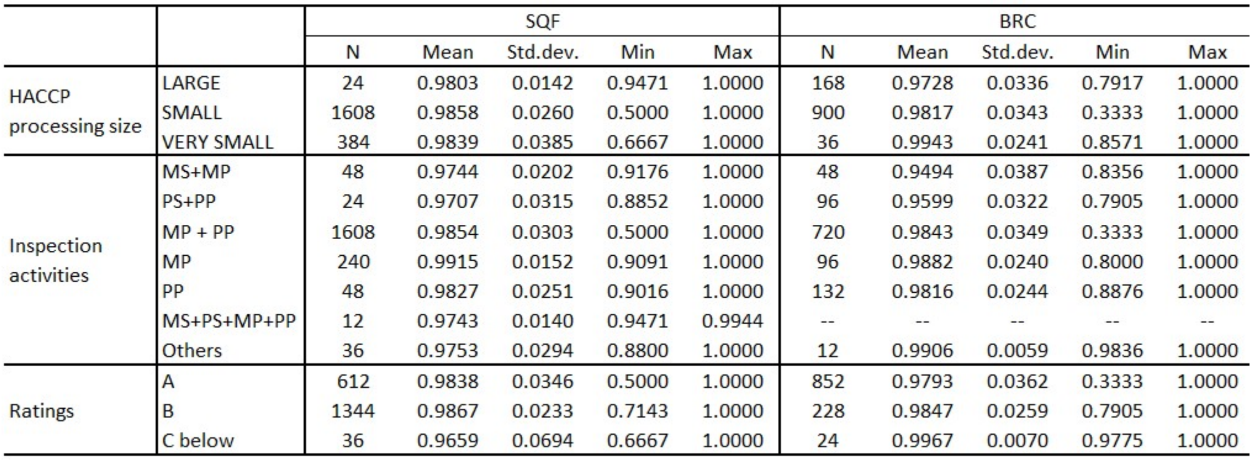
and α_4 is significantly different from zero for SPS CR at 10% significance level. However, I did not see a significant difference in HACCP CR around initial certification. Therefore, the results indicate that on average SQF certified plants may show improvement in SSOP and SPS tasks after the initial certification audit.

In contrast, for BRC certification, Figure 6.3b shows that all the point estimates of coefficients αs are not significantly different from zero at 10% significance level and I did not observe an increasing trend of point estimates after the initial certification event. Thus, there is no supporting evidence that BRC initial certification has a significant impact on plants' overall sanitary CR.

There are multiple reasons why I may observe a different average initial audit effect for SQF & BRC certified plants. First, it is possible that some differences in SQF and BRC certification codes and procedures might cause differences in estimation results. As we can see from Table 2.3 and 2.4, BRC has a more lenient rule to get an A above rating compared to SQF because BRC allows for a higher number of minor non-conformities for plants to get A and above ratings. As a result, it is not surprising for us to observe in Figure 4.3 that the proportion of top ratings in the BRC certification is much higher than in the SQF. However, when comparing the average CR of the BRC and SQF certified plants before the initial certification audit in the initial balanced sample restricted to either the same size or production activities or ratings, in general, I observe that the BRC certified plants have a lower CR than the SQF certified plants (see Table 6.4). Thus, due to the less stringent rules to get good grades for BRC certification, it is possible that those selected for BRC certification might only use it as a signal tool to get an easier top rating and BRC certification procedures will not have an improvement effect for these plants.

Second, unlike what was stated above, it may not matter whether the certification is SQF or BRC per se. It is the plant's heterogeneous response to the initial certification audit that causes the estimation differences for SQF and BRC. The plant compositions of SQF and BRC initial certification balanced samples are quite different and plants of different characteristics may respond to initial audit differently regardless of SQF or BRC certification. As shown in Table 6.1, the BRC sample has a bigger portion of large size or poultry plants compared to SQF. Therefore, it could be the heterogeneous response among different types of plants (HACCP size and activity) and sample composition that lead to the different estimated average treatment effects for SQF and BRC. I can potentially test this by restricting both

Table 6.4: 12 month average CR of all routine sanitary tasks in SQF/BRC initial balanced sample before the initial audit



		SQF					BRC				
		N	Mean	Std.dev.	Min	Max	N	Mean	Std.dev.	Min	Max
HACCP processing size	LARGE	24	0.9803	0.0142	0.9471	1.0000	168	0.9728	0.0336	0.7917	1.0000
	SMALL	1608	0.9858	0.0260	0.5000	1.0000	900	0.9817	0.0343	0.3333	1.0000
	VERY SMALL	384	0.9839	0.0385	0.6667	1.0000	36	0.9943	0.0241	0.8571	1.0000
Inspection activities	MS+MP	48	0.9744	0.0202	0.9176	1.0000	48	0.9494	0.0387	0.8356	1.0000
	PS+PP	24	0.9707	0.0315	0.8852	1.0000	96	0.9599	0.0322	0.7905	1.0000
	MP + PP	1608	0.9854	0.0303	0.5000	1.0000	720	0.9843	0.0349	0.3333	1.0000
	MP	240	0.9915	0.0152	0.9091	1.0000	96	0.9882	0.0240	0.8000	1.0000
	PP	48	0.9827	0.0251	0.9016	1.0000	132	0.9816	0.0244	0.8876	1.0000
	MS+PS+MP+PP	12	0.9743	0.0140	0.9471	0.9944	--	--	--	--	--
	Others	36	0.9753	0.0294	0.8800	1.0000	12	0.9906	0.0059	0.9836	1.0000
Ratings	A	612	0.9838	0.0346	0.5000	1.0000	852	0.9793	0.0362	0.3333	1.0000
	B	1344	0.9867	0.0233	0.7143	1.0000	228	0.9847	0.0259	0.7905	1.0000
	C below	36	0.9659	0.0694	0.6667	1.0000	24	0.9967	0.0070	0.9775	1.0000

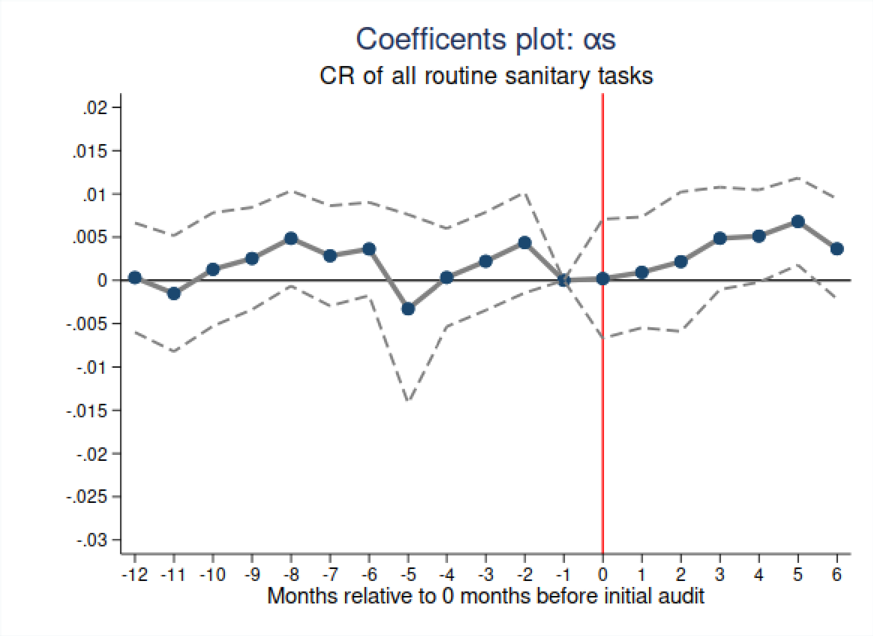
Note: MS = Meat Slaughter, MP = Meat Process, PS = Poultry Slaughter, PP= Poultry Process

SQF and BRC samples to only SMALL MP+PP plants and comparing their results. [4] As shown in Figure 6.5, after restraining the SQF & BRC samples to only SMALL MP+PP plants, for SQF initial certification, I still observe an upper trend of point estimates up to event month +5 with $\alpha_4 = 0.0051$ significantly different from zero at the 10% significance level and $\alpha_5 = 0.0068$ significantly different from zero at the 5% significance level. However, overall the point estimates of SQF after-event dummies are not bigger than the pre-event dummies anymore. For BRC, I still did not see a significant impact of initial certification on CR when I focused only on SMALL MP + PP plants. From the comparison of the results, it is hard to get conclusive answers regarding our hypothesis. Nonetheless, it seems that the difference in the estimated effect of average initial certification audit can not be fully explained by the differences in sample compositions of SQF and BRC.

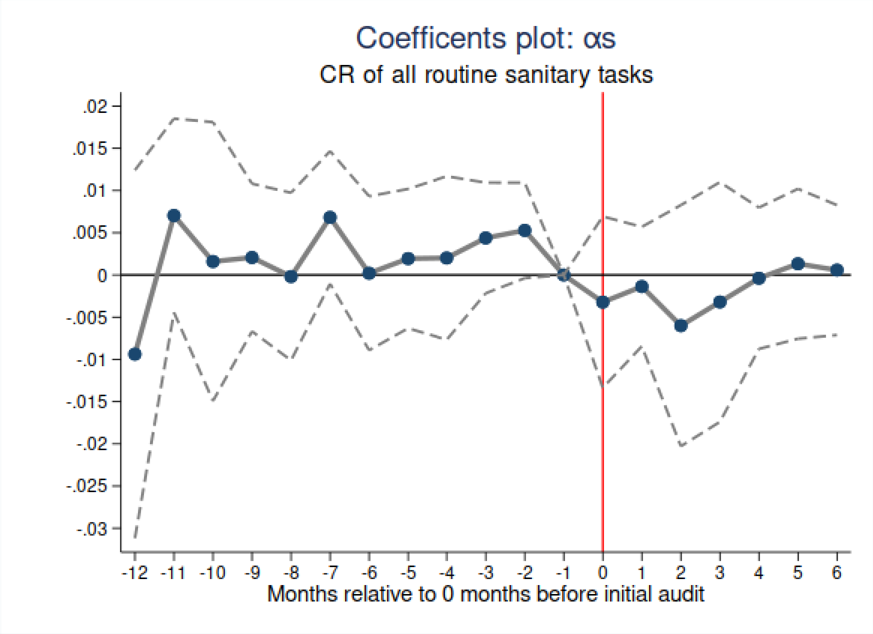
---

[4] Small size MP+PP plants are the most common type of plants in both of the samples. Other types of plants suffer from small observations in the initial audit sample of SQF or BRC or both

(a) SQF, initial certification



(b) BRC, initial certification



Note: the result is based on fixed effect estimator from regression 6.1 and SMALL MP+PP Plants in SQF/BRC initial certification balanced sample. Blue dots represent the point estimates of pre- and post-event dummies. The dotted lines represents the 95% confidence interval of each point estimate.

Figure 6.5: Effect of SQF & BRC initial certification on SMALL MP + PP plant's overall food safety practice level

6.4.2 Re-certification audit

The baseline results (regression 6.4 with re-certification audit balanced sample) of how SQF and BRC certified plants respond to re-certification audits (βs) are shown in column (1) of Table 6.8 and Table 6.9 respectively. Both of the tables are divided into 3 sections: the first top section reports the point estimate and standard error of β_j, $j = -5, -4, ...5$, the middle section reports the number of observations, R^2, and the p-values of 2 joint hypotheses 6.5 and 6.6, and the bottom section reports the sample and regression version used. To intuitively look at the results, Figure 6.6a & Figure 6.6b graphically plot the βs, dynamic effect of the SQF and BRC re-certification audit on the average levels of food safety practice for SQF & BRC certified plants.

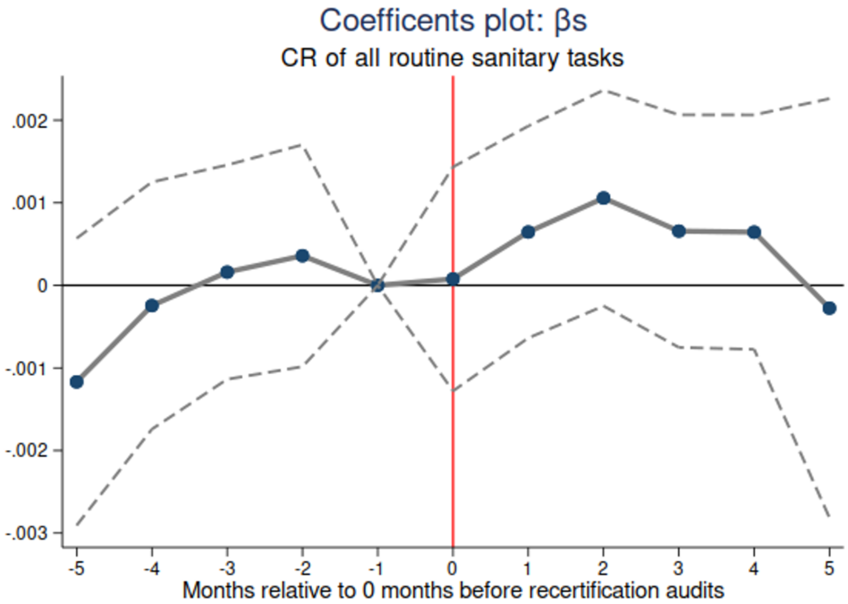
Figure 6.6a shows that there is no significantly different from zero effect of SQF re-certification audit on plant's CR overall routine sanitary tasks before or after the re-certification event at 5% (or 10%) significance level. Each of the coefficients before and after the audit event is not significantly different from zero, and so does the joint hypothesis (see the p-values of "pre-event coefficients all zero" and "post-event coefficients all zero" in Table 6.8 column (1)). Though I do observe the overall pattern of point estimates going up when the event time gets closer to the re-certification audit time and trending down when the event time is way after the re-certification audit time (i.e., at event time -5, -4, and $+5$ are below zero), the point estimates are not precisely estimated.

In Figure 6.6b, similarly, for BRC re-certification audits, I can observe the ramping up and down trends of point estimators βs but only $\beta_5 = -0.0014$ is significantly different from zero at 5% significance level. It means that compared to one month before the re-certification audit, the CR of all routine sanitary tasks 5 months before the re-certification audit is 0.14% lower, which is 6.86% of the mean NCR of all routine sanitary tasks one month before the re-certification audit (2.04%). To further investigate, what aspects of sanitary tasks might cause the ramping up and down pattern of βs, and the significant negative effect at event time -5, I plotted the estimated re-certification audit effects on CR of Pre-Op SSOP, Op

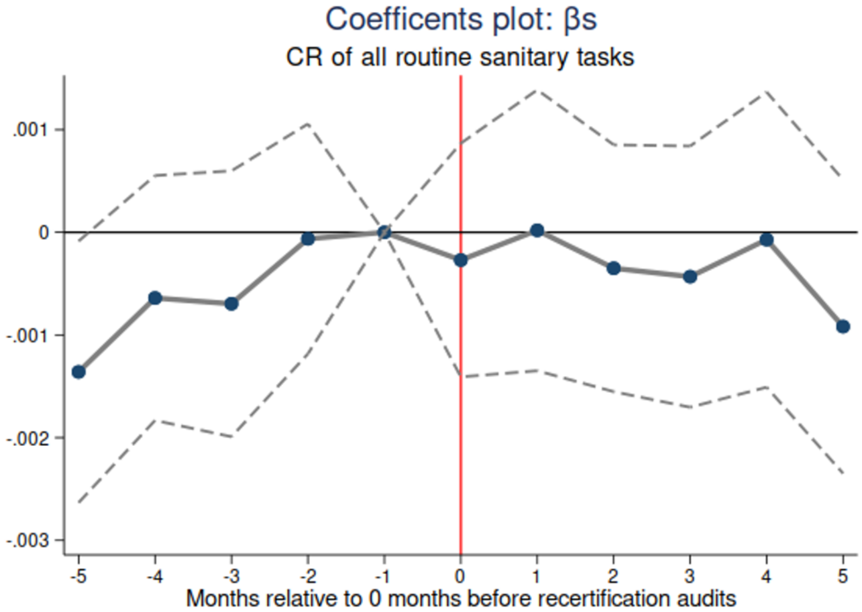
SSOP, SPS, and HACCP tasks in Figure 6.7 respectively. The point estimates of Op SSOP and HACCP task CR more or less follow the ramping up and down pattern and β_{-5} with Op SSOP CR as the outcome variable is negative and significantly different from zero at 5% significance level. Therefore, considering the overall pattern on the point estimates of re-certification audit effects and the small imprecise estimation of the effect, Some caution is warranted interpreting these results: there might be evidence of plants performing better on Op SSOP tasks only close to the re-certification, but the effect is very small and not precisely estimated.

Overall, I do not observe a large impact of re-certification audits on certified plants' CR if there is any. Though the micro pattern on the point estimates for both SQF and BRC re-certification audit effect may indicate a slight ramping up and down CR of overall routine sanitary tasks before and after the re-certification audits, they are estimated with a lot of noise.

(a) SQF, re-certification audits



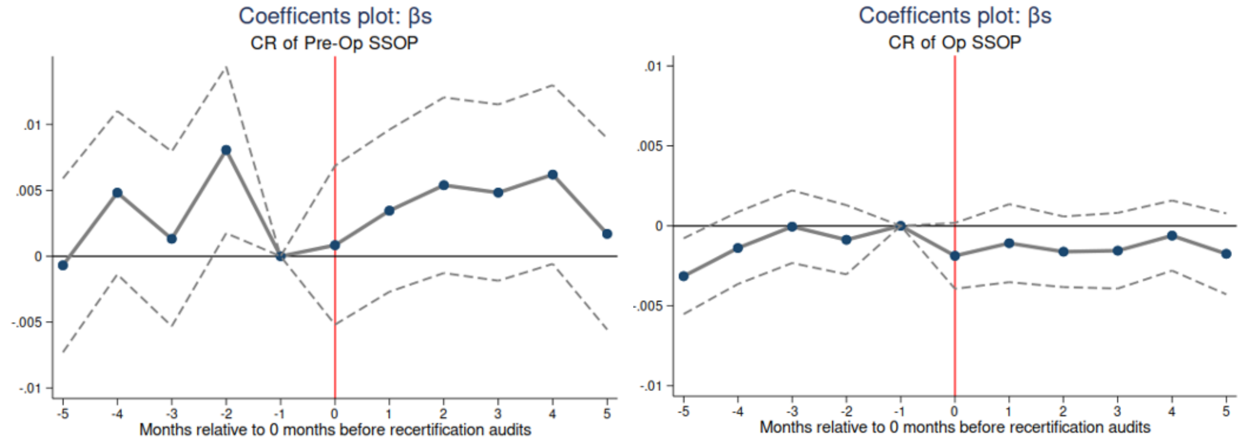
(b) BRC, re-certification audits



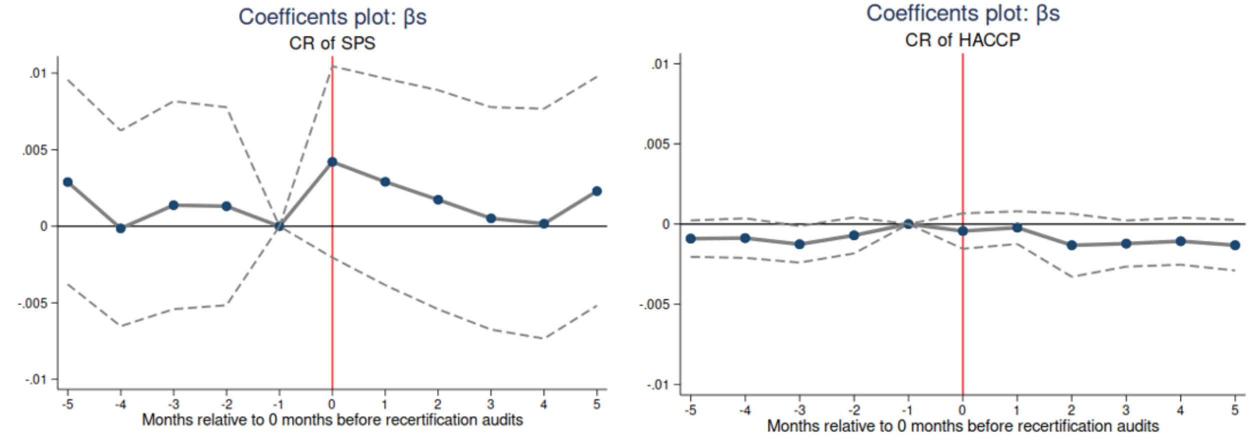
Note: the result is based on fixed effect estimator from regression 6.4 and SQF/BRC re-certification balanced sample. Blue dots represent the point estimates of pre- and post-event dummies. The dotted lines represents the 95% confidence interval of each point estimate.

Figure 6.6: Effect of SQF & BRC re-certification audits on plant's overall food safety practice level

(a) Pre-Op SSOP (b) Op SSOP



(c) SPS (d) HACCP



Note: the results are based on fixed effect estimator from regression 6.4 using different outcome variables as indicated in the caption of the sub-figures and BRC re-certification balanced sample. Blue dots represent the point estimates of pre- and post-event dummies. The dotted lines represents the 95% confidence interval of each point estimate.

Figure 6.7: Effect of BRC re-certification audits on plant's food safety practice level from 4 different aspects

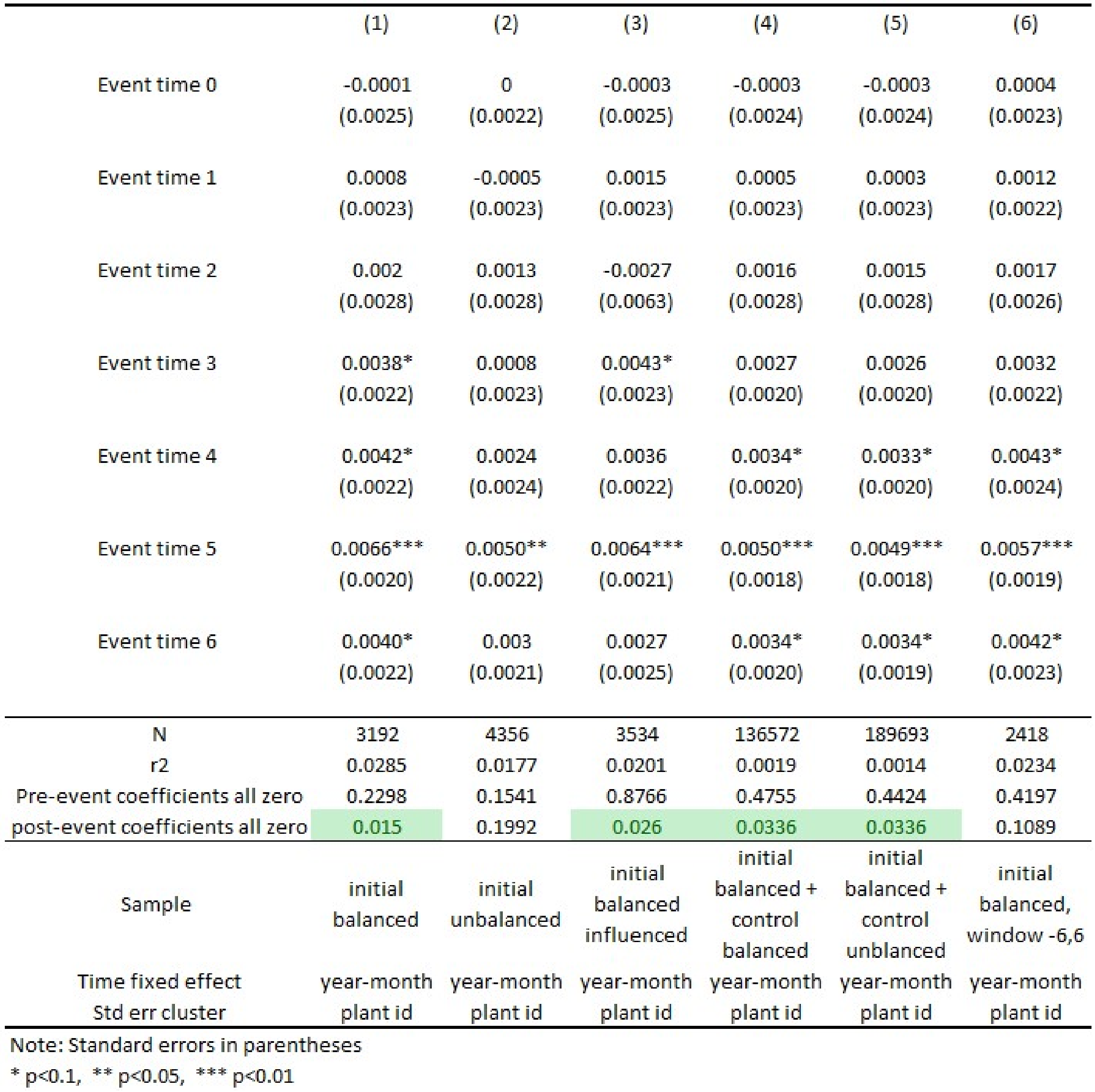
6.5 Robustness check

6.5.1 Initial certification audits

For robustness checks on the initial certification effect, I re-estimated the regression 6.1 using different versions of SQF/BRC initial certification samples introduced in section 6.1.1: initial unbalanced, initial balanced influenced, initial balanced + control balanced, initial balanced + control unbalanced, and initial balanced with window -6 and 6 samples. The results are shown in Table 6.5 and 6.6 respectively for SQF and BRC initial certification audit. I can still see the gradual ramping effect of the SQF initial certification audit for different samples. There is still no evidence that the initial BRC certification audit has any effect on the routine sanitary CR of the BRC certified plants. Thus, our baseline result is robust to the different event study samples mentioned and the number of months included in the pre-event window.

In addition, Table 6.7 displays the results of the initial certification effect using the initial balanced samples with different specifications. Specifically, I present the results using year and month fixed effect instead of the year-month effect, or clustering the standard error at the FSIS circuit number level instead of the plant level. As we can see, the baseline result is also robust to different specifications for both the SQF and the BRC initial certification audit.

Table 6.5: Regression results of SQF initial audit effect on CR of all routine sanitary tasks using different samples

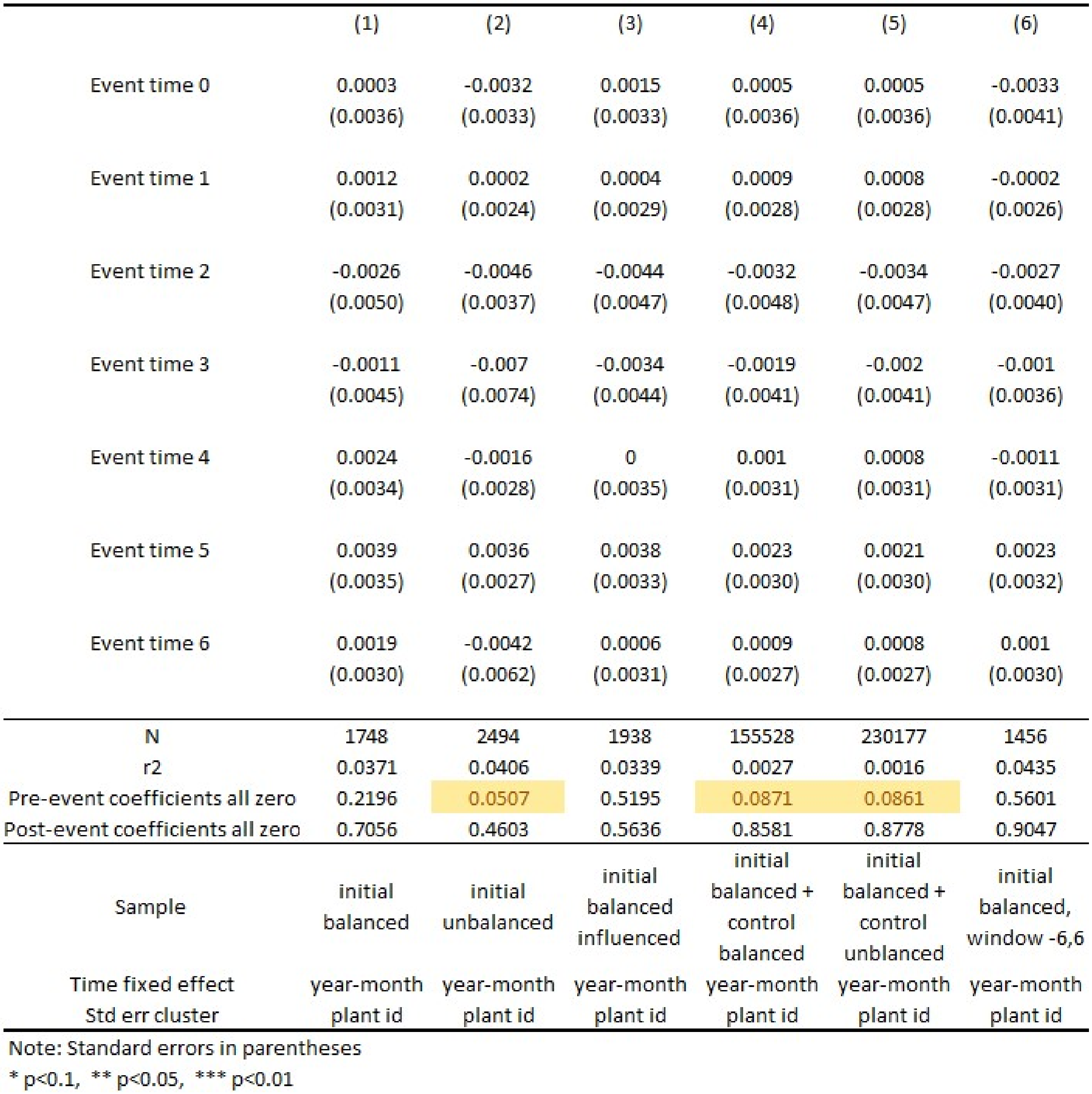


	(1)	(2)	(3)	(4)	(5)	(6)
Event time 0	-0.0001	0	-0.0003	-0.0003	-0.0003	0.0004
	(0.0025)	(0.0022)	(0.0025)	(0.0024)	(0.0024)	(0.0023)
Event time 1	0.0008	-0.0005	0.0015	0.0005	0.0003	0.0012
	(0.0023)	(0.0023)	(0.0023)	(0.0023)	(0.0023)	(0.0022)
Event time 2	0.002	0.0013	-0.0027	0.0016	0.0015	0.0017
	(0.0028)	(0.0028)	(0.0063)	(0.0028)	(0.0028)	(0.0026)
Event time 3	0.0038*	0.0008	0.0043*	0.0027	0.0026	0.0032
	(0.0022)	(0.0023)	(0.0023)	(0.0020)	(0.0020)	(0.0022)
Event time 4	0.0042*	0.0024	0.0036	0.0034*	0.0033*	0.0043*
	(0.0022)	(0.0024)	(0.0022)	(0.0020)	(0.0020)	(0.0024)
Event time 5	0.0066***	0.0050**	0.0064***	0.0050***	0.0049***	0.0057***
	(0.0020)	(0.0022)	(0.0021)	(0.0018)	(0.0018)	(0.0019)
Event time 6	0.0040*	0.003	0.0027	0.0034*	0.0034*	0.0042*
	(0.0022)	(0.0021)	(0.0025)	(0.0020)	(0.0019)	(0.0023)
N	3192	4356	3534	136572	189693	2418
r2	0.0285	0.0177	0.0201	0.0019	0.0014	0.0234
Pre-event coefficients all zero	0.2298	0.1541	0.8766	0.4755	0.4424	0.4197
post-event coefficients all zero	0.015	0.1992	0.026	0.0336	0.0336	0.1089
Sample	initial balanced	initial unbalanced	initial balanced influenced	initial balanced + control balanced	initial balanced + control unblanced	initial balanced, window -6,6
Time fixed effect	year-month	year-month	year-month	year-month	year-month	year-month
Std err cluster	plant id	plant id	plant id	plant id	plant id	plant id

Note: Standard errors in parentheses

* p<0.1, ** p<0.05, *** p<0.01

Table 6.6: Regression results of BRC initial audit effect on CR of all routine sanitary tasks using different samples

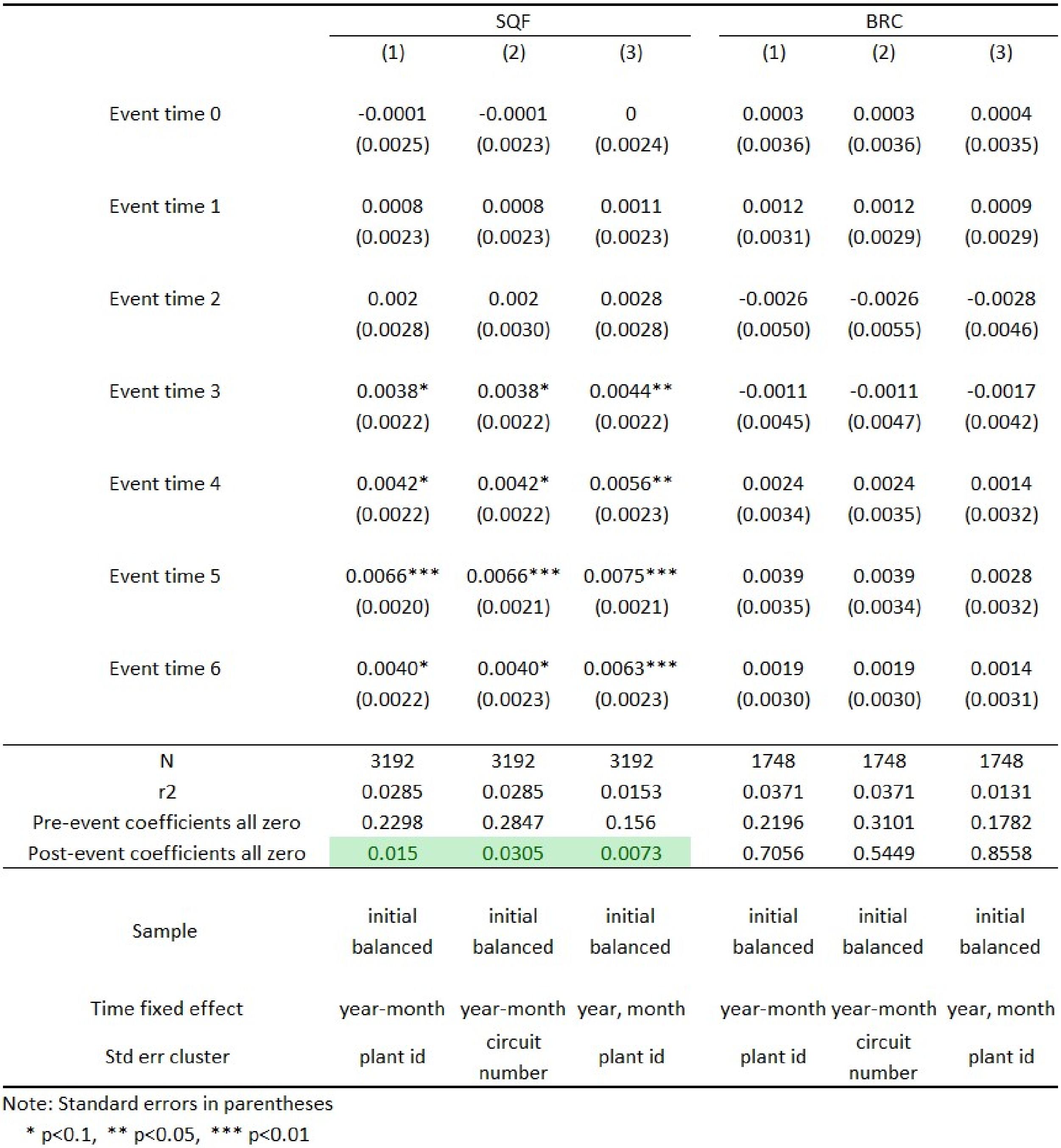


	(1)	(2)	(3)	(4)	(5)	(6)
Event time 0	0.0003	-0.0032	0.0015	0.0005	0.0005	-0.0033
	(0.0036)	(0.0033)	(0.0033)	(0.0036)	(0.0036)	(0.0041)
Event time 1	0.0012	0.0002	0.0004	0.0009	0.0008	-0.0002
	(0.0031)	(0.0024)	(0.0029)	(0.0028)	(0.0028)	(0.0026)
Event time 2	-0.0026	-0.0046	-0.0044	-0.0032	-0.0034	-0.0027
	(0.0050)	(0.0037)	(0.0047)	(0.0048)	(0.0047)	(0.0040)
Event time 3	-0.0011	-0.007	-0.0034	-0.0019	-0.002	-0.001
	(0.0045)	(0.0074)	(0.0044)	(0.0041)	(0.0041)	(0.0036)
Event time 4	0.0024	-0.0016	0	0.001	0.0008	-0.0011
	(0.0034)	(0.0028)	(0.0035)	(0.0031)	(0.0031)	(0.0031)
Event time 5	0.0039	0.0036	0.0038	0.0023	0.0021	0.0023
	(0.0035)	(0.0027)	(0.0033)	(0.0030)	(0.0030)	(0.0032)
Event time 6	0.0019	-0.0042	0.0006	0.0009	0.0008	0.001
	(0.0030)	(0.0062)	(0.0031)	(0.0027)	(0.0027)	(0.0030)
N	1748	2494	1938	155528	230177	1456
r2	0.0371	0.0406	0.0339	0.0027	0.0016	0.0435
Pre-event coefficients all zero	0.2196	0.0507	0.5195	0.0871	0.0861	0.5601
Post-event coefficients all zero	0.7056	0.4603	0.5636	0.8581	0.8778	0.9047
Sample	initial balanced	initial unbalanced	initial balanced influenced	initial balanced + control balanced	initial balanced + control unblanced	initial balanced, window -6,6
Time fixed effect	year-month	year-month	year-month	year-month	year-month	year-month
Std err cluster	plant id	plant id	plant id	plant id	plant id	plant id

Note: Standard errors in parentheses

* p<0.1, ** p<0.05, *** p<0.01

Table 6.7: Regression results of SQF and BRC initial audit effect on CR of all routine sanitary tasks using various specifications



	SQF			BRC		
	(1)	(2)	(3)	(1)	(2)	(3)
Event time 0	-0.0001	-0.0001	0	0.0003	0.0003	0.0004
	(0.0025)	(0.0023)	(0.0024)	(0.0036)	(0.0036)	(0.0035)
Event time 1	0.0008	0.0008	0.0011	0.0012	0.0012	0.0009
	(0.0023)	(0.0023)	(0.0023)	(0.0031)	(0.0029)	(0.0029)
Event time 2	0.002	0.002	0.0028	-0.0026	-0.0026	-0.0028
	(0.0028)	(0.0030)	(0.0028)	(0.0050)	(0.0055)	(0.0046)
Event time 3	0.0038*	0.0038*	0.0044**	-0.0011	-0.0011	-0.0017
	(0.0022)	(0.0022)	(0.0022)	(0.0045)	(0.0047)	(0.0042)
Event time 4	0.0042*	0.0042*	0.0056**	0.0024	0.0024	0.0014
	(0.0022)	(0.0022)	(0.0023)	(0.0034)	(0.0035)	(0.0032)
Event time 5	0.0066***	0.0066***	0.0075***	0.0039	0.0039	0.0028
	(0.0020)	(0.0021)	(0.0021)	(0.0035)	(0.0034)	(0.0032)
Event time 6	0.0040*	0.0040*	0.0063***	0.0019	0.0019	0.0014
	(0.0022)	(0.0023)	(0.0023)	(0.0030)	(0.0030)	(0.0031)
N	3192	3192	3192	1748	1748	1748
r2	0.0285	0.0285	0.0153	0.0371	0.0371	0.0131
Pre-event coefficients all zero	0.2298	0.2847	0.156	0.2196	0.3101	0.1782
Post-event coefficients all zero	0.015	0.0305	0.0073	0.7056	0.5449	0.8558
Sample	initial balanced	initial balanced	initial balanced	initial balanced	initial balanced	initial balanced
Time fixed effect	year-month	year-month	year, month	year-month	year-month	year, month
Std err cluster	plant id	circuit number	plant id	plant id	circuit number	plant id

Note: Standard errors in parentheses
* p<0.1, ** p<0.05, *** p<0.01

6.5.2 Re-certification audits

For robustness checks on the re-certification audit effect, I re-estimated regression 6.4 using different versions of SQF/BRC re-certification samples introduced in Section 6.1.2: unbalanced re-certification, balanced re-certification + control balanced, balanced re-certification + control unbalanced samples. The results are shown in columns (1) - (4) of Table 6.8 and 6.9 for the SQF and BRC re-certification audit. Also, I showed the estimation results with various specifications of regression 6.4: various versions of time and individual fixed effect and cluster level of standard errors in column (5) - (8).

For SQF re-certification audit results, the ramping up and down patterns of point estimates are similar across different samples and specifications: generally speaking, point estimates are negative when the event time is several months before or after the re-certification audit time, and non-negative or close to zero in the months close to re-certification audit time. Again, the point estimates might indicate that CR of all routine sanitary tasks ramp up towards the re-certification audit time and ramp down after. However, only in the unbalanced re-certification sample, I observe a significant negative estimation of the effect at event months -5 and +5 with 10% significance. Thus, our baseline result is generally robust to different sample constructions. The imprecise estimation of the re-certification audit effect suggests that we should take the micro pattern with a grain of salt.

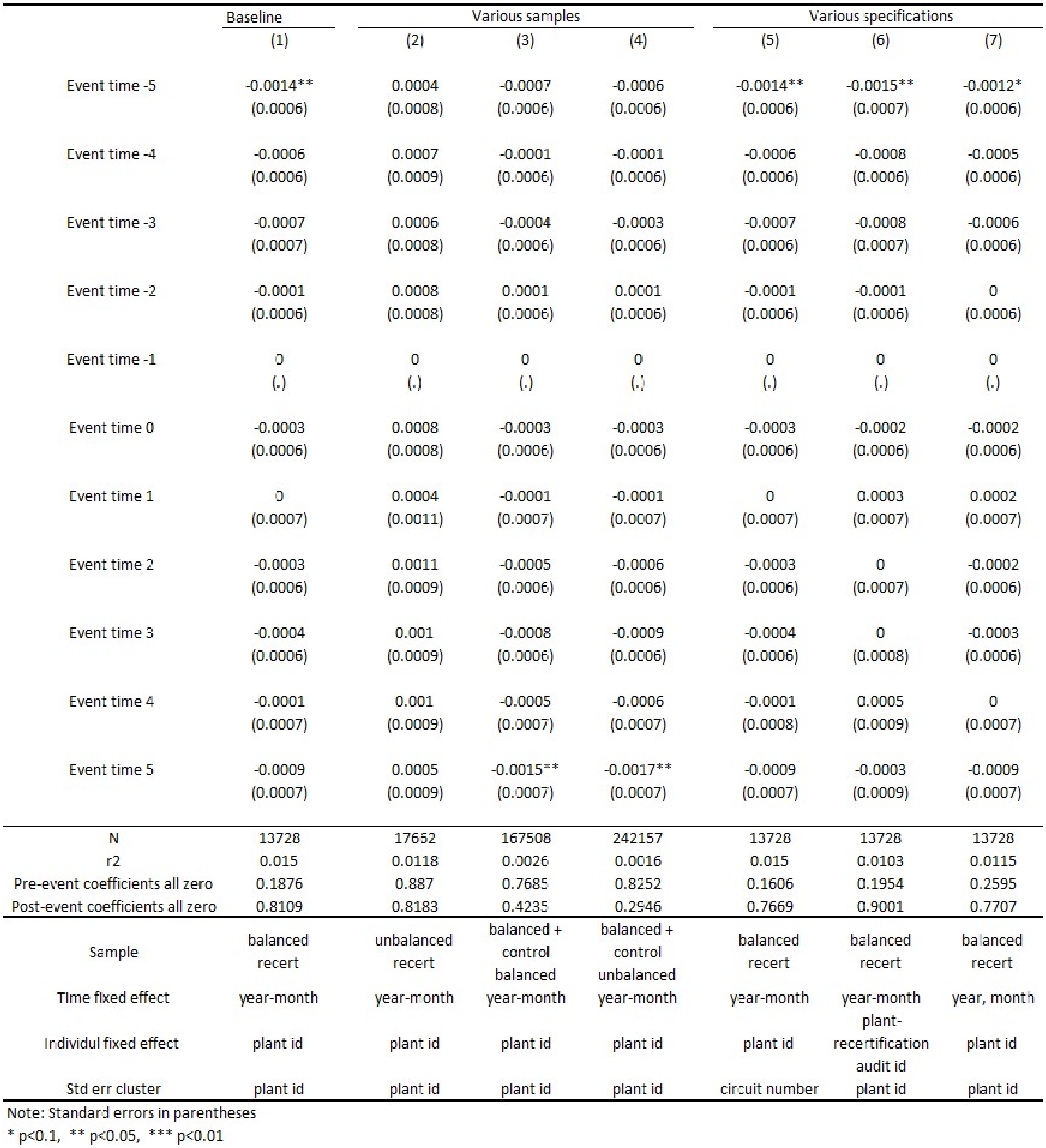
For BRC re-certification, I observe the overall routine sanitary CR ramping up and down pattern of point estimates across all samples and specifications except for the unbalanced re-certification sample result. Although the point estimates of β_{-5} for the balanced re-certification + control samples are not significantly different from zero at 5% significance as the baseline results, it is still negative but with a smaller magnitude. Instead, the point estimates of β_5 is significantly negative and with a slightly bigger magnitude than the baseline results. Therefore, overall, our results are consistent across different samples and specifications.

Table 6.8: Regression results of SQF re-certification audit effect on CR of all routine sanitary tasks using various samples and specifications

	Baseline	Various samples			Various specifications		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Event time -5	-0.0012	-0.0033**	-0.0012	-0.0012	-0.0012*	-0.0022**	-0.0011
	(0.0009)	(0.0013)	(0.0009)	(0.0009)	(0.0007)	(0.0010)	(0.0009)
Event time -4	-0.0002	-0.0008	0	0	-0.0002	-0.001	0
	(0.0008)	(0.0007)	(0.0007)	(0.0007)	(0.0009)	(0.0008)	(0.0007)
Event time -3	0.0002	-0.0002	0.0004	0.0004	0.0002	-0.0003	0.0003
	(0.0007)	(0.0006)	(0.0007)	(0.0007)	(0.0006)	(0.0007)	(0.0007)
Event time -2	0.0004	-0.0005	0.0005	0.0005	0.0004	0.0001	0.0005
	(0.0007)	(0.0007)	(0.0007)	(0.0007)	(0.0006)	(0.0007)	(0.0007)
Event time -1	0	0	0	0	0	0	0
	(.)	(.)	(.)	(.)	(.)	(.)	(.)
Event time 0	0.0001	-0.0002	0	0	0.0001	0.0003	0.0001
	(0.0007)	(0.0006)	(0.0007)	(0.0007)	(0.0006)	(0.0007)	(0.0007)
Event time 1	0.0006	0.0002	0.0007	0.0006	0.0006	0.0011	0.0008
	(0.0007)	(0.0006)	(0.0006)	(0.0006)	(0.0007)	(0.0007)	(0.0007)
Event time 2	0.0011	0.0005	0.0011*	0.001	0.0011*	0.0017**	0.0012*
	(0.0007)	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0007)	(0.0007)
Event time 3	0.0007	0	0.0007	0.0006	0.0007	0.0016**	0.0008
	(0.0007)	(0.0007)	(0.0007)	(0.0007)	(0.0006)	(0.0008)	(0.0007)
Event time 4	0.0006	-0.0005	0.0008	0.0006	0.0006	0.0018**	0.0009
	(0.0007)	(0.0008)	(0.0007)	(0.0007)	(0.0006)	(0.0008)	(0.0007)
Event time 5	-0.0003	-0.0021*	-0.0001	-0.0003	-0.0003	0.0011	0
	(0.0013)	(0.0012)	(0.0012)	(0.0012)	(0.0011)	(0.0012)	(0.0012)
N	11671	16171	112371	142436	11671	11671	11671
r2	0.0066	0.0049	0.0014	0.0013	0.0066	0.0069	0.004
Pre-event coefficients all zero	0.535	0.1195	0.4114	0.3815	0.3081	0.1093	0.4983
Post-event coefficients all zero	0.6912	0.4122	0.6249	0.6669	0.6126	0.2734	0.5025
Sample	balanced recert	unbalanced recert	balanced + control balanced	balanced + control unbalanced	balanced recert	balanced recert	balanced recert
Time fixed effect	year-month	year-month	year-month	year-month	year-month	year-month	year, month
Individul fixed effect	plant id	plant id	plant id	plant id	plant id	plant-recertification audit id	plant id
Std err cluster	plant id	plant id	plant id	plant id	circuit number	plant id	plant id

Note: Standard errors in parentheses
* p<0.1, ** p<0.05, *** p<0.01

Table 6.9: Regression results of BRC re-certification audit effect on CR of all routine sanitary tasks using various samples and specifications



	Baseline	Various samples			Various specifications		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Event time -5	-0.0014**	0.0004	-0.0007	-0.0006	-0.0014**	-0.0015**	-0.0012*
	(0.0006)	(0.0008)	(0.0006)	(0.0006)	(0.0006)	(0.0007)	(0.0006)
Event time -4	-0.0006	0.0007	-0.0001	-0.0001	-0.0006	-0.0008	-0.0005
	(0.0006)	(0.0009)	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)
Event time -3	-0.0007	0.0006	-0.0004	-0.0003	-0.0007	-0.0008	-0.0006
	(0.0007)	(0.0008)	(0.0006)	(0.0006)	(0.0006)	(0.0007)	(0.0006)
Event time -2	-0.0001	0.0008	0.0001	0.0001	-0.0001	-0.0001	0
	(0.0006)	(0.0008)	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)
Event time -1	0	0	0	0	0	0	0
	(.)	(.)	(.)	(.)	(.)	(.)	(.)
Event time 0	-0.0003	0.0008	-0.0003	-0.0003	-0.0003	-0.0002	-0.0002
	(0.0006)	(0.0008)	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)
Event time 1	0	0.0004	-0.0001	-0.0001	0	0.0003	0.0002
	(0.0007)	(0.0011)	(0.0007)	(0.0007)	(0.0007)	(0.0007)	(0.0007)
Event time 2	-0.0003	0.0011	-0.0005	-0.0006	-0.0003	0	-0.0002
	(0.0006)	(0.0009)	(0.0006)	(0.0006)	(0.0006)	(0.0007)	(0.0006)
Event time 3	-0.0004	0.001	-0.0008	-0.0009	-0.0004	0	-0.0003
	(0.0006)	(0.0009)	(0.0006)	(0.0006)	(0.0006)	(0.0008)	(0.0006)
Event time 4	-0.0001	0.001	-0.0005	-0.0006	-0.0001	0.0005	0
	(0.0007)	(0.0009)	(0.0007)	(0.0007)	(0.0008)	(0.0009)	(0.0007)
Event time 5	-0.0009	0.0005	-0.0015**	-0.0017**	-0.0009	-0.0003	-0.0009
	(0.0007)	(0.0009)	(0.0007)	(0.0007)	(0.0007)	(0.0009)	(0.0007)
N	13728	17662	167508	242157	13728	13728	13728
r2	0.015	0.0118	0.0026	0.0016	0.015	0.0103	0.0115
Pre-event coefficients all zero	0.1876	0.887	0.7685	0.8252	0.1606	0.1954	0.2595
Post-event coefficients all zero	0.8109	0.8183	0.4235	0.2946	0.7669	0.9001	0.7707
Sample	balanced recert	unbalanced recert	balanced + control balanced	balanced + control unbalanced	balanced recert	balanced recert	balanced recert
Time fixed effect	year-month	year-month	year-month	year-month	year-month	year-month	year, month
Individul fixed effect	plant id	plant id	plant id	plant id	plant id	plant-recertification audit id	plant id
Std err cluster	plant id	plant id	plant id	plant id	circuit number	plant id	plant id

Note: Standard errors in parentheses

* p<0.1, ** p<0.05, *** p<0.01

6.6 Limitations

How to estimate the causal response of anticipated and repeating events is still unresolved in the econometrics literature. In this paper, I attempted to adapt and apply the classic event study method in this new setting. I estimated the effects of initial certification and a series of re-certification audits on certified plants' food safety practices by carefully constructing the event study samples and conducting multiple robustness checks shown in Appendix 6.5. However, there are several caveats to this study.

First, traditional event study relies on the classic parallel trends assumption and no anticipation effect (or the existence of a period when the treated are not affected by the treatment). In our setting, certified plants are self-selected into the initial certification treatment and can choose the timing. Though I tried to control for the selection due to time-invariant characteristics and possible common shocks across plants such as seasonality and found no evidence to reject the parallel trend and no anticipation assumptions, one can still criticize the unclean research design due to other plant specific time variant confounders. In terms of the repeated re-certification audit event analysis, due to the potential anticipation response before the re-audits and short pre-event window, it relies on the untestable assumption, a parallel trend had the re-certification audit not happened, to identify the pre-event and after-event responses of certified plants to the repeated re-certification audit events.

Second, the other caveat of applying two-way fixed effect event study estimators in our setting is that staggered treatments and heterogeneous effects among different cohorts can cause the estimator to not recover the Average Treatment Effect on the Treated (ATT), even when the parallel assumption holds (De Chaisemartin and D'Haultfoeuille, 2022; Sun and Abraham, 2021; Goodman-Bacon, 2021). One potential future work is to adapt the heterogeneity-robust event study estimators (Callaway and Sant'Anna, 2021) to our repeated treatment setting to estimate a cohort-specific treatment effect. One of the difficulties lies in how to properly define the counterfactual of repeated treatment.

Third, this paper is unable to study the effect of surveillance audits because the timing of a surveillance event is too close to the previous and subsequent re-certification audit events. One way to incorporate surveillance audit events is to apply the multiple event study analysis (Schmidheiny and Siegloch, 2019), but it is not without any cost. Basically, on top of the previously stated assumptions in the classic event study, it requires the effect of each event (initial, re-certification, and surveillance audit) to be additive.

Fourth, this analysis is unable to use sampling test results to directly reflect the change in a plant's food safety level around certification audits. Although the result of the phytosanitary sampling tests is a more direct and interesting way to measure the outcomes of plant food safety, due to the sparsity of data around certification audits, it is not feasible to use it in the analysis of event studies, which requires a relatively long time window for the outcome variables.

Last but not least, this paper is unable to give a definite answer for why we observe a different estimated initial audit effect for SQF and BRC certification. Though I attempt to give two possibilities, it is by no means exhaustive or conclusive. Why certain plants choose SQF, but others chose BRC certification is beyond the scope of this paper. Future work can be done on how plants choose among certifications.

6.7 Conclusions

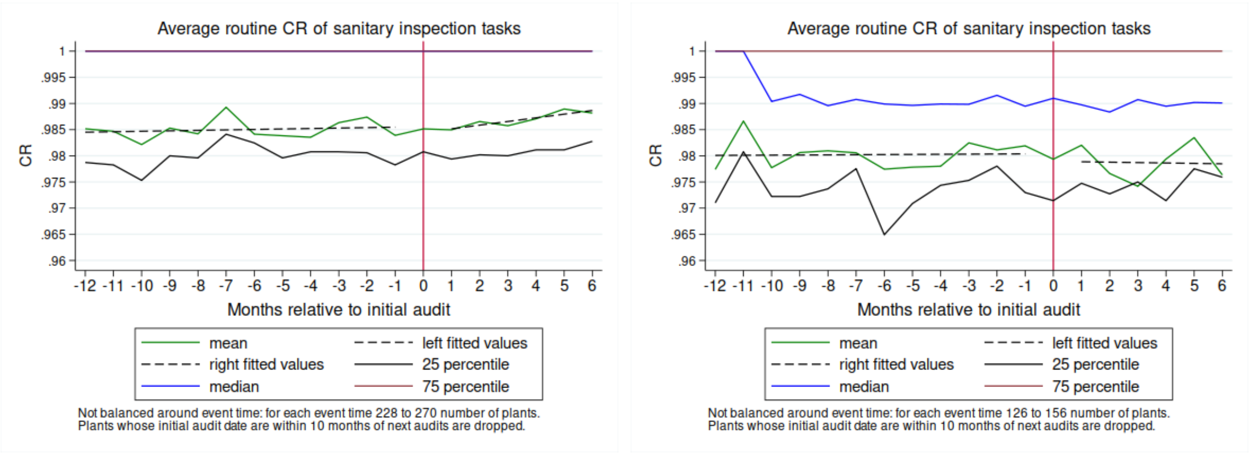
This chapter studies the dynamic effect of the initial certification and re-certification audits of SQF and BRC on the level of food safety practices of certified plants. By exploiting the variations of certification audit dates and FSIS routine sanitary task CR, I adapted and applied the two-way fixed effect event study estimator to estimate certified plants' CR response to initial certification audits and annually repeating re-certification audits.

I find a gradual increase in the average CR of routine sanitary tasks for SQF certified plants after the initial certification. The improvement is mainly due to better performance in SSOP and SPS tasks. However, for BRC certified plants, I do not observe any improvement in routine sanitary CR during the initial certification process. The potential reasons for the discrepancy between the average treatment effect on treated for SQF and BRC initial certification audits could be that (1) BRC has less stringent rules to give out good grades and plants choose to use it as an easy way to signal their underlying food safety condition instead of improving their food safety practices; (2) the combination of plant's heterogeneous response on initial certification audit and the different plant compositions of SQF and BRC initial certification event study sample cause the differences in average treated effect on the treated. These explanations are by no means conclusive or exhaustive and need further research.

The results of re-certification audit events are similar for SQF and BRC certification. Compared to initial certification, the impact of re-certification on the plant's CR is limiting. I do observe a micro pattern of plants' CR ramping up and down before and after re-certification audits, which implies plants could temporarily "perform" better for the re-certification audits without fundamentally improving their food safety practice level in the long run. However, the estimated effect is imprecise. Therefore, the results are only suggestive and should be interpreted with caution.

6.A Appendix

6.A.1 Raw evidence of initial certification audit using initial unbalanced sample



Not balanced around event time: for each event time 228 to 270 number of plants.
Plants whose initial audit date are within 10 months of next audits are dropped.

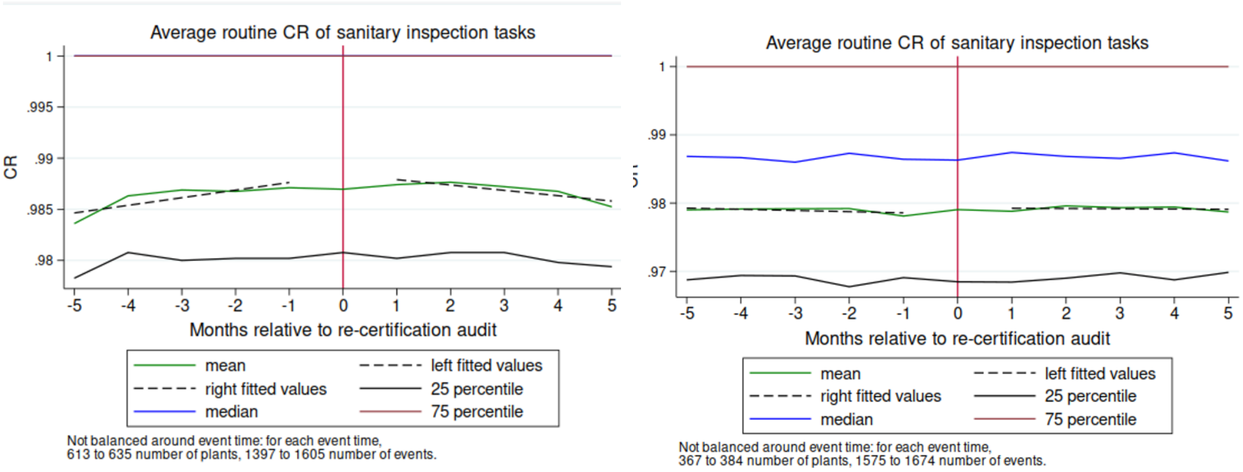
Not balanced around event time: for each event time 126 to 156 number of plants.
Plants whose initial audit date are within 10 months of next audits are dropped.

(a) SQF, initial audits

(b) BRC,initial audits

Figure 6.A.1: Average and percentiles of routine sanitary inspection CR before and after initial certification audits

6.A.2 Raw evidence of re-certification audit using re-certification audit unbalanced sample



(a) SQF, re-certification audits

(b) BRC, re-certification audits

Figure 6.A.2: Average and percentiles of routine sanitary inspection CR before and after re-certification audits

Part II

Private lawsuits and food safety

Chapter 7

Introduction

Food safety remains a significant public health concern in the United States. According to the 2011 estimates by the Centers for Disease Control and Prevention (CDC), there were an estimated 48 million cases of foodborne illness annually, leading to 128,000 hospitalizations and 3,000 deaths (CDC, 2011). The United States Department of Agriculture (USDA)'s Economic Research Service (ERS) estimates that the annual economic cost of foodborne illnesses resulting from the 15 most common pathogens is over 17.5 billion dollars in 2018 dollars (USDA-ERS, 2021). To safeguard the nation's food supply, USDA allocated nearly $1.2 billion in Fiscal Year 2022 to support over 8,600 personnel in the Food Safety and Inspection Service (FSIS), who ensure meat, poultry, and egg products are safe, wholesome and properly labeled at over 6,500 processing, slaughter, and import establishments across the U.S (USDA, 2022).

While public regulations provide one avenue for inducing food safety and proper food labeling, private lawsuits provide another. Through private lawsuits, consumers can hold companies accountable for selling defective products that cause personal harm through product liability claims. A study of 511 foodborne illness jury trials between 1979 and 2014 revealed that plaintiffs received a median award of $32,264 (Mahdu, 2015). Consumers also have the option to sue companies for misbranding or mislabeling their food products. Consumer advocacy groups have filed more than 150 food labeling class action lawsuits against food and beverage companies between 2011 and 2014 (Negowetti, 2014). The number of food labeling lawsuits has been steadily increasing every year and hit a high of 220 lawsuits in 2020 (Jacobs, 2021). Given the magnitude and complexity of food safety and labeling issues, understanding how private lawsuits influence public regulatory activities and plants' food safety practices is important. [1]

This study focuses on the impact of food safety and food labeling-related lawsuits on food safety practices through interaction with governmental regulatory behavior and private companies' behavior. There are two interesting questions that are going to be answered: (1) Do lawsuits crowd in or crowd out public regulatory activities, i.e. monitoring? (2) Do lawsuits incentivize firms to improve food safety practices?

For FSIS, lawsuits can crowd in or crowd out public regulatory activities. Lawsuits draw more attention to violating firms, which may in turn increase regulatory intensity. On the other hand, for firms facing active private litigation, government agents may allocate scarce resources toward other activities. In the context of environmental regulation, Langpap and Shimshack (2010) find that private citizen suits crowd in public monitoring but significantly crowd out public sanctions.

The impact of lawsuits on firm behavior is unclear. Risks such as litigation costs, plaintiff compensation, reputation, and sales loss from lawsuits can stimulate activity beyond that needed for facility-level compliance. However, it is also possible that threats of lawsuits are not credible or frequent enough to invest in or change behavior in order to mitigate the adverse effect of food safety issues. Facing the high information and transaction costs, individual plaintiffs are likely to have weak incentives to pursue litigation. Furthermore, product liability insurance and food safety certification can distort legal incentives to produce safer food (Buzby and Frenzen, 1999; Mahdu, 2015). Companies may become less incentivized to

---

[1] In this study, the term "plants' food safety practices" refers to the general practices involved in handling food products, such as ensuring sanitary conditions, practicing humane animal handling, and adhering to labeling practices. Additionally, unless otherwise specified, the term "food safety lawsuits" is used as a general term to refer to legal cases related to foodborne illness, mislabeling, misbranding, or inhumane handling of animals.

invest in and implement food safety measures, knowing that they have insurance to cover the costs of litigation and compensation and they can use the certification as a legal defense in case of a lawsuit. The effort a company puts into food safety is the result of balancing the benefits of taking food safety precautions and the expected costs, such as litigation costs. Whether private lawsuits incentivize firms to produce safer products is an empirical question.

The interaction between private lawsuits, regulatory activities, and firm behavior is not well studied in the food safety setting. In this study, I systematically collected food safety lawsuit data and link it to FSIS inspected plants, in order to construct a rich dataset of public and private monitoring activity in food safety. The unique dataset makes this empirical study possible. I leverage the variation in the timing of the lawsuit filing date to estimate the general and specific lawsuit events on FSIS monitoring intensity and the defendant plant's food safety practices. I find that food safety lawsuits have a crowd-out effect on the regulatory intensity in inspection tasks that are directly relevant to the food safety disputed issues in the lawsuit. However, there is no statistically significant impact of food safety lawsuits on the food safety practices of the defendant plants. This study is the first research to empirically tackle this important issue. It contributes to the food safety literature by adding to the general understanding of the causes and consequences of private lawsuits on government regulation and private firm behavior.

The rest of Part II is organized as follows. I start by introducing the background and institutional knowledge of FSIS regulatory activities and food safety lawsuits. Then, I describe the data collection process and present summary statistics of the analysis sample. Next, I show the main empirical framework and estimators used to study the impact of lawsuits. Finally, I present the results and conclusions.

Chapter 8

Background

8.1 FSIS monitoring activities

The Food Safety and Inspection Service (FSIS) is the USDA regulatory branch that is responsible for ensuring that the commercial supply of meat, poultry, and egg products in the United States is safe, healthy, and correctly labeled and packaged. FSIS employs approximately 8,000 in-plant and other front-line personnel in more than 6000 federally inspected slaughter and processing establishments, in laboratories, and in commerce nationwide (USDA, 2019).

To describe the general practices of handling food products, this study employs the term "food safety practice" in a comprehensive manner, encompassing practices that ensure sanitary conditions, humane animal handling, and adherence to labeling guidelines. To monitor plants' food safety practices, each month, inspectors are given a list of inspection tasks generated by the Public Health Information System (PHIS) based on establishments' profile information such as the HACCP processing category and products, with important information such as inspection frequency and priority. Inspectors have the flexibility to determine which tasks need to be added or removed and when to perform the tasks. There are two types of inspection tasks: routine tasks and directed tasks. Routine tasks are performed routinely, continuously, or planned under normal conditions. Directed inspection tasks are performed as needed, such as sampling tasks and export certification tasks. Directed tasks occur in several situations: (1) after establishments have positive pathogen sample results, (2) when requested by FSIS headquarter personnel or district office personnel, or supervisors based on identified needs, (3) when poultry establishments have multiple slaughter lines and need additional tasks to accomplish a zero-tolerance verification task twice per shift for each slaughter line, (4) when inspectors need to verify corrective actions of establishments, and (5) when initiated by inspectors based on the "stumble-on" conditions observed in the establishments. If inspectors increase their regulatory intensity, one would expect them to conduct more inspector initiated directed tasks to impose more stringent monitoring activities, compared to the routine regular inspection tasks. This would lead to an increase in the share of directed tasks. Therefore, in general, the ratio of the number of directed and routine inspection tasks could be a proxy of the FSIS inspector's monitoring intensity.

Noncompliance records (NR) are generated if establishments do not comply with routine or directed inspection tasks. NR serves as a notification and documentation of firm noncompliance with regulatory standards. After receiving NRs, firms need to take corrective actions to reach regulatory requirements. Inspectors then verify their corrective actions and close the NRs. Thus, to measure a plant's food safety practice level, one can use the compliance rate of sanitary routine inspection tasks (1 - total number of open and closed routine inspection NRs/total number of routine inspection tasks). Directed inspection tasks occur less frequently compared to routine inspection tasks and are subject to different scenarios for different plants, making the compliance rate of directed tasks noisy and less comparable across plants. Therefore, the compliance rate of directed tasks is less ideal to reflect the level of average food safety practices of plants.

8.2 Food safety lawsuits

In this study, unless otherwise specified, the term "food safety lawsuits" is used as a general term to refer to legal cases related to FSIS inspection activities including issues on food safety, mislabeling, or inhumane handling and slaughter of livestock.

In the United States, food safety incidents can lead to product liability lawsuits against responsible companies. Strict product liability, negligence, and breach of warranty are the three main legal causes of action in these cases (Gursoy, 2019; Connally, 2009; Buzby and Frenzen, 1999). While foodborne illness claims make up the majority of food safety lawsuits, there has been a recent increase in lawsuits over misleading food labels. However, according to New York Times, these lawsuits have had little success due to possible deceptive marketing practices by large food companies (Jacobs, 2021).

The possibility of lawsuits and reputation effects can create incentives for firms to improve the safety and labeling of their food products. Legal liability for food safety damages can result in significant penalties, including fines, settlements, and loss of license. When information about a company's food safety lawsuit becomes public, the company may also face reputational damage and increased monitoring from government, business partners, or consumers, even if the company ultimately prevails in the lawsuit. These outcomes can encourage plants to enhance their overall food safety practices after a lawsuit. In the context of environmental regulations, a study by Keohane, Mansur and Voynov (2009) found that coal-fired power plants facing a higher probability of being sued were more likely to reduce their emissions compared to plants with lower risks of lawsuits.

Food safety lawsuits, especially those related to food poisoning, pose challenges in identifying the cause of illness, particularly outside of outbreak situations, and may have high transaction costs for plaintiffs. As a result, only a small fraction of food safety cases, less than 0.01%, end up in litigation (Buzby and Frenzen, 1999), with most claims resolved through settlements or negotiations. This suggests that food safety cases are not often publicly visible, and plants may have limited legal incentives to improve food safety. Moreover, product liability insurance and third-party food safety certification may weaken the deterrent effect of lawsuits. Product liability insurance can shift the cost of rare product liability lawsuits to a company's operating expenses, leading a company to prioritize higher insurance premiums over efforts to produce safer food. Food safety certification can also be used as a defense to influence trial outcomes or legal compensation to plaintiffs (Mahdu, 2015). Therefore, it is important to empirically examine whether food safety lawsuits have a deterrent effect and lead firms to improve their food safety practices.

Although the relationship between private lawsuits and public regulation is not well studied in the food safety context, the substitutional or complemental effect of lawsuits on public monitoring and enforcement activities has been researched in other settings, such as environmental regulation. Langpap (2007)'s economic model suggests that public enforcement may increase or decrease its intensity compared to the situation without private lawsuits, depending on the relative costs of public and private enforcement and the changes in public inspection costs due to private lawsuits. Langpap and Shimshack (2010) provide empirical evidence on the causal effect of the impact of private environmental lawsuits on public regulation by analyzing the Clean Water Act public agency enforcement and monitoring actions. They find that private citizen suits crowd in public monitoring but crowd out public sanctions. This study aims to tackle the empirical question in the food safety setting by combining food safety lawsuits and FSIS inspection activity data.

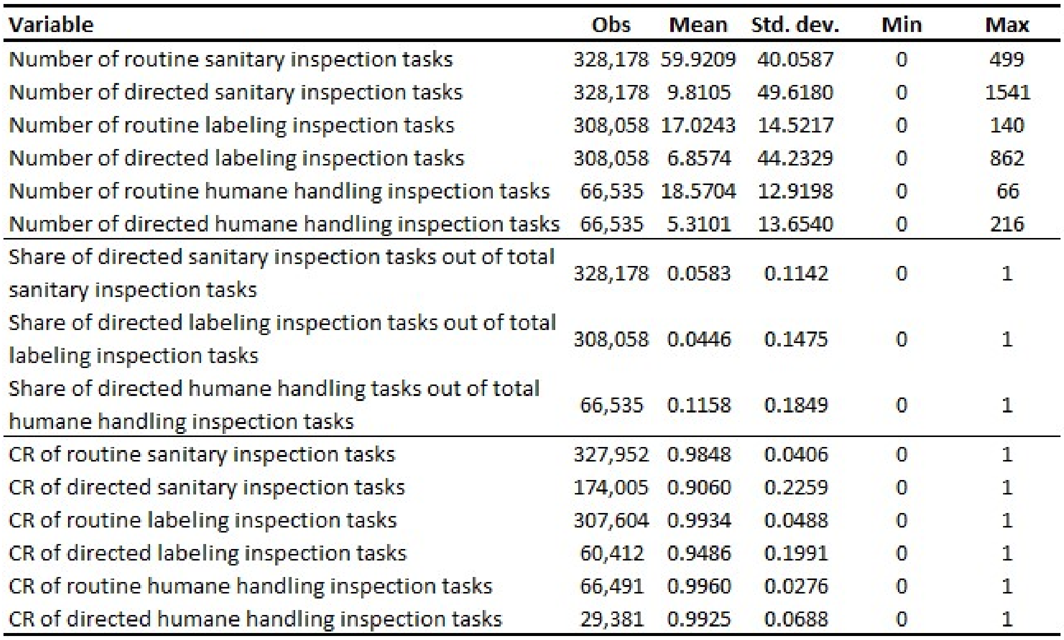
Chapter 9

Data and summary statistics

To study the effect of lawsuits on public regulatory activities and firm behavior in the food safety setting, I combined two unique data sources: (1) private food safety lawsuits and (2) FSIS inspection activities. I use the proportion of direct inspection tasks related to sanitation, labeling, and humane handling out of all routine and directed tasks to gauge the level of regulatory intensity for each of these categories. Additionally, I use the compliance rates (CRs) of routine sanitary inspection tasks, routine labeling inspection tasks, and humane handling inspection tasks to assess the level of food safety practices in each of these areas.

The public regulatory data used in this study was obtained from the USDA FSIS administrative dataset, which provides detailed information on the monitoring activities of FSIS inspectors in all FSIS-inspected meat and poultry plants. This dataset includes plant-level inspection dates and types, such as routine or directed inspection tasks on issues like sanitary, labeling, humane handling, and others, covering the period from July 2012 to December 2017. The data are aggregated to the plant-month level, with over 30 million distinct inspection tasks and more than 500,000 distinct noncompliance records. Routine inspection tasks are based on the facility's HAACP plan, which is approved by FSIS. The inspection tasks are aggregated based on task code and description to create three categories: sanitary, labeling, and humane handling inspection tasks. For each category, the total number of

Table 9.1: Summary statistics on 3 types of FSIS inspection tasks at the establishment-month level



Variable	Obs	Mean	Std. dev.	Min	Max
Number of routine sanitary inspection tasks	328,178	59.9209	40.0587	0	499
Number of directed sanitary inspection tasks	328,178	9.8105	49.6180	0	1541
Number of routine labeling inspection tasks	308,058	17.0243	14.5217	0	140
Number of directed labeling inspection tasks	308,058	6.8574	44.2329	0	862
Number of routine humane handling inspection tasks	66,535	18.5704	12.9198	0	66
Number of directed humane handling inspection tasks	66,535	5.3101	13.6540	0	216
Share of directed sanitary inspection tasks out of total sanitary inspection tasks	328,178	0.0583	0.1142	0	1
Share of directed labeling inspection tasks out of total labeling inspection tasks	308,058	0.0446	0.1475	0	1
Share of directed humane handling tasks out of total humane handling inspection tasks	66,535	0.1158	0.1849	0	1
CR of routine sanitary inspection tasks	327,952	0.9848	0.0406	0	1
CR of directed sanitary inspection tasks	174,005	0.9060	0.2259	0	1
CR of routine labeling inspection tasks	307,604	0.9934	0.0488	0	1
CR of directed labeling inspection tasks	60,412	0.9486	0.1991	0	1
CR of routine humane handling inspection tasks	66,491	0.9960	0.0276	0	1
CR of directed humane handling inspection tasks	29,381	0.9925	0.0688	0	1

tasks, compliance rates (CRs), and the share of directed tasks were calculated. Table 9.1 reports the average of these monthly measures of all establishments. Approximately 98-99% of routine sanitary inspection tasks meet standards, with average CRs of over 99% for routine inspection tasks of labeling and humane handling. Directed tasks have lower CRs than routine tasks, as expected. The share of directed sanitary and labeling inspection tasks is around 5%, while the share of directed humane handling inspection tasks is 12%. More details of the plant characteristics in the FSIS dataset can be found in Part I, Section 4.1.

I built a private lawsuit dataset related to FSIS regulatory activities such as food safety, labeling, and animal handling issues by conducting systematic searches in the LexisNexis law database. I restricted cases that are decided between 2012 and 2018, and then used different search terms such as “foodborne illness”, “mislabeled”, and an extensive list of possible relevant terms to build an inventory of potentially relevant cases. The search results and the date of the searches are reported in Appendix A.1.

After reviewing 2,168 cases, I identified 17 distinct cases that meet the following criteria for merging with the FSIS dataset later: (1) The initial lawsuit filing date is between July

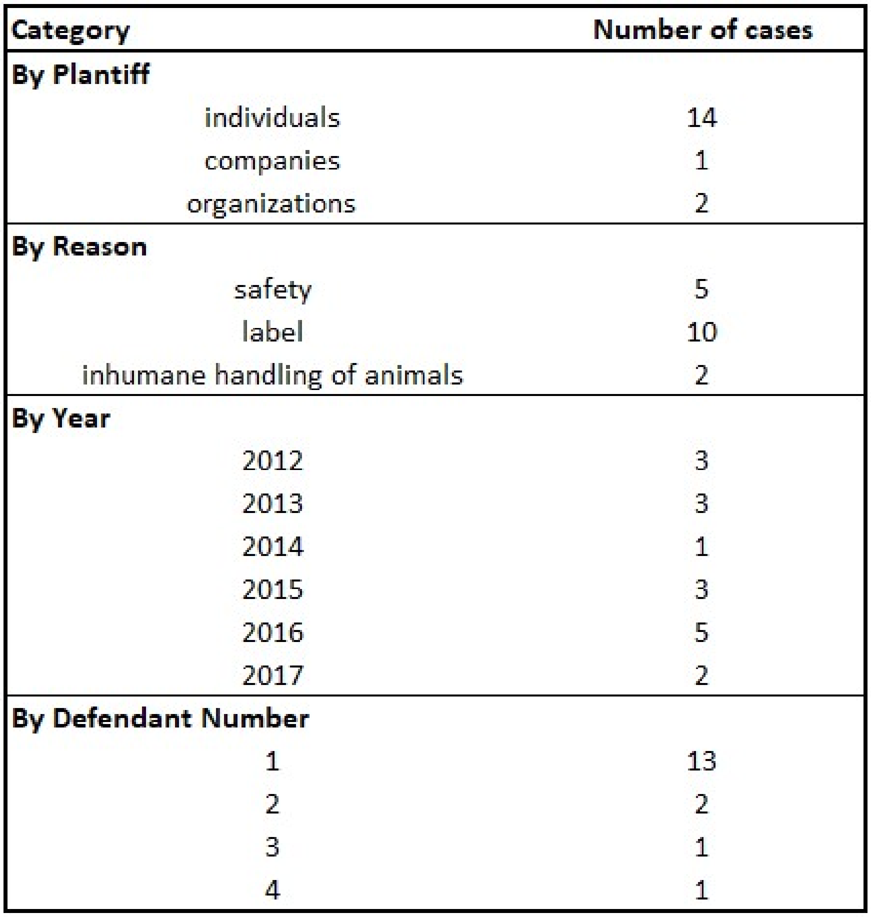
2012 and December 2017. (2) One or all defendants are firms that have been inspected by the FSIS. (3) The plaintiffs are not the government, but private entities such as companies, organizations, and individuals. (4) The lawsuits are related to meat and poultry product issues regulated by FSIS, such as safety or labeling or treatment of animals at plants.

For each of the final 17 lawsuits, I downloaded and reviewed the relevant documents to extract information, such as the processing plant location and the initial filing date if the case was an appeal. The resulting lawsuit dataset includes information such as the case name, defendant firm names, number of defendants, category of plaintiffs, and the tag for the relevant issue, such as safety, mislabeling, or humane handling. I also gathered the filing date of the lawsuit from the Bloomberg Law Database or the content of the lawsuit. Since the filing date is the earliest date that a defendant receives notice of the petition, I treated it as the lawsuit event date.

I merged the lawsuit dataset with the FSIS inspection dataset through the company name and/or location by month. If the lawsuit document mentioned the specific processing plants relevant to the case, then only these plants in the FSIS dataset are matched with the lawsuit with an event date on the filing date of the lawsuit. However, most of the lawsuits are not plant-specific. If that is the case, then all plants of the defendant firms in the FSIS dataset are matched to the lawsuit. As shown in Appendix A.2, there are 17 distinct civil lawsuits affecting 20 distinct defendants (companies) with 14 of them inspected by FSIS. Only 4 lawsuits could be traced to the plant level instead of the firm level, which is a challenge for our analysis. Table 9.2 shows a summary of the selected cases by categories. As can be seen, most of the plaintiffs are individuals; the most common reasons firms are sued are related to mislabeling; most of the cases have one defendant, and there is no specific trend of the number of cases each year.

Our sample covers July 2012 to December 2017 and includes 6394 distinct FSIS inspected meat and poultry plants. There are 122 plants associated with at least one food safety lawsuit (safety or mislabeling or humane handling issue) during the analysis time window. Campbell

Table 9.2: Summary of select lawsuits by different categories



Category	Number of cases
By Plantiff	
individuals	14
companies	1
organizations	2
By Reason	
safety	5
label	10
inhumane handling of animals	2
By Year	
2012	3
2013	3
2014	1
2015	3
2016	5
2017	2
By Defendant Number	
1	13
2	2
3	1
4	1

Soup Company plants are subject to 3 lawsuit events; plants of Whole Foods Markets and 2 Foster Farm plants are subject to 2 lawsuit events; 112 plants are subject to 1 lawsuit event; the rest of FSIS plants are not matched with any published lawsuits of the aforementioned type.

Chapter 10

Methodology

The goal of the study is to test the crowd-in or crowd-out effect of private lawsuits on FSIS regulatory behavior and to estimate the net effect of food safety lawsuits on a firm's food safety practice behavior. To do this, I adopted the event study approach. As described in Chapter 9, FSIS's monitoring intensity at different tasks can be measured by the share of directed tasks toward specific tasks; the plant's food safety practice level at different tasks is measured by the CRs of different routine inspection tasks. I use the food safety lawsuit filing date as the event time. [1] FSIS plants with lawsuit filing dates are regarded as treated units, and the rest of FSIS plants are control units. Identification is based on the exogenous timing of lawsuit filing dates after controlling for potential confounders. However, one drawback of the research design is that most lawsuits settle before going to court, and the decision to file a lawsuit is often the result of a breakdown in negotiations between the parties. This might cause anticipatory behavior in plants and inspectors. However, without knowing how long the anticipatory period could be and it may vary among different lawsuits, I choose to not address the issue by setting the event time at an arbitrary month before the filing time without making further assumptions. Also, setting the event time too far away from the filing date can cause some lawsuit events to run out of the pre-event period for estimation. With

---

[1] For plants with multiple lawsuits, I set the first lawsuit date as the event date. One can also make assumptions about how long it takes the effect to be stable and see whether it is appropriate to treat the subsequent events as separate events and include them in the sample.

this potential flaw in mind, instead, I will test for parallel trend assumption and interpret the estimation results with caution if the event study did not pass the parallel trend tests.

Before diving into the details of the event study method, it is worth mentioning the differences in the two parameters of interest that I plan to estimate below: the general and specific lawsuit effects on FSIS regulatory and plant behaviors. The general lawsuit effect refers to the case where no matter what type of lawsuit a plant is filed against (i.e., safety, labeling, or humane handling), the lawsuit event has a general impact on all types of FSIS regulatory intensities and plant food safety practices. However, the specific lawsuit effect captures the impact of lawsuits on monitoring intensity and firm behavior at the same type of issues the lawsuits address. For example, a labeling lawsuit may not have an impact on how FSIS inspectors monitor the sanitary tasks of the defendant plants or influence the sanitary food safety practices of the defendant plants, but it will have an impact on the monitoring intensity and plant food safety practices in labeling tasks. I will first estimate the general effect and then the specific effect.

I use equation 10.1 to model the general lawsuit effects on FSIS monitoring intensity and food safety practices of the sued plants.

$$y_{it} = \alpha_{-6+} L_{it}^{-6+} + \sum_{j=-5, j\neq -1}^{5} \alpha_j L_{it}^{j} + \alpha_{6+} L_{it}^{6+} + \gamma_i + \lambda_t + \epsilon_{it} \tag{10.1}$$

where y_{it} is one of the following measurements at plant i in time t: the CR of routine sanitary inspection tasks, CR of humane handling inspection tasks, CR of labeling inspection tasks, share of directed sanitary inspection tasks out of total sanitary inspection tasks, share of directed humane handling tasks out of total humane handling inspection tasks, or share of directed labeling inspection tasks out of total labeling inspection tasks. L_{it}^{j} is equal to 1 when the plant i at calendar time t is j month away from its lawsuit filing date. $-6+$ means 6 months before the lawsuit filing date (the 6th month is included). $+6$ means 6 months after the filing date (the 6th month is included). Thus, L_{it}^{j}, where $j = -6+, 6+$ is a bin indicator. λ_t is time-fixed effect, γ_i is the plant-fixed effect. The robust standard error is clustered at the plant level, allowing for the correlation of errors over time but not between individual plants.

Under the following assumptions: parallel trend, no anticipation effect, homogenous treatment effect, and the effect stays constant after 6 months (Sun and Abraham, 2021; Callaway and Sant'Anna, 2021), the two-way fixed effect estimators (TWFE) for α_js are consistent with the main parameters of interests, the average treatment effect on treated (ATT). It captures the dynamic general effects of the lawsuits on the various outcomes of interests. However, it is plausible that the assumptions are too strong in our setting, especially the third assumption. Lawsuits might have heterogenous effects across time and plants. In this case, the two-way fixed effect estimator is a weighted sum of ATTS where some weights may be negative (Goodman-Bacon, 2021). An alternative estimation method for weighted ATT that is robust to the heterogeneity of treatment effects is the interaction-weighted estimator (IW) proposed by Sun and Abraham (2021). Considering the plants whose filing date is in the same month as a cohort, the ATT weights are equal to each cohort's share in the relevant periods. The IW estimator is consistent for this weighted ATT under the first two assumptions. The sample for estimation includes all the control units, the FSIS plants that are not observed to be sued in our sample period, and a balanced sample of treated plants around their event time with a [-12,12] time window. I report the estimated results of both estimators: TWFE and IW estimator. The standard errors are clustered at the plant level.

Although a food safety lawsuit may not have a general effect on all aspects of inspection tasks, it can still have a specific impact on the aspects of regulatory monitoring behavior or plant food safety practices that are directly relevant to the lawsuit. To this end, I transform the sample data mentioned above in the following way and estimate the specific lawsuit effect using equation 10.2: (1) For all treated plants, the CR or share of the directed task variable is set to be the CR of the specific tasks or share of the specific directed tasks related to the type of lawsuit a treated plant faces. In total, there are three types of lawsuits (see

Table A.2), which are directly relevant to sanitary, labeling, and humane handling inspection tasks respectively. The lawsuit type or the relevant inspection type (1-to-1 correspondence) is denoted by k, where k = safety, labeling, humane handling. For example, if the plant i is sued for safety reasons (k = safety), then the outcome variable to measure the plant i's behavior relevant to the lawsuit at month$t$, $yikt$, is the CR of routine sanitary tasks. The same rule applies to the variable, share of directed tasks. (2) For control plants, each observation is duplicated 3 times to make sure that the 3 types of CRs for control plants are included in the estimation sample to help identify inspection task type-specific time trends. Each control plant has 3 rows of data with outcome variable CR y_{ikt} equal to CR of routine sanitary tasks, labeling tasks, and humane handling tasks. Following (1) & (2), the original estimation sample mentioned above is transformed from plant-month level to plant-month-type level.

$$y_{ikt} = \alpha_{-6+} L_{ikt}^{-6+} + \sum_{j=-5, j\neq -1}^{5} \alpha_j L_{ikt}^{j} + \alpha_{6+} L_{ikt}^{6+} + \gamma_i + \lambda_{kt} + \epsilon_{ikt} \tag{10.2}$$

where y_{ikt} is the CR or share of directed tasks corresponding to plant i's lawsuit/inspection type k for plant i at month t as explained above. The dummy variables for the type-specific, lawsuits L_{ikt}^{j}s, are defined the same way as mentioned in the general effect estimation model. λ_{kt} is time-by-type fixed effect, γ_i is plant-type fixed effect. The error terms of the same plant and different types of inspection tasks are correlated, so the standard error is clustered at the plant level.

Our parameters of interest are the average specific lawsuit treatment effect. Under stringent assumptions as explained in the general lawsuit effect setting, the TWFE estimator is consistent with the ATT I am interested in. To allow the results to be robust to treatment heterogeneity, I also report the IW estimator. Since for the same plant, the regulatory intensity or plants' food safety practices on different types of tasks are probably correlated, I cluster the standard errors to the plant level.

To recover the general or specific lawsuit effects, both estimators require parallel and no anticipation assumptions. They will be tested by joint hypothesis testing (10.3). Failing to reject the hypothesis test can lend some credibility to the identification assumptions.

$$H_0 : \alpha_{-6+} = \alpha{-5} = \alpha{-4} = ... = \alpha{-2} = 0 \qquad (10.3)$$

Chapter 11

Results and discussion

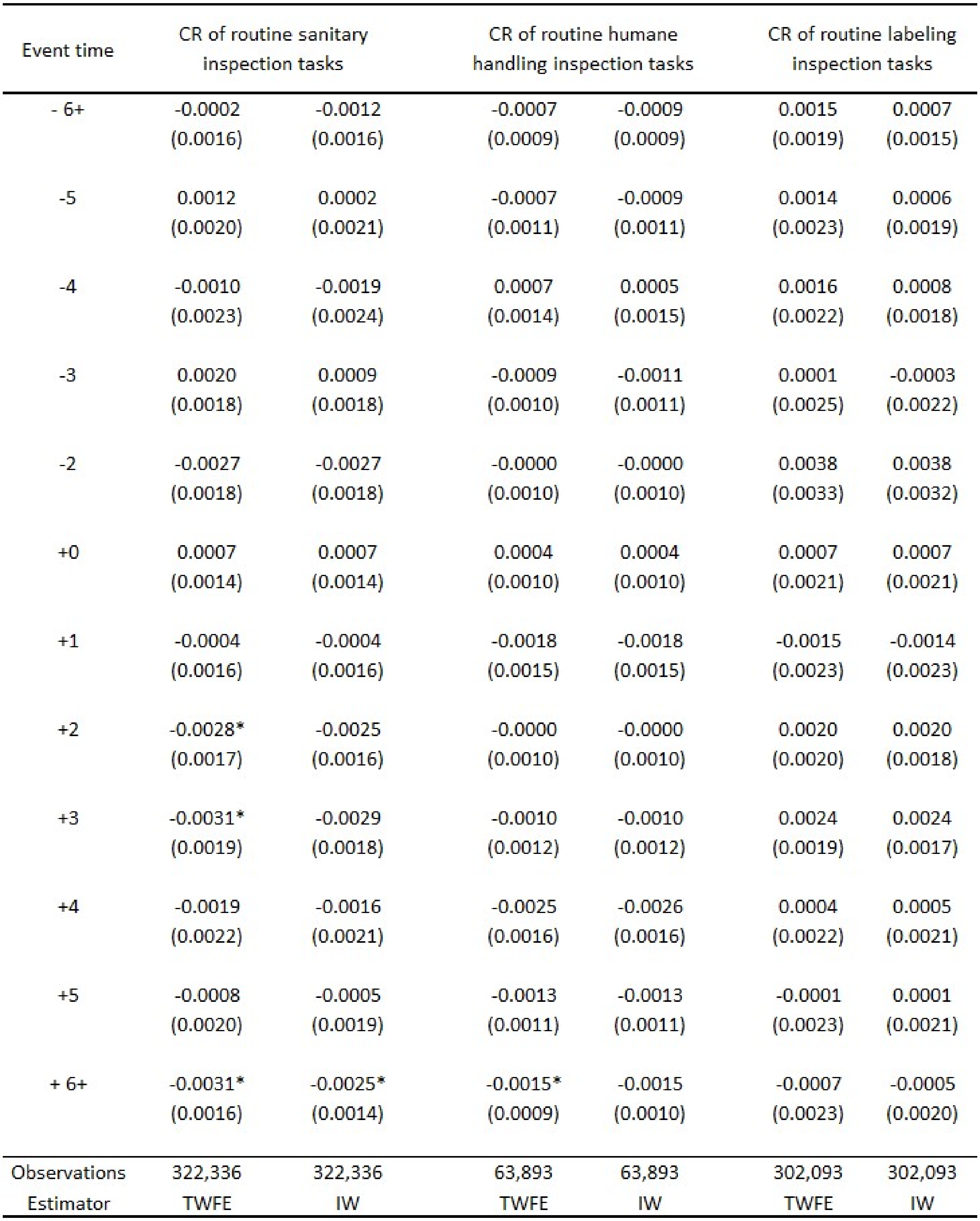
This section reports the regression results of the lawsuit's general and specific effect on plants' food safety practices and FSIS regulatory intensity using both TWFE and IW estimators. I do not find that food safety lawsuits have a statistically significant impact on the food safety practice of defendant plants. However, I do see a crowd-out effect of food safety lawsuits on FSIS inspectors' monitoring intensity in lawsuit-relevant inspection tasks.

Table 11.1 summarizes the results of general lawsuit impacts on plants' CRs of sanitary, humane animal handling, and labeling inspection tasks (see equation (10.1)). First, regardless of the CR type, our data do not reject the parallel and no anticipation effect assumption. The coefficients of pre-event dummies are not significantly different from zero. Second, the TWFE and IW estimators of the ATT generally have the same sign and similar magnitudes. In general, I do not observe a significantly different from zero effect of lawsuits on any CR of FSIS inspection tasks at the 5% significance level. I do observe a small negative effect of lawsuits on the CR of routine sanitary inspection tasks 6 months after the lawsuit filing month at the 10% significance level of both TWFE and IW estimators.

Table 11.2 shows the results of general lawsuit effects on FSIS monitoring intensity of sanitary, humane handling, and labeling tasks. From the first two columns, we can see that compared to one month before the filing month, FSIS inspectors' sanitary regulatory intensity might respond to the incoming lawsuits earlier than the actual filing date. It violates the no anticipation assumption. Even when I set 6 or 12 months before the lawsuit filing month, I still see a significantly different from zero coefficients for the pre-event time dummies. Due to the nature of the research setting and the length of the analysis sample, I am not able to address the issue of the violation of the no anticipation effect assumption. However, based on the magnitude of the point estimates, we can see that in general, they ramp down from pre-event time to post-event time, and more coefficients before post-event dummies are not statistically significant from zero. It could indicate a general "crowd-out" effect of the lawsuit for FSIS sanitary regulatory intensity, but I suggest taking the interpretation with a grain of salt. With regard to the FSIS regulatory intensity of humane handling and labeling inspection tasks, in general, I could not reject the no anticipation and parallel trend assumption and I do not observe a statistically significantly different from zero general lawsuit effect.

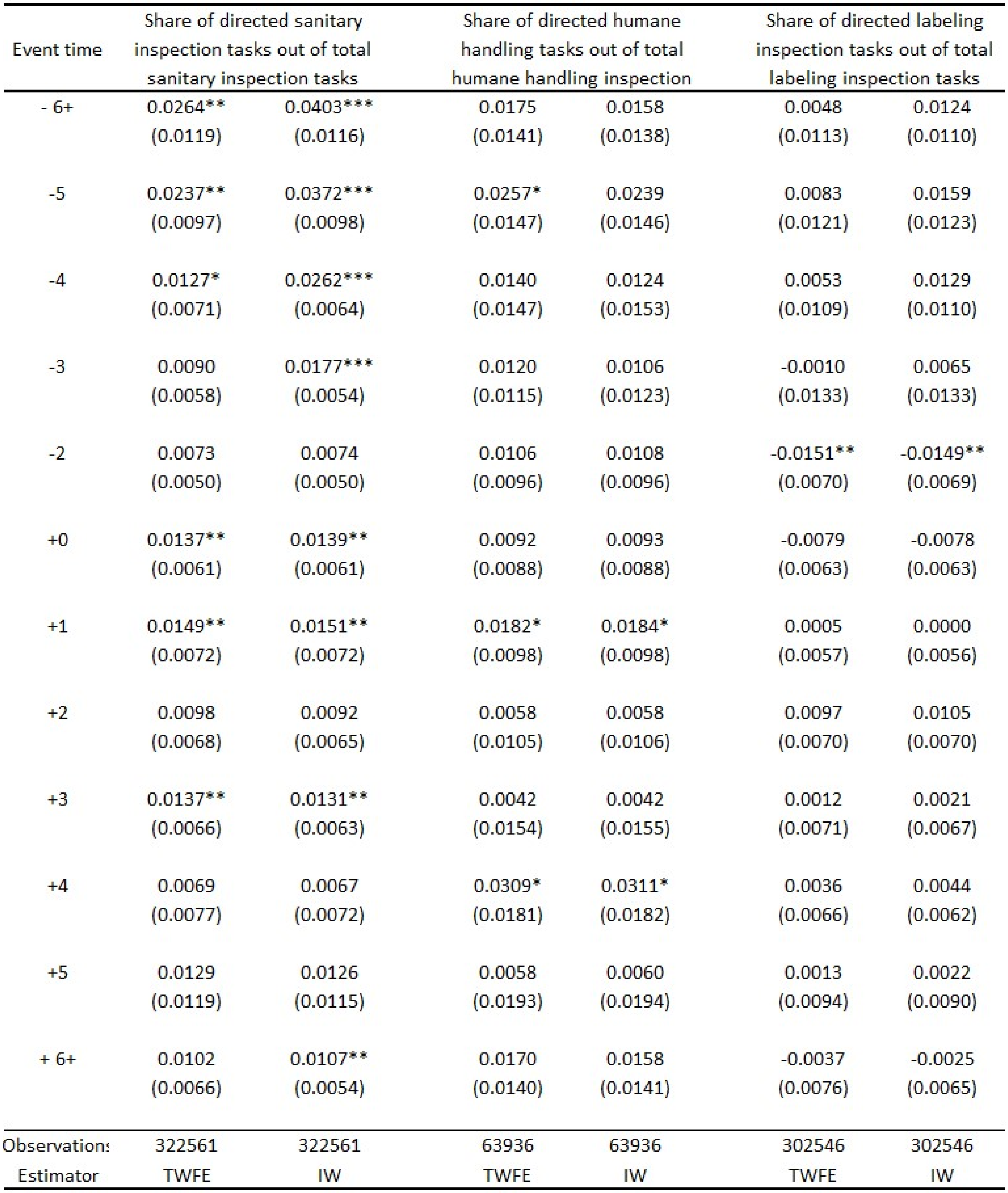
Table 11.3 presents the specific lawsuit effect on lawsuit-relevant food safety practice behaviors and FSIS monitoring intensity in the first two columns and last two columns respectively. Both the TWFE and heterogenous-robust IW estimators have the same sign and similar magnitudes for the parameters of interest, the dynamic effects for plants' food safety practices, and FSIS regulatory intensity. I do not find evidence in the data to reject our key assumptions. The estimated results of the parameters for the after-event dummies in the first column imply that there is no evidence that plants' lawsuit-relevant food safety practices respond to the filing of the lawsuit at a 5%/10% significance level. However, I do see a decrease in regulatory intensity for FSIS inspectors 4 months after the lawsuit filing month and the effect could be transitory. There is evidence that private lawsuits could crowd out the regulatory intensity of these lawsuit-relevant inspection tasks to help inspectors focus their limited resources in other places.

Table 11.1: Regression results of lawsuit's general impact on plants' food safety practices using different estimators



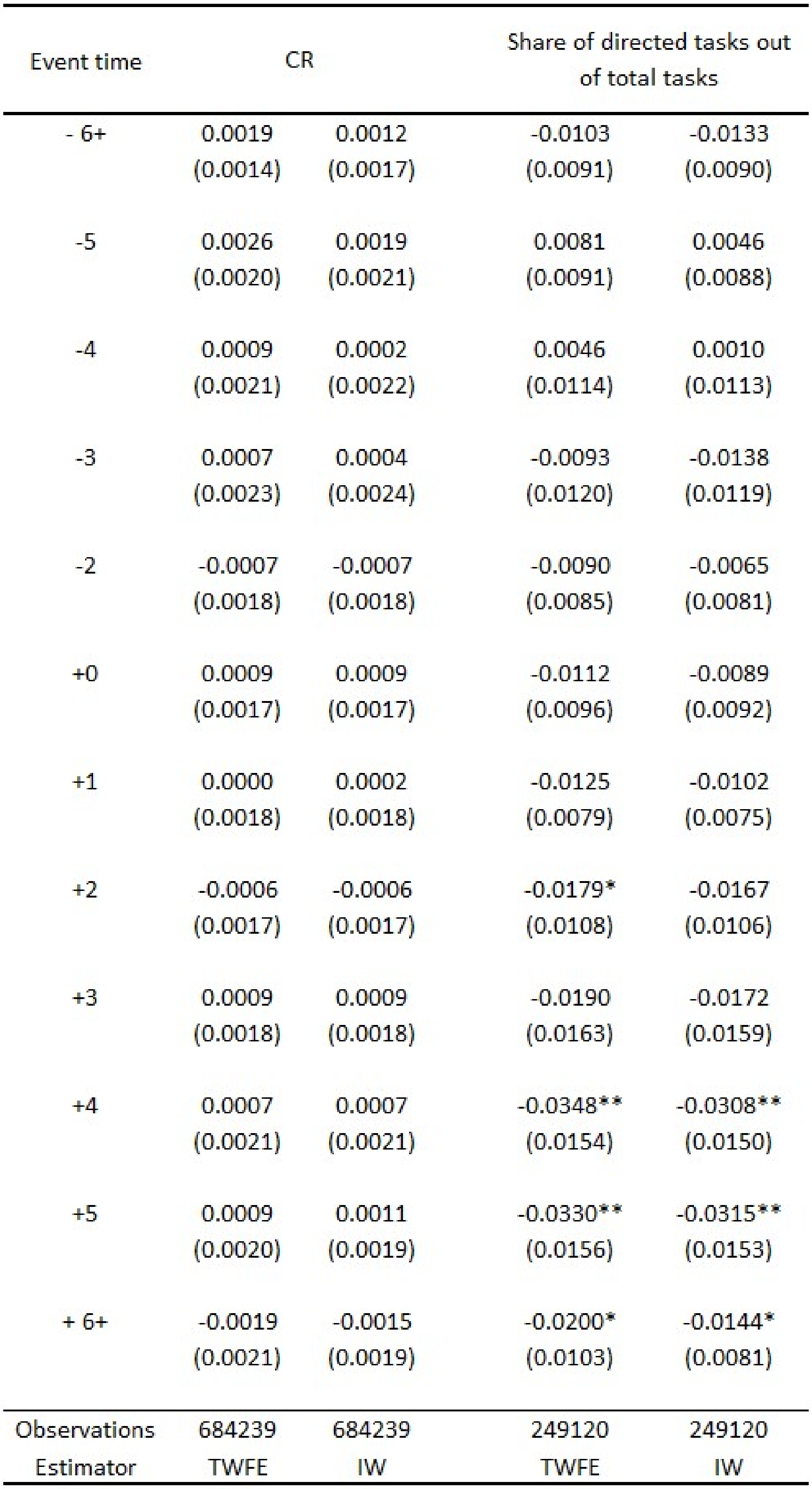
Event time	CR of routine sanitary inspection tasks		CR of routine humane handling inspection tasks		CR of routine labeling inspection tasks	
- 6+	-0.0002	-0.0012	-0.0007	-0.0009	0.0015	0.0007
	(0.0016)	(0.0016)	(0.0009)	(0.0009)	(0.0019)	(0.0015)
-5	0.0012	0.0002	-0.0007	-0.0009	0.0014	0.0006
	(0.0020)	(0.0021)	(0.0011)	(0.0011)	(0.0023)	(0.0019)
-4	-0.0010	-0.0019	0.0007	0.0005	0.0016	0.0008
	(0.0023)	(0.0024)	(0.0014)	(0.0015)	(0.0022)	(0.0018)
-3	0.0020	0.0009	-0.0009	-0.0011	0.0001	-0.0003
	(0.0018)	(0.0018)	(0.0010)	(0.0011)	(0.0025)	(0.0022)
-2	-0.0027	-0.0027	-0.0000	-0.0000	0.0038	0.0038
	(0.0018)	(0.0018)	(0.0010)	(0.0010)	(0.0033)	(0.0032)
+0	0.0007	0.0007	0.0004	0.0004	0.0007	0.0007
	(0.0014)	(0.0014)	(0.0010)	(0.0010)	(0.0021)	(0.0021)
+1	-0.0004	-0.0004	-0.0018	-0.0018	-0.0015	-0.0014
	(0.0016)	(0.0016)	(0.0015)	(0.0015)	(0.0023)	(0.0023)
+2	-0.0028*	-0.0025	-0.0000	-0.0000	0.0020	0.0020
	(0.0017)	(0.0016)	(0.0010)	(0.0010)	(0.0020)	(0.0018)
+3	-0.0031*	-0.0029	-0.0010	-0.0010	0.0024	0.0024
	(0.0019)	(0.0018)	(0.0012)	(0.0012)	(0.0019)	(0.0017)
+4	-0.0019	-0.0016	-0.0025	-0.0026	0.0004	0.0005
	(0.0022)	(0.0021)	(0.0016)	(0.0016)	(0.0022)	(0.0021)
+5	-0.0008	-0.0005	-0.0013	-0.0013	-0.0001	0.0001
	(0.0020)	(0.0019)	(0.0011)	(0.0011)	(0.0023)	(0.0021)
+ 6+	-0.0031*	-0.0025*	-0.0015*	-0.0015	-0.0007	-0.0005
	(0.0016)	(0.0014)	(0.0009)	(0.0010)	(0.0023)	(0.0020)
Observations	322,336	322,336	63,893	63,893	302,093	302,093
Estimator	TWFE	IW	TWFE	IW	TWFE	IW

Table 11.2: Regression results of lawsuit's general impact on FSIS regulatory intensity using different estimators



Event time	Share of directed sanitary inspection tasks out of total sanitary inspection tasks		Share of directed humane handling tasks out of total humane handling inspection		Share of directed labeling inspection tasks out of total labeling inspection tasks	
- 6+	0.0264**	0.0403***	0.0175	0.0158	0.0048	0.0124
	(0.0119)	(0.0116)	(0.0141)	(0.0138)	(0.0113)	(0.0110)
-5	0.0237**	0.0372***	0.0257*	0.0239	0.0083	0.0159
	(0.0097)	(0.0098)	(0.0147)	(0.0146)	(0.0121)	(0.0123)
-4	0.0127*	0.0262***	0.0140	0.0124	0.0053	0.0129
	(0.0071)	(0.0064)	(0.0147)	(0.0153)	(0.0109)	(0.0110)
-3	0.0090	0.0177***	0.0120	0.0106	-0.0010	0.0065
	(0.0058)	(0.0054)	(0.0115)	(0.0123)	(0.0133)	(0.0133)
-2	0.0073	0.0074	0.0106	0.0108	-0.0151**	-0.0149**
	(0.0050)	(0.0050)	(0.0096)	(0.0096)	(0.0070)	(0.0069)
+0	0.0137**	0.0139**	0.0092	0.0093	-0.0079	-0.0078
	(0.0061)	(0.0061)	(0.0088)	(0.0088)	(0.0063)	(0.0063)
+1	0.0149**	0.0151**	0.0182*	0.0184*	0.0005	0.0000
	(0.0072)	(0.0072)	(0.0098)	(0.0098)	(0.0057)	(0.0056)
+2	0.0098	0.0092	0.0058	0.0058	0.0097	0.0105
	(0.0068)	(0.0065)	(0.0105)	(0.0106)	(0.0070)	(0.0070)
+3	0.0137**	0.0131**	0.0042	0.0042	0.0012	0.0021
	(0.0066)	(0.0063)	(0.0154)	(0.0155)	(0.0071)	(0.0067)
+4	0.0069	0.0067	0.0309*	0.0311*	0.0036	0.0044
	(0.0077)	(0.0072)	(0.0181)	(0.0182)	(0.0066)	(0.0062)
+5	0.0129	0.0126	0.0058	0.0060	0.0013	0.0022
	(0.0119)	(0.0115)	(0.0193)	(0.0194)	(0.0094)	(0.0090)
+ 6+	0.0102	0.0107**	0.0170	0.0158	-0.0037	-0.0025
	(0.0066)	(0.0054)	(0.0140)	(0.0141)	(0.0076)	(0.0065)
Observations	322561	322561	63936	63936	302546	302546
Estimator	TWFE	IW	TWFE	IW	TWFE	IW

Table 11.3: Regression results of lawsuit's specific impact on FSIS regulatory intensity and plants' food safety practices using different estimators



Event time	CR		Share of directed tasks out of total tasks	
- 6+	0.0019	0.0012	-0.0103	-0.0133
	(0.0014)	(0.0017)	(0.0091)	(0.0090)
-5	0.0026	0.0019	0.0081	0.0046
	(0.0020)	(0.0021)	(0.0091)	(0.0088)
-4	0.0009	0.0002	0.0046	0.0010
	(0.0021)	(0.0022)	(0.0114)	(0.0113)
-3	0.0007	0.0004	-0.0093	-0.0138
	(0.0023)	(0.0024)	(0.0120)	(0.0119)
-2	-0.0007	-0.0007	-0.0090	-0.0065
	(0.0018)	(0.0018)	(0.0085)	(0.0081)
+0	0.0009	0.0009	-0.0112	-0.0089
	(0.0017)	(0.0017)	(0.0096)	(0.0092)
+1	0.0000	0.0002	-0.0125	-0.0102
	(0.0018)	(0.0018)	(0.0079)	(0.0075)
+2	-0.0006	-0.0006	-0.0179*	-0.0167
	(0.0017)	(0.0017)	(0.0108)	(0.0106)
+3	0.0009	0.0009	-0.0190	-0.0172
	(0.0018)	(0.0018)	(0.0163)	(0.0159)
+4	0.0007	0.0007	-0.0348**	-0.0308**
	(0.0021)	(0.0021)	(0.0154)	(0.0150)
+5	0.0009	0.0011	-0.0330**	-0.0315**
	(0.0020)	(0.0019)	(0.0156)	(0.0153)
+ 6+	-0.0019	-0.0015	-0.0200*	-0.0144*
	(0.0021)	(0.0019)	(0.0103)	(0.0081)
Observations	684239	684239	249120	249120
Estimator	TWFE	IW	TWFE	IW

Chapter 12

Conclusion

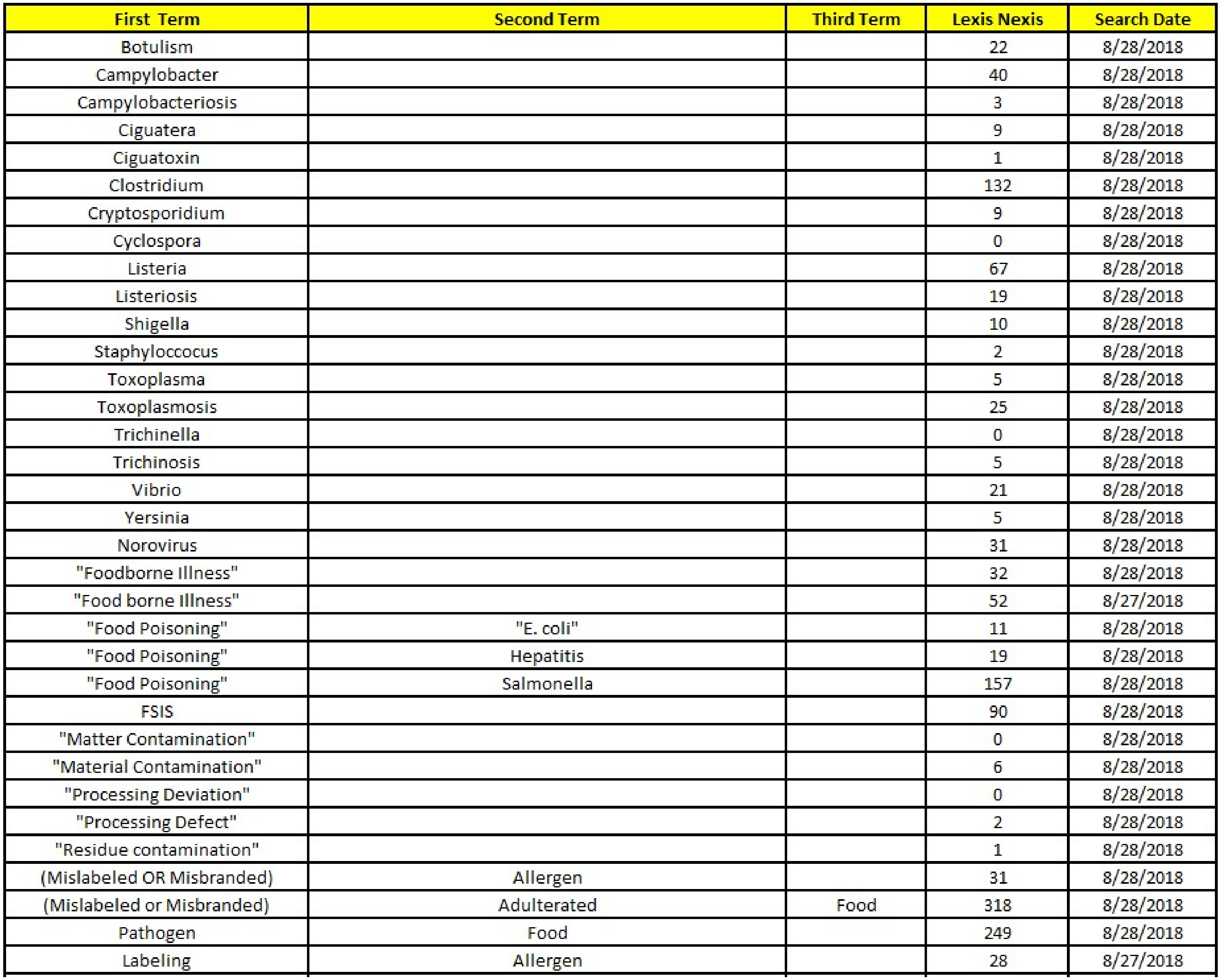
Part II of the dissertation measures the impact of private lawsuits regarding food safety, labeling, and humane animal handling on public regulatory intensity and food safety practices of defendant plants. First, I find that food safety lawsuits have a net crowd-out effect on the regulatory intensity in inspection tasks that are directly relevant to the food safety disputed issues in the lawsuit. Contrary to what some might think, namely that lawsuits draw more regulatory attention to the less conforming plants leading to an increase in monitoring intensity for the non-compliance area of the defendant plant, our empirical analysis finds that private lawsuits might help direct regulators' limited attention and resources to other places and decrease the monitoring intensity of the issues brought up in the lawsuit. However, I cannot confidently conclude whether food safety lawsuits have a general crowd-out effect on all monitoring activities of the defendant plant, which is a good future research area. Second, I do not find that food safety lawsuits have a statistically significantly different from zero impact on defendant plants' food safety practices.

Our results provide the first micro-level empirical evidence of the effects of private food safety lawsuits. It expands the knowledge of the interactions of lawsuits, public regulation, and plant behaviors in the context of food safety issues in the US meat and poultry industry.

Chapter A

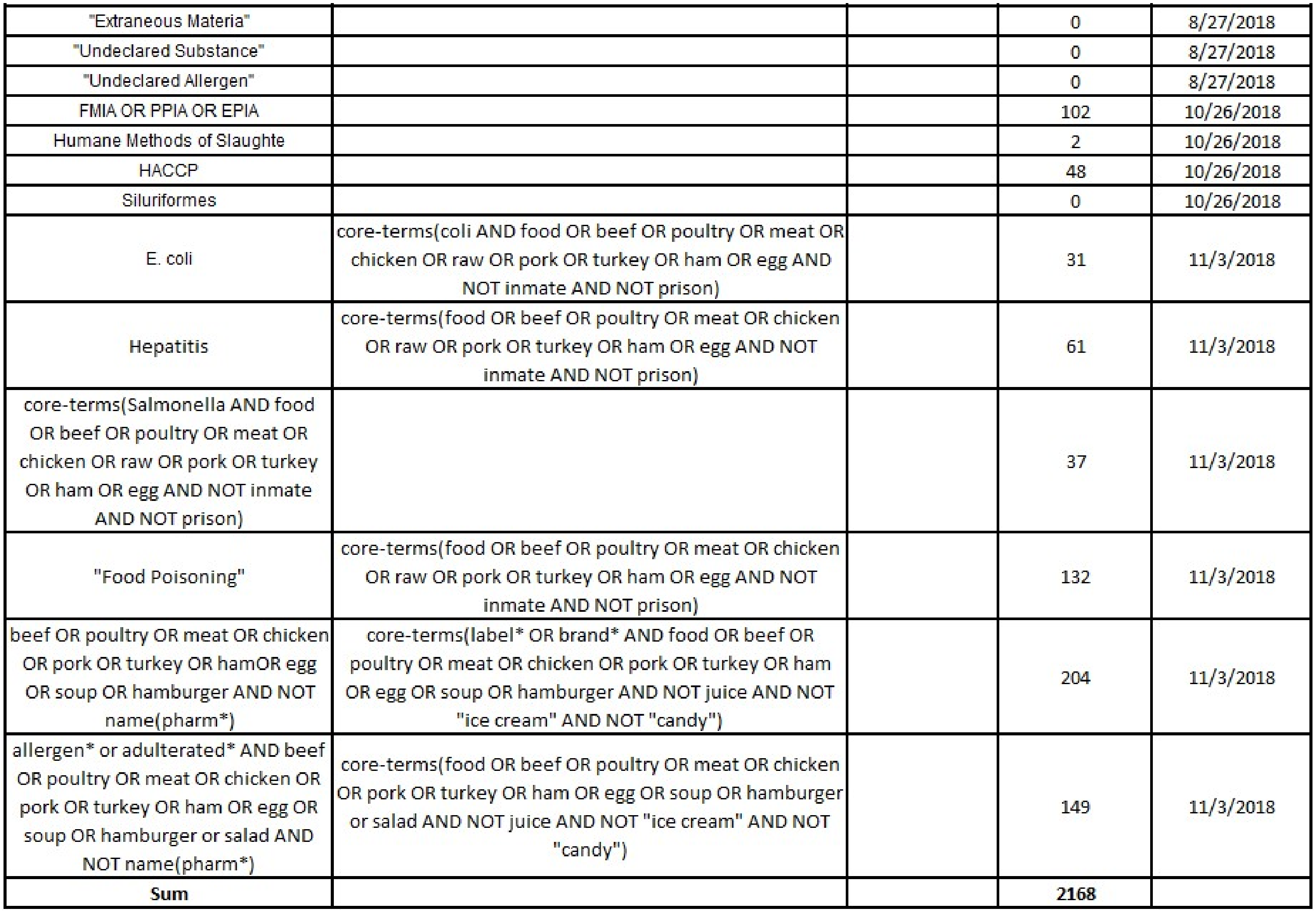
Appendix

Table A.1: Lexis Nexis search results by November 2018



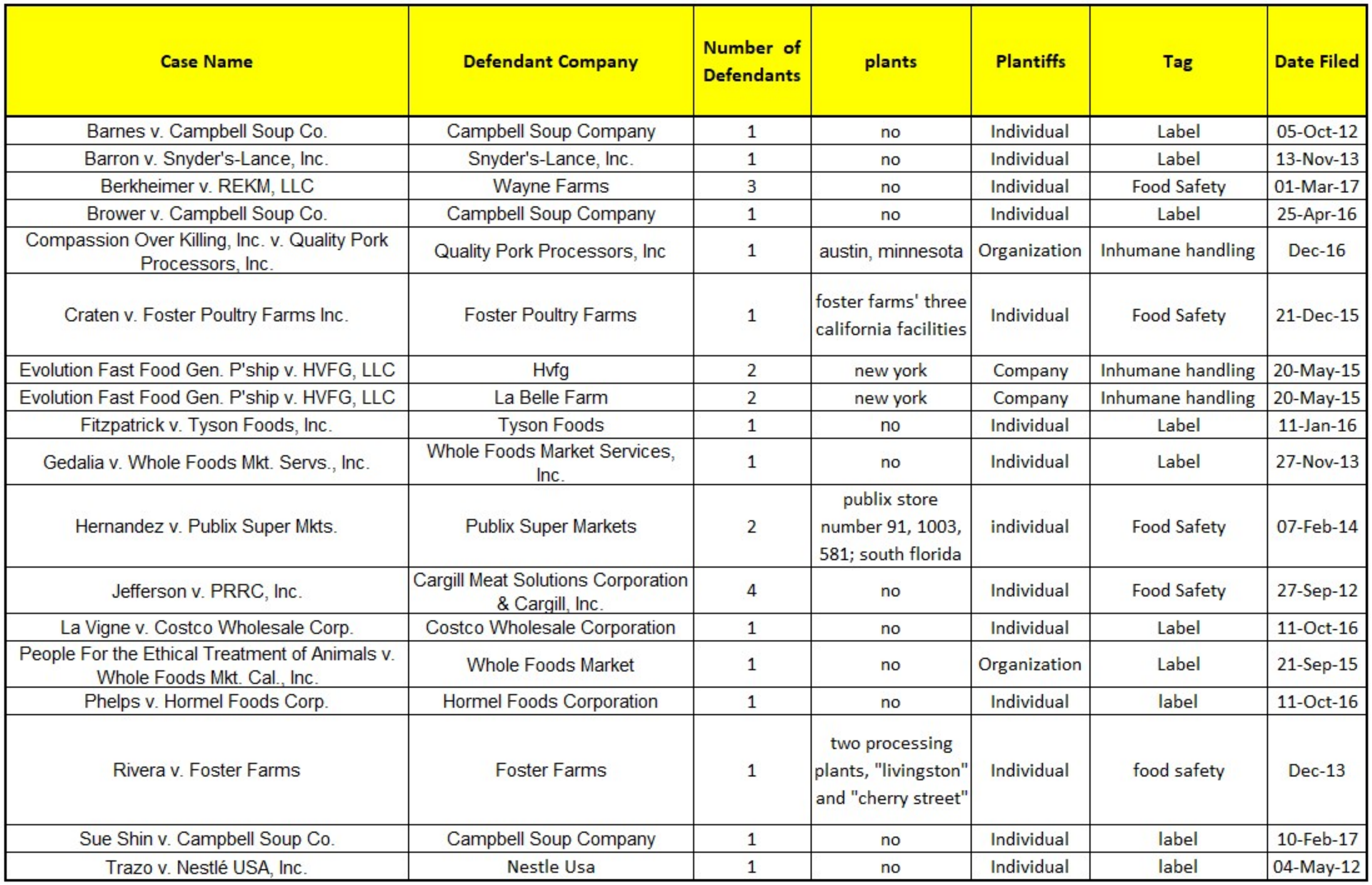
First Term	Second Term	Third Term	Lexis Nexis	Search Date
Botulism			22	8/28/2018
Campylobacter			40	8/28/2018
Campylobacteriosis			3	8/28/2018
Ciguatera			9	8/28/2018
Ciguatoxin			1	8/28/2018
Clostridium			132	8/28/2018
Cryptosporidium			9	8/28/2018
Cyclospora			0	8/28/2018
Listeria			67	8/28/2018
Listeriosis			19	8/28/2018
Shigella			10	8/28/2018
Staphyloccocus			2	8/28/2018
Toxoplasma			5	8/28/2018
Toxoplasmosis			25	8/28/2018
Trichinella			0	8/28/2018
Trichinosis			5	8/28/2018
Vibrio			21	8/28/2018
Yersinia			5	8/28/2018
Norovirus			31	8/28/2018
"Foodborne Illness"			32	8/28/2018
"Food borne Illness"			52	8/27/2018
"Food Poisoning"	"E. coli"		11	8/28/2018
"Food Poisoning"	Hepatitis		19	8/28/2018
"Food Poisoning"	Salmonella		157	8/28/2018
FSIS			90	8/28/2018
"Matter Contamination"			0	8/28/2018
"Material Contamination"			6	8/28/2018
"Processing Deviation"			0	8/28/2018
"Processing Defect"			2	8/28/2018
"Residue contamination"			1	8/28/2018
(Mislabeled OR Misbranded)	Allergen		31	8/28/2018
(Mislabeled or Misbranded)	Adulterated	Food	318	8/28/2018
Pathogen	Food		249	8/28/2018
Labeling	Allergen		28	8/27/2018

Table continued



"Extraneous Materia"			0	8/27/2018
"Undeclared Substance"			0	8/27/2018
"Undeclared Allergen"			0	8/27/2018
FMIA OR PPIA OR EPIA			102	10/26/2018
Humane Methods of Slaughte			2	10/26/2018
HACCP			48	10/26/2018
Siluriformes			0	10/26/2018
E. coli	core-terms(coli AND food OR beef OR poultry OR meat OR chicken OR raw OR pork OR turkey OR ham OR egg AND NOT inmate AND NOT prison)		31	11/3/2018
Hepatitis	core-terms(food OR beef OR poultry OR meat OR chicken OR raw OR pork OR turkey OR ham OR egg AND NOT inmate AND NOT prison)		61	11/3/2018
core-terms(Salmonella AND food OR beef OR poultry OR meat OR chicken OR raw OR pork OR turkey OR ham OR egg AND NOT inmate AND NOT prison)			37	11/3/2018
"Food Poisoning"	core-terms(food OR beef OR poultry OR meat OR chicken OR raw OR pork OR turkey OR ham OR egg AND NOT inmate AND NOT prison)		132	11/3/2018
beef OR poultry OR meat OR chicken OR pork OR turkey OR hamOR egg OR soup OR hamburger AND NOT name(pharm*)	core-terms(label* OR brand* AND food OR beef OR poultry OR meat OR chicken OR pork OR turkey OR ham OR egg OR soup OR hamburger AND NOT juice AND NOT "ice cream" AND NOT "candy")		204	11/3/2018
allergen* or adulterated* AND beef OR poultry OR meat OR chicken OR pork OR turkey OR ham OR egg OR soup OR hamburger or salad AND NOT name(pharm*)	core-terms(food OR beef OR poultry OR meat OR chicken OR pork OR turkey OR ham OR egg OR soup OR hamburger or salad AND NOT juice AND NOT "ice cream" AND NOT "candy")		149	11/3/2018
Sum			**2168**	

Table A.2: Selected lawsuits for event studies



Case Name	Defendant Company	Number of Defendants	plants	Plantiffs	Tag	Date Filed
Barnes v. Campbell Soup Co.	Campbell Soup Company	1	no	Individual	Label	05-Oct-12
Barron v. Snyder's-Lance, Inc.	Snyder's-Lance, Inc.	1	no	Individual	Label	13-Nov-13
Berkheimer v. REKM, LLC	Wayne Farms	3	no	Individual	Food Safety	01-Mar-17
Brower v. Campbell Soup Co.	Campbell Soup Company	1	no	Individual	Label	25-Apr-16
Compassion Over Killing, Inc. v. Quality Pork Processors, Inc.	Quality Pork Processors, Inc	1	austin, minnesota	Organization	Inhumane handling	Dec-16
Craten v. Foster Poultry Farms Inc.	Foster Poultry Farms	1	foster farms' three california facilities	Individual	Food Safety	21-Dec-15
Evolution Fast Food Gen. P'ship v. HVFG, LLC	Hvfg	2	new york	Company	Inhumane handling	20-May-15
Evolution Fast Food Gen. P'ship v. HVFG, LLC	La Belle Farm	2	new york	Company	Inhumane handling	20-May-15
Fitzpatrick v. Tyson Foods, Inc.	Tyson Foods	1	no	Individual	Label	11-Jan-16
Gedalia v. Whole Foods Mkt. Servs., Inc.	Whole Foods Market Services, Inc.	1	no	Individual	Label	27-Nov-13
Hernandez v. Publix Super Mkts.	Publix Super Markets	2	publix store number 91, 1003, 581; south florida	individual	Food Safety	07-Feb-14
Jefferson v. PRRC, Inc.	Cargill Meat Solutions Corporation & Cargill, Inc.	4	no	Individual	Food Safety	27-Sep-12
La Vigne v. Costco Wholesale Corp.	Costco Wholesale Corporation	1	no	Individual	Label	11-Oct-16
People For the Ethical Treatment of Animals v. Whole Foods Mkt. Cal., Inc.	Whole Foods Market	1	no	Organization	Label	21-Sep-15
Phelps v. Hormel Foods Corp.	Hormel Foods Corporation	1	no	Individual	label	11-Oct-16
Rivera v. Foster Farms	Foster Farms	1	two processing plants, "livingston" and "cherry street"	Individual	food safety	Dec-13
Sue Shin v. Campbell Soup Co.	Campbell Soup Company	1	no	Individual	label	10-Feb-17
Trazo v. Nestlé USA, Inc.	Nestle Usa	1	no	Individual	label	04-May-12

References

Akerlof, George A. 1970. “The Market for "Lemons": Quality Uncertainty and the Market Mechanism.” *The Quarterly Journal of Economics*, 84(3): 480–500.

Albersmeier, Friederike, Holger Schulze, Gabriele Jahn, and Achim Spiller. 2009. “The reliability of third-party certification in the food chain: From checklists to risk-oriented auditing.” *Food Control*, 20(10): 927–935.

Anderson, Michael L. 2008. “Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects.” *Journal of the American Statistical Association*, 103: 1481–1495.

Anders, Sven, Diogo Souza-Monteiro, and Elodie Rouviere. 2010. “Competition and credibility of private third-party certification in international food supply.” *Journal of International Food and Agribusiness Marketing*, 22(3): 328–341.

Bai, Jie. 2018. “Melons as Lemons: Asymmetric Information, Consumer Learning and Quality Provision.” *Working paper.*

Bai, Jie, Ludovica Gazze, and Yukun Wang. 2022. “Collective reputation in trade: Evidence from the Chinese dairy industry.” *Review of Economics and Statistics*, 104(6): 1121–1137.

Bar, Talia, and Yuqing Zheng. 2018. "Choosing Certifiers: Evidence from the British Retail Consortium Food Safety Standard." *American Journal of Agricultural Economics*, 101(1): 74–88.

Benjamini, Yoav, Abba M Krieger, and Daniel Yekutieli. 2006. "Adaptive linear step-up procedures that control the false discovery rate." *Biometrika*, 93(3): 491–507.

Benjamini, Yoav, and Yosef Hochberg. 1995. "Controlling the false discovery rate: a practical and powerful approach to multiple testing." *Journal of the Royal statistical society: series B (Methodological)*, 57(1): 289–300.

Bonferroni, Carlo E. 1935. "Il calcolo delle assicurazioni su gruppi di teste." *Studi in onore del professore salvatore ortu carboni*, 13–60.

Boys, Kathryn A, Julie A Caswell, and Sandra A Hoffmann. 2015. "the Business of Safety Food: an Assessment of the Global Food Safety Certification Industry." *Selected Paper prepared for presentation at the 2015 Agricultural & Applied Economics Association and Western Agricultural Economics Association Annual Meeting, San Francisco, CA, July 26-28.*

Buzby, Jean C, and Paul D Frenzen. 1999. "Food safety and product liability." *Food Policy*, 24(6): 637–651.

Callaway, Brantly, and Pedro H.C. Sant'Anna. 2021. "Difference-in-Differences with multiple time periods." *Journal of Econometrics*, 225(2): 200–230. Themed Issue: Treatment Effect 1.

Castka, Pavel, Daniel Prajogo, Amrik Sohal, and Andy C.L. Yeung. 2015. "Understanding firms selection of their ISO 9000 third-party certifiers." *International Journal of Production Economics*, 162: 125–133.

Caswell, Julie A, Kathryn A Boys, Alyssa A Danilow, and Kathryn E Lynch. 2017. "Food Certification Industry Capacity and Ability to Comply with FSMA Final Rule on Accredited Third-Party Certification." *Selected Paper prepared for presentation at the 2017 Agricultural & Applied Economics Association Annual Meeting, July 30-August 1, Chicago, Illinois.*

CDC. 2011. "Foodborne illness estimates help us understand an important public health problem." *https://www.cdc.gov/foodborneburden/estimates-overview.html*.

Connally, Elizabeth Haws. 2009. "Good food safety practices: Managing risks to reduce or avoid legal liability."

Crandall, Philip G., Andy Mauromoustakos, Corliss A. O'Bryan, Kevin C. Thompson, Frank Yiannas, Kerry Bridges, and Catherine Francois. 2017. "Impact of the global food safety initiative on food safety worldwide: Statistical analysis of a survey of international food processors." *Journal of Food Protection*, 80(10): 1613–1622.

De Chaisemartin, Clément, and Xavier D'Haultfoeuille. 2022. "Difference-in-differences estimators of intertemporal treatment effects." National Bureau of Economic Research.

Dranove, David, and Ginger Zhe Jin. 2010. "Quality disclosure and certification: Theory and practice." *Journal of Economic Literature*, 48(4): 935–963.

Duflo, Esther, Michael Greenstone, Rohini Pande, and Nicholas Ryan. 2013. "Truth-Telling by Third-Party Auditors and the Response of Polluting Firms: Experimental Evidence from India." *Quarterly Journal of Economics*, 1499–1545.

Goodman-Bacon, Andrew. 2021. "Difference-in-differences with variation in treatment timing." *Journal of Econometrics*, 225(2): 254–277. Themed Issue: Treatment Effect 1.

Gursoy, Dogan. 2019. “Foodborne illnesses: An overview of hospitality operations liability.” *Journal of Hospitality*, 1(1): 41–49.

Hatanaka, Maki, Carmen Bain, and Lawrence Busch. 2005. “Third-party certification in the global agrifood system.” *Food Policy*, 30(3): 354–369.

Henson, Spencer, and Thomas Reardon. 2005. “Private agri-food standards : Implications for food policy and the agri-food system.” 30: 241–253.

Holm, Sture. 1979. “A simple sequentially rejective multiple test procedure.” *Scandinavian journal of statistics*, 65–70.

Hu, Lijiao, and Yuqing Zheng. 2019. “The Impacts of Food Safety Certification on Produceràs Safety outcomes.” *Selected Paper prepared for presentation at the 2019 Agricultural & Applied Economics Association Annual Meeting, Atlanta, GA, July 21-23.*

Jacobs, Andrew. 2021. “Lawsuits Over "Misleading" Food Labels Surge as Groups Cite Lax U.S. Oversight.” *The New York Times.*

Jahn, Gabriele, Matthias Schramm, and Achim Spiller. 2005. “Reliability of Certification : Quality Labels as a Consumer Policy Tool.” *Journal of Consumer Policy*, 28: 53–73.

Jin, Ginger, and Phillip Leslie. 2003. “Quality : Evidence From Restaurant Hygiene.” *Quarterly Journal of Economics*, 118(2): 409–451.

Keohane, Nathaniel O, Erin T Mansur, and Andrey Voynov. 2009. “Averting regulatory enforcement: Evidence from new source review.” *Journal of Economics & Management Strategy*, 18(1): 75–104.

Langpap, Christian. 2007. “Pollution abatement with limited enforcement power and citizen suits.” *Journal of Regulatory Economics*, 31(1): 57–81.

Langpap, Christian, and Jay P Shimshack. 2010. “Private citizen suits and public enforcement: Substitutes or complements?” *Journal of Environmental Economics and Management*, 59(3): 235–249.

Mahdu, Omchand. 2015. “Penalties for foodborne illness: jury decisions and awards in foodborne illness lawsuits.” PhD diss. Virginia Tech.

Marinovic, Iván, Andrzej Skrzypacz, and Felipe Varas. 2018. “Dynamic certification and reputation for quality.” *American Economic Journal: Microeconomics*, 10(2): 58–82.

MarketsandMarkets. 2020. “Food Certification Market by Type, Application, Risk Category, and Region - Global Forecast for 2025.” *https://www.marketsandmarkets.com/Market-Reports/food-certification-market-82063283.html*.

Mérel, Pierre, Ariel Ortiz-Bobea, and Emmanuel Paroissien. 2021. “How big is the "lemons" problem? Historical evidence from French wines.” *European Economic Review*, 138: 103824.

Negowetti, Nicole. 2014. “Food Labeling Litigation: Exposing Gaps in the FDA’s Resources and Regulatory Authority.” *Brookings Institution*, 1–31.

Ollinger, Michael, and John Bovay. 2018. “Pass or Fail: Economic Incentives to Reduce Salmonella Contamination in Ground Beef Sold to the National School Lunch Program.” *American Journal of Agricultural Economics*, 100(2): 414–433.

Ollinger, Michael, and John Bovay. 2019. “Producer Response to Public Disclosure of Food-Safety Information.” *American Journal of Agricultural Economics*, 0(0): 1–16.

Pouliot, Sebastien, and H. Holly Wang. 2018. “Information, Incentives, and Government Intervention for Food Safety.” *Annual Review of Resource Economics*, 10(1): 83–103.

Romano, Joseph P, and Michael Wolf. 2005*a*. “Exact and approximate stepdown methods for multiple hypothesis testing.” *Journal of the American Statistical Association*, 100(469): 94–108.

Romano, Joseph P, and Michael Wolf. 2005*b*. “Stepwise multiple testing as formalized data snooping.” *Econometrica*, 73(4): 1237–1282.

Romano, Joseph P, and Michael Wolf. 2016. “Efficient computation of adjusted p-values for resampling-based stepdown multiple testing.” *Statistics & Probability Letters*, 113: 38–40.

Rosenthal, Isadore, and Howard Kunreuther. 2010. “Roles for Third Parties in Implementing USDA Food Safety and Inspection Service (FSIS)’ s Food Safety Process Management Programs.” *Risk Management and Decision Processes Center, The Wharton School of the University of Pennsylvania.*

Schmidheiny, Kurt, and Sebastian Siegloch. 2019. “On event study designs and distributed-lag models: Equivalence, generalization and practical implications.”

Spence, Michael. 1973. “Job Market Signaling.” *The Quarterly Journal of Economics*, 87(3): 355–374.

Sun, Liyang, and Sarah Abraham. 2021. “Estimating dynamic treatment effects in event studies with heterogeneous treatment effects.” *Journal of Econometrics*, 225(2): 175–199. Themed Issue: Treatment Effect 1.

Tanner, Bob. 2000. “Independent assessment by third-party certification bodies.” *Food Control*, 11(5): 415–417.

USDA. 2019. “How many food inspectors are employed by the Food Safety and Inspection Service?” `https://ask.usda.gov/s/article/How-many-food-inspectors-are-employed-by-FSIS`.

USDA. 2022. “2022 Budget Summary.” `chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.usda.gov/sites/default/files/documents/2022-budget-summary.pdf`.

USDA-ERS. 2021. “Cost Estimates of Foodborne Illnesses.” `https://www.ers.usda.gov/data-products/cost-estimates-of-foodborne-illnesses.aspx#.VDW27r4mUfy`.

Varas, Felipe, Iván Marinovic, and Andrzej Skrzypacz. 2020. “Random Inspections and Periodic Reviews: Optimal Dynamic Monitoring.” *The Review of Economic Studies*, 87(6): 2893–2937.

Veazie, Peter J. 2006. “When to combine hypotheses and adjust for multiple tests.” *Health Services Research*, 41: 804–818.

Weinroth, Margaret D., Aeriel D. Belk, and Keith E. Belk. 2018. “History, development, and current status of food safety systems worldwide.” *Animal Frontiers*, 8(4): 9–15.

Westfall, Peter H, and S Stanley Young. 1993. *Resampling-based multiple testing: Examples and methods for p-value adjustment.* Vol. 279, John Wiley & Sons.

Zheng, Yuqing, and Talia R. Bar. 2019. “Certifier Competition and Audit Grades.” *SSRN Electronic Journal*, 1–39.