Public Finance Review
1-32

DOI: 10.1177/1091142117715112
journals.sagepub.com/home/pfr
SAGE

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

Rachel Meltzer[1], Michah W. Rothbart[2], Amy Ellen Schwartz[2], Thad Calabrese[3], Diana Silver[4], Tod Mijanovich[5], and Meryle Weinstein[6]

Abstract

Grading schemes are an increasingly common method of quality disclosure for public services. Restaurant grading makes information about food safety

[1] Milano School, The New School, New York City, NY, USA
[2] Maxwell School, Syracuse University, Syracuse, NY, USA
[3] Wagner School, New York University, New York City, NY, USA
[4] College of Global Public Health, New York University, New York City, NY, USA
[5] Steinhardt School, New York University, New York City, NY, USA
[6] Institute for Education and Social Policy, Steinhardt School, New York University, New York City, NY, USA

Corresponding Author:
Rachel Meltzer, Milano School, The New School, 72 Fifth Avenue, Room 616, New York City, NY 10011, USA.
Email: meltzerr@newschool.edu

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.

Keywords
public grades, restaurant revenues, public resources

Public grading schemes are a popular way to concisely convey the quality of certain services. Municipalities across the United States grade the performance of public schools, street cleanliness is frequently scored and graded, and the Straphangers Campaign in New York City (NYC) produces a "report card" ranking the performance of each subway 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, empowering restaurant goers to choose safer restaurants and reducing their 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 to their bottom line. Does it increase compliance-related costs? Does it affect restaurants' sales? And, is the municipality just using the policy to enhance its own revenues? In this article, we shed light on all of these questions for the largest municipality with a restaurant-grading regime in the United States and present the first estimates of such a policy's aggregate financial impacts.

We use fine-grained administrative data that track food safety compliance and sales activity for the universe of restaurants in NYC and its bordering Westchester and Nassau counties, over multiple years, to assess the broader financial consequences of the restaurant grading policy. 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 prepolicy period continue, inspection scores (which reflect the number and severity of food safety violations) eventually improve following the implementation of public grading. More specifically, initial inspection scores increase (i.e., are worse) by about 2.5 points (about 10 percent of the pregrading mean) upon implementation but then decline (improve) roughly 1.5 points per quarter thereafter. Further, fines increase immediately after the start of the grading policy (by between US$65 and US$100 per inspection, or about 6 to 10 percent of the mean fine before grading) but decline thereafter such that any gain is reversed by the second quarter postimplementation (and further reduced each quarter after that).

The impacts on sales revenues, however, are less stark. While simple pre–post analyses indicate increasing revenue immediately after policy implementation, the relationship reverses when compared to control groups of ungraded food and entertainment establishments. However, allowing the postgrading effect to vary nonlinearly over time, and accounting for a long-term convergence of revenues across graded and ungraded establishments, attenuates this effect. While the difference in revenues is insignificant in the full sample, we see statistically significant differences when we narrow our analyses to areas in NYC that border suburban counties without any grading regime. These results suggest little overall revenue effect municipality-wide, but, smaller revenue increases for restaurants near ungraded restaurants in another county.

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 such information creates information asymmetries, whereby the restaurant operator knows the sanitation conditions inside the establishment and the consumer knows only what is easily visible. Consumers can also be informed by a business' reputation or personal experiences related to the establishment's food safety. These sources are 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 restaurant decisions based on incomplete information, restaurants have reduced incentives to address less obvious sanitation issues, and the incidence of foodborne illness may be inefficiently high. Public grading policies aim to address these information asymmetries (and associated health risks) by making sanitation inspection information more readily available to consumers and reducing their search costs. Specifically, 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). Since rating is mandatory and imposed by the government, presumably an objective third party, the information is likely to be viewed as more consistent and trustworthy than information volunteered by the restaurant or a private rating company with either monetary or reputational incentives (Dranove and Jin 2010). In addition, posting 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 as products of microdecisions by consumers. We make the reasonable assumption (supported by conversations with professionals and officials involved in the grading regime) that, in the NYC case, restaurant operating costs changed minimally; therefore, any response in revenues should capture shifts in consumer behavior.[1] Here, we discuss how, depending on the nature of these microdecisions, aggregate financial effects are theoretically ambiguous. In the most optimistic scenario, posting grades will induce existing restaurant consumers to sort away from establishments with lower grades and toward those with higher grades. Also, it will induce consumers to increase the frequency of their restaurant going. 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 food safety compliance, and sales will improve on average.

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 there is little (or no) increase in restaurant patronage and grading may simply trigger a resorting of existing patrons. In that case, food safety compliance may still improve over time, but sales would, on average, change little (if at all), as spending is reallocated from restaurants with lower grades to those with higher grades. Second, there may be little effect if the posted grades do not alter preconceptions about food safety. In this case, loyal patrons may prioritize information gathered from their firsthand experiences with the restaurant 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.

Background

NYC's Restaurant Grading Policy

The NYC Department of Health and Mental Hygiene (DOHMH) has long inspected NYC restaurants to ensure proper food safety practices, levying fines for violations and closing restaurants with public health hazards. In the early 2000s, they introduced a website where restaurants' violations were made available. The letter grading policy began in July 2010, following the same inspection criteria and scoring system, but assigning each restaurant a letter grade (A, B, or C) and requiring restaurants to post grades near the restaurant's entrance. Letters are printed in large bold font and are required to be near eye level and at the front entrance so that passersby can easily discern them (a posted grade example is shown in Online Appendix A). DOHMH also added the grades for each inspection to its website.

Inspection scores are the sum of violation points assigned during inspections, with lower scores reflecting more hygienic conditions. The number of points depends on the health risk posed to the public, classified in three categories: public health hazards, critical violations, and general violations. Additional points depend upon severity of the violation.[2] The scoring rubric used before and after the policy change is exactly the same. Following the policy change, however, DOHMH translated inspection scores into letter grades as follows: an inspection score of 13 or below earns a grade of A, between 14 and 27 a B, and 28 or higher a C.[3] 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 reinspection (for initial inspection scores above 13) or at the initial inspection if an A is earned. In addition to determining the grade, the inspection outcome affects the time to the next inspection visit. A's restaurants are inspected annually, B's twice a year, and C's three times a year. Inspection visits are unannounced within these known intervals.

Restaurants do have the right to due process. They may challenge violations (and, therefore, inspection scores, fines, and grades) at a tribunal administered by an independent agency, the Office of Administrative Trials and Hearings (OATH). Silver et al. (2015) find that restaurants earning B and C grades at inspection are much more likely to have scores reduced (improved) through adjudication in the postperiod than they were in the preperiod, resulting in substantially better grades. Thus, challenging grades in court seems to be an important tool for restaurants motivated to post A grades. 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.

Both before and after the implementation of grading, the type and count of inspection violations determine fines assessed. Fines range from US$200 to US$2,000 per violation and are assessed at an adjudication hearing at the discretion of a hearing officer.[4] After January 18, 2011, however, no fines were assessed for restaurants receiving an A grade at inspection; therefore, A's restaurants have no fines for much of the period following implementation.

Empirical Literature Review

Impact studies of public grading largely focus on the effects of grades and how differentiated information affects relevant outcomes.[5] In health, studies of health plan rankings typically find that higher-ranked plans see increases in market share, but the impact is lower when they are hard to understand or provide little new information to the consumer (e.g., Wedig and Tai-Seale 2002; Dafny and Dranove 2008). In education, many districts grade schools on 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 affect house prices (Figlio and Lucas 2004). In addition, Hastings and Weinstein (2008) show parents use mandated school quality information to move their children to higher performing schools.

Empirical research on food quality disclosure, and grading in particular, is scarce. While the mechanisms of disclosure might be similar to other contexts, the magnitude and transient nature of the decision is at a different scale than for health-care plans or schools (Ippolito and Mathios 1990). Studies assessing the impact of food content disclosure on consumer behavior, in contrast, find that consumers increasingly opt for the "healthier" option when the disclosure of relevant information becomes mandatory (Ippolito and Mathios 1990; Mathios 2000). These do not, however, evaluate a standardized grading regime like restaurant grading. Ho (2012) analyzes publicly available restaurant grading data for NYC, exploring the link between inspection scores in one period and the next. He finds that prior scores predict future grades poorly. Wong et al. (2015) provide new evidence of improved compliance since the beginning of NYC's public grading program and offer survey evidence of 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).

There are three impact studies of restaurant grading regimes.[6] Two (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) use ordinary least squares (OLS) and difference-in-difference 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. Simon et al. (2005) also provide evidence of reduced hospitalizations due to foodborne illnesses in Los Angeles County, compared to California overall. Jin and Leslie also suggest that 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 (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 restaurant sales (and associated sales taxes) and decreases 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.

Data and Measures

Our analytical approach is multipronged. We employ several metrics to capture the grading program's effect, two data sets, and alternative identification strategies to exploit, where possible, more detailed data on the NYC grading program. Data were drawn in two samples: for NYC restaurants only and for a larger sample including restaurant and other ungraded food and entertainment establishments for NYC and two suburban counties for comparison.

NYC Restaurant Grading and Compliance Data

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

We use scores assessed at initial inspection to capture food safety compliance. Importantly, inspections take place without advanced notice, and inspectors are randomly assigned to their visits (and do not visit the same site for reinspections).[8] Initial scores will reflect an unanticipated response from restaurants with respect to food safety compliance.

Finally, we use data on fines levied to assess the program's public revenue generation (and conversely, the financial burden on restaurants). All dollar values are adjusted using urban Consumer Price Index (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 "preperiod" and "postperiod," 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 postperiod.

Sales Revenue and Tax Data

NYC 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 restaurants according to the following protocol: (1) matching

Table 1. Restaurant Descriptive Statistics, New York City Sample.

Restaurant Characteristics	Prepublic Grading	Postpublic 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

Note: Inspections include initial and reinspections. Final inspections include all inspections in the preperiod, initial A inspections in the postperiod, and reinspections for those initially receiving B or C in the postperiod. 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.

DOHMH restaurant data and NYC DOF sales data using employer identification numbers (EINs) and (2) 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 a set of ten restaurants randomly assigned to the same group for each quarter year in the pregrading and postgrading 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]

NYC, 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 coordinated with the State's Department of Taxation and Finance.[13] This involved a separate data request, with slightly different grouping parameters (subject to the same confidentiality requirements to group the tax filer data into bins of ten). 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 North American Industry Classification System (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: (1) suburban establishments with similar NAICS codes as "graded" restaurants in NYC (henceforth referred to as ungraded restaurants) and (2) 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 Online Appendix C 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 data set 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 the ten-year sample period from 2007 to 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.

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.

Inspection Scores and Fines

We use inspection scores and fines to measure (1) food safety compliance and (2) fiscal burdens.[16] Fines are also a good measure of the policy's direct fiscal impact on the city since the new grading regime imposes little or no new costs.[17] We begin with a standard pre–post model as follows:

$$y_{it} = \beta_0 + \boldsymbol{Grading\_Post}_{it}{}'\beta_1 + \mathbf{X}_i{}'\beta_2 + \beta_3 Pre\_Post_{it} + \gamma_i + \delta_t + \varepsilon_{it}. \tag{1}$$

Here, y is restaurant outcomes including inspection scores and fines. ***Grading_Post*** is a set of three variables that capture the implementation of the policy: *Post*, *Post_trend*, and *Post_trend*^{2}. *Post* takes a 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 policy implementation. *Post_trend* and *Post_trend*^{2} are the interaction of *Post* with linear and quadratic time trends, allowing the effect to vary over time; **X** is a vector of restaurant characteristics including cuisine, service, and venue type; *Pre_Post* is a linear (and, in some 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) substituting restaurant fixed effects, μ_i, for γ_i and $\boldsymbol{X}_i$.

When grading started, restaurants were not uniformly exposed to the new inspection regime; grades had to be posted only after the first postperiod graded inspection. We exploit this variation in grade posting during the rollout period in an alternative specification, where we limit the sample to restaurant-quarter observations in the policy's first year. Thus, we compare "early posters" to "late posters" as follows:

$$y_{it} = \beta_0 + \beta_1 Post\_Rollout_{it} + \mathbf{X}_i'\beta_2 + \gamma_i + \delta_t + \varepsilon_{it}. \tag{2}$$

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

Sales Revenues

As described above, we compare "graded" NYC restaurants to ungraded restaurants in suburban counties and to ungraded food and entertainment establishments in NYC and suburban counties. We include this third group of establishments, which are never subject to grading, 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:

$$y_{gq} = \boldsymbol{Grading\_Post}_{gq}{}'\tau_1 + \mathbf{Restaurant}_{gq}{}'\tau_2 + \boldsymbol{Rest\_Post}_{gq}{}'\tau_3 + \tau_4 Pre\_Post_{gq} + \gamma_{cr} + \delta_q + \varepsilon_{gq}. \tag{3}$$

Here, y_{gq} is the group's average restaurant sales in quarter q. As above, ***Grading_Post*** is a set of three variables: Post, Post_trend, and *Post_trend*^{2}. *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_trend*^{2} 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. ungraded food and entertainment establishments, the latter of which is omitted as the reference category). These variables will also help control for reputational differences across different types of establishments. We also include interaction terms, included in ***Rest_Post***, whose coefficients capture the postgrading effect on "graded" NYC restaurants and suburban restaurants relative to ungraded food and entertainment establishments. Therefore, τ_3 identifies the effect from any new information provided by the posted grades, above and beyond the influence of sector, geography or pregrading reputation. Finally, γ_{cr} and δ_q are county-NAICS and quarter fixed effects, respectively.[18] We cluster the standard errors by county-NAICS to mitigate spatial autocorrelation across proximate establishments.[19]

As above, we specify an alternative, rollout model, exploiting the detailed panel data on sales revenues for NYC graded restaurants alone:

$$y_{gq} = \tau_1 Post\_Rollout_{gq} + \mathbf{X}'_{gq}\tau_2 + \gamma_g + \delta_q + \varepsilon_{gq}, \tag{4}$$

where y_{gq} is group mean sales in quarter q and *Post_Rollout* is the average share of days in quarter q a grade placard is posted for restaurants in group g. The remaining variables are identical to those defined above for inspection score and fines models but aggregated to the group level.

Results

Inspection Scores

We first discuss results from the pre–post analysis, estimating the grading policy's impact on food safety compliance, as measured by inspection scores from 2007 to 2013. Table 2 shows initial inspection score results.

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 preperiod. When we include additional controls, such as seasonal and zip code 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 code fixed effects and time-invariant restaurant characteristics) and 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 postgrading 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 effect of the grading policy indicates less initial compliance, the slope effect over time indicates that compliance has improved since public grading began. 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.[20]

As an alternative specification, we exploit the policy's rollout period to identify the impact of posting a grade on inspection scores. One concern

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

Variables	(1)	(2)	(3)	(4)	(5)
Post	−1.33*** (0.09)	1.21*** (0.17)	2.41*** (0.20)	1.23*** (0.20)	2.44*** (0.34)
Post × Linear Trend				−0.33*** (0.04)	−1.46*** (0.14)
Post × Linear Trend^2					−0.03* (0.01)
Linear Trend		−0.22*** (0.01)	−0.28*** (0.02)	−0.10*** (0.03)	0.55*** (0.10)
Linear Trend^2					0.07*** (0.01)
Seasonal FE	N	Y	Y	Y	Y
Restaurant Characteristics	N	Y	N	N	N
Restaurant FE	N	N	Y	Y	Y
Constant	24.15*** (0.07)	16.10*** (4.16)	22.68*** (0.13)	23.40*** (0.16)	24.41*** (0.22)
Inspections	159,588	159,588	116,228	116,228	116,228
Restaurants	41,362	41,362	20,641	20,641	20,641
R^2	0.00	0.06	0.30	0.30	0.30

Note: Robust standard errors clustered by restaurant in parentheses. FE = fixed effect.
*$p < 0.1$.
**$p < 0.05$.
***$p < 0.01$.

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

The results for the rollout regression analysis are displayed in Table 3, and we begin with the most parsimonious model, controlling for restaurant characteristics and time trends. Initial scores go down; specifically, in the models with restaurant fixed effects, inspection scores for restaurants exposed to the grading policy (compared to those not yet exposed) go down by just under one point.[21] We recognize that there could be a period of adjustment, even during the rollout period. To test for this, we replicate the rollout analysis, allowing the effect of the graded inspection to vary across time. These results are displayed in Online Appendix E. The immediate effect of the graded inspection is positive for initial inspections, and over the course of the rollout 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 inspection scores are 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 Online

Table 3. Regression Results, Rollout Estimation, Inspection Scores, and Fines.

	Initial Inspection Score		Fines	
Variables	(1)	(2)	(3)	(4)
Graded inspection	−3.553*** (.184)	−0.949*** (.159)	50.92*** (10.55)	−132.36*** (13.92)
Quarter of grading policy				
2	0.696*** (.054)	0.379*** (.053)	95.59*** (9.50)	163.41*** (10.11)
3	0.543*** (.091)	−0.071 (.066)	102.87*** (12.86)	235.97*** (15.45)
4	0.635*** (.112)	0.334*** (.073)	99.76*** (11.11)	279.12*** (15.13)
Restaurant FE	N	Y	N	Y
Constant	25.981*** (.109)	25.536*** (.059)	358.596*** (5.546)	333.149*** (5.741)
Observations	122,886	122,886	94,752	94,752
R^2	.007	.848	.003	.551

Note: Robust standard errors in parentheses. FE = fixed effect.
*p < .1.
**p < .05.
***p < .01.

Appendix F, which also show an improvement (i.e., decline) in inspection scores as the program progressed.[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 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 it is unlikely that inspectors are colluding to systematically reduce scores, other than based on an observed improvement in food safety compliance.[23] 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. Third, a similar distribution of initial scores (see Online Appendix G), before and after grading, suggests that any inspector-driven grade inflation was, at worst, minimal.

Fines

Next, we consider how the grading policy affects fines. 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. Starting with a simple pre–post model in the first column of Table 4, 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 declined US$271 per inspection. When we add in restaurant controls, a linear time trend, and zip code and seasonal fixed effects, the magnitude on the *Post* coefficient declines substantially but still remains negative: fines were reduced by about US$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 4, our preferred models show that upon policy implementation, fines increase (by about US$65 or US$110 per inspection, depending on whether or not a nonlinear posttrend 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 inspection scores, which also go down over the first year of the policy.

Table 4. Regression Results, Impact on Inspection Fines, Pre–Post Estimation.

Variables	(1)	(2)	(3)	(4)	(5)
Post	−270.72*** (5.77)	−61.59*** (10.79)	88.67*** (10.54)	65.26*** (10.52)	110.57*** (16.51)
Post × Linear Trend				−55.98*** (1.84)	−131.17*** (7.01)
Post × Linear Trend^2					0.77 (0.67)
Linear Trend		−14.35*** (0.87)	−22.16*** (0.89)	11.10*** (1.45)	42.24*** (4.99)
Linear Trend^2					3.34*** (0.49)
Seasonal FE	N	Y	Y	Y	Y
Restaurant Characteristics	N	Y	N	N	N
Restaurant FE	N	N	Y	Y	Y
Constant	1,141.52*** (5.03)	245.74* (137.72)	947.74*** (7.04)	1,081.79*** (8.44)	1,132.0*** (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^2	.01	.05	.28	.29	.29

Note: Robust standard errors clustered by restaurant in parentheses. FE = fixed effect.
*$p < .1$.
**$p < .05$.
***$p < .01$.

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 beyond the first six months.

Again, we repeat the analysis using the rollout sample, comparing restaurants that were exposed to the grading regime earlier in the rollout period to those that were exposed later. The third column of table 3 displays the results for the model without restaurant fixed effects; restaurants exposed earlier pay higher fines on average than those exposed later—about US$51 more per inspection. Reestimating with restaurant fixed effects so that we can compare fines within restaurants and the sign of the coefficient on *Post_Rollout* flips to a negative sign. Therefore, in our preferred model, we estimate fines decline US$132 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.

Falsification tests for inspection scores and fines

For the pre–post analyses of inspection scores and fines, we conduct a placebo test to assess the timing of level and slope changes relative to the grading policy implementation. We reestimate the impact, assigning a placebo policy start date one year prior to the actual policy start, because the policy change 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 Online Appendix H.[24]

In the case of inspection scores, we see that the false start date is associated with an immediate decline in 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 linear trend line. The true policy start date (while still controlling for the false start date) is associated with statistically significant declines for initial scores and then a continued decline over the postperiod.[25] 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, 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.

Sales Revenues

We now turn to the sample including NYC and bordering counties for 2007 through 2016. The first column of Table 5 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 reported a nearly US$10,000 increase in sales.[26] 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 *Rest_Post*, our impact estimate, is negative and statistically insignificant. Compared to food and entertainment establishments not subject to grading, "graded" restaurants saw no significant change in revenues after the policy implementation. We also note that the coefficient on *Restaurant* is positive and significant (albeit marginally), which captures, at least in part, pregrading differences (between "graded" restaurants and ungraded establishments in NYC).

One concern with the difference-in-difference estimate is that ungraded food and entertainment establishments could be subject to different macroeconomic trends than the "graded" restaurants. In addition, the estimates of *Rest_Post* could be biased by cross contamination of the grading effect onto ungraded 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 5, we use the sample including NYC and bordering counties but retain only restaurants. This model compares revenue outcomes for NYC "graded" restaurants to those for ungraded restaurants in Nassau and Westchester. While the coefficient on *Rest_Post* 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.

The next set of results (shown in Table 6) includes a second counterfactual. We specify the model as a "triple difference"—where graded NYC

Table 5. Regression Results, Impact on Sales by Quarter, Difference-in-Difference.

Variables	(1)	(2)	(3)
Rest × Post	9,888.22* (4,855.19)	−2,363.78 7,706.707)	3,494.83 (7,518.347)
Restaurant		56,691.90* (27,931.331)	
Post		11,466.94 (6,658.478)	6,324.63 (5,824.012)
Seasonal FE	Y	Y	Y
County FE	Y	Y	Y
Includes Suburbs	N	N	Y
Constant	346,500.53*** (21,612.73)	269,051.04*** (48,250.49)	345,939.48*** (21,331.12)
Observations	804,053	1,212,558	1,019,032
R^2	.034	.022	.012

Note: Robust standard errors clustered by county-NAICS in parentheses.
*p < .1.
**p < .05.
***p < .01.

Table 6. Regression Results, Impact on Sales by Quarter, Triple Difference-in-Difference.

Variables	(1)	(2)	(3)	(4)
Rest × Post				
NYC	10,337.39 (11,294.69)	21,983.01 (15,931.95)	17,385.42** (7,110.68)	15,224.04** (5,194.44)
Suburbs	7,631.83 (11,799.32)	38,453.69** (15,886.70)	19,561.50*** (3,359.05)	24,479.21*** (4,959.45)
Rest × Post Linear Trend				
NYC		3,312.77** (1,333.45)	−1,367.26 (5,333.27)	−4,416.55 (9,054.88)
Suburbs		3,452.09 (2,160.76)	4,802.69 (4,116.35)	790.47 (5,342.32)
Rest × Post Linear Trend^2				
NYC			−38.64 (215.39)	−315.33 (343.78)
Suburbs			−589.92 (395.07)	−712.36 (542.78)
Post	1,323.71 (10,427.72)			
NYC postlinear trend		714.01 (1,856.84)	8,439.84 (7,081.94)	6,468.87 (7,000.81)
NYC postlinear trend^2			−316.31 (234.99)	−277.33 (232.54)
County FE	Y	N	N	N
Business-type FE	Y	N	N	N
County-business–type FE	N	Y	Y	Y
Seasonal FE	Y	N	N	N
Quarter FE	N	Y	Y	Y
County-specific trends	N	Y	Y	Y
Border counties only	N	N	N	Y
Constant	276,225.43*** (49,067.64)	207,073.21*** (55,060.56)	186,436.26*** (57,007.36)	79,800.24** (27,852.27)
Observations	1,525,330	1,525,330	1,525,330	780,910
R^2	.01	.01	.01	.002

Note: NYC = New York City; FE = fixed effect.

restaurants are compared both to similarly classified ungraded restaurants in suburban counties and to ungraded food and entertainment establishments in the NYC and suburban counties. We adopt a model that is otherwise similar to the NYC difference-in-difference model. Two new variables are *NYC_Rest_Post* and *Suburb_Rest_Post*. Their coefficients capture the effect of the grading policy on revenues, relative to ungraded food and entertainment establishments. In column one, the change in revenues for graded NYC restaurants is not statistically different from that for ungraded food and entertainment establishments (indicated by the insignificant coefficient on *NYC_Rest_Post*); a *t*-test against the coefficient on *Suburb_Rest_Post* indicates the difference is not statistically significant. In the second column of table 6, we add county-specific time trends and control for a linear trend in revenues (both before and after the implementation of grading). First, we see that the magnitudes on the *NYC_Rest_Post* and *Suburb_Rest_Post* coefficients increase substantially, suggesting that there is county-specific variation over time that was pushing those estimates down. Second, the postgrading effect on NYC restaurants is still positive and insignificant (relative to the ungraded food and entertainment establishments). Finally, the coefficient on *NYC_Rest_Post* is statistically different than that on *Suburb_Rest_Post*, 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 US$16,500 or 8 to 10 percent 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 nonfood entertainment establishments, there is no significant difference in the posttrends 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.[27]

Finally, we estimate the impact using a model that allows the postgrading effect to vary nonlinearly over time (displayed in the third column of table 6). 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 US$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 postgrading trends for sales, the nonlinear posttrend 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 a convergence of sales (this is visible in Online Appendix J). Therefore, any immediate response in revenues is muted when we account for nonlinear revenue adjustments over the longer term.[28]

We test the robustness of our results in two ways. First, in order to narrow the comparison space, we restrict the sample to NYC counties that are adjacent to the suburban counties. It is conceivable that microregional shocks to the consumption or operation 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 shown in the final column of table 6. Results are consistent with the full-sample results: 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 US$9,250 (about 10 percent of mean sales for restaurants in the sample) but remains statistically significant in the presence of nonlinear trends.[29] This suggests consumers may be sorting away from graded restaurants in NYC to those in Nassau or Westchester counties (which may be adjusting 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 rollout model using the panel of NYC-only restaurants (see Table 7). In the model with group fixed effects, the *Post* coefficient is negative and statistically insignificant. These results suggest no significant revenue effect on those exposed to grading early compared to those exposed later, in the first year.[30] However, the negative sign on the grading impact coefficient is consistent with the findings from the triple difference models where the increase in revenues for graded NYC restaurants was smaller than that for ungraded suburban restaurants. Two explanations for the insignificance of this estimate are possible. First, it may reflect the loss of statistical power, and therefore precision, in the reduced sample. Second, it could reflect a lack of clear information to consumers in the first months of the policy. Among early exposed restaurants, only those earning A posted grades in the first months of the postperiod. The time

Table 7. Regression Results, Impact on Sales by Quarter, Food and Beverage Rollout Sample.

Variables	(1)	(2)
Graded inspection	31,552.42 (39,710.18)	−4,485.51 (6,686.31)
Quarter of grading policy		
2	−7,763.47 (10,710.80)	1,866.53 (2,298.15)
3	−26,199.94 (23,558.12)	−4,923.04 (4,723.71)
4	−16,070.56 (31,101.40)	11,821.03*** (5,603.23)
Group FE	N	Y
Constant	212,102.95*** (5,506.50)	212,970.37*** (1,023.34)
Observations	3,800	3,800
Restaurant quarters	39,188	39,188
R^2	.002	.981

Note: Robust standard errors clustered by group in parentheses. FE = fixed effect.
*p < .1.
**p < .05.
***p < .01.

between initial inspection and reinspection is two to four weeks, and between reinspection and adjudication, it is between four to six 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.

Conclusion

Cities have long inspected restaurants for sanitary conditions, but the public disclosure of that information is a relatively new phenomenon. The motivation for a restaurant grading policy is to make restaurant sanitary conditions more transparent and easily understood as a means to reduce 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 behavior, restaurants (and municipalities) could bear economic repercussions as well.

We systematically test these predictions using 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, improved sanitary conditions (as measured by inspection scores) and reduced public revenues collected through fines.

The impact on sales revenues is more nuanced. For NYC restaurants overall, revenues rise following the policy (as for other kinds of establishments) but less than restaurants not subject to grading. This suppressed revenue growth disappears once we account for nonlinear trends in revenues over time. Furthermore, models that estimate effects during the rollout year show no significant revenue change for restaurants exposed early to the grading regime, compared to those exposed later. Finally, restricting the sample to only those NYC counties bordering the suburban counties shows a slower growth in revenues for NYC "graded" restaurants compared to ungraded restaurants outside of the city, which is both statistically and financially meaningful. This suggests consumers may substitute away from graded restaurants toward ungraded restaurants. 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 time trends echo the full-sample results, suggesting a similar convergence in revenues in these border counties.

The health goals, related to food safety compliance, do seem 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; results consistently indicate a period of adjustment, which does not last longer than six months. This is consistent with consumers and businesses taking time to act upon the new information provided by grades. Further, while a reduction in fines can be a boon to businesses (assuming compliance costs do not exceed the savings in fines), it constitutes a revenue reduction for the city. According to our preferred estimates, this fiscal loss amounts to about US$1.2 million for a typical fiscal quarter for the municipality or about 10 percent of total fines levied in the typical quarter leading up to the grading policy. This loss is in addition to any increased administrative costs for running the program (which are estimated to be somewhere between US$245 and US$320 per inspection, averaging approximately US$2.3 million in total annually to the city[31]). 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.

Beyond overall welfare, we might also be concerned about the distributional effects. While the current analysis obscures variation across restaurants over time, related papers (using NYC and Los Angeles data) find meaningful differences in economic performance across restaurants with different grades: restaurants posting As are less likely to close, owe fewer fines, and earn more revenue compared to B restaurants (Schwartz et al. 2015; Jin and Leslie 2003). Thus, the relative benefits and burdens of the policy differ across restaurants. Some restaurants may more easily absorb the costs of managing higher stakes inspections and benefit 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.

Acknowledgments

We gratefully acknowledge contributions from the New York City Department of Finance and the New York State Department of Finance. Special thanks to Karen Schlain, Zachary Papper, Steven Robinson, William Raleigh, and James Stevens.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The authors thank the Robert Wood Johnson Foundation for generous support for this research.

Supplemental Material

The online appendices are available at online.

Notes

1. Department of Health and Mental Hygiene (DOHMH) informed us that violations are most frequently due to habits and training and not capital disrepair.
2. See New York City (NYC) Department of Health and Mental Hygiene (2010, 2012) and Schwartz et al. (2015) for detailed explanations of how points and grades are assigned (http://www1.nyc.gov/assets/doh/downloads/pdf/rii/blue-book.pdf).
3. Restaurants can also be temporarily closed if they pose a large public safety risk.
4. An exception occurs if the grade is accepted, in which case a smaller fine is paid by the restaurant operator.

5. For a comprehensive review of the literature on quality disclosure, see Dranove and Jin 2010.
6. We do not focus on the incidence of foodborne illness. While important, it is extremely difficult to empirically link it to the prevalence of posted inspection grades (Filion and Powell 2009). It is difficult for restaurant patrons to correctly identify the foodborne illness and attribute its source (Jones and Angulo 2006; Mead et al. 1999; Fein, Lin, and Levy 1995); moreover, it relies on their reporting the illness, which is done inconsistently (Jones and Angulo 2006; Mead et al. 1999).
7. Restaurant characteristics are recorded at the last inspection only, so they do not vary with time (they are not tracked by DOHMH).
8. We replicate all analyses using final inspection scores, which will, to some extent, reflect learning or adjustment by restaurants. The results from analyses using final inspection scores are substantively the same as those using the initial inspection scores. Results from regressions using final inspection scores are available in Online Appendix B.
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 percent of gross sales—4.875 percent for New York State and 4.0 percent 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 US$300,000 or less of sales in the previous quarter may remit sales taxes to New York State quarterly, while restaurants with more than US$300,000 of sales in the previous quarter remit monthly to the state.
10. To address attrition and entry, we stratify the sample by quarters of operation and then assign restaurants to groups of ten within each stratum.
11. A small number of groups have eleven rather than ten restaurants in ensure all restaurants are included.
12. The policy is implemented in the middle of the second sales tax quarter in 2011. Our NYC analytic sample includes data observed from the fourth quarter of 2008 to the third quarter of 2013.
13. Sales and tax data from NYC's Department of Finance is originally sourced from this state data.
14. This means that we are capturing the universe of food-service establishments that are very likely subject to NYC's grading regime. Therefore, we have an intent-to-treat group rather than a clean treatment group. While this overinclusion 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 have replicated analyses using only the

subset of NAICS codes that definitively apply to food service and restaurant establishments (those coded as 722—) and are certainly subject to grading. The results are substantively the same as those using the full sample of establishments.

15. 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.
17. This is a reasonable assumption that is corroborated by accounts from DOHMH.
18. We group observations (and cluster standard errors) by two-, three- and four-digit NAICS codes; for restaurants, we group those coded 772— and 445— separately; for 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 replicate these models with less restrictive geographic controls, and results are substantively the same. The finer spatial controls help minimize the effect of reputation on grading estimates since consumers are more likely to be exposed to similar reputational information in closer spatial proximity (Jin and Leslie 2009).
19. We are less concerned about geographic codependence of the inspections themselves since inspectors are randomly assigned to restaurants and not on a geographic basis.
20. The results in Online Appendix B show larger declines for final inspection scores immediately following the policy change. This might reflect improved food safety conditions after feedback or general learning from initial inspections, which is consistent with declines in both initial and final scores over time (but at a decreasing rate) after policy implementation.
21. Again, score improvements are greater for final inspections.
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.
23. For a more comprehensive discussion of this issue in the case of the NYC grading regime, see Schwartz et al. (2015).
24. We run more parsimonious models, including only *Post_false* and *Post_trend_false*, but for the 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 postgrading regimes. We further specify models with quadratic and cubic time trends, finding similar results.

25. As shown in Online Appendix B, the decline for final inspection scores is substantially larger and also declines over the postperiod.
26. 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 US$8,000 and US$10,700, depending on the specification.
27. 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 Online Appendix I for a visual).
28. This is consistent with the fact that the share of "A" grades has grown over time, diminishing the value-added of the posted grade.
29. 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 subsample) and rerun identical specifications. The results remain substantively the same.
30. We also test for differential effects in the rollout period, but none of the results are significant.
31. 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 NYC Office of Management and Budget (OMB) Budget Function Analysis, and actual inspections come from DOHMH reporting to OMB (New York City Office of Management and Budget 2014).

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* 39:790–821.

Dranove, David, and Ginger Z. Jin. 2010. "Quality Disclosure and Certification: Theory and Practice." *Journal of Economic Literature* 48:935–63.

Fein, Sara B., C. T. Jordan Lin, and Alan S. Levy. 1995. "Foodborne Illness: Perceptions, Experience, and Preventive Behaviors in the United States." *Journal of Food Protection* 58:1405–11.

Figlio, David N., and Maurice E. Lucas. 2004. "What's in a Grade? School Report Cards and the Housing Market." *American Economic Review* 94:591–604.

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* 20:287–97.

Hastings, Justine S., and Jeffrey M. Weinstein. 2008. "Information, School Choice, and Academic Achievement: Evidence from Two Experiments." *The Quarterly Journal of Economics* 123:1373–414.

Ho, Daniel E. 2012. "Fudging the Nudge: Information Disclosure and Restaurant Grading." *Yale Law Journal* 112:574–688.

Ippolito, Pauline M., and Alan D. Mathios. 1990. "The Regulation of Science-based Claims in Advertising." *Journal of Consumer Policy* 13:413–45.

Jin, Ginger Z., and Phillip Leslie. 2003. "The Effect of Information on Product Quality: Evidence from Restaurant Hygiene Grade Cards." *The Quarterly Journal of Economics* 118:409–51.

Jin, Ginger Z., and Phillip Leslie. 2009. "Reputational Incentives for Restaurant Hygiene." *American Economic Journal: Microeconomics* 1:237–67.

Jones, Timothy F., and Frederick J. Angulo. 2006. "Eating in Restaurants: A Risk Factor for Foodborne Disease?" *Clinical Infectious Diseases* 43:1324–28.

Mathios, Alan D. 2000. "The Impact of Mandatory Disclosure Laws on Product Choices: An Analysis of the Salad Dressing Market." *The Journal of Law and Economics* 43:651–78.

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* 5:607–25.

New York City Department of Health and Mental Hygiene. 2010. *Self-Inspection Worksheet for Food Service Establishments*. http://www.nyc.gov/html/doh/downloads/pdf/rii/self-inspection-worksheet.pdf. Accessed September 4, 2015.

New York City Department of Health and Mental Hygiene. 2012. *How We Score and Grade*. http://www.nyc.gov/html/doh/downloads/pdf/rii/how-we-score-grade.pdf. Accessed September 4, 2015.

New York City Office of Management and Budget. 2014. *Tax Revenue Forecasting Documentation*. http://www.nyc.gov/html/omb/downloads/pdf/methodology_2014_08.pdf. Accessed September 4, 2015.

Rockoff, Jonah, and Leslie J. Turner. 2010. "Short-run Impacts of Accountability on School Quality." *American Economic Journal: Economic Policy* 2:119–47.

Schwartz, Amy E., 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." Unpublished manuscript, Syracuse University.

Silver, Diana, Michah W. Rothbart, Jin Y. Bae, Amy E. Schwartz, and Tod Mijanovich. 2015. Do the stakes in food safety inspections change outcomes for restaurants in court? The case of New York City's restaurant grading initiative. Presented at the APHA 2015 Annual Meeting and Expo in Chicago, IL. American Public Health Association.

Simon, Paul A., Phillip Leslie, Grace Run, Ginger Z. Jin, Roshan Reporter, Arturo Aguirre, and Jonathan E. Fielding. 2005. "Impact of Restaurant Hygiene Grade Cards on Foodborne-disease Hospitalization in Los Angeles County." *Journal of Environmental Health* 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* 21:1031–48.

Winters, Marcus A., and Joshua M. Cowen. 2012. "Grading New York: Accountability and Student Proficiency in America's Largest School District." *Educational Evaluation and Policy Analysis* 34:313–27.

Wong, Melissa R., Wendy McKelvey, Kazuhiko Ito, Corinne Schiff, J. Bryan Jacobson, and Daniel Kass. 2015. "Impact of Letter Grade Program on Restaurant Sanitary Conditions and Diner Behavior in New York City." *American Journal of Public Health* 105:81–87.

Author Biographies

Rachel Meltzer is an assistant professor in the Milano School of International Affairs at The New School.

Michah W. Rothbart is an assistant professor in the Maxwell School of Syracuse University.

Amy Ellen Schwartz is the Daniel Patrick Moynihan Professor of Public Affairs at the Maxwell School of Syracuse University.

Thad Calabrese is an associate professor in the Wagner School at New York University.

Diana Silver is an associate professor in the College of Global Public Health at New York University.

Tod Mijanovich is a research associate professor in the Steinhardt School at New York University.

Meryle Weinstein is the acting Director of the New York University Institute for Education and Social Policy and a research assistant professor in the Steinhardt School at New York University.