Spring 4-2017

What are the Financial Implications of Public Quality Disclosure? Evidence from New York City's Restaurant Food Safety Grading Policy

Rachel Meltzer
The New School, meltzerr@newschool.edu

Michah W. Rothbart
Syracuse University, mwrothba@syr.edu

Amy Ellen Schwartz
Syracuse University, amyschwartz@syr.edu

Thad Calabrese
New York University, thad.calabrese@nyu.edu

Diana Silver
New York University, drs1@nyu.edu

See next page for additional authors

Author(s)/Creator(s)
Rachel Meltzer, Michah W. Rothbart, Amy Ellen Schwartz, Thad Calabrese, Diana Silver, Tod Mijanovich, and Meryle Weinstein

CENTER FOR POLICY RESEARCH
THE MAXWELL SCHOOL

WORKING PAPER SERIES

What are the Financial Implications of Public Quality Disclosure? Evidence from New York City's Restaurant Food Safety Grading Policy

Rachel Meltzer, Michah W. Rothbart, Amy Ellen Schwartz, Thad Calabrese, Diana Silver, Tod Mijanovich, and Meryle Weinstein

Paper No. 202
April 2017

ISSN: 1525-3066

Maxwell
Syracuse University
Center for Policy Research
426 Eggers Hall
Syracuse University
Syracuse, NY 13244-1020
(315) 443-3114/ email: ctrpol@syr.edu
http://www.maxwell.syr.edu/CPR_Working_Papers.aspx

CENTER FOR POLICY RESEARCH –Spring 2017

Leonard M. Lopoo, Director
Professor of Public Administration and International Affairs (PAIA)

Associate Directors

Margaret Austin
Associate Director, Budget and Administration

John Yinger
Trustee Professor of ECON and PAIA
Associate Director, Metropolitan Studies

SENIOR RESEARCH ASSOCIATES

Badi Baltagi, ECON
Robert Bifulco, PAIA
Leonard Burman, PAIA
Thomas Dennison, PAIA
Alfonso Flores-Lagunes, ECON
Sarah Hamersma, PAIA
Madonna Harrington Meyer, SOC
William Horrace, ECON
Yilin Hou, PAIA
Jeffrey Kubik, ECON
Yoonseok Lee, ECON
Amy Lutz, SOC
Yingyi Ma, SOC
Katherine Michelmore, PAIA
Jerry Miner, ECON
Cynthia Morrow, PAIA
Jan Ondrich, ECON
David Popp, PAIA
Stuart Rosenthal, ECON
Michah Rothbart, PAIA
Rebecca Schewe, SOC
Amy Ellen Schwartz, PAIA/ECON
Perry Singleton, ECON
Michael Wasylenko, ECON
Peter Wilcoxen, PAIA

GRADUATE ASSOCIATES

Alejandro Alfaro Aco, PAIA
Emily Cardon, PAIA
Ziqiao Chen, PAIA
Jena Daggett, PAIA
Alex Falevich, ECON
Wancong Fu, ECON
Emily Gutierrez, PAIA
Jeehee Han, PAIA
Boqian Jiang, ECON
Hyunseok Jung, ECON
Khatami, Seyedvahid, IR
Yusun Kim, PAIA
Ling Li, ECON
Michelle Lofton, PAIA
Judson Murchie, PAIA
Brian Ohl, PAIA
Jindong Pang, ECON
Krushna Ranaware, SOC
Laura Rodriquez-Ortiz, PAIA
Fabio Rueda De Vivero, ECON
David Schwegman, PAIA
Shulin Shen, ECON
Iuliia Shybalkina, PAIA
Kelly Stevens, PAIA
Saied Toossi, PAIA
Rebecca Wang, SOC
Xirui Zhang, ECON

STAFF

Kathleen Nasto, Administrative Assistant
Candi Patterson, Computer Consultant
Mary Santy, Administrative Assistant
Laura Walsh, Administrative Assistant
Katrina Wingle, Administrative Specialist

Abstract

Grading schemes are an increasingly common method of quality disclosure for public services. Restaurant grading makes information about food safety practices more readily available and may reduce the prevalence of foodborne illnesses. However, it may also have meaningful financial repercussions. Using fine-grained administrative data that tracks food safety compliance and sales activity for the universe of graded restaurants in New York City and its bordering counties, we assess the aggregate financial effects from restaurant grading. Results indicate that the grading policy, after an initial period of adjustment, improves restaurants' food safety compliance and reduces fines. While the average effect on revenues for graded restaurants across the municipality is null, the graded restaurants located geographically closer to an ungraded regime experience slower growth in revenues. There is also evidence of revenue convergence across graded and ungraded restaurants in the long-term.

JEL No. D12, H22, H27

Keywords: Public Grades, Restaurant Revenues, Public Resources

Authors: Rachel Meltzer, Milano School, The New School; Michah W. Rothbart, Maxwell School, Syracuse University; Amy Ellen Schwartz, Maxwell School, Syracuse University; Thad Calabrese, Wagner School, New York University; Diana Silver, Steinhardt School, New York University; Tod Mijanovich, Wagner School, New York University; and Meryle Weinstein, Institute for Education and Social Policy, Steinhardt School, New York University

Acknowledgement: The authors gratefully acknowledge contributions from the NYC Department of Health and Mental Hygiene and the NYC Department of Finance, and financial support from the Robert Wood Johnson Foundation. Special thanks to Dan Kass, Wendy McKelvey, Bryan Jacobson, Melissa Wong, Corinne Schiff, Karen Schlain, Zachary Papper, Steven Robinson, William Raleigh and James Stevens.

I. Introduction

Grading schemes have become a popular way to accessibly and concisely convey the quality of public services. They are a good example of an information-based policy, which aims to influence, or "nudge" (Thaler and Sunstein 2009), the user's behavior in a particular direction rather than mandating or directly incentivizing it. Municipalities across the U.S. grade the performance of public schools, street cleanliness is frequently scored and graded, and the Straphangers Campaign in New York City even produces a "report card" ranking the performance of each subway train line. One of the more recent grading initiatives applies to restaurants' food safety compliance, a policy that has taken hold in cities across the globe (Filion and Powell 2009). The underlying intuition is clear: the grades succinctly and conspicuously summarize information on sanitary conditions, by an independent third-party regulator, empowering restaurant goers to choose safer restaurants, reducing the exposure to foodborne illnesses. However, its reception has not been uniformly positive: restaurant owners, in particular, have pushed back forcefully against the policy. Their main concern is over the potential impact on their bottom line. Does it increase compliance-related costs? Does it help or harm restaurants' sales? And, more broadly, what are the financial implications for the municipality? In this paper, we shed light on all of these questions for the largest municipality with a restaurant-grading regime in the U.S. and present the first estimates of such a policy's aggregate financial impacts.

We use fine-grained administrative data that tracks the food safety compliance and sales activity for the universe of restaurants in New York City (NYC) and its bordering counties, over multiple years, to assess the broader financial consequences of the restaurant grading policy in NYC. Results indicate that NYC's grading policy affected both restaurant sanitary conditions and fines levied. While there appears to be a short period of adjustment, during which trends from the pre-policy period continue, both initial and final inspection scores (which reflect the number and severity of food safety violations) eventually decline following the implementation of public grading. More specifically, initial inspection scores increase by about 2.5 points (about 10% of the pre-grading mean) upon policy implementation, but then decline at about 1.5 points per quarter thereafter. Further, fines increase immediately after the start of the grading policy (by between $65 and $100 per inspection, or about 6% - 10% of the mean fine before grading), but decline thereafter such that any gain is reversed by the second quarter post-implementation (and further reduced each quarter after that).

The impacts on sales revenues, however, are less stark. While simple pre-post analyses of restaurants in New York City indicate positive revenue effects immediately after the policy's implementation, this effect reverses when we add control groups of ungraded food and entertainment establishments. However, when we allow the post-grading effect to vary non-linearly over time, and account for a long-term convergence of revenues across graded and ungraded establishments, the initial suppression of revenues for NYC restaurants subject to a grading regime attenuates. While the difference in revenues is insignificant in the full sample, it remains highly significant when we narrow in on the part of NYC that borders the suburban counties without any grading regime. These results suggest that while the average effect on revenues across the municipality is null, the graded restaurants located geographically closer to an ungraded regime see smaller increases in revenues.

II. Public Restaurant Grading and Quality Disclosure

There is an established history of government either mandating quality disclosure or conducting its own inspections, especially when the health or safety of consumers is at issue. While restaurants have long been inspected and monitored by the government, the results of those inspections, while publicly available, have not always been readily accessible. The obscurity of that information creates information asymmetries, whereby the restaurant operator knows the sanitation conditions inside the establishment and the consumer knows only what is easily visible on site at the restaurant. Consumers can also be informed by a business' reputation or personal experiences related to the establishment's food safety. These sources are also imperfect, however, as it is difficult for the consumer to link ex post health symptoms with the restaurant's product, especially if the restaurant is infrequently patronized (Dranove and Jin 2010). Without explicit quality disclosure, consumers make decisions about their eating habits based on incomplete information and the restaurants have fewer incentives to change behaviors around less discernible sanitation issues. Theory suggests, then, that consumers should be affected by excess incidents of foodborne illness. Public grading policies aim to address these information asymmetries (and the subsequent health risks) by making the sanitation inspection information more readily available to the consumer, in a way that minimizes his/her search costs. Specifically, the letter grading presents a format that is systematic, easily understood and comparable, and with a clear ranking or "mapping" of grades onto food-safety ratings (Thaler and Sunstein 2009; Dranove and Jin 2010). The fact that the rating is mandatory and imposed by the government, a seemingly objective third-party, also makes the information seem more consistent and trustworthy than were it to come from the restaurant itself or even a private rating company with either monetary or reputational incentives (Dranove and Jin 2010). In addition, the posting of the letter grade in plain sight, at the point of purchase, makes the information particularly salient and minimizes the effort required to gather and process it (Thaler and Sunstein 2009).

We consider the implications of grading policies on restaurants' aggregate financial outcomes, which we understand to be a product of a series of micro-decisions on the part of consumers. We make the reasonable assumption (supported by conversations with professionals and officials involved in the grading regime) that in the case of NYC restaurant operating costs changed in negligible ways; therefore, any response in revenues should capture shifts in consumer behavior.[1] Here, we discuss how, depending on the nature of these micro-decisions, the aggregate financial effects are theoretically ambiguous. In the most optimistic scenario, the posting of grades will do two things. First, it will influence existing restaurant consumers to sort away from the establishments with lower grades and towards those with higher grades. Second, it will also induce consumers, who did not patronize restaurants before (or at least not as frequently), to dine out (presumably more so at the establishments with higher grades). The restaurants will, in turn, adjust sanitary practices to improve compliance and earn higher grades, either in response to or in expectation of a change in consumer behavior. This suggests that food safety compliance will improve and sales will increase for the average restaurant.

[1] DOHMH claims that the types of actions that lead to violations most frequently are based on habits and training, and not those requiring capital repairs. They suggested that we were unlikely to see impacts on capital investments because those were not the type of violations that were most frequent.

We consider this first scenario an optimistic, or upper bound, condition, as the grading policy's impact could be dampened in two important ways. First, it could be that no or few new consumers enter the restaurant market and the grading may simply trigger a re-sorting of existing restaurant patrons. In that case, food safety compliance may still improve over time, but sales would, on average, exhibit no or little change, as spending is reallocated from restaurants with lower grades to those with higher grades. Second, expected outcomes may be minimized if the new information provided by the posted grades does not alter preconceptions about food safety. In this case, loyal patrons may prioritize information gathered from their first-hand experiences with the restaurant over the posted grade and continue their patronage in the same manner as before. For similar reasons, a restaurant's broader reputation could lessen the effect of the information provided by the posted grade (Dranove and Jin 2010; Jin and Leslie 2009). Captive patrons, such as those without any other dining options nearby, may also not process the posted grade in the same way. Under these conditions, changes in both food safety compliance and revenues could be attenuated: the restaurants may be less motivated to invest in improving the posted grade so that any grade-induced sorting would be less evident in sales activity.

III. Background

A. New York City's Restaurant Grading Policy

The NYC Department of Health and Mental Hygiene (DOHMH) has long inspected the City's restaurants to ensure proper food safety practices, fining restaurants for violations and closing restaurants with public health hazards. In the early 2000's, they introduced a website where anyone could view the restaurants' violations. The letter grading policy, which began July 2010, followed the same inspection criteria and scoring system, but started assigning each restaurant a letter grade (A, B, or C) and required restaurants to post the grades near the restaurant's entrance. These letters are printed in large bold font and are required to be near eye-level and at the front entrance—passersby can easily discern them, even from across the street (a sample of a posted grade is shown in Appendix A). DOHMH also added the grades for each inspection to its website. The stated goal of the public grading law was to improve restaurant sanitary practices and decrease the incidence of restaurant-attributable foodborne illness in NYC.

Inspection scores are calculated as the sum of violation points assigned during inspections. The points for a particular violation depend on the health risk it poses to the public, and the level of public health risk falls into three categories: public health hazards, critical violations and general violations. Additional points are added to each violation to reflect the severity of the violation.[2] Points from violations are then aggregated to generate the final inspection score, with lower scores reflecting more hygienic conditions. The scoring rubric used after the policy change is exactly the same as that used prior to 2010. Following the policy change, however, DOHMH translated inspection scores into letter grades as follows: an inspection score of 13 or below is assigned a grade of *A*; a score between 14-27 a *B*; and a score of 28 or higher are assigned a *C*.[3] That said, if an initial inspection yields a score in the *B* or *C* grade range, the grade is not viewed as final. Instead, the restaurant is inspected again within one month. Thus, a final grade is assigned either after a re-inspection (for those with initial inspection scores above

[2] See NYC Department of Health and Mental Hygiene (2010) for detailed explanation of the relationship between violation severity and violation points assessed and http://www1.nyc.gov/assets/doh/downloads/pdf/rii/blue-book.pdf and Schwartz et a. 2015 for more detail on how DOHMH translates scores into grades.

[3] Restaurants can also be temporarily closed if they pose a large public safety risk.

13) or at the initial inspection if an *A* is earned. In addition to determining the publicly posted grade, the inspection outcome affects the time to the next inspection visit. *A* restaurants are visited annually for food safety inspections; *B*'s are inspected twice a year; and *C*'s are inspected three times a year. These inspection visits are unannounced within these known longer intervals.

Restaurants are also given the right to due process. They may challenge their violations (and, therefore, inspection scores, fines, and grades) at a tribunal administered by an independent agency, the Office of Administrative Trials and Hearings (OATH). Cases regarding violations received during the inspection process are adjudicated by hearing examiners (acting as judges), typically lawyers hired by OATH. Silver et. al. (2017) find that restaurants earning *B* and *C* grades at inspection are much more likely to have inspection scores reduced (improved) through the adjudication process in the post-period than they were in the pre-period, resulting in substantially better grades (Silver, et. al. 2017). These findings suggest that challenging grades in court is an important tool for restaurants motivated to post *A* grades in their window. Restaurants DO have the right to post a placard that reads "Grade Pending" in lieu of posting *B* or *C* grades until they have their case heard at the OATH tribunal; this practice may reduce consumer certainty about restaurant food safety compliance.

Throughout the study period (both before and after the implementation of restaurant grading), the type and count of inspection violations determines fines assessed. Fines range from $200-$2,000 per violation and are assessed at a restaurant's adjudication hearing at the discretion of a hearing officer.[4] It is important to note that after January 18, 2011, restaurants

[4] An exception to this is if the grade is accepted and a lower fine is paid by the restaurant operator.

receiving an *A* grade at inspection were not fined for any inspection violations; therefore *A* restaurants do not incur any fines for much of the period following the introduction of the grading policy.

B. Empirical Literature Review

Grading, as a means of conveying information about the quality of services or goods, is used in other policy contexts, including public education and health care.[5] Impact studies of public grading largely focus on the effects of the grades themselves and how differentiated information affects relevant outcomes. In the case of ranking health plans, most studies find that higher-ranked plans see increases in their market share, but that the ratings have less of an effect when they are hard to understand or when they do not provide new information to the consumer (for example, Wedig and Tai-Seale 2002 and Dafny and Dranove 2008). In the case of education, many districts grade schools on their effectiveness (e.g., improvements in test scores) and make these grades publicly available. There is some evidence that schools with lower grades have short-term improvement in aggregate student achievements (Rockoff and Turner, 2010; Winters and Cowen, 2012) and that the information provided by school grades affects housing prices above and beyond information provided by test scores (Figlio and Lucas, 2004). In addition, Hastings and Weinstein (2008) show that parents use mandated information on school quality to move their children to higher performing schools.

The empirical research on food quality disclosure, and grading in particular, is scarce. While the mechanisms of disclosure might be similar to other contexts, one could argue that the

[5] For a comprehensive review of the literature on quality disclosure see Dranove and Jin 2010.

magnitude and transient nature of the decision is at a different scale than that related to health care plans or schools (Ippolito and Mathios 1990). There is a set of studies that assess the impact of food content disclosure on consumer behavior; these studies find that consumers do increasingly opt for the "healthier" option when the disclosure of relevant information becomes mandatory (Ippolito and Mathios 1990; Mathios 2000). These are not instances, however, of a standardized grading regime; restaurant grading provides such a scenario. Ho (2012) analyzes publicly available restaurant grading data for NYC, exploring the extent to which inspection scores in one period predict future scores: he observes that prior scores predict less than 2% of future grades. Wong et al. (2015) provide new evidence of improved compliance since the beginning of NYC's public grading program, and offers survey evidence of the program's high approval ratings among New Yorkers. This is consistent with other surveys that have demonstrated that consumers use public inspection results to inform their dining decisions (Filion and Powell 2009).

As for impact studies of restaurant grading regimes, there are three.[6] Two studies (Jin and Leslie, 2003; Simon et al., 2005) focus on the effects of the Los Angeles health inspection letter grade system, which started requiring posted letter grades in 1998. Jin and Leslie (2003)

---

[6] We do not focus on the incidence of foodborne illness as an outcome of interest. While important it is extremely difficult to link it to the prevalence and use of posted inspection grades. Indeed, the correlation between inspection scores and foodborne disease outbreaks is inconsistent (Filion and Powell 2009) and empirically hard to identify. It relies entirely on restaurant patrons correctly identifying the foodborne illness and attributing its source, which is difficult to do for a number of reasons (duration of latency, expectations about food safety, proclivities towards gastrointestinal symptoms) (Jones and Angulo 2006; Mead et. al. 1999; Fein et. al. 1995;); moreover, it relies on their reporting the illness, which we know is done inconsistently (Jones and Angulo 2006 Mead et. al. 1999). For all of these reasons, we focus instead on food safety compliance and economic impacts (for both the restaurants and municipal finances).

use OLS and difference-in-differences regression analyses to estimate the effect of the Los Angeles letter grades program on inspection scores, restaurant revenues, and foodborne illness hospitalizations. They find that posted grades improve restaurant inspection scores, that restaurant revenues respond to hygiene quality signals (i.e. better grades), and that foodborne-disease hospitalizations decrease in Los Angeles County following the implementation of the public letter grade program. The study by Simon et. al. (2005) also provides evidence of reduced hospitalizations due to foodborne illnesses in Los Angeles County, compared to California overall. Jin and Leslie also suggest that the improvements in health outcomes cannot be explained by consumption choices alone, but are also likely a result of restaurant hygiene improvements. The most recent study, by Schwartz et. al. (2015), focuses on the impact of individual grades on restaurant food inspection compliance and economic activity and finds that a better grade increases a restaurant's sales (and associated sales taxes) and decreases the amount of fines assessed and the probability of the restaurant's closure. These results are also consistent with the expectation that public restaurant grading provides new information for consumers' dining decisions.

IV. Data and Measures

Our analytical approach is multi-pronged. We employ several metrics to capture the grading program's effect, two different data sets, and alternative identification strategies to exploit, where possible, more detailed data on the NYC grading program. For NYC only, we use detailed information on restaurants, including their food-safety compliance (e.g., inspection scores, posted grades) and sales revenues. These data were pulled in two different ways: for NYC restaurants only, and again for a larger sample including information on restaurant and other ungraded food and entertainment establishments for NYC and two additional suburban counties for comparison.

A. NYC Restaurant Grading and Compliance Data

We use data on the restaurant-grading program from the NYC Department of Health and Mental Hygiene (DOHMH), the city agency tasked with administering and monitoring food safety compliance. These data include restaurant characteristics, zip codes, inspection dates and scores, adjudication dates, grades assigned, and fines assessed. Restaurant characteristics include number of seats, number of employees, an indicator for chain restaurant (at least 15 locations nationwide), and a series of variables indicating cuisine offered, service type, and venue type.[7] Table 1 shows descriptive statistics for the restaurants in our NYC sample. The mean restaurant has 3.25 final inspections over the study period, employs 6.2 workers and has about 29.6 seats. Just under 11% of restaurants in the sample are chains.

We use scores assessed at each initial and final inspection to capture food safety compliance. Importantly, both initial and re-inspections take place without advanced notice and inspectors are randomly assigned to their visits (and do not come back to the same site for re-inspections). Initial scores will reflect the unanticipated response on the part of the restaurants and the final inspection scores will, to some extent, reflect learning or adjustment by restaurants. This difference is confirmed in the data (and displayed in Figure 1), where we see a starker shift in final inspection scores (compared to pre-grading scores) towards lower values; the distribution

[7] A full list of cuisine, service, and venue types is available upon request. We use data on restaurant characteristics recorded at the last inspection only, so these characteristics do not vary with time. Unfortunately, we cannot access information on restaurant characteristics over time, as they are not tracked by DOHMH.

of initial scores is not identical, but certainly more similar, before and after the implementation of grading.[8]

Finally, we use data on fines to assess the program's public revenue generation (and conversely, the financial burden on restaurants) – focusing on fines levied. All dollar values are adjusted using urban CPI to real 2013 dollars.

The NYC grading and compliance data span December 1, 2007 through February 28, 2013, two and a half years before and after the implementation of public grading (hereafter referred to as "pre-period" and "post-period", respectively). This sample includes 159,588 initial inspections and 167,045 final inspections of 41,362 restaurants in all, including 29,864 restaurants operating and graded in the post-period.

B. Sales Revenue and Tax Data

New York City Sample

We obtain reported quarterly sales for all NYC restaurants from the city's Department of Finance (NYC DOF).[9] Due to statutory restrictions on data sharing, we could not access filer-level information. Instead, NYC DOF aggregated the data in order to ensure the confidentiality of the

---

[8] This finding, in addition to the fact that inspectors are randomly assigned to restaurants and do not conduct repeat visits, will help to mitigate against concerns that inspectors are systematically inflating scores post-grading to achieve higher grades. For a more comprehensive discussion of this issue in the case of the NYC grading regime, please see Schwartz et al. 2015.

[9] These values are also adjusted to 2013 dollars. NYC restaurants are required to collect sales tax on food and beverage sales at a rate of 8.875% of gross sales - 4.875% for New York State and 4.0% for NYC. The State collects the entire sales tax from restaurants and remits the City's portion of sales tax revenue in the following month. Restaurants with $300,000 or less of sales in the previous quarter may remit sales taxes to New York State quarterly, while restaurants with more than $300,000 of sales in the previous quarter remit monthly to the State.

restaurants according to the following protocol: (i) matching DOHMH restaurant data and NYC DOF sales data using Employer Identification Numbers (EINs), and (ii) aggregating the EIN-level sales reports into randomly-assigned groups, to create a group-level data set.[10] In the resulting group-level data set, each observation contains summary data for the set of 10 restaurants randomly assigned to the same group, or "bin", for each quarter-year in the pre-grading and post-grading periods.[11] The data set is a panel of restaurant-groups, each of which can be followed throughout the study period. The random assignment of the restaurants into the groups mitigates any bias caused either by geographic clustering of grades or subsequent spillover effects. The summary statistics for each group-quarter include means and standard deviations of sales. Our sample for the NYC sales analyses includes 2,288 groups and 24,464 observations.[12]

New York City, Long Island and Westchester Sample

We construct a control group of establishments, using sales data from the two counties bordering NYC (Nassau and Westchester) that were never subject to a grading policy during the study period. Since we required data outside of NYC's jurisdiction, we needed to coordinate with

[10] To address attrition and entry, we stratify the sample by quarters of operation and then assign restaurants to groups of ten within each strata. Thus, the 5,145 restaurants operating in all 20 quarters of our study period were randomly assigned to 509 groups of 10 and five groups of 11; the 149 restaurants operating in all but the last quarter were grouped in five groups of 10 and 9 groups of 11; the 244 operating in all but the first were grouped in 20 groups of 10 and 4 groups of 11. They continue this process, sequentially, until all restaurants are assigned to groups, homogeneous in their quarters of operation and no group (and no observation) ever provides information on fewer than ten establishments.

[11] A small number of groups have 11 rather than 10 restaurants in order to make sure all restaurants are included.

[12] The policy is implemented in the middle of the 2nd sales tax quarter in 2011. Our NYC analytic sample includes data observed from the 4th quarter of 2008 to the 3rd quarter of 2013.

the State's Department of Taxation and Finance (NYS DOF).[13] This involved a separate data request, with slightly different grouping parameters (we were subject to the same confidentiality requirements and had to, again, group the tax filer data into bins of 10). First, we distinguish the types of food establishments that would be subject to grading from those that earn commercial revenues (such as grocery stores or entertainment venues) but are never subject to a grading regime. Unlike the NYC data, we were unable to merge the DOHMH restaurant data directly with the State's finance data. We instead identified the subset of NAICS codes that appear among the graded establishments in DOHMH's restaurant data and used these to identify the restaurants in NYC that should be subject to grading; these establishments are henceforth referred to as "graded" restaurants.[14] We then use two comparison groups: (i) suburban establishments with similar NAICS codes as "graded" restaurants in NYC (henceforth referred to as ungraded restaurants) and (ii) ungraded food and entertainment establishments (with distinct NAICS codes than those subject to grading) unlikely to have been graded, in NYC and the suburban counties. See Appendix B for a detailed breakdown. Second, we grouped the tax filers by their reported county of operation and type of establishment (i.e. NAICS code), by quarter. Thus, we create a

[13] It is important to note that the sales and tax data from NYC's Department of Finance is originally sourced from the state data as well. Therefore, the underlying data is the same across the two samples; the meaningful difference is in how the individual filer data was grouped.
[14] This means that we are capturing the universe of food-service establishments that are very likely (but not definitively) subject to NYC's grading regime. Therefore, we set up something akin to an intent-to-treat group rather than a clean treatment group. While this over-inclusion of "graded" establishments could attenuate our estimates, we mitigate against this by pulling an identical set of NAICS codes for the comparison establishments in the suburban counties and conducting a difference-in-difference estimation. In addition, we've replicated analyses using only the subset of NAICS codes that definitively apply to food service and restaurant establishments (those coded as 722---), which are certainly subject to grading. The results are substantively the same as those using the full sample of establishments; they are available from the authors upon request.

dataset where the level of observation is the group-quarter and each observation includes the means and standard deviations of sales. Our sample for the expanded NYC sales analyses includes 1,525,330 group-quarter observations over a the 10-year sample period from 2007-2016. This includes over 800,000 observations of NYC restaurants that are "graded," or most likely to be subject to the grading law based on their NAICS classification.

V. Empirical Strategy

Our empirical strategy relies on a difference-in-difference specification, comparing establishments subject to the grading policy to those exempt from the policy, before and after the implementation of the policy. For some outcomes, we augment this specification to include two simultaneous control groups. Restaurants in NYC are continuously inspected (and scored) throughout the study period, but only after the start of the grading policy are the inspection results made conspicuous via the posted grade.[15] Estimates of the policy effect, therefore, capture the impact of new information provided through the posted grade.

A. Inspection Scores and Fines

We use inspection scores and fines to measure (i) food safety compliance behaviors, and (ii) fiscal burdens.[16] Fines are also a good measure of the policy's direct fiscal impact on the City,

[15] We know that restaurants in Nassau and Westchester counties are inspected for food safety compliance and the results of these inspections are made available via https://health.data.ny.gov.

[16] We would also like to estimate changes in the likelihood of restaurant closures, but we are limited in how precisely we can identify the timing of closure. We identify restaurant closure as whether or not it is still operating at the time of inspection; since inspections occur irregularly, we are unable to precisely pinpoint the timing of closure, introducing considerable bias into our estimates. To address this issue, we would have to exclude all restaurants closing in the year

since the new grading regime imposes little or no new costs.[17] We begin with a standard pre-post model, as follows:

$$(1) \quad y_{it} = \beta_0 + \boldsymbol{Grading\_Post}_{it}{}' \boldsymbol{\beta_1} + \boldsymbol{X_i}{}' \boldsymbol{\beta_2} + \beta_3 Pre\_Post_{it} + \gamma_i + \delta_t + \varepsilon_{it}$$

Here, *y* reflects restaurant outcomes, including inspection scores and fines. ***Grading_Post*** is a set of three variables, which capture the implementation of the grading policy: *Post, Post_trend* and *Post_trend²*. *Post* takes on the value of 0 prior to the start of the grading policy (for $t<0$) and 1 thereafter (for $t \geq 0$); the coefficient on *Post* captures the initial effect of the policy's implementation. *Post_trend* and *Post_trend²* are created by interacting *Post* with linear and quadratic time trends, and allow the effect to change over time; ***X*** is a vector of restaurant characteristics including cuisine, service, and venue type; *Pre_Post* is a linear (and, in certain specifications, nonlinear) time trend; γ and δ are zip code and seasonal fixed effects, respectively; and ε is an error term. In an alternative model specification, we estimate model (1) controlling for restaurant fixed effects, μ_i, instead of γ_i and X_i.

When grading started, restaurants were not uniformly exposed to the new inspection regime (i.e. did not have to post a grade until they had a graded inspection following the policy change). We exploit the variation in grade posting during the roll-out period in an alternative specification, where we limit the sample to restaurant-quarter observations during the first year of the policy's legislative start, or the roll-out period. Here we are comparing the "early posters" to the "late posters" and this model takes on the following form:

$$(2) \quad y_{it} = \beta_0 + \beta_1 Post\_Rollout_{it} + \mathbf{X'}_i \boldsymbol{\beta_2} + \gamma_i + \delta_t + \varepsilon_{it}$$

---

during which the policy is implemented (approximately 20% of our sample) or during the last year of the panel (approximately 20% of our sample).

[17] This is a reasonable assumption that is corroborated by accounts from DOHMH.

Again, y is a restaurant-specific outcome (inspection scores and fines) and *Post_Rollout* takes on the value of 1 if the restaurant has posted a grade placard by the beginning of the quarter t and 0 otherwise. The remaining variables are identical to those defined above.

B. Sales Revenues

As described above, we compare "graded" NYC restaurants to ungraded restaurants in the suburban counties and to ungraded food and entertainment establishments in NYC and the suburban counties.[18] We include this third group of establishments, which are never subject to grading, in order to capture sector-specific trends. Therefore, we end up with the intent-to-treat group ("graded" restaurants in NYC) and two counterfactuals (ungraded restaurants in the suburban counties and ungraded food establishments and entertainment venues). We set up a triple difference-in-difference model, using the grouped sales data. The fully-specified model is as follows:

$$(3)\ y_{gq} = \boldsymbol{Grading\_Post_{gq}}'\boldsymbol{\tau_1} + \boldsymbol{Restaurant_{gq}}'\boldsymbol{\tau_2} + \boldsymbol{Rest\_Grading_{gq}}'\boldsymbol{\tau_3} + \tau_4 Pre\_Post_{gq} + \gamma_{cr} + \delta_q + \varepsilon_{gq}$$

Here, y_{gq} is the group's average restaurant sales in quarter q. As above, ***Grading_Post*** a set of three variables, *Post, Post_trend*, and *Post_trend²*. As above, *Post* takes on the value of 1 if quarter q is after the start of the grading policy and 0 otherwise; *Post_trend* and *Post_trend²* are interactions between *Post* and linear and quadratic time trends, respectively. We add to this model a vector, ***Restaurant***, to control for the degree to which establishments are subject to the NYC grading policy (i.e. NYC "graded" restaurants vs. suburban ungraded restaurants vs.

[18] Again, a table with the assignment of NAICS codes for grade-able restaurants and ungraded food establishments and entertainment venues is presented in Appendix B.

ungraded food and entertainment establishments, the latter of which is omitted as the reference category). These variables will also help to control for reputational differences across the different types of establishments. We also include interaction terms, included in ***Rest_Grading***, whose coefficients capture the post-grading effect on "graded" NYC restaurants and suburban restaurants relative to ungraded food and entertainment establishments. Therefore, $\boldsymbol{\tau_3}$ identifies the effect from any new information provided by the posted grades, above-and-beyond the influence of sector, geography or pre-grading reputation. Finally, γ_{cr} and δ_q are county-NAICS and quarter fixed effects, respectively.[19] We note that while point estimates are of the mean impact on restaurants, the standard errors are larger than if we observed individual restaurant sales. We also cluster the standard errors by county-NAICS to mitigate spatial autocorrelation across proximate establishments.[20]

And as above, we also specify an alternative, roll-out model, exploiting the detailed panel data on sales revenues for NYC graded restaurants alone:

$$(4)\ y_{gq} = \tau_1 Post\_Rollout_{gq} + \boldsymbol{X'_{gq}\tau_2} + \gamma_g + \delta_q + \varepsilon_{gq}$$

Where y_{gq} is the group's mean sales in quarter q and $Post\_Rollout$ is the average share of days in quarter q of posting a grade placard for restaurants in group g. As above, grade placards can

[19] Specifically, we group observations (and cluster their standard errors) by two-, three- and four-digit NAICS codes: in case of restaurants (graded and ungraded) we group those coded 772--- and 445--- separately; in the case of ungraded food/entertainment establishments we group 71----, 4451--, 4452-- and 4453—separately. These groupings are consistent with how the sales data were processed and grouped. We do replicate these models with less restrictive geographic controls and the results are substantively the same. The finer spatial controls will also help in minimizing the effect of reputation on the grading estimates, since consumers are more likely to be exposed to similar reputational information in closer spatial proximity (Jin and Leslie 2009).

[20] We are less concerned about geographic co-dependence of the inspections themselves, since inspectors are randomly assigned to restaurants and not on a geographic basis.

read *A, B, C,* or *Grade Pending*, because treatment begins with the first grades assigned at inspection. Insofar as *Grade Pending* is an unclear signal to consumers, sales results will be attenuated to zero. Again, these estimates are consistent -- but inefficient -- as compared to estimates derived from models run on restaurant-level observations, as described earlier. The remaining variables are identical to those defined above for the inspection score and fines models, but aggregated to the group level.

VI. Results

A. Inspection Scores

We first discuss results from the pre-post analysis, estimating the grading policy's impact on the restaurant's food safety compliance, as measured by inspection scores from 2007-2013. Table 2 shows initial inspection score results, which provide the strictest test of improved food safety compliance. Table 3 shows final inspection score results, which test the extent to which further compliance improvement occurs after poor initial inspection performance.

The first column of Table 2 shows that after the implementation of the grading policy, initial inspection scores decline (i.e. health conditions improve) by about 1.3 points on average per inspection. This is about 6% percent of the sample mean in the pre-period. When we include additional controls, such as seasonal and ZIP fixed effects, restaurant characteristics and a pre-post trend line, the coefficient on *Post* turns positive and decreases slightly in magnitude, suggesting that compliance improves over time, but does not improve precisely at the time of policy implementation. The coefficient on *Post* remains positive and significant when we include restaurant fixed effects (instead of ZIP fixed effects and time-invariant restaurant characteristics) and then when we include a linear *Post_trend*. The results from this model indicate that upon policy implementation initial inspection scores go up by about 1.2 points, but decline over time by about .33 per quarter, implying that mean initial inspection scores improve starting about one year after policy implementation. When we add in *Post_trend*^{2} in the final column, the magnitude of the post-grading effect increases to 2.4 and the ensuing change is clearly nonlinear: after the initial bump up in scores, there is a decline of about 1.5 points per quarter that flattens out over time. In sum, while the initial effect of the grading policy indicates less compliance, the slope effect over time indicates that compliance has improved since public grading began. On net, the overall effect is improved compliance within the first two quarters of implementation that also continues through the end of the sample period. In general, we note that the other covariates display generally expected signs: initial inspection scores are lower (better) for chains and uncorrelated with number of seats and number of workers. There is also variation in scores depending on cuisine.

Table 3 shows the same models, using final inspection scores. Just as above, the first column displays a significant and negative coefficient on *Post*, but much larger in magnitude: final inspection scores decline by about 7.6 points on average. As we add in controls to the model, the coefficient on *Post* remains negative and decreases slightly in magnitude, and in the final, preferred models, inspection scores decline by about 4 points on average upon implementation of the grading policy. When we include only the linear *Post_trend,* scores continue to decline over time, at about .29 per quarter. The final column includes a nonlinear *Post*_trend^2, and there is, again, evidence of a nonlinear response over time: initially, scores go down by 5 and continue to decline (at 1.1 points per quarter), but at a declining rate. Since final inspection scores reflect food safety conditions after feedback or general learning from initial inspections, it is not surprising that the immediate effect (i.e. the coefficient on *Post*) is a reduction in scores. Both initial and final scores decline over time (but at a decreasing rate) after policy implementation, suggesting improved food safety compliance and ongoing learning on the part of restaurants during the first six months of the policy. In general, we note that the other covariates display generally expected signs: final inspection scores are lower (better) for chains and uncorrelated with the number of workers and number of seats. There is also variation in scores depending on cuisine and the directions of those relationships are generally consistent with the relationships observed in the results for initial inspection scores.

As an alternative specification, we exploit the policy's roll-out period to identify the impact of posting a grade on inspection scores. One concern with this approach is that the restaurants exposed earlier to the policy were systematically different than those exposed later. We assess the differences in early- and late-graded restaurants across a range of observed restaurant characteristics and sanitary conditions—displayed in Appendix C. In general, we find no meaningful difference between the early- and late-inspections, and fail to reject the null of group equivalence in a joint-significance F-test.[21] This mitigates some concerns of selection bias, based on observed characteristics (which we assume are at least somewhat correlated with unobserved characteristics), and we proceed with the assumption that the roll-out of the program was random, conditional on the restaurants' observed characteristics. We also control for restaurant-level characteristics in the regression models and, in some specifications,

[21] We also find little or no difference between the sample for the roll-out analysis and the larger sample for the pre-post analysis.

restaurant fixed effects for a within-restaurant comparison over time, further reducing unobserved heterogeneity across early- and late-graded restaurants.

The results for the roll-out regression analysis are displayed in Table 4, and we begin with the most parsimonious model, controlling for restaurant characteristics and time trends. We show results for both initial and final inspection score results. Both initial and final scores go down, although the declines are bigger for the latter measure, which is consistent with the results from the pre-post sample. Again, this is likely due to some combination of restaurant learning and improved compliance over somewhat shorter re-inspection windows. In the models with restaurant fixed effects, initial inspection scores for restaurants exposed to the grading policy (compared to those not yet exposed) go down by just under 1 point. Similarly, final inspection scores decline almost 4 points per inspection. We recognize that there could be a period of adjustment, even during the roll-out period. To test for this, we replicate the roll-out analysis, allowing the effect of the graded inspection to vary across time. These results are displayed in Appendix D. The immediate effect of the graded inspection is positive for initial inspections and negative for final ones; over the course of the roll-out period, this effect progressively becomes more negative (i.e. scores are improving). Thus, by the end of the first year of the grading policy, mean initial and final inspection scores are both lower than they were before public grading. Again, this is consistent with the findings from the pre-post analysis.

Altogether, the results for inspection scores indicate a period of adjustment on the part of the restaurants, which initially see a slight bump up in initial inspection scores and then a steady decline over time. The initial increase in scores could mean two things. First, it suggests that restaurants were changing their food safety compliance behaviors in response to the policy (and the feedback from the inspections), but that it took time for it to manifest itself in the actual restaurants' conditions. The initial scores (and therefore food safety conditions) could have been more reflective of restaurants' conditions prior to the start of the grade-posting policy, and later inspection scores a product of their response to the change in policy. This pattern is supported by the descriptive statistics displayed in Table 5, which also show an improvement (i.e. decline) in inspection scores as the program progressed. During the period before public grading, restaurants earned inspection scores of 24.6 on average. In the first five quarters after public grading, mean final inspection scores improved to 18.2. This initial improvement was driven by improved compliance during re-inspections, as the initial scores are about level with the pre-implementation scores. In the next five quarters, average final inspection scores further improved to 15.6. This second improvement was driven in part by improved compliance on initial inspections (which on average went down to 22.2).[22]

A second explanation for initially increasing and then declining scores relates to inspector behavior—that they had incentives to improve scores under the new grading regime regardless of actual food safety compliance. While we cannot test this directly, there are three reasons we think this mechanism is unlikely (and that any improvement in scores predominantly reflects an improvement in food safety compliance). First, inspectors are randomly assigned to their site visits, and therefore restaurants are dealing with different individuals for the initial and re-inspections—it is unlikely that the randomly assigned inspector for re-inspection would be colluding with the initial inspector to systematically reduce final inspection scores, other than

[22] We find similar changes in mean inspection scores for "continuously operating" restaurants that operate for two and half years before and two and half years after public grading.

based on an observed improvement in food safety compliance. Second, conversations with food safety practice and epidemiological experts at the DOHMH confirm that the same inspection procedures, rubrics and trainings were used before and after the policy and that inspectors were trained to maintain the same standards.[23] Third, the similar distribution of initial scores, before and after grading, suggests that any inspector-driven grade inflation was, at worst, minimal.

B. Fines

Next we consider how the grading policy affects fines; this is one potential financial benefit for the City (and conversely, a financial burden for restaurants). Some claimed that the policy was an excuse for the City to collect more revenues from inspected establishments; we test the validity of this claim here. To start, we consider Figure 2, which shows mean restaurant fines by quarter. While fines per restaurant increase in the year immediately following program implementation, this extends a pre-existing trend (that temporarily discontinues in the second quarter of 2011, during program implementation). Quarterly fines reach a peak of $675 per operating restaurant in the first quarter of 2012 and then decline steadily, reaching pre-program levels by the third quarter of 2013 ($353 in fines for the average restaurant). The question is

[23] We recognize that there is selection into which restaurants are re-inspected. The re-inspected restaurants get a second draw from the score distribution, which mechanically leads to score "inflation." This is not inspectors' score inflation, but rather the design of the program (and should still reflect food safety improvement). The new policy did include the hiring of more inspectors, but there is no evidence to suggest that the new inspectors were more lenient than the older ones; and again, they were still being randomly assigned to inspections. In addition, while the higher stakes of the posted grades could change the nature of the interaction between inspectors and restaurants, it is unlikely that this shift would be so systematic as to drive the effects we observe.

whether the post-grading trend is significantly different than what would have continued otherwise.

Starting with a simple pre-post model in the first column of Table 6, we see that fines declined after the policy's implementation. The coefficient on *Post* is negative and highly significant and indicates that on average fines went down by $271 per inspection. When we add in restaurant controls, a linear time trend, and ZIP and seasonal fixed effects, the magnitude on the *Post* coefficient goes down substantially, but still remains negative: fines reduced by about $62 per inspection after the grading policy's implementation. When we instead rely on restaurant fixed effects, the coefficient on *Post* flips its sign to positive, suggesting that fines decline over time, but the decline does not coincide with the precise timing of the policy change. In the final columns of Table 6, our preferred models show that upon policy implementation, fines increase (by about $65 or $110 per inspection, depending on whether or not a nonlinear post-trend is included), but they decline precipitously (and linearly) over time, such that by the second quarter after implementation any increase in fines had been reversed. This immediate increase in fines is consistent with the short-term increase in initial inspection scores, which also goes down over the first year of the policy. Again, this suggests learning and adjustments on the part of restaurants. Altogether, these results suggest that the grading policy did not increase fine-driven revenues for the City, which initially increase but drop during the first six months and end up lower than before the policy change.

Again, we repeat the analysis using the roll-out sample, comparing those restaurants that were exposed to the grading regime earlier in the rollout period to those that were exposed later. The first column of Table 7 displays the results for the model without restaurant fixed effects;

those restaurants exposed earlier to the grading regime pay higher fines on average than those exposed later—about $44 more per inspection. We then add in restaurant fixed effects so that we can compare fines within restaurants, and the magnitude of the coefficient on *Post_Rollout* goes down and flips to a negative sign. Therefore, in our preferred model, we estimate fines decline $31 per quarter during the first year of the policy. This decline is consistent with that observed in the larger pre-post sample, which also produced fine declines over the course of the initial year.

C. Falsification Tests for Inspection Score and Fines

For the pre-post analyses on inspection scores and fines, we implement a placebo test to assess the timing of level and slope changes relative to the grading policy's implementation. To do this, we replicate the estimation, assigning a placebo policy start date one year prior to the actual policy start date. We find that the program was already being discussed in the popular press one year prior to the program's actual start, and so we use this date. The results from this analysis are displayed in Table 8.[24]

In the case of initial and final scores, we see that the false start date is associated with an immediate decline in scores (which is statistically insignificant for final scores), but a large increase in the three quarters thereafter. Thus, inspection scores are actually worse overall in the period immediately preceding policy implementation than would be detected with a simple

[24] We run more parsimonious models, including only *Post_false* and *Post_trend_false*, but for purposes of brevity we display only the more comprehensive models since they better control for all the possible points of inflection during the pre- and post-grading regimes. We further specify models with quadratic and cubic time trends, finding similar results, which are available upon request of the authors.

linear trend-line. The true policy start date (while still controlling for the false start date) is associated with statistically significant declines for both initial and final scores (the latter one about ten times larger), and then a continued decline over the post-period. The discontinuity in intercept and reversal in slope both suggest that any decline in inspection scores is in fact associated with the grade posting and not a continuation of a prior trend or expectation.

For fines, we conduct the same test and see that the false start date is associated with an increase in fines and a positive slope thereafter. The actual policy start date, however, indicates a drop in fines and a subsequent decline (that rather quickly reverses any prior increase in fines). Again, this reversal suggests that any initial bump up in fines (as observed in the pre-post analysis) could be driven by trends prior to the actual start date, and that the grading policy itself is associated with a drop in fines.

D. Sales Revenues

We now turn to our broader sample, which includes NYC, Westchester and Nassau counties for 2007-2016. Again, as a starting point, we look at unadjusted trends over the course of the study period. Figure 3 shows mean sales by quarter for the 14 quarters before public grading and the 26 quarters after, stratified by exposure to grading and location. We see that mean sales follow parallel trends for "graded" NYC restaurants and ungraded suburban restaurants before the implementation of the grading policy in mid-2010; there is, however, a slight divergence immediately after the policy starts, and then a re-convergence in the second half of 2014. We also note that mean sales for suburban ungraded food and entertainment establishments exhibit more volatility than those for the restaurants, and are consistently higher than those for similarly classified establishments in NYC.

To test the extent to which this trend persists in the presence of restaurant, geographic and temporal controls, we consider the regression results in Tables 9 and 10, which display the results for the pre-post and difference-in-difference models. The first column of Table 9 shows a simple pre-post model, on only the sample of the NYC food establishments subject to the grading policy (the "graded restaurants"). Controlling only for county and seasonal variation, we see that, after the policy started, restaurants witnessed an increase in sales revenues of nearly \$10,000.[25] Without a control group of ungraded establishments, it is unclear whether or not this increase in revenues is due to the new grading regime or to trends that would have continued even in the absence of the policy. Therefore, in the second column, we add in a control group of ungraded establishments (including food and entertainment services, still in NYC only) and estimate a difference-in-difference. The coefficient on *Restaurant_Grading* is our impact estimate of interest, and it is negative and statistically insignificant. Compared to food and entertainment establishments that were not subject to grading, "graded" restaurants saw no significant change in revenues after the policy's implementation. We also note that the coefficient on *Restaurant* is positive and significant (albeit marginally), which captures, at least in part, pre-grading differences (between "graded" restaurants and ungraded establishments in NYC).

[25] This finding is consistent with what we find when running a similar model on the panel of grouped NYC restaurants. For that sample, we find that revenues increase between \$8,000 and \$10,700, depending on the specification. These results are available from the authors upon request.

One concern with the current difference-in-difference estimate is that we do not have the right counterfactual, since the ungraded food and entertainment establishments could be subject to different macroeconomic trends than those for the "graded" restaurants. In addition, the estimates of *Restaurant_Grading* could be biased by cross-contamination of the grading effect onto un-graded establishments. Specifically, the posted grades on certain establishments could shift consumers' general assessment and internalization of hygiene for all kinds of food and entertainment establishments. Therefore, we expand our sample to include counties bordering NYC that were not exposed to any grading regime, but were arguably subject to similar regional trends and shocks.

In the final column of Table 9, we use the sample including NYC and bordering suburban counties, but retain only restaurants. This model compares revenue outcomes for NYC "graded" restaurants against those for ungraded restaurants in Nassau and Westchester counties. While the coefficient on *Restaurant_Grading* turns positive, it remains insignificant. Altogether, these findings suggest two things: first, the selection of the control group matters (as indicated by the change in signs), and, second, compared to a range of reasonably similar establishments, the graded restaurants did not experience any significant revenue changes after the implementation of the grading policy.

For the next set of results (displayed in Table 10), we include a second counterfactual. We specify the model as a triple difference-in-difference, where graded NYC restaurants are compared to both similarly classified ungraded restaurants in Westchester and Nassau counties and to ungraded food and entertainment establishments in the NYC and suburban counties. We start with a model that is otherwise similarly specified as the one for the NYC difference-in-difference. The new variables of interest are *NYC_Rest_Graded* and *Suburb_Rest_Graded*, and their coefficients capture the effect of the grading policy on revenues, relative to ungraded food and entertainment establishments. In the first column, the change in revenues for graded restaurants in NYC is not statistically different from that for ungraded food and entertainment establishments (as indicated by the insignificant coefficient on *NYC_Rest_Graded*); a t-test against the coefficient on *Suburb_Rest_Graded* indicates that their difference is not statistically significant either. In the second column of Table 10 we add in county-specific time trends and control for a linear trend in revenues (both before and after the implementation of the grading regime). First, we see that the magnitudes on the *NYC_Rest_Graded* and *Suburb_Rest_Graded* coefficients increase substantially, suggesting that there is county-specific variation over time that was pushing those estimates down. Second, the post-grading effect on NYC restaurants is still positive and insignificant (relative to the ungraded food and entertainment establishments). Finally, the coefficient on *NYC_Rest_Graded* is statistically different than that on *Suburb_Rest_Graded*, suggesting that, on average, "graded" restaurants in NYC experienced a relatively smaller increase in revenues compared to similarly classified restaurants in the suburban counties that were not subject to grading. The fact that this difference is significant, while the difference with ungraded food and entertainment establishments is not, adds credence to a grading-induced suppression of sales. Moreover, this suppression is substantively meaningful at about $16,500, or 8-10% of the typical restaurant's revenues in the sample.

We now turn to a second set of estimates, *NYC_Rest_Posttrend* and *Suburb_Rest_Posttrend*, which allow the grading effect to vary linearly over time. While NYC "graded" restaurants see a significant increase in revenues over time relative to non-food entertainment establishments, there is no significant difference in the post-trends between NYC and Suburban restaurants; this suggests that, when we assume a linear trend for the six years after the implementation of the grading policy, the initial difference in revenues is sustained.[26]

Finally, we estimate the impact using a model that allows the post-grading effect to vary non-linearly over time (displayed in the third column of Table 10). A number of the findings change. First, the magnitude of the effect declines substantially for "graded" NYC restaurants relative to ungraded suburban restaurants (to about $2,000). Second, and more importantly, this difference is no longer statistically significant (although the difference, relative to ungraded food and entertainment establishments, becomes significant). In addition, while there is no statistical difference in the linear post-grading trends for sales, the nonlinear post-trend for the "graded" NYC restaurants is significantly different from that for the ungraded suburban restaurants. Both linear trends flatten out over time (more so for suburban restaurant sales), which results in the convergence of sales that was initially evident in Figure 3. Therefore, any immediate response in revenues is muted when we account for nonlinear revenue adjustments over the longer term.[27]

We test for the robustness of these findings in two ways. First, in order to narrow the comparison space, we restrict the sample to the counties in NYC that are adjacent to the suburban counties. It is conceivable that micro-regional shocks to the consumption or operation

[26] We check whether or not our estimates are influenced by a change in the composition of establishments over time (and specifically a higher likelihood of closure among "graded" NYC restaurants), and we see no evidence of this in the data. In fact, we see a slightly higher growth in establishments in the "graded" NYC group. See Appendix E for a visual.

[27] This finding is consistent with reports from officials at DOHMH, who observe an increasing number of "A" grades over time, such that the value-added of the posted food safety information could disappear over time.

of restaurants could still bias the estimates, and this specification will better control for such a threat. In addition, it is plausible that the spatially proximate communities on either side of the NYC border are more similar. These results are displayed in the final column of Table 10. The results are consistent with those from the full sample: while revenues grow across the board, the increase is relatively smaller for “graded” NYC restaurants compared to ungraded suburban restaurants. The magnitude of the difference is smaller, about $9,250 (about 10 percent of the mean sales for restaurants in the sample), but remains statistically significant in the presence of non-linear trends.[28] This finding suggests that consumers could be sorting away from the graded restaurants in NYC to those across the border in Nassau and Westchester counties (restaurants across the border could also be adjusting their behavior to signal better hygiene); for the citywide sample, consumers could also be sorting away from lower graded restaurants to higher graded ones within NYC, which would result in the neutral fiscal effect that we observe.

Second, we estimate a roll-out model on the panel of NYC-only restaurants; these results are displayed in Table 11. In the fully specified model (i.e. with group fixed effects), the coefficient on *Post* is negative and statistically insignificant. These results suggest that within the first year of the program, there was no significant revenue effect on those exposed earlier to the grading regime compared to those exposed later.[29] However, the sign on the grading impact coefficient

---

[28] One might also expect some cross-border contamination, if consumers are now internalizing restaurant food safety signals differently or if ungraded restaurants in the suburban counties change their behaviors to respond to an increased awareness around food safety. We exclude the NYC counties that are geographically proximate to the suburban counties (retaining only Manhattan, Brooklyn and Staten Island in the NYC sub-sample) and re-run identical specifications. The results remain substantively the same.

[29] We also tested for differential revenue effects across the roll-out period, but none of these results were statistically significant. These results are available from the authors upon request.

is negative, which is consistent with the findings from the triple difference-in-difference models where the increase in revenues for graded NYC restaurants was smaller than that for ungraded suburban restaurants. The lack of significance for this estimate can be explained in two ways. First, it is possible that we lose power, and therefore precision, in our reduced-sample estimation. Second, it could also reflect a lack of clear information to consumers in the first months of the policy. Among early exposed restaurants, only those earning *A*'s posted grades in the first couple of months post-grading (due to timing of grade posting requirements in the inspection cycle). The time between initial inspection and re-inspection is 2 to 4 weeks and between re-inspection and adjudication, between 4 to 6 weeks; treated restaurants could post nothing during the first window and *Grade Pending* during the second window if they did not earn an *A* initially. It is not clear whether or not this distinction was meaningful enough to influence dining choices, and therefore revenues, during the initial months of the policy.

VII. Conclusion

Cities have long inspected restaurants for their sanitary conditions, but the public disclosure of that information for consumers to incorporate into dining decisions is a relatively new phenomenon. The motivation for a restaurant grading policy is to make the sanitary conditions of a establishment more transparent and easily understood as a means of reducing the incidence of foodborne illnesses. Therefore, it is very much a health policy. However, if theory is correct, and consumers use this information to change their behaviors, restaurants (and the municipal fisc) could bear economic repercussions as well.

We systematically test these predictions using a collection of rich data on restaurants' food safety compliance and sales activities in NYC and neighboring counties, both before and after the implementation of NYC's grading policy. Our results suggest that NYC's restaurant grading policy, after an initial adjustment period, improves sanitary conditions (as measured by inspection scores) and reduces public revenues collected through fines.

The impact on sales revenues, however, is more nuanced. For restaurants in NYC overall, revenues go up following the policy (as they do for other kinds of establishments), but to a lesser degree than restaurants not subject to a grading regime. This suppressed revenue growth, however, goes away once we account for nonlinear trends in revenues over time. Furthermore, models that estimate effects during the policy's roll-out year show no significant revenue change for restaurants that are first exposed to the grading regime, compared to those that are later exposed (although the sign of the effect still does indicate smaller revenues). Finally, when we restrict the sample to only those NYC counties that border the suburban counties, the relatively slower growth in revenues for NYC "graded" restaurants (compared to ungraded restaurants outside of the city) is both statistically and financially meaningful. This suggests that consumers are perhaps substituting away from the graded restaurants towards those without grades, favoring lack of information on food safety over explicit information on poor hygiene. Restaurants in the ungraded regime might be adjusting their practices as well, to signal better food safety. While not significant, the signs and magnitudes of the trends over time echo those from the full-sample analysis, suggesting a similar convergence in revenues in these border counties.

The health goals, as they relate to the restaurant's food safety compliance, do seem to be addressed through the improvement of inspection scores. And the fiscal effects are neutral for the city overall, neither improving nor depressing restaurants' revenues (and the taxes they in turn generate). In addition, violation fines decline. However, these outcomes were not achieved immediately; the results consistently show that there is a period of adjustment, but one that does not last longer than six months. This makes sense, as consumers need to internalize the new information provided by the grades and the businesses need to then respond in their operations. Further, while a reduction in fines can be a boon to businesses (assuming their compliance costs do not exceed the savings in fines), it constitutes a revenue reduction for the City. According to our preferred estimates (and using the more conservative number of initial inspections only), this fiscal loss amounts to about $1.2 million for a typical fiscal quarter for the municipality, in addition to any increased administrative costs for running the program (which are estimated to be somewhere between $245 and $320 per inspection, averaging approximately $2.3 million in total annually to the City[30]). The magnitude of these costs relative to the potential benefits, however, is not entirely obvious without considering the potential health care savings from reduced incidences of foodborne illnesses.

Apart from the City's overall welfare, we should also be concerned about the distributional effects of such a policy. While the current analysis obscures any variation across restaurants over time, related papers (using NYC and Los Angeles data) find that there are meaningful differences in economic performance across restaurants with different grades: restaurants that post *A*s are less likely to close, owe fewer fines and bring in more revenues

[30] The average cost per inspection is calculated by dividing all spending in DOHMH's Food Safety budget by the reported number of actual restaurant inspections. Budget figures come from OMB Budget Function Analysis, and actual inspections come from DOHMH reporting to OMB. One reason costs may rise is that the City had to hire and train additional inspectors to implement increased use of reinspections.

compared to *B* restaurants (Schwartz et. al. 2015; Jin and Leslie 2003). Therefore, the relative benefits and burdens of the policy differ across restaurants. Certain restaurants may be able to more easily absorb the costs of managing higher stakes inspections and will likely benefit more from improved compliance. Likewise, depending on how these restaurants cluster across space, neighborhoods within cities could be differentially affected by the policy. Our border analysis suggests that this indeed could be the case.

VIII. References

Dafny, Leemore, and David Dranove. 2008. "Do report cards tell consumers anything they don't already know? The case of Medicare HMOs." *The Rand Journal of Economics* 39(3): 790-821.

Dranove, D., & Jin, G. Z. 2010. Quality disclosure and certification: Theory and practice. *Journal of Economic Literature*, 48(4): 935-963.

Fein, S. B., Lin, C. T. J., & Levy, A. S. 1995. Foodborne illness: perceptions, experience, and preventive behaviors in the United States. *Journal of Food Protection*, 58(12): 1405-1411.

Figlio, David N., and Maurice E. Lucas. 2004. "What's in a grade? School report cards and the housing market." *American Economic Review*, 94(3): 591-604.

Filion, Katie, and Douglas A. Powell. 2009. "The use of restaurant inspection disclosure systems as a means of communicating food safety information." *Journal of Foodservice*, 20(6): 287-297.

Freebairn, John W. "Grading as a market innovation." Review of Marketing and Agricultural Economics 35.3 (1967): 147-162.

Freebairn, John W. "The value of information provided by a uniform grading system." Australian Journal of Agricultural Economics 17.2 (1973): 127-139.

Hastings, J. S., & Weinstein, J. M. 2008. Information, school choice, and academic achievement: Evidence from two experiments. *The Quarterly Journal of Economics*, 123(4): 1373-1414.

Ho, D.E. (2012). Fudging the Nudge: Information Disclosure and Restaurant Grading. *Yale Law Journal,* Vol. 112(3): 574-688.

Ippolito, P. M., & Mathios, A. D. 1990. The regulation of science-based claims in advertising. *Journal of Consumer Policy*, 13(4): 413-445.

Jin, G.Z., and Leslie, P. 2003. The Effect of Information on Product Quality: Evidence from Restaurant Hygiene Grade Cards. *The Quarterly Journal of Economics*, Vol. 118 (2): 409-451.

Jin, G. Z., & Leslie, P. 2009. Reputational incentives for restaurant hygiene. *American Economic Journal: Microeconomics*, 1(1): 237-267.

Jones TF, Angulo FJ. 2006. Eating in restaurants: a risk factor for foodborne disease? *Clin Infect Dis.*, 43(10):1324-1328.

Mathios, A. D. 2000. The impact of mandatory disclosure laws on product choices: An analysis of the salad dressing market. *The Journal of Law and Economics*, 43(2): 651-678.

Mead, Paul S., Laurence Slutsker, Vance Dietz, Linda F. McCaig, Joseph S. Bresee, Craig Shapiro, Patricia M. Griffin, and Robert V. Tauxe. 1999. Food-related illness and death in the United States. *Emerging infectious diseases*, 5(5): 607-625.

New York City Department of Health and Mental Hygiene. (2010). Self-Inspection Worksheet for Food Service Establishments. Retrieved from http://www.nyc.gov/html/doh/downloads/pdf/rii/self-inspection-worksheet.pdf

New York City Department of Health and Mental Hygiene. 2012. How We Score and Grade. Retrieved from http://www.nyc.gov/html/doh/downloads/pdf/rii/how-we-score-grade.pdf.

New York City Office of Management and Budget. 2014. *Tax Revenue Forecasting Documentation*. Retrieved from http://www.nyc.gov/html/omb/downloads/pdf/methodology_2014_08.pdf

Rockoff, J.E., and Turner, L.J. (2010). Short-Run Impacts of Accountability on School Quality. *American Economic Journal: Economic Policy*, Vol. 2 (4):119-147.

Schwartz, Amy Ellen, Michah W. Rothbart, Thad Calabrese, Zachary Papper, Tod Mijanovich, Rachel Meltzer, Diana Silver and Meryle Weinstein. 2015. “The Impact of Restaurant Letter Grades on Taxes and Sales: Micro Evidence from New York City.” Working paper.

Silver, D., Rothbart, M., Bae, J., Schwartz, A.E., Mijanovich, T. 2015. *Public Restaurant Grades and Inspection Adjudication in New York City*. Manuscript in preparation.

Simon, P.A., Leslie, P., Run, G., Jin, G.Z., Reporter, R., Aguirre, A., Fielding, J.E. (2005). Impact of Restaurant Hygiene Grade Cards on Foodborne-Disease Hospitalization in Los Angeles County. *Journal of Environmental Health*, Vol. 67: 32-36.

Thaler, Richard H. and Cass R. Sunstein. 2009. *Nudge: Improving Decisions About Health, Wealth, and Happiness*. New York: Penguin Books.

Wedig, Gerard J., and Ming Tai-Seale. 2002. "The effect of report cards on consumer choice in the health insurance market." *Journal of Health Economics* 21(6): 1031-1048.

Winters, M.A., and Cowen, J.M. (2012). Grading New York: Accountability and Student Proficiency in America’s Largest School District. Educational Evaluation and Policy Analysis. http://epa.sagepub.com/content/early/2012/05/14/0162373712440039

Wong, M.R., McKelvey, W., Ito, K., Schiff, C., Jacobson, B., and Kass, D. (2015). Impact of Letter Grade Program on Restaurant Sanitary Conditions and Diner Behavior in New York City. *American Journal of Public Health* 105(3): 81-87.

Table 1. Restaurant Descriptive Statistics, New York City Sample

	Pre-Public Grading	**Post-Public Grading**
Number		
Inspections	3.3	6.2
Final Inspections	3.3	3.2
Workers	6.5	6.7
Seats	29.6	29.5
Cuisine		
American	0.22	0.24
Chinese	0.09	0.11
Pizza	0.04	0.06
Latin	0.04	0.04
Café/Coffee/Tea	0.03	0.04
Others	0.38	0.51
Missing	0.20	0.00
	1.00	1.00
Service		
Takeout-Limited Eat in	0.35	0.39
Wait Service	0.15	0.18
Wait and Counter Service	0.11	0.17
Takeout Only	0.08	0.08
Counter Service	0.07	0.12
Others	0.06	0.07
Missing	0.20	0.00
	1.00	1.00
Chain	0.10	0.10
Annual Closure Rate	0.16	0.12
N	30,405	34,917

Notes: Inspections include initial and re-inspections. Final inspections include all inspections in the pre-period, initial *A* inspections in the post-period, and re-inspections for those initially receiving *B* or *C* in the post-period. Workers, seats, cuisine, service, and chain reflect restaurant characteristics at the most recent restaurant inspection and are time-invariant variables. Annual closure rate is the fraction of open restaurants closing each year.

Table 2. Regression Results, Impact on Initial Inspection Scores, Pre-Post Estimation

VARIABLES	(1)	(2)	(3)	(4)	(5)
Post	-1.33***	1.21***	2.41***	1.23***	2.44***
	(0.09)	(0.17)	(0.20)	(0.20)	(0.34)
Post*Linear Trend	__	__	__	-0.33***	-1.46***
				(0.04)	(0.14)
Post*Linear Trend^2					-.03*
					(0.01)
Linear Trend	__	-0.22***	-0.28***	-0.10***	0.55***
		(0.01)	(0.02)	(0.03)	(0.10)
Linear Trend^2					.07***
					(0.01)
Seasonal FE	N	Y	Y	Y	Y
Rest. Char.	N	Y	N	N	N
Restaurant FE	N	N	Y	Y	Y
Constant	24.15***	16.10***	22.68***	23.40***	24.41***
	(0.07)	(4.16)	(0.13)	(0.16)	(0.22)
Inspections	159588	159588	116228	116228	116228
Restaurants	41362	41362	20641	20641	20641
R-squared	0.00	0.06	0.30	0.30	0.30

Notes: Robust standard errors clustered by restaurant in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 3. Regression Results, Impact on Final Inspection Scores, Pre-Post Estimation

VARIABLES	(1)	(2)	(3)	(4)	(5)
Post	-7.62***	-5.08***	-4.25***	-4.05***	-4.94***
	(0.08)	(0.16)	(0.18)	(0.18)	(0.29)
Post*Linear Trend	__	__	__	-0.29***	-1.10***
				(0.03)	(0.12)
Post*Linear Trend^2					-.07***
					(0.01)
Linear Trend	__	-0.20***	-0.18***	-0.05*	.68***
		(0.01)	(0.02)	(0.03)	(0.09)
Linear Trend^2					.08***
					(0.01)
Seasonal FE	N	Y	Y	Y	Y
Rest. Char.	N	Y	N	N	N
Restaurant FE	N	N	Y	Y	Y
Constant	24.55***	13.27***	22.57***	23.12***	24.18***
	(0.07)	(2.33)	(0.12)	(0.14)	(0.19)
Inspections	167,045	167,045	125,036	125,036	125,036
Restaurants	40,554	40,554	20,634	20,634	20,633
R-squared	0.06	0.11	0.31	0.31	0.31

Notes: Robust standard errors clustered by restaurant in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

Table 4. Regression Results, Impact on Inspection Scores, Rollout Estimation

	Initial Inspection Score		**Final Inspection Score**	
VARIABLES	(1)	(2)	(3)	(4)
Graded Inspection	-3.553***	-0.949***	-3.382***	-3.778***
	(0.184)	(0.159)	(0.181)	(0.160)
Quarter Of Grading Policy				
2	0.696***	0.379***	1.217***	1.573***
	(0.054)	(0.053)	(0.071)	(0.070)
3	0.543***	-0.071	0.574***	1.233***
	(0.091)	(0.066)	(0.123)	(0.093)
4	0.635***	0.334***	0.756***	2.051***
	(0.112)	(0.073)	(0.150)	(0.113)
Restaurant FE	N	Y	N	Y
Constant	25.981***	25.536***	20.219***	19.791***
	(0.109)	(0.059)	(0.099)	(0.066)
Observations	122,886	122,886	93,788	93,788
R-squared	0.007	0.848	0.011	0.763

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 5. Inspection Scores and Count of Inspections by Treatment Period, Pre-Post Sample

		Inspections: All Operating Restaurants	Inspections: Continuously Operating Restaurants
Pre		25.07 (76,231)	23.50 (29,804)
Quarters Post			
1-5	Initial	25.31 (41,933)	24.26 (17,723)
	Final Inspection Score	21.87 (27,874)	20.88 (11,743)
6-10	Initial	22.46 (46,180)	21.62 (18,409)
	Final Inspection Score	19.52 (29,135)	18.78 (11,544)

Includes pre-adjudicated inspection scores. Mean score shown on top; number of inspections shown parenthetically. Final Inspection Score is the mean restaurant inspection score for final inspections each cycle (all A-graded initial inspections and re-inspections of restaurants that do not get an A grade on initial inspection). Continuously Operating Restaurants are open for every quarter of the sample period.

Table 6. Regression Results, Impact on Inspection Fines, Pre-Post Estimation

VARIABLES	(1)	(2)	(3)	(4)	(5)
Post	-270.72***	-61.59***	88.67***	65.26***	110.57***
	(5.77)	(10.79)	(10.54)	(10.52)	(16.51)
Post*Linear Trend	__	__	__	-55.98***	-131.17***
				(1.84)	(7.01)
Post*Linear Trend^2					0.77
					(0.67)
Linear Trend	__	-14.35***	-22.16***	11.10***	42.24***
		(0.87)	(0.89)	(1.45)	-5
Linear Trend^2					3.34***
					(0.49)
Seasonal FE	N	Y	Y	Y	Y
Rest. Char.	N	Y	N	N	N
Restaurant FE	N	N	Y	Y	Y
Constant	1,141.52***	245.74*	947.74***	1,081.79***	1,132.0***
	(5.03)	(137.72)	(7.04)	(8.44)	(11.44)
Inspections	233,642	233,642	172,098	172,098	172,098
Restaurants	41,362	41,362	20,641	20,641	20,640
R-squared	0.01	0.05	0.28	0.29	0.29

Notes: Robust standard errors clustered by restaurant in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

Table 7. Regression Results, Impact on Fines by Quarter, Rollout Estimation

VARIABLES	(1)	(2)
Graded Inspection	50.918***	-132.357***
	(10.547)	(13.919)
Quarter Of Grading Policy		
2	95.591***	163.405***
	(9.496)	(10.108)
3	102.872***	235.971***
	(12.863)	(15.448)
4	99.762***	279.118***
	(11.105)	(15.132)
Restaurant FE	N	Y
Constant	358.596***	333.149***
	(5.546)	(5.741)
Observations	94,752	94,752
R-squared	0.003	0.551

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 8. Falsification Test, Inspection Scores, Fines, Revenues, and Sales Taxes, Pre-Post With 1 Year Lead

VARIABLES	Initial Score	Final Score	Fines
Policy Start			
Post	-0.64**	-6.76***	-138.92***
	(0.31)	(0.27)	(15.86)
Post*Linear Trend	-1.38***	-1.27***	-98.96***
	(0.13)	(0.12)	(7.01)
One Year Before			
Post	-0.76**	-0.43	88.83***
	(0.38)	(0.33)	(20.04)
Post*Linear Trend	1.30***	1.27***	53.91***
	(0.14)	(0.12)	(7.37)
Linear Trend	-0.34***	-0.34***	-0.32
	(0.06)	(0.05)	(2.85)
Seasonal FE	Y	Y	Y
Restaurant/Group FE	Y	Y	Y
Constant	23.15***	22.59***	989.98***
	(0.18)	(0.15)	(9.39)
Inspections	109,197	120,261	159,617
Restaurants	20,480	20,482	20,485
Observations	--	--	--
Rest-Quarters	--	--	--
R-squared	0.31	0.32	0.32

Robust standard errors clustered by restaurant in parentheses.

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

Table 9. Regression Results, Impact on Sales by Quarter, Difference-in-Difference

VARIABLES	(1)	(2)	(3)
Grading_Post	9,888.22*	11,466.94	6,324.63
	(4855.19)	(6658.48)	(5824.01)
Restaurant		56,691.90*	
		-27,931.33	
Rest_Grading		(2363.78)	3494.83
		(7706.71)	(7518.35)
Seasonal FE	Y	Y	Y
County FE	Y	Y	Y
Constant	346,500.53***	269,051.04***	345,939.48***
	(21612.73)	(48250.49)	(21331.12)
Observations	804,053	1,212,558	1,019,032
R-squared	0.034	0.022	0.012

Notes: Robust standard errors clustered by county-NAICS in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 10. Regression Results, Impact on Sales by Quarter, Triple Difference-in-Difference

VARIABLES	(1)	(2)	(3)	(4)
Grading_Post	1,323.71			
	(10427.72)			
Restaurant_NYC	48,765.38			
	(28799.77)			
Restaurant_Suburbs	-36,580.01			
	(26055.05)			
NYC_Rest_Graded	10,337.39	21,983.01	17,385.42**	15,224.04**
	(11294.69)	(15931.95)	(7110.68)	(5194.44)
Suburb_Rest_Graded	7631.83	38,453.69**	19,561.50***	24,479.21***
	(11799.32)	(15886.70)	(3359.05)	(4959.45)
NYC_Rest_Post_trend		3,312.77**	-1,367.26	-4,416.55
		(1333.45)	(5333.27)	(9054.88)
NYC_Rest_Post_trend^2			-38.64	-315.33
			(215.39)	(343.78)
Suburb_Rest_Post_trend		3452.09	4802.69	790.47
		(2160.76)	(4116.35)	(5342.32)
Suburb_Rest_Post_trend^2			-589.92	-712.36
			(395.07)	(542.78)
NYC_Post_trend		714.01	8439.84	6468.87
		(1856.84)	(7081.94)	(7000.81)
NYC_Post_trend^2			-316.31	-277.33
			(234.99)	(232.54)
NYC_Rest_Prepost_trend		-2,981.61**	-319.55	3132.46
		(1225.03)	(4053.26)	(6528.92)
NYC_Rest_Prepost_trend^2			134.92	388.66
			(243.21)	(433.32)
Suburb_Rest_Prepost_trend		-1,658.54	2727.12	5306.53
		(1646.23)	(2755.09)	(5701.21)
Suburb_Rest_Prepost_trend^2			308.08	483.36
			(237.58)	(434.71)
Manhattan_Restaurant		113,435.91**	123,895.66**	
		(55067.07)	(53081.06)	
Bronx_Restaurant		-149,372.11***	-138,909.01***	(24061.85)
		(51053.85)	(50013.71)	(17127.22)
Brooklyn_Restaurant		-145,897.73***	-135,430.95**	
		(50979.06)	(49987.64)	
Queens_Restaurant		-121,880.45**	-111,412.13**	3505.30
		(50889.46)	(49986.32)	(14586.02)
StatenIsland_Restaurant		-106,994.04**	-96,525.94*	
		(51415.43)	(50334.91)	

VARIABLES	(1)	(2)	(3)	(4)
Nassau_Restaurant		-16,768.20	3,152.55	111,873.41***
		(50521.79)	(49038.05)	(13238.04)
Westchester_Restaurant		-87,245.35*	-67,317.25	41,401.11**
		(51056.07)	(49662.37)	(15272.96)
Bronx_Ungraded		-153,607.40***	-153,650.87***	-41,501.26***
		(49767.62)	(49786.72)	(11508.28)
Brooklyn_Ungraded		-146,622.67***	-146,649.53***	
		(49482.83)	(49497.41)	
Queens_Ungraded		-112,248.91**	-112,264.42**	
		(50391.71)	(50404.59)	
StatenIsland_Ungraded		-118,834.21**	-118,814.04**	
		(49398.97)	(49402.34)	
Nassau_Ungraded		49,691.02	65,340.53	170,850.48***
		(51234.38)	(51776.35)	(20752.49)
Westchester_Ungraded		-6,561.37	9,078.67	114590.59
		(78110.87)	(81036.46)	(66472.72)
Seasonal FE	Y	N	N	N
County FE	Y	N	N	N
Quarter FE	N	Y	Y	Y
County trend	N	Y	Y	Y
Constant	276,225.43***	207,073.21***	186,436.26***	79,800.24**
	(49067.64)	(55060.56)	(57007.36)	(27852.27)
Observations	1,525,330	1,525,330	1,525,330	780,910
R-squared	0.01	0.01	0.01	0.002

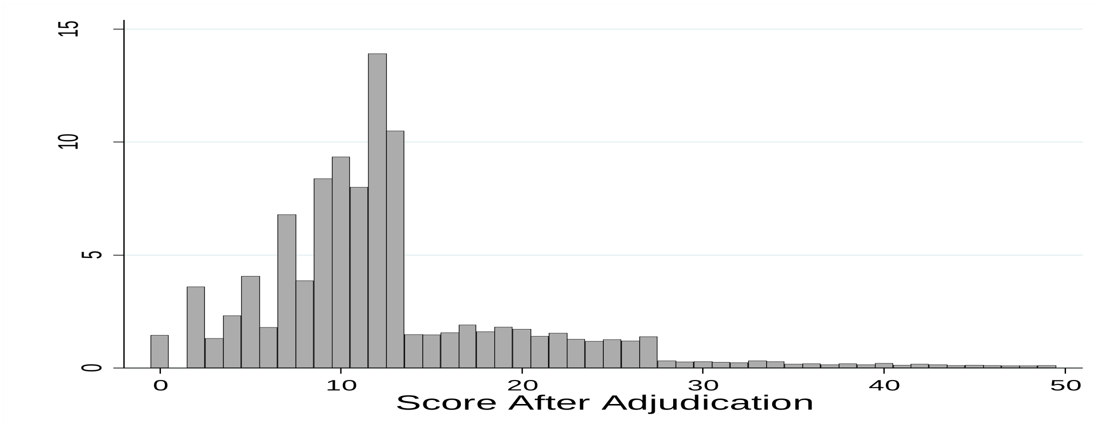
Table 11. Regression Results, Impact on Sales by Quarter, Food and Beverage Rollout Sample

VARIABLES	(1)	(2)
Graded Inspection	31,552.417	-4,485.497
	(39,710.175)	(6,686.311)
Quarter Of Grading Policy		
2	-7,763.466	1,866.532
	(10,710.803)	(2,298.146)
3	-26,199.939	-4,923.041
	(23,558.118)	(4,723.712)
4	-16,070.555	11,821.031***
	(31,101.395)	(5,603.230)
Group FE	N	Y
Constant	212,102.946***	212,970.365***
	(5,506.497)	(1,023.340)
Observations	3800	3800
Restaurant-Quarters	39,188	39,188
R-squared	0.002	0.981

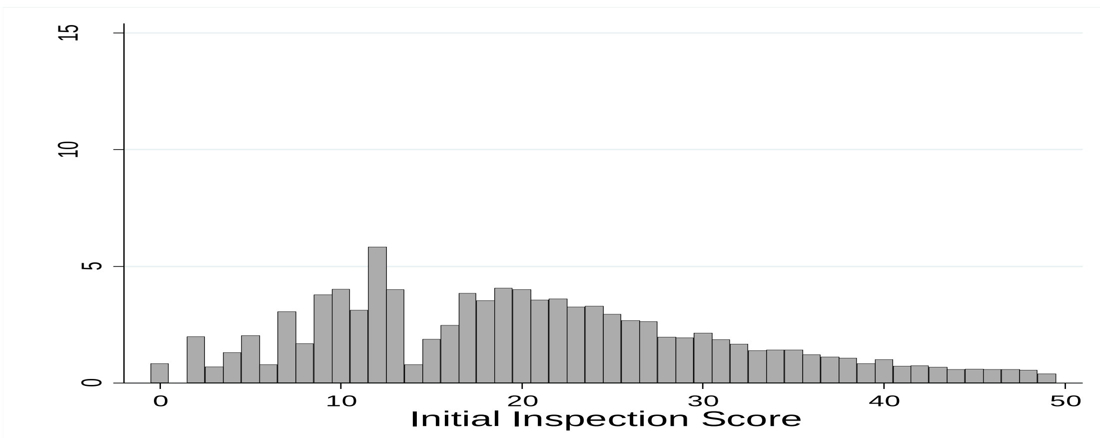
Notes: Robust standard errors clustered by group in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Figure 1: Distribution of inspection scores

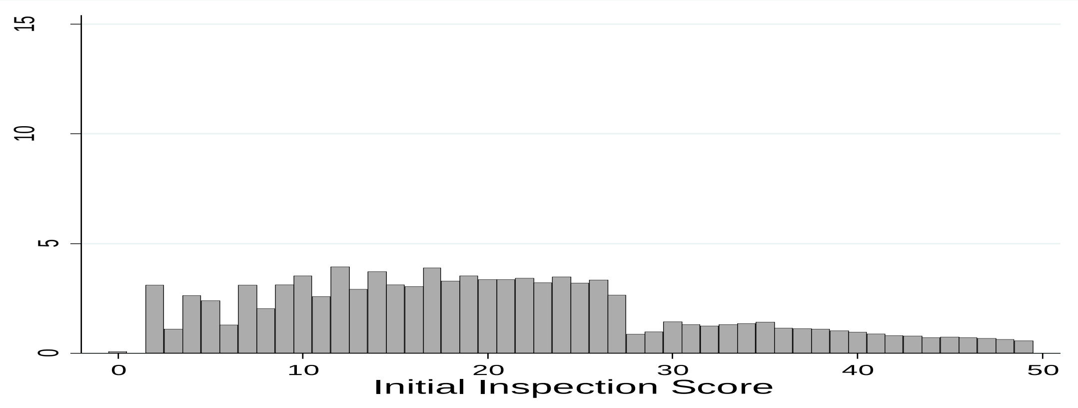
Panel A. Final Score Distribution, Second Year of Grading



Panel B. Initial Inspection Score Distribution, Second Year of Grading



Panel C. Initial Inspection Score Distribution, Two Years Before Grading



0
5
10
15
0
10
20
30
40
50
Initial Inspection Score

Figure 2. Average Fines by Quarter, Operating Restaurants

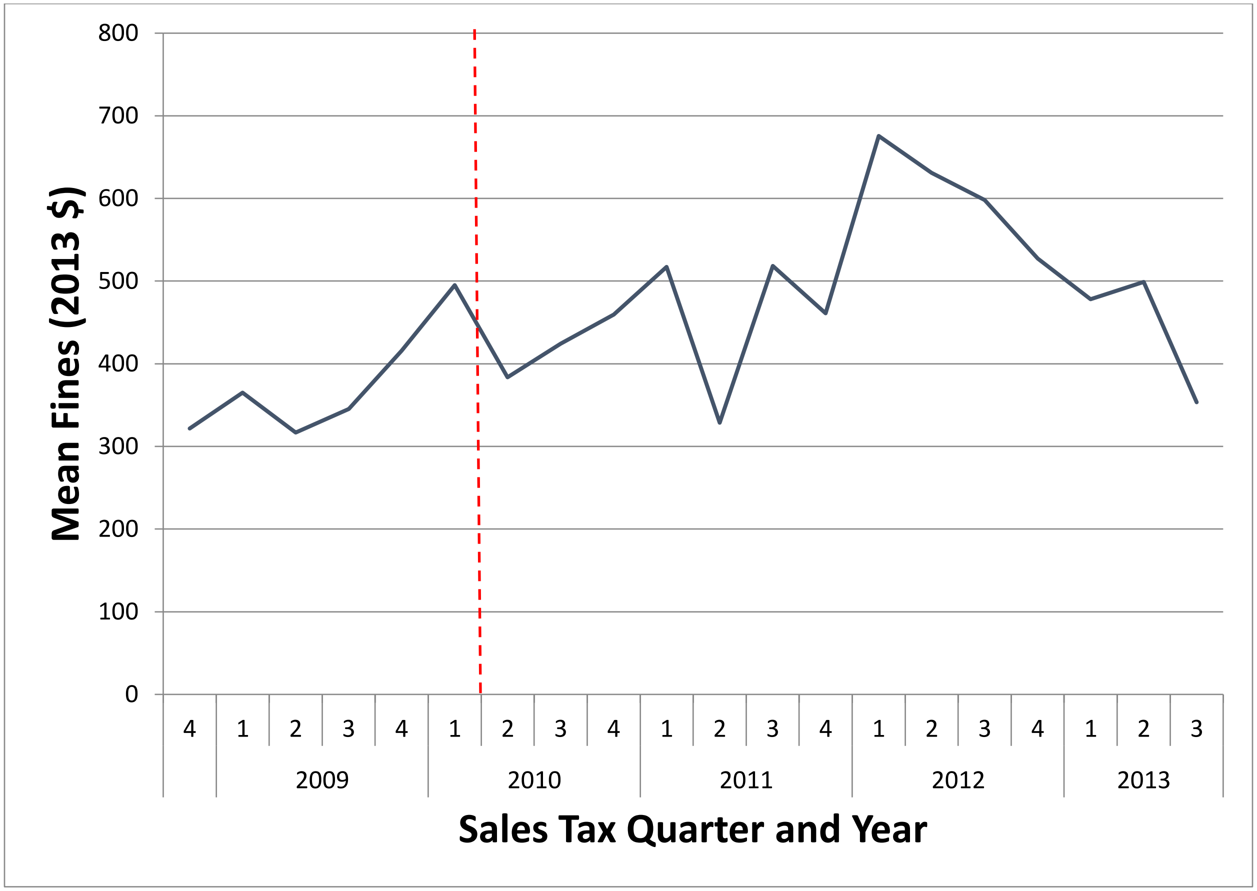
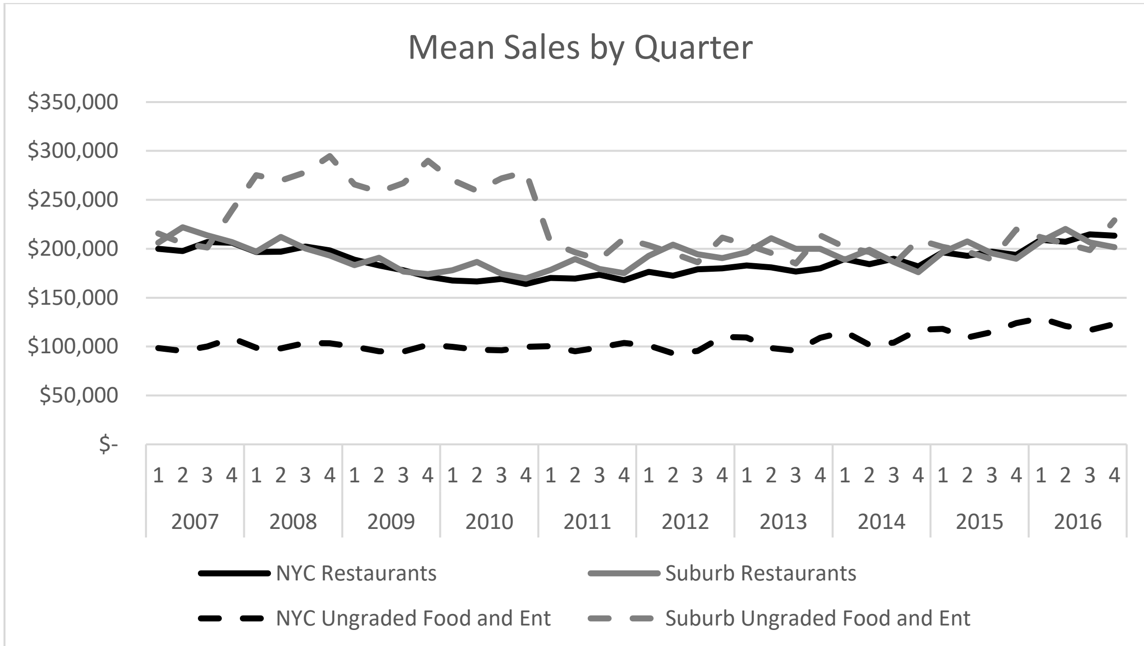


Figure 3. Average Sales, By Graded and Ungraded Establishments



Appendix A. Sample Posted B Grade

Appendix B. NAICS Codes for Establishment Groupings

"Graded" NYC Restaurants		Ungraded Suburban Restaurants		Ungraded Food and Entertainment Establishments	
NAICS	**Business/Service**	**NAICS**	**Business/Service**	**NAICS**	**Business/Service**
722---	Food Services and Drinking Places	722---	Food Services and Drinking Places	4451--*	Grocery Stores
445299	All Other Specialty Food Stores	445299	All Other Specialty Food Stores	4452--*	Specialty Food Stores
445291	Baked Goods Stores	445291	Baked Goods Stores	4453--	Beer, Wine, and Liquor Stores
445120	Convenience Stores	445120	Convenience Stores	7139--	Other Amusement and Recreation Industries
				712---	Museums, Historical Sites, and Similar Institutions
				711--	Performing Arts, Spectator Sports, and Related Industries

*These exclude the NAICS codes included in the restaurant groups, i.e. 445299, 445291, 445120.

Appendix C. Mean Treatment and Control Group Characteristics, Rollout Sample By Quarter

	Jun – Aug, 2010		Sept – Nov, 2010		Dec – Feb, 2011		Mar – May, 2011	
	Treatment	Control	Treatment	Control	Treatment	Control	Treatment	Control
Characteristics:								
Borough								
Manhattan	39.6%	41.4%	40.3%	41.8%	41.5%	40.8%	41.6%	40.6%
Bronx	9.4	9.8	10.2	9.4	10.0	9.0	9.6	9.3
Brooklyn	24.8	22.8	23.2	22.6	22.9	22.6	22.9	22.0
Queens	25.4	21.8	23.2	21.8	22.2	22.7	22.2	23.5
Staten Island	0.9	4.2	3.1	4.4	3.4	4.9	3.7	4.5
Chain	15.6%	12.3%	13.6%	12.1%	13.8%	9.9%	13.5%	7.8%
Workers	5.9	7.3	6.8	7.5	7.0	7.5	7.2	6.6
Building:								
Assessed Value	7,322,578	9,503,942	9,108,455	9,557,440	9,151,351	9,293,507	9,061,502	10,200,000

	Jun – Aug, 2010		Sept – Nov, 2010		Dec – Feb, 2011		Mar – May, 2011	
	Treatment	Control	Treatment	Control	Treatment	Control	Treatment	Control
Building Type								
Office/Commercial	9.2%	7.2%	8.0%	6.9%	8.0%	5.6%	7.6%	4.7%
Retail/Commercial	34.7	34.7	35.5	34.1	34.4	34.8	34.5	34.0
Mixed Retail	41.0	43.3	42.7	43.5	43.8	42.8	43.7	41.8
Other Commercial	4.5	4.5	4.0	4.8	4.1	5.4	4.3	6.4
Residential	9.0	8.2	7.9	8.4	8.3	8.2	8.2	8.7
Government/Public	1.6	2.2	1.9	2.4	1.5	3.4	1.7	4.4
Joint Significance	F(12, 937)	=0.80	F(12, 937)	=1.08	F(12, 937)	=0.71	F(11, 937)	=1.39

	Jun – Aug, 2010		Sept – Nov, 2010		Dec – Feb, 2011		Mar – May, 2011	
	Treatment	Control	Treatment	Control	Treatment	Control	Treatment	Control
	Prob > F	= 0.6523	Prob > F	= 0.3779	Prob > F	=0.7279	Prob > F	=0.1738
Initial Insp. Score	23.5	24.3	26.0	22.9	25.0	21.8	24.1	22.6
Final Insp. Score	18.0	20.1	20.9	19.2	20.5	18.3	19.9	18.8
Fines per Quarter	$341	$250	$329	$202	$285	$175	$262	$170
N	1,560	17,326	7,597	11,118	13,098	5,755	16,514	2,717
N with Building Code	1,082	12,122	5,364	7,817	9,333	4,000	11,742	1,858

Notes: Restaurants in the treatment group have their first graded inspection by the end of the fiscal quarter. Control group restaurants do not have a graded inspection until after the quarter ends. Each observation is a restaurant-quarter.

Appendix D. Regression Results, Impact on Inspection Scores and Fines by Quarter, Rollout Sample

VARIABLES	Initial Inspection	Final Inspection	Fines
Graded Inspection			
In Quarter 1:	1.897***	-2.137***	153.759***
	(0.677)	(0.379)	(21.950)
In Quarter 2:	0.810***	-2.377***	26.014
	(0.180)	(0.177)	(17.981)
In Quarter 3:	-0.693***	-3.616***	-147.830***
	(0.170)	(0.173)	(20.183)
In Quarter 4:	-2.071***	-5.561***	-275.079***
	(0.182)	(0.216)	(18.357)
Quarter Of Grading Policy			
2	0.074	1.243***	130.444***
	(0.052)	(0.072)	(10.915)
3	-0.133**	1.179***	251.674***
	(0.063)	(0.094)	(19.331)
4	0.974***	3.409***	387.343***
	(0.076)	(0.151)	(17.798)
Restaurant FE	Y	Y	Y
Constant	25.483***	19.736***	326.148***
	(0.059)	(0.067)	(6.021)
Observations	122,886	93,451	94,752
R-squared	0.849	0.764	0.553

Notes: Robust standard errors clustered by restaurant in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Impact estimates of graded inspections on initial inspection scores, final inspection scores, and fines by quarter. Graded Inspection captures if the most recent inspection occurs after the restaurant grade program begins in July 2010. Quarters Post * Graded is an interaction of the number of quarters after the grading policy is implemented and share of a quarter a restaurant has been treated. In Quarter 1, 2, 3, 4 are a vector of interactions between the quarter of observation and share of quarter with a graded inspection. Sample includes all restaurants continuously operating from the quarter before grading to five quarters following grading.

Appendix E. Composition of establishments, over time