PAR PUBLIC ADMINISTRATION REVIEW

ASPA American Society for Public Administration

Michah W. Rothbart
Amy Ellen Schwartz
Syracuse University
Thad D. Calabrese
New York University

Zachary Papper
American Express
Todor Mijanovich
New York University

Rachel Meltzer
The New School
Diana Silver
New York University

Research Article

What a Difference a Grade Makes: Evidence from New York City's Restaurant Grading Policy

Abstract: *Can governments use grades to induce businesses to improve their compliance with regulations? Does public disclosure of compliance with food safety regulations matter for restaurants? Ultimately, this depends on whether grades matter for the bottom line. Based on 28 months of data on more than 15,000 restaurants in New York City, this article explores the impact of public restaurant grades on economic activity and public resources using rigorous panel data methods, including fixed-effects models with controls for underlying food safety compliance. Results show that A grades reduce the probability of restaurant closure and increase revenues while increasing sales taxes remitted and decreasing fines relative to B grades. Conversely, C grades increase the probability of restaurant closure and decrease revenues while decreasing sales taxes remitted relative to B grades. These findings suggest that policy makers can incorporate public information into regulations to more strongly incentivize compliance.*

Evidence for Practice

- Integrating consumers into the regulatory framework can create new and substantial incentives for improved compliance.
- Providing consumers with salient information can increase compliant behavior without stronger implementation of punitive regulatory instruments.
- Public grades, which provide consumers salient information, may be an effective regulatory approach in arenas with commonly established standards of practice.
- Consumer-based regulatory approaches have important effects on both the level and mix of public revenues, which governments should consider in evaluating the costs and benefits of information-based regulation.

In July 2010, the New York City Department of Health and Mental Hygiene (DOHMH) began requiring restaurants to post summary results of food safety inspections in the form of letter grades (A, B, or C) in a conspicuous location near the restaurant's entrance. In 2012, New York City mayor Michael R. Bloomberg touted the success of the letter grades, claiming that they had beneficial health and economic effects: declines in reported cases of *Salmonella* and hospitalizations due to foodborne illness, improvements in compliance with food safety regulations, and increases in total restaurant sales (9.3 percent or $800 million) (City of New York, Office of the Press Secretary 2012). However, this enthusiasm was not shared by the restaurant industry, which charged that the new public grading program hampered business. Andrew Rigie, executive vice president of New York City's chapter of the New York State Restaurant Association, contended that "if you define success as taxing small-business owners and making their lives miserable, then letter grades have been a complete success" (Saul 2012).

Thus, the debate about whether grading policies are appropriate and effective regulations to improve practice in the private sector continues. Grading reforms are intended to improve compliance with food safety regulations by offering consumers information as an enforcement mechanism. Public restaurant grades have a clear intuitive appeal: grades provide easy-to-access information about restaurant food safety, allowing consumers to make informed decisions about where to eat ("vote with their feet"), directing business to restaurants with high grades and perhaps lower risks of foodborne illness.[1] The magnitude of the effect is unknown, depending, in part, on the distribution of letter grades and behavioral responses by consumers and restaurant owners.

The extent to which the grading policy creates incentives for restaurants to improve compliance—or, conversely, simply operates as a tax by increasing the costs of operating businesses without improving compliance—is an empirical question explored in this article. If grading policies were incredibly successful,

Michah W. Rothbart is assistant professor of public administration and international affairs in the Maxwell School of Syracuse University and senior research associate in the Center for Policy Research. His research focuses on public finance and financial management, particularly in the fields of education and food policy. He holds a doctorate in public administration from New York University's Robert F. Wagner School of Public Service.
E-mail: mwrothba@syr.edu

Amy Ellen Schwartz is Daniel Patrick Moynihan Chair in Public Affairs and professor of economics and public administration in the Maxwell School of Syracuse University and professor emeritus of public policy, education, and economics at New York University. She is editor of *Education Finance and Policy* and former president of the Association for Education Finance and Policy. Her research spans education policy, urban economics, and public finance. She holds a doctorate in economics from Columbia University.
E-mail: amyschwartz@maxwell.syr.edu

Thad D. Calabrese is associate professor in the Robert F. Wagner Graduate School of Public Service and currently leads the finance specialization at the school. His research interests span topics in financial management of government, not-for-profit, and health care organizations. He holds a doctorate in public administration from New York University's Robert F. Wagner School of Public Service.
E-mail: thad.calabrese@nyu.edu

Zachary Papper is director of analytics for global financial crimes compliance at American Express. He previously served as director of the Data Intelligence Group in the Tax Policy Division of the New York City Department of Finance, where he concurrently worked during his efforts on this project. He holds a PhD in molecular medicine from the Wayne State University School of Medicine.
E-mail: zpapper@gmail.com

Public Administration Review,
Vol. 00, Iss. 00, pp. 00. © 2019 by
The American Society for Public Administration.
DOI: 10.1111/puar.13091.

Todor Mijanovich is research associate professor at NYU Steinhardt. He is a program evaluator and health policy researcher who studies the effects of policies and programs that serve disadvantaged populations. He holds a doctorate in public administration from New York University's Robert F. Wagner School of Public Service.
E-mail: tod.mijanovich@nyu.edu

Rachel Meltzer is associate professor of urban policy and chair of the Master of Science in Public and Urban Policy program at the Milano School of Policy, Management and Environment at the New School. Her research is broadly concerned with urban economies and how market and policy forces can shape disparate outcomes across neighborhoods. She focuses on issues related to housing, land use, economic development, and local public finance.
E-mail: meltzerr@newschool.edu

Diana Silver is associate professor of public health policy in New York University's College of Global Public Health. Her research focuses on the design, implementation, and evaluation of public health policies at the state and local levels. She holds a doctorate in public administration from New York University's Robert F. Wagner School of Public Service.
E-mail: diana.silver@nyu.edu

then the vast majority of restaurants would get A grades and fine revenues would be trivial because of high rates of compliance. To be clear, an *intended* consequence of these policies is for fines to become an unimportant revenue source, because consumer behavior would enforce regulations. Thus, the policy will only be successful if the information is used, affects consumer demand, and, ultimately, improves compliance. In the absence of grades, municipalities generally still inspect restaurants, fine them for noncompliance, and close them if they present an unaddressed public health hazard. The question for policy makers, then, is the extent to which providing consumers information at the point of consumption changes their behavior above and beyond traditional regulations (such as fines).

This study uses detailed longitudinal data on taxes, fines, and health inspections for all New York City restaurants over a five-year period to gain insight into the impact of grades on the economic activity of restaurants and, by extension, the city's tax and fine revenues. The analysis explores one key feature of the program: the "return" to restaurants and the city for good inspection grades, comparing the returns to compliance with the size of traditional enforcement mechanisms such as fines.

A series of panel data analyses (including fixed-effects models and Wald estimators) are used to estimate the impact of posting an A grade (versus a B and/or C) on economic activity (revenues and closures) and payments to the city (fines and sales taxes) while controlling for underlying food safety compliance scores, a range of restaurant characteristics, and spatial and temporal fixed effects. Note that this method identifies the impact of receiving an A (or a C) as the "treatment," relative to otherwise similar restaurants that receive a B (in fixed-effects models, we estimate impacts within restaurants over time). These results shed light on the financial incentives for restaurants to improve their food safety compliance and the fiscal consequences of grades for the public sector.

The results suggest that, indeed, grades matter. Receiving an A grade has positive effects, increasing a restaurant's sales ($28,000–$53,000 annually) and sales taxes collected ($1,400–$2,200 annually), decreasing the level of fines assessed (more than $2,000 annually), and decreasing the probability of closure (3–5 percentage points). Conversely, receiving a C grade has negative effects, decreasing sales (and taxes) and increasing the probability of closure.

Together, the results indicate that grading is effective at changing consumers' and restaurant operators' behaviors. In sum, grades provide incentives to improve compliance, affect restaurant sales and closures, and affect the level and mix of city revenues (i.e., total revenues and revenue sources).

The rest of the article is organized as follows. It begins with a brief history of the restaurant grade policy in New York City. The next two sections present the previous literature and data and measures. The empirical strategy is outlined, and then the results are shown. The article concludes with a summary of findings, implications, and avenues for future research.

Restaurant Grading in New York City

Food safety regulations in the United States are guided by the U.S. Food and Drug Administration's model food code but vary across jurisdictions and typically are enforced at the state or local level, with each jurisdiction choosing which regulatory instruments to use. The DOHMH has long inspected New York City restaurants to ensure proper food safety practices, fining restaurants for violations and closing restaurants with unfixed public health hazards. Prior to grading, all restaurants were inspected annually, and violations found were posted on the DOHMH's website in a searchable database. Starting in July 2010, DOHMH began summarizing inspection results (using an otherwise identical scoring system) using letter grades (A, B, or C), which restaurants were required to post in a conspicuous location near the restaurant's entrance. The DOHMH also posted each restaurant's grade on its website.

Even prior to grading, inspections occurred regularly, with inspectors randomly assigned, and the precise timing of inspections was random within a window of approximately two months. Thus, grading simply made the inspection results more accessible by moving information from an offsite, online database to the point of purchase and creating more easily interpretable discrete grades (A, B, or C).

Inspection Scores

Inspection scores are the sum of violation points assigned during inspections. The points for a violation depend on the health risk it poses to the public, falling into three categories:

1. *Public health hazards,* such as failing to keep food at the proper temperature (minimum 7 points per violation)
2. *Critical violations,* such as serving salad without proper washing (minimum 5 points per violation)
3. *General violations,* such as not properly sanitizing cooking utensils (minimum 2 points per violation)[2]

Additional points can be added to each violation to reflect the extent of the violation (on a scale of 1 to 5).

Thus, lower inspection scores reflect greater levels of food safety compliance than higher scores. Inspection scores are assessed in the same manner (and for the same violations) as they were before the grading policy.

Grade Assignment

The DOHMH assigns A grades for scores of 13 and below, B grades for scores of 14–27, and C grades for scores of 28 or higher.[3] If an initial inspection leads to an inspection score in the B or C grade range (14 points or higher), then the restaurant is inspected again no sooner than seven days later, and a final grade is not assigned until after the reinspection. Final inspections, which comprise initial A inspections and all reinspections, yield a provisional grade. In addition to reinspections, inspection scores (and grades) can be lowered (improved) through an adjudication process.[4] During the period between final inspection and adjudication, a restaurant can post a sign reading "Grade Pending" in lieu of its B or C grade. Restaurants only post their provisional grade (B or C) or "Grade Pending"; they do not post both placards. In addition to posting grades, the DOHMH inspects restaurants receiving an A at an initial inspection annually, those in the B range twice per year, and those in the C range every four months. Figure 1 shows a simplified diagram of the inspection process.

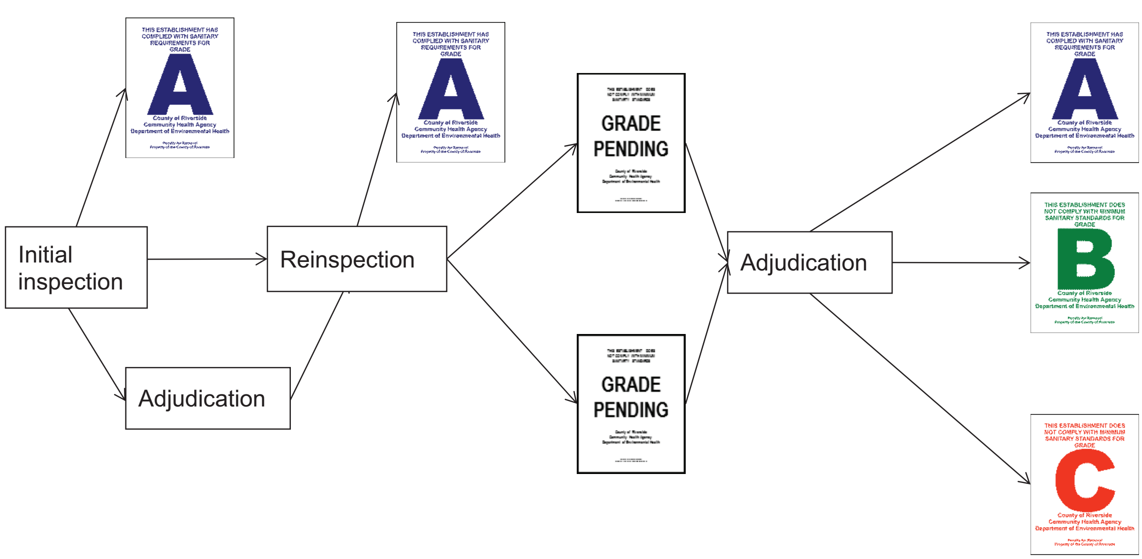
Fines

Traditionally, fines were the key regulatory lever used by the DOHMH to enforce food safety compliance. The type and count of inspection violations determine the level of fines assessed. Fines range from $200 to $2,000 per violation and are assessed at a restaurant's adjudication hearing at the discretion of a hearing officer—unless the grade is accepted and a lower fine is paid by the restaurant operator. Beginning January 19, 2011, restaurants assigned an A grade at inspection are exempt from fines. As a result, restaurants with an A inspection do not incur any fines for much of the post-inspection period. By design, this process creates a sharp increase in fines for B-graded inspections compared with otherwise similar A-graded inspections. Those assigned an A grade through adjudication, however, are still fined based on the number and severity of violations. In all other respects, fines are assessed in the same manner (and for the same violations) as they were before the grading policy.

Literature Review on Public Grading

Local, county, or state health departments across the United States regularly inspect food establishments in an effort to prevent foodborne diseases, of which at least 40 percent are attributed to restaurants (Jones et al. 2004). These inspection regimes face the "regulation dilemma," meaning they can enforce compliance through cooperative or conflictual incentive structures (Potoski and Prakash 2004). Conflictual regimes focus on deterrence: restaurants are routinely inspected, and all violations are punished. These deterrence regulatory mechanisms tend to lead to restaurant complaints about costs and an adversarial relationship with health departments. Cooperative regimes tend toward more flexibility and positive incentives for compliance (Potoski and Prakash 2004).

Public disclosure of restaurant hygiene is an increasingly common and popular regulatory instrument (Filion and Powell 2009). Public disclosure of food safety compliance may solve an information asymmetry problem, providing consumers with information and, perhaps, changing their purchasing behavior (Dranove and Jin 2010; Thaler and Sunstein 2008). Traditional restaurant inspections largely relied on deterrence to improve food safety, but the public rarely was aware of these inspection results. As noted by Bovaird (2007), effective public policies and regulation increasingly are a negotiation among different stakeholders. While older conflictual restaurant inspection regimes were top-down processes, public disclosure—by making hygiene scores readily available—better permits consumers to vote with their feet, effectively serving a coproduction role in the regulatory regime. That is, while public disclosure regulatory instruments still levy fines on restaurants for violations, the real costs (and benefits) of the regulation are changes in consumer behavior. Further, the publicness of grade disclosure incentivizes restaurant compliance, potentially reducing



Note: Fines are assessed for violations at each inspection that does not lead to an A grade (assessed for A inspections for the first six months of grading).

Figure 1 Simplified Model of the Inspection Cycle after Grading

the adversarial relationship between health departments and restaurants and increasing partnerships between the two parties. Such partnerships also permit business and government to develop methods and codes that improve outcomes such as the reduction in foodborne illnesses (King 2013). Despite its theoretical appeal and attention in the political arena, there is little research on the topic.

Two key empirical studies (Jin and Leslie 2003; Simon et al. 2005) estimated the effects of the Los Angeles health inspection letter grade policy, which began in 1998, and found consumer behavior changed and some evidence that foodborne illness decreased. Jin and Leslie (2003) used ordinary least squares and difference-in-differences regression analyses, finding that grades improved restaurant inspection scores, restaurant revenues increased in response to good grades, and foodborne disease hospitalizations decreased in Los Angeles. Simon et al. (2005) used ordinary least squares regressions, comparing Los Angeles with the rest of California and finding a decrease in foodborne illness hospitalizations that was sustained for at least three years.

Previous work also suggests that a restaurant's reputation could moderate the effect of posted grades (Dranove and Jin 2010; Jin and Leslie 2009). Sources of reputation—such as food quality or consumer-observed hygiene practices—may mute the impact of grades. Thus, a well-specified model should include controls for quality (such as restaurant characteristics or fixed effects) and for underlying food safety compliance (observed poor hygiene, such as rodents, may harm reputation even in the absence of grades). Unlike previous work, which failed to control for restaurant food safety practice, the models in this article isolate the effect of grades. Evidence is presented from models that omit inspection scores (consistent with previous work), models that control for inspection scores linearly, and models that restrict the sample to restaurants with very similar inspection scores.

More recently, Wong et al. (2015) found that public grading increased the probability of a restaurant scoring in the A range during the initial inspection and offered survey evidence of the program's high approval ratings among New Yorkers. Similarly, Filion and Powell (2009) found that similar disclosure policies were quite popular and informed consumer decision making when applied to restaurant hygiene. Taken together, these results suggest that grades provide significant incentives for restaurants to improve compliance (akin to the cooperative self-policing regulatory regimes described in Potoski and Prakash 2004)—incentives explored directly in this article.[5]

Finally, in a time of increasing competition for public resources, understanding the potential effects of these public health initiatives on municipal government revenues is critical and yet largely unexplored. Previous work by Meltzer et al. (2019) found little evidence of impacts (positive or negative) from the New York City grading policy on restaurant and bar sales. While sales increased, so did sales in similar establishments just outside the city. Does that mean the effects reported by Meltzer et al. (2019) reflect gains by As and Bs and losses by Cs? Or, perhaps, do the null results reflect no changes for any restaurants? The answers to these questions have important implications for equity and tax incidence. This article examines the impact of public grades on the level and mix of revenues paid to local municipal governments to explore the consequences of grades and the magnitude of behavioral responses to this still-new regulatory approach.

Data and Measures

This study uses richly detailed, longitudinal inspection and restaurant data from the DOHMH matched by Employer Identification Number (EIN) with longitudinal sales tax data from the New York City Department of Finance (DOF). This section describes the data from the DOHMH and DOF and explains how the data are matched and aggregated in groups to preserve the privacy of restaurant sales information.

DOHMH Data

The first analytic sample includes the universe of final DOHMH food safety inspections from July 27, 2010, through February 28, 2013 (the two and a half years following the implementation of public grading).[6] Records include restaurant characteristics and zip codes, inspection and adjudication dates, inspection scores, grades, and fines. The sample includes all 82,977 inspections for which one of 29,742 restaurants received a provisional grade.

Final inspection scores capture food safety compliance and are used by DOHMH to grade restaurants. On average, restaurants have 3.2 final inspections between July 27, 2010, and February 28, 2013 (table 1). Table 2 provides summary statistics for the inspection scores, which improve (i.e., points decline) over time. This result provides evidence that the policy meets one of its intended goals: to improve compliance with food safety regulations, as measured by inspection scores.[7] Before public grading, average final inspection scores were 25.1. In the first five quarters after public grading, average final inspection scores improved to 18.2, driven by improved compliance during

Table 1 Restaurant Descriptive Statistics

Number	**Mean**
Inspections	6.2
Final inspections	3.2
Workers	6.7
Seats	29.5
Cuisine	**Share**
American	0.24
Chinese	0.11
Pizza	0.06
Latin	0.04
Café/coffee/tea	0.04
Other*	0.51
Service	
Takeout-limited eat-in	0.39
Wait service	0.18
Wait and counter service	0.17
Takeout only	0.08
Counter service	0.12
Other*	0.07
Chain	0.10
Annual closure rate	0.12
N	34,917

Notes: Inspections include initial inspections and reinspections. Final inspections include initial A-graded inspections and reinspections for those initially receiving B or C grades. Workers, seats, cuisine, service, and chain are observed once per restaurant based on inspector reports. Annual closure rate is the share of open restaurants that close each year.
*Other includes 76 additional cuisine types and 8 additional service types.

Table 2 Inspection Scores Statistics, by Treatment Period

		All Restaurants	Continuously Operating
Before public grading		25.1 (75,767)	23.5 (29,668)
Quarters after public grading			
1–5	Initial	25.1 (42,391)	24.0 (17,667)
	Final inspection score	18.2 (40,451)	17.4 (17,305)
6–10	Initial	22.3 (43,017)	21.5 (17,103)
	Final inspection score	15.7 (41,329)	15.1 (16,718)

Notes: Lower scores indicate more hygienic restaurant conditions. Data are pre-adjudicated inspection scores. The mean score is shown on top; the number of inspections is shown in parentheses. Final inspection score includes all A-graded inspections and reinspections of restaurants that do not get an A grade on initial inspection. An inspection score of 13 or lower leads to an A grade. A final inspection score of 14–27 leads to a B grade, and restaurants can post "Grade Pending" until adjudication. A final inspection score of more than 27 leads to a C grade, and restaurants can post "Grade Pending" until adjudication. Continuously operating restaurants are open for two and half years before and after the introduction of public grading.

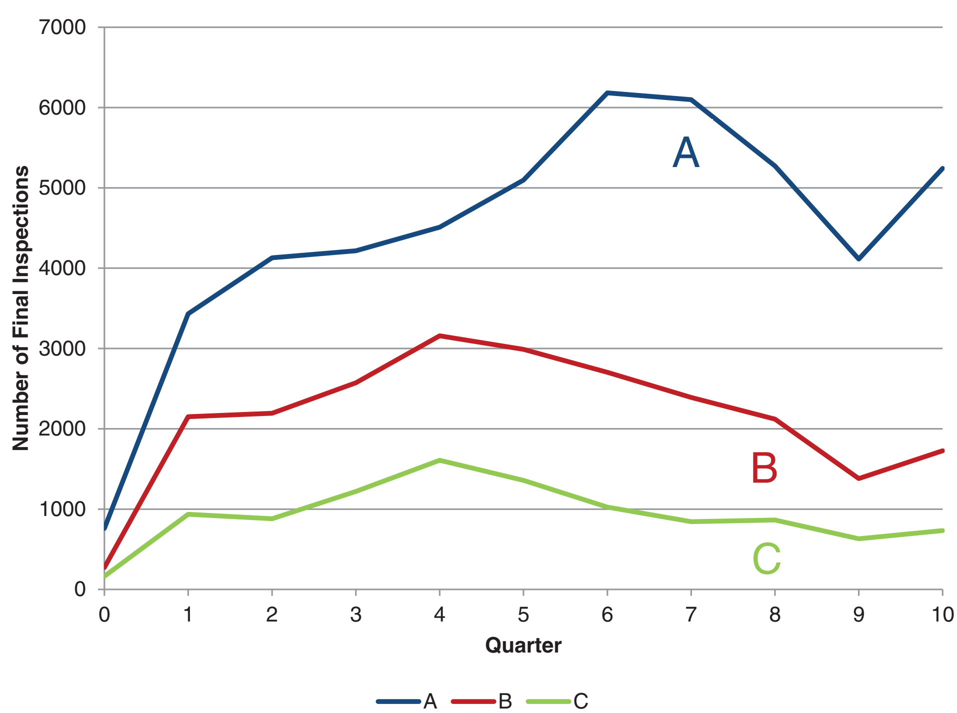
reinspections. In the next five quarters, average final inspection scores further improved to 15.7, driven partly by improved compliance on *initial inspection scores*. The set of "continuously operating" restaurants, which are in operation for two and half years before and two and half years after public grading, show similar improvements.

Grades summarize final inspection scores into discrete measures of food safety compliance (A, B, and C). *Provisional grades* are grades based on final inspection scores, which may be contested through adjudication. Provisional grades include A, B, and C grades. *Posted grades* reflect the grades that consumers see at the point of sale (the "treatment"). Posted grades include A, B, C grades and "Grade Pending" and reflect *adjudicated scores,* which can provide improvements on final inspection scores through a third-party tribunal. Posted grades are always at least as good as provisional grades, because final inspection scores can only improve or remain the same during an adjudication hearing.

Both provisional grades and posted grades improve over time (figures 2 and 3, respectively), consistent with final inspection score trends shown in table 2. Restaurants increasingly earn A grades and post those grades, perhaps reflecting the improved food safety compliance intended with the policy. For example, only about 65 percent of restaurants post A grades in the fifth quarter of public grading, but more than 80 percent do by the tenth quarter.[8] These aggregate trends mask changing grades within restaurants over time; only 50.2 percent of restaurants earn the same grade in every observed period, while 49.8 percent change grades (and the vast majority of restaurants go from ungraded to graded at some point during the sample period).

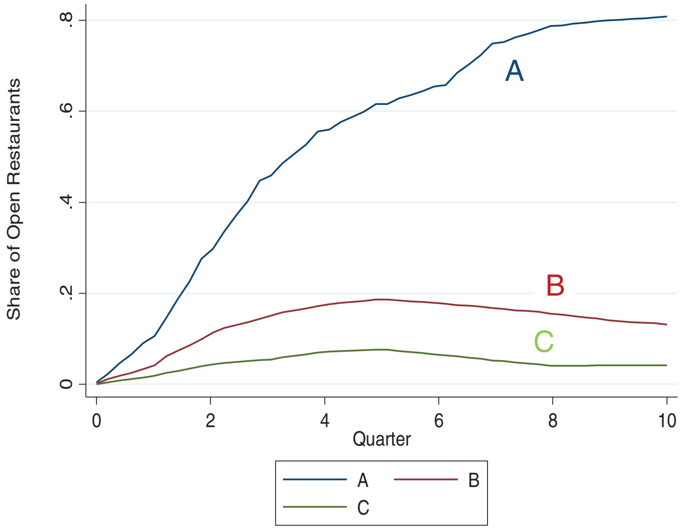
A restaurant is considered closed or "out of business" (*OOB*) if it is reported as not operating for three straight inspection attempts (on different days and at different times of day). The OOB date is the first failed inspection attempt (the first day the restaurant is observed not operating). For each final inspection, an indicator variable (OOB) takes a value of 1 if a restaurant goes out of business within 365 days of an inspection and 0 otherwise.[9] OOB timing typically lags the true closure date, unless the restaurant closed on the exact day of the first inspection attempt. Twelve percent of restaurants go out of business each year (table 1).

Fines are final (post-adjudication) fines assessed for each restaurant inspection (all dollar values adjusted using the urban Consumer Price Index to real 2013 dollars). Fines average about $500 per



Notes: The number of A-graded inspections increases over the first 10 quarters of the program, with a small dip in the quarter of and following Hurricane Sandy. The number of B- and C-graded inspections increases for the first four quarters and then begins to decline.

Figure 2 Provisional Grades Awarded by Quarters after Grading



Notes: The share of restaurants with A grades posted in the window increases over the first 10 quarters of the program and reaches 80 percent by the end of the sample period. The share of restaurants with B and C grades posted in the window increases for the first five quarters and then begins to decline.

Figure 3 Grades Posted in Windows by Quarters after Grading, Rollout of Treatment

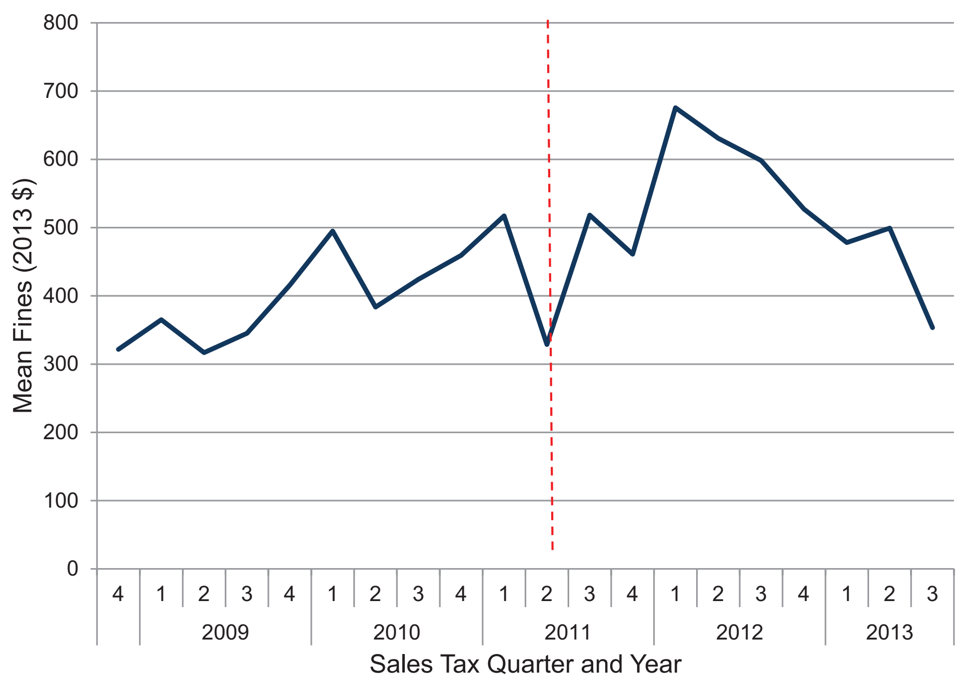
restaurant per quarter (figure 4). While mean fines per restaurant-quarter increased in the year following program implementation, this extended a preexisting trend (temporarily discontinued during program implementation in the second quarter of 2011). Fines reach a peak of $675 per operating restaurant in the first quarter of 2012 and then decline steadily, reaching pre-program levels by the third quarter of 2013 ($353).[10]

Other observed restaurant characteristics include *number of seats, number of employees,* an indicator for *chain* (at least 15 locations nationwide), and *cuisine, service type,* and *venue type*. Table 1 shows descriptive statistics for the restaurants in our sample. On average, a restaurant has 3.2 final inspections, employs 5.9 workers, and has 29.5 seats. Chains make up 11 percent of restaurants. The DOHMH defines more than 80 cuisine types; the most common cuisines are American, Chinese, and pizza. There are 13 service types, including wait service only, wait service and counter service, takeout only, and so on. Venue type summarizes a restaurant's setting (26 total) including diner, arena-stadium concession stand, bar/pub/brewery (food served), nightclub, restaurant (with bar), and restaurant (no bar).[11] Making full use of available administrative data, restaurant characteristics reflect the last inspection only, and as a result, they do not vary over time in our data set. Descriptive statistics for time-varying variables—namely, inspection scores, grades, fines, and sales—are shown in table 2 and depicted graphically in figures 2–5.

DOF Data

The second sample includes graded food and beverage establishments with any sales from July 27, 2010, through November 31, 2012. Some DOHMH-graded establishments are not primarily restaurants and derive a large share of sales from nonrestaurant activity.[12] To reduce statistical noise, we use a subsample of primarily food and beverage providers as identified by their North American Industry Classification System (NAICS) code.[13]

DOF data include reported quarterly sales and sales tax liabilities (hereafter, *sales* and *sales taxes*) from graded establishments.[14] Figure 5 shows mean sales by quarter, which are quite high ($175,000–$200,000 per quarter) and mask a high level of heterogeneity across restaurants. Mean sales increase slightly after



Notes: Mean fines by quarter. Average quarterly fines range from about $300 to $700 over the studied period. Average fines increase substantially in the first year after the restaurant grades law and then steadily decline from there. Average fines levied citywide are at pre-program levels on a quarterly basis starting in the middle of the 2013 sales tax year.

Figure 4 Average Fines by Quarter, Operating Restaurants, before and after Grading

grading, which could result from public grades or from other macro-level trends.[15]

Sales taxes mirror sales revenue, ranging from $8,000 to $9,000 per quarter, with a slight (statistically significant) rise over time. Like mean sales, mean sales taxes mask substantial heterogeneity. Like fines, increasing sales taxes increases public revenues.

Zero revenue takes a value of 1 if establishments have no revenue in a fiscal quarter and 0 otherwise. Zero revenue indicates closure but with less precision than OOB.[16] Estimates of impacts on zero revenue serve as a robustness check to other closure estimates.

Building class is a vector of indicator variables constructed from the DOF's Real Property Assessment Database (RPAD). Building class controls for locational or use characteristics, which may be associated with revenue generation and/or grades. There are six building class types: office commercial, retail commercial, mixed use retail, other commercial, residential, and government/public.[17] Restaurant BBL is matched to building class for each year.

Matching Tax and Inspection Data

We match date-specific DOHMH data to quarterly DOF data on EINs by aggregating restaurant inspection data (including grades and inspection scores) by quarter. Quarterly inspection scores and grades variables include mean inspection score, inspection score at the beginning and the end of the quarter, share of days with each grade, and grade at the beginning and end of the quarter (for both provisional grades and posted grades). The matched sample includes 15,899 restaurants or bars observed over nine quarters.

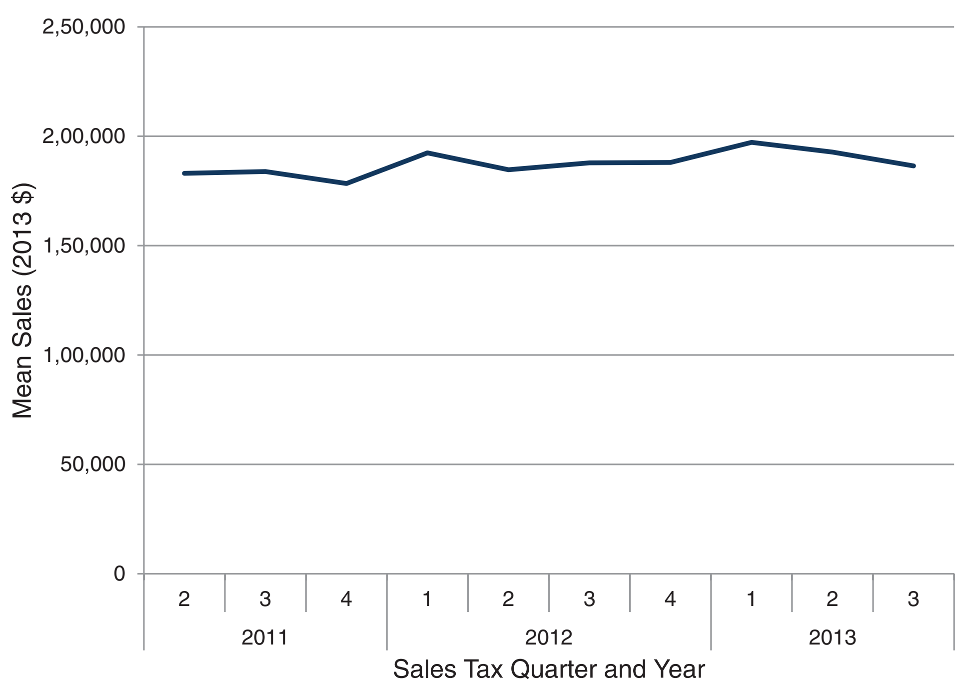
Aggregating by Groups/Privacy

Confidentiality rules prohibit the DOF from providing establishment-level data to outside researchers. Data were aggregated in groups of 10 randomly assigned restaurants; each observation provides data for a set of 10 restaurants randomly assigned to the same group.[18] To address attrition and entry, the DOF first stratified the sample based on quarters of operation and then assigned groups within these.[19]

The second data set is organized by group-quarter and includes sales and tax information and summary inspection results. The data provide variables summarizing the sales and tax activity in each group, including quarterly means and standard deviations of sales, log(sales), sales taxes, and log(sales taxes). The data also include quarterly means and standard deviations of inspection scores, number of seats and workers, daily mean provisional grades and posted grades, as well as the share in each grade category at the beginning and the end of the quarter, in each zip code, operating in each building class and with each cuisine, venue, and service type. The second analytic sample includes 9,182 observations in 1,538 groups (made up of 15,899 restaurants or bars).[20] Analyses using a third sample, which includes all graded establishments (including those not primarily food and beverage) and a separate random assignment procedure, produce results consistent with those reported in this article.[21]

Empirical Strategy

We use a variety of panel data methods comparing fines, closures, sales, and sales taxes across restaurants with A, B, and C grades. We begin with a pooled analysis using restaurant controls and quarter-by-year fixed effects. We then add zip code fixed effects, followed by additional controls for final inspection scores or,



Notes: Average restaurant sales increase following the implementation of the public grades program in the second quarter of fiscal year 2011 (July 2010). Mean sales during this period range from about $175,000 per quarter to close to $200,000 per quarter. Mean sales mask large levels of heterogeneity across restaurants, including heterogeneous impacts of grades that this article explores.

Figure 5 Average Sales, Operating Food and Beverage Entities, after Grading

alternatively, restricting the sample to restaurants near the grade cut points. Finally, restaurant fixed effects are added to control for unobserved differences in restaurant quality. Thus, we use three distinct strategies to address differences in food safety practices and reputations across restaurants: (1) control for underlying food safety compliance scores used to assign grades, (2) restrict the sample to restaurants near and on either side of the grade assignment threshold, and (3) employ restaurant fixed effects.

Fines and Closures

The analysis begins with a regression linking restaurant outcomes to grades and other variables:

$$y_{it} = \beta_0 + \beta_1 A_{it} + \beta_2 C_{it+} \beta_3 Score_{it} + X_i' \beta_4 + \gamma_i + \delta_t + \varepsilon_{\text{it}}, \quad (1)$$

where y is a restaurant-specific outcome (i.e., OOB, fines); A (C) is an indicator variable that takes a value of 1 if restaurant i is awarded an A (C) grade—its provisional grade—for an inspection in period t; *Score* is the final inspection score; X is the vector of observed restaurant characteristics; γ and δ are zip code and quarter-by-year fixed effects, respectively; and ε is an error term with the usual properties. The omitted category in these and subsequent analyses is a B grade. In alternative specifications, we add restaurant fixed effects, μ_i, dropping γ_i and X_i. In that case, model 1 yields estimates of the impact of A (C) grades identified by variation in *Score* and the restaurant fixed effects. The terms β_1 and β_2 are unbiased estimates, if restaurants cannot manipulate their provisional grade near the thresholds and there is a linear relationship between *Score* and y_{it}. As stated previously, the first is likely and credible because the DOHMH randomly assigns inspectors to restaurants each cycle, the timing of inspections are random within two-month windows, and we use provisional grades to address manipulation through the adjudication process. The second is plausible because the correlation between *Score* and y_{it} is nearly zero, conditional on grade.

Put differently, one might be concerned that managerial skill or effort (or some other unobserved characteristic) affects both grades and y_{it}. Surely a highly skilled manager may be better able to earn an A and run a restaurant that is less likely to go out of business, even without public grades. Even so, model 1 yields unbiased estimates of the impact of the public grades if managerial skill is adequately controlled for with the variables in vector X (seats, number of employees, chain, cuisine, service type, and venue type) and restaurant location (zip code fixed effects). Alternatively, in models with restaurant fixed effects, grades are conditionally independent of any skill (or other characteristic) that does not vary over time within a restaurant during the sample period; bias can only derive from skill or effort that improves or declines over time. Even more importantly, both versions of model 1 include the time-varying inspection score, which controls to some extent for effort or skill. Finally, increases in effort or skill that change over time across the city are addressed with quarter-by-year fixed effects.

Altogether, the key assumption of the model is that unobserved characteristics of restaurants that decrease fines or probability of OOB (such as manager skill) are captured by either the underlying inspection score (which determines provisional grades mechanically), the restaurant fixed effect, the quarter-by-year fixed effects, or the restaurant location and observed characteristics. If this assumption holds, then model 1 yields unbiased estimates of the returns to A (or C) grades.

Later, we relax the linearity assumption, estimating the impact using a Wald estimator (restricting the sample to inspection scores near the grade assignment thresholds). That is, we use *Score* as an instrumental variable for A (C) and restrict our sample to a subset of inspections just above and just below the A (or C) grade cut point. Here, for the restricted sample, *Score* is an instrumental variable that "just identifies" provisional grade. Our identification assumption is that, among inspections with scores within 1 or 2 points of each other, *Score* is correlated with A (C) but uncorrelated with other variables related to OOB and fines. Under those conditions, differences in these outcomes can be attributed to the grades.

Sales and Sales Taxes

We estimate the impact of posted grades on sales and taxes using model 2:

$$\begin{aligned} y_{gq} &= \tau_1 A_{gq} + \tau_2 C_{gq} + \tau_3 GP^B_{gq} + \tau_4 GP^C_{gq} + \tau_5 Score_{gq} + X'_{gq} \tau_6 \\ &\quad + \gamma_g + \delta_q + \varepsilon_{gq}, \end{aligned} \quad (2)$$

where y_{gq} is the group's average daily restaurant sales or sales taxes in quarter q[22]; A_{gq} (C_{gq}) is the mean share of days in quarter q that restaurants in group g hold an A (C) grade; GP^B_{gq} (GP^C_{gq}) is the mean share of days that the restaurants have a B (C) grade but may post "Grade Pending"; $Score_{gq}$ is a group's average inspection score over the course of the quarter, weighted by restaurant-days; X_{gq} is a vector of mean restaurant characteristics and building class; and γ_g and δ_q are group and quarter-by-year fixed effects, respectively. Model 2 yields estimates of the impact of A (C) grades on daily sales, τ_1 and τ_2, which are unbiased if restaurant i's average grade in group g in quarter q only affects restaurant i's sales and not the sales of other restaurants in group g.[23] This assumption is plausible because of the random assignment of restaurants to groups and the inclusion of the quarter-by-year fixed effects, δ_q.

The impact using provisional grades is estimated using the following model:

$$\begin{aligned} y_{gq} &= \tau_1 PROVA_{gq} + \tau_2 PROVC_{gq} + \tau_3 PROVScore_{gq} \\ &\quad + \tau_4 X_{gq} + \delta_q + \varepsilon_{iq}, \end{aligned} \quad (3)$$

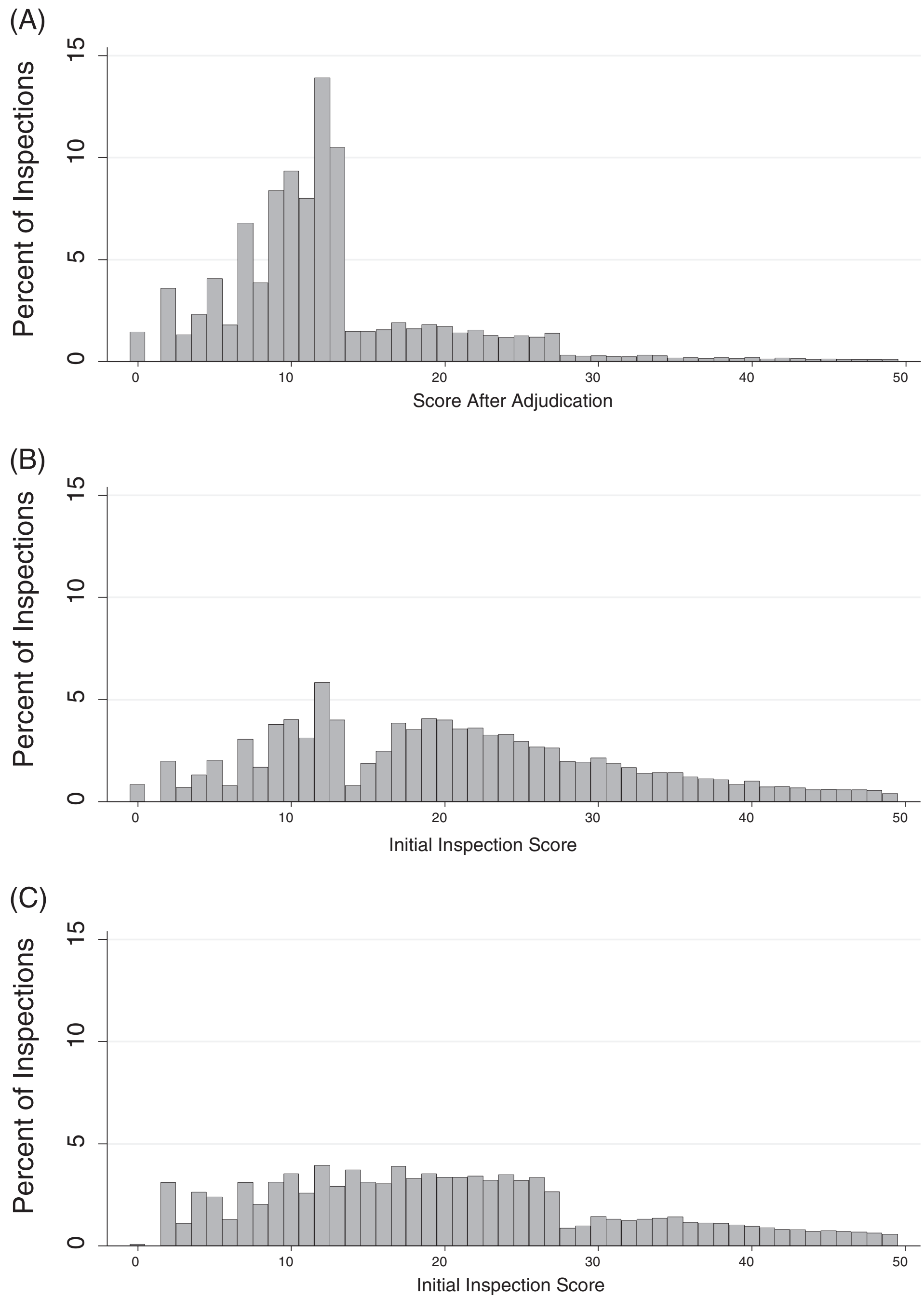
where y_{iq} is the restaurant's mean daily sales or sales taxes in quarter q and *PROVA, PROVC,* and *PROVScore* denote restaurant provisional grades and score at the beginning of quarter q.[24] The terms τ_1 and τ_2 represent the impact of an A or C grade assigned during inspection on daily sales (the effect of provisional grades assigned during inspection). These coefficients may be smaller than those for posted grades because only the posted grades are salient at the point of sale. Estimates will be unbiased if restaurant i's grade does not affect the sales of others in group g and if restaurants cannot manipulate provisional grades.

Again, one might be concerned that managerial skill or effort (or other unobserved characteristics) affects both grades and y_{gq}. It is likely that a highly skilled manager may be more able to earn an A grade and earn more revenues, in the absence of public grades. Even so, model 3 yields unbiased estimates of the impact of the public grades if managerial skill is adequately controlled for with the variables in vector X and the locations of restaurants in group *g*. Alternatively, in models with group fixed effects, grades are conditionally independent of any skill (or other characteristic) that is constant over time within restaurants; bias can only derive from skill or effort that improves or declines over time.

Even more importantly, both versions of model 3 include time-varying inspection scores, which control to some extent for effort or skill. Both versions of model 3 also include quarter-by-year fixed effects, which control for citywide changes in effort or skill over time. Taken together, the key assumption of the model is that unobserved characteristics that increase sales (and, by extension, sales taxes) are captured by either underlying inspection scores (which determines provisional grades mechanically), the group fixed effect, the quarter-by-year fixed effects, or restaurant locations and other observed characteristics. If this assumption holds, then model 3 yields unbiased estimates of the returns to A (or C) grades.

Managerial Gaming

The models described here provide unbiased estimates of the impact of grades if the assignment of grades near the grade threshold is conditionally random. That is, the assignment of grades for nearly identical final inspection scores reflects small differences in food



Panel A. Post-Adjudication 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.

Figure 6 Inspection Score Distributions, before and after Grading

safety compliance or noise (such as random inspector assigned) rather than some managerial skill to improve grades associated with the ability to generate more business. The likelihood of restaurant manipulation is minimized by program design; inspections are unannounced and randomly assigned a specific time within an inspection window, and inspectors rotate among restaurants and are randomly assigned to a specific inspection.[25] The program is designed so that final inspection scores accurately reflect food safety compliance rather than gaming.[26] As a result, for such gaming to bias our estimates, restaurant manipulation would have to occur at the point of inspection, because inspection timing is randomly assigned within a window, and gaming would require success with a large enough share of randomly assigned inspectors to justify the effort. Further, models with restaurant fixed effects control for manipulation skill that does not vary over time.[27]

The adjudication process is one aspect of the program design that seems vulnerable to managerial gaming. In a forthcoming article, Silver, Rothbart and Bae explore whether restaurant challenges to inspection scores reflect managerial skill and resources in the adjudication process in addition to food safety practice. Thus, preferred models use an intent-to-treat framework, using provisional grades assigned through the inspection process (provisional grades), which are unaffected by adjudication.[28] If consumers only respond to posted grades, then these estimates are conservative (attenuated toward zero), because some provisional grades improve during adjudication. While posted grades is arguably a better measure of what consumers actually observe, provisional grades is our preferred measure because of concerns that restaurants may manipulate posted grades through the adjudication process.

Results

Impact on Closures

Table 3 shows estimates of the impact of A and C grades on restaurant closure (OOB). As shown in column 1, A-graded restaurants are 4.2 percentage points less likely to close than those earning a B. Models that include a linear control for scores (column 2) yield estimates of a 2.6 percentage point decrease in the probability of closure. Column 3, including controls for restaurant characteristics and zip code fixed effects, yields statistically and substantively similar impact estimates. Results are also robust to including restaurant fixed effects (column 4). Thus, an A grade decreases the probability of closure by about 2.6 percentage points compared to a B.

Does a C grade hurt? As shown in table 3 (column 1), C-graded restaurants are 4.9 percentage points more likely to close than Bs. Our preferred model controls for inspection score and restaurant fixed effects; in column 4, a C is 2.1 percentage points more likely to close than a B, a substantial increase relative to the overall closure rate of 12 percent.

Next, results in table 4 relax the linearity assumption, using *Score* as an instrumental variable for *A* (*C*) to provide Wald estimates of the impact of A and C grades on closure. The preferred model (columns 1 and 4) limits the sample to scores one point above and below the cut point. The estimate in column 1 shows that restaurants earning A grades are 4.9 percentage points less likely to close within a year than those earning B grades. Column 4 shows that restaurants earning C grades are 4.2 percentage points more likely to close (only marginally significant). The point estimates are not sensitive to increasing the bandwidth to 2 points or to including restaurant characteristics and zip code fixed effects.[29]

Impact estimates using zero revenue to measure closure yield similar results, demonstrating that findings are robust to a more conservative measure. Estimated effects of group-level provisional grades on zero revenue are qualitatively similar to the inspection-

Table 3 Effect of Restaurant Grade on Closures, Panel Data Estimates

	1	2	3	4
A	–0.042***	–0.026***	–0.027***	–0.026***
	(0.002)	(0.004)	(0.003)	(0.004)
C	0.049***	0.021***	0.016***	0.021***
	(0.004)	(0.006)	(0.006)	(0.006)
Inspection score		0.001***	0.001***	0.001***
		(0.000)	(0.000)	(0.000)
Constant	0.104***	0.076***	0.115*	0.076***
	(0.008)	(0.009)	(0.065)	(0.009)
Quarter-year FE	Y	Y	Y	Y
Rest. Char.	N	N	Y	N
Zip FE	N	N	Y	N
Restaurant FE	N	N	N	Y
# of Inspections	82,977	82,977	82,977	82,977
Restaurants	34,917	34,917	34,917	34,917

Notes: Lower scores indicate more hygienic restaurant conditions. Robust standard errors, adjusted for within-restaurant clusters, are shown in parentheses. ***$p<.01$.; **$p<.05$.; *$p<.1$. Closure = 1 if a restaurant is permanently closed within the next four fiscal periods. Columns 2, 3, and 4 include a control for the inspection's score. Column 3 includes restaurant controls for chain restaurants, number of workers, number of seats, and a set of indicator variables for restaurant cuisine, service type, and venue type, as well as zip code fixed effects. Restaurant controls are time invariant and measured in the most recent restaurant inspection. Column 4 includes a restaurant fixed effect and excludes time-invariant restaurant and location controls. The reference group is inspections assigned a B grade.

Table 4 Effect of Restaurant Grade on Closures, Wald Estimates

	A–B			C–B		
	1	2	3	4	5	6
A	–0.049***	–0.040***	–0.048***	—	—	—
	(0.017)	(0.017)	(0.010)			
C	—	—	—	0.042*	0.039*	0.043***
				(0.023)	(0.024)	(0.015)
Constant	0.116***	0.281*	0.115***	0.114***	–0.223	0.112***
	(0.017)	(0.165)	(0.010)	(0.006)	(0.209)	(0.005)
Q-Y FE	Y	Y	Y	Y	Y	Y
Rest. Char.	N	Y	N	N	Y	N
Zip FE	N	Y	N	N	Y	N
# Inspections	7,387	7,387	17,113	2,921	2,921	5,398
Restaurants	6,812	6,812	13,609	2,710	2,710	4,744
Bandwidth	1 point	1 point	2 points	1 point	1 point	2 points

Notes: Robust standard errors, adjusted for within-restaurant clusters, are shown in parentheses. ***$p<.01$.; **$p<.05$.; *$p<.1$. Columns 1, 2, 4, and 5 restrict the sample to inspections 1 point above and 1 point below the grade cutoff. Columns 3 and 6 restrict the sample to inspections 2 points above and 2 points below the grade cutoff. The optimal bandwidth for a local linear RD estimate of an A inspection effect, which minimizes mean squared error, as in Imbens and Kalyanaraman (2012), is 1.293 points. The optimal bandwidth for a local linear RD estimate of a C inspection effect is 1.939 points. The estimated effects in local linear models are exactly equal to columns 1 and 4, respectively. Columns 2 and 5 include restaurant controls for chain restaurants, number of workers, number of seats, and a set of indicator variables for restaurant cuisine, service type, and venue type, as well as zip code fixed effects. Restaurant controls are time invariant and measured in the most recent restaurant inspection. The reference group is inspections assigned a B grade.

level estimates on OOB. An A is 3.6 percentage points less likely to lead to closure, and a C is 6.6 percentage points more likely to lead to closure than a B.[30]

Finally, a falsification test shows estimates of the impact of "grades" on OOB in the period before public grades (assigned using the same grading criteria). Results in appendix C and appendix D in the Supporting Information show that the "impact" on closure is smaller before grading, and it is insignificant at the 95 percent level for all but one model (column 4 in appendix C). These results, statistically indistinguishable from zero, suggest that other impact estimates are a result of the grades themselves rather than consumer-observed differences in food safety practices or other confounders.

Effect on Fines

Table 5, column 1 shows that A grades yield fines $518.46 lower than otherwise similar restaurants that earned a B. Results are insensitive to widening the bandwidth (column 2). As shown in column 3, C grades do not yield fines appreciably different from similar B-graded restaurants. While differences are statistically significant in column 4, the difference is equivalent to the de facto observed price of two additional inspection points (the linear estimate is $40 per point), which is the size of the bandwidth and suggests that this results from differences in inspection scores. Estimates from local linear regression models yield similar results (see appendix B).[31] In sum, A grades reduce fines, but no similar effect is observed between B and C grades.

Effects on Sales and Sales Taxes

Estimates in column 1 of table 6 show that posted A grades increase and posted C grades decrease sales.[32] (The impact of C grades is statistically insignificant but qualitatively meaningful.) An A grade is associated with a $145 increase in mean daily sales compared to a B. For a C, sales decrease by $104 daily, but this is not statistically significant in this model specification. The estimated effect for "Grade Pending" is not significantly different from a posted B grade.[33] Perhaps consumers perceive "Grade Pending" as equivalent to a B grade, or they do not understand it consistently enough to change systematic consumption patterns. Further, during the period that restaurants may post "Grade Pending," there is no significant difference in sales between Bs and Cs.

Column 2 of table 6 shows estimates using provisional grades rather than posted grades. As increase restaurant sales by about $78 a day (roughly $28,000 annually), and Cs decrease restaurant sales by about $122 per day ($45,000 annually).[34] These estimates are similar to posted grades, but they are more precisely estimated.

Columns 3 and 4 of table 6 show estimates of the impact of grades on sales, controlling for mean inspection score and group fixed effects. Inspection scores are uncorrelated with sales after controlling for grades; increased sales from improved food safety practice are driven by improved grades. The independent effects of posted grades on sales (column 3) are quite large and similar to those presented above. Again, "Grade Pending" mutes the effect of grades.

Column 4 shows estimates of the effects of provisional grades. As increase restaurant sales by about $83 a day, and Cs decrease restaurant sales by about $144 per day (roughly $30,000 and $53,000 annually, respectively).[35] Again, coefficients are similar to posted grades, but they are more precisely estimated and preferred because they do not rely on assumptions that restaurants are unable to manipulate grades during the adjudication process. In sum, grades affect sales, but underlying inspection scores do not.

Table 5 Effect of Restaurant Grade on Inspection-Level Fines ($), Wald Estimate

	A–B		C–B	
	1	**2**	**3**	**4**
A	–518.46***	–553.00***	—	—
	(34.53)	(18.19)		
C	—	—	–17.40	100.37**
			(53.29)	(39.01)
Constant	1129.91***	1157.09***	1016.17***	1127.49***
	(46.00)	(29.75)	(124.21)	(93.89)
Q-Y FE	Y	Y	Y	Y
Rest. Char.	N	N	N	N
Zip FE	N	N	N	N
# Inspections	7,387	17,113	2,921	5,398
Restaurants	6,812	13,609	2,710	4,744
Bandwidth	1 point	2 points	1 point	2 points

Notes: Robust standard errors, adjusted for within-restaurant clusters, are shown in parentheses. ***$p<.01$.; **$p<.05$.; *$p<.1$. Columns 1 and 3 restrict the sample to inspections 1 point above and 1 point below the grade cutoff. Columns 2 and 4 restrict the sample to inspections 2 points above and 2 points below the grade cutoff. The optimal bandwidth for a local linear RD estimate, which minimizes mean squared error, as in Imbens and Kalyanaraman (2012), is 3.77 for columns 1 and 2 and 10.35 for columns 3 and 4. The estimated effects in local linear models are qualitatively similar (–609.21 and significant for A; 22.32 and insignificant for C). The reference group is inspections assigned a B grade.

Table 6 Effect of Restaurant Grades on Sales ($), Group Fixed-Effects Models

	1	2	3	4
A	144.71***	77.61**	123.33**	82.86**
	(51.58)	(31.91)	(55.06)	(35.77)
C	–103.77	–122.26***	–113.04	–143.65***
	(124.06)	(43.83)	(126.41)	(53.67)
Grade Pending				
B	1.54	—	6.40	—
	(67.38)		(67.59)	
C	–20.00	—	13.12	—
	(84.37)		(88.38)	
Inspection score	—	—	–0.68	0.80
			(1.47)	(1.44)
Ungraded	45.26	58.65	61.40	73.34
	(63.43)	(44.87)	(63.60)	(45.31)
Building class FE	Y	Y	Y	Y
Quarter-year FE	Y	Y	Y	Y
Group FE	Y	Y	Y	Y
Constant	2,347.31***	2,393.44***	2,396.65***	2,380.98***
	(449.56)	(448.36)	(449.78)	(449.27)
Observations	9,182	9,182	9,182	9,182
Groups	1,538	1,538	1,538	1,538
R^2	0.98	0.98	0.98	0.98

Notes: Robust standard errors, adjusted for within-group clusters, are shown in parentheses. ***$p<.01$.; **$p<.05$.; *$p<.1$. Table shows the estimated impact of restaurant grades on daily sales. Columns 1 and 3 show estimates of the impact of average posted grades on sales. Columns 2 and 4 show estimates of the impact of provisional grades earned at inspection by the beginning of the quarter on sales. A and C represent the share of a group with an A or C grade, respectively. Because estimates are reported on the means of all variables, these are estimates of impacts on a single restaurant. Grade Pending is the share of the group with the option to post either "Grade Pending" or the grade indicated in the window. Columns 3 and 4 include controls for inspection score and all models control for building class, group fixed effects (which control for time-invariant group characteristics such as share serving each cuisine type), and quarter-by-year fixed effects. The reference group is restaurants posting B grades.

Table 7 Effect of Restaurant Grades on Sales Taxes ($), Group Fixed Effects Models

	1	2
A	5.93**	4.08***
	(2.44)	(1.58)
C	–5.07	–5.97**
	(5.59)	(2.37)
Grade Pending		
B	0.83	—
	(2.99)	
C	0.41	—
	(3.91)	
Inspection score	–0.01	0.05
	(0.07)	(0.06)
Ungraded	2.13	2.61
	(2.81)	(2.00)
Building class FE	Y	Y
Quarter-Year FE	Y	Y
Group FE	Y	Y
Constant	107.84***	107.30***
	(19.89)	(19.88)
Observations	9,182	9,182
Groups	1,538	1,538
R^2	0.98	0.98

Notes: Robust standard errors, adjusted for within-group clusters, are shown in parentheses. ***$p<.01$.; **$p<.05$.; *$p<.1$. Table shows estimated impact of restaurant grades on daily sales taxes. Column 1 shows estimates of the impact of average posted grades on sales taxes. Column 2 shows estimates of the impact of provisional grade earned at inspection by the beginning of the quarter on sales taxes. A and C represent share of a group with an A or C grade, respectively. Because estimates are reported on the means of all variables, these are estimates of impacts on a single restaurant. Grade Pending is the share of the group with the option to post either "Grade Pending" or the grade indicated in the window. All models control for inspection score, building class, group fixed effects (which control for time-invariant group characteristics such as share serving each cuisine type), and quarter-by-year fixed effects. The reference group is restaurants posting B grades.

Table 7 shows the sales tax implications of restaurant grades are consistent with the impact on sales. A grades increase sales taxes remitted, C grades decrease them, and there is no statistical or substantively meaningful difference in the sales taxes remitted between B- and C-graded restaurants during the "Grade Pending" period. Preferred estimates, shown in column 2, suggest beginning a quarter with an A leads to about $4 more in sales taxes remitted daily than beginning the quarter with a B ($1,500 annually). Conversely, beginning a quarter with a provisional C leads to remitting about $6 less in sales taxes every day (roughly $2,200 annually).[36] These are small effects each day, but imply that the city should expect substantially more annual sales taxes from A-graded restaurants than otherwise similar B-graded restaurants.

Discussion and Conclusions

This article offers evidence that public restaurant grades have important financial implications for both the private and public sector. It contributes to the growing literature evaluating the consequences of public grades, offering estimates of the impact on the restaurant's economic activity—controlling for food safety compliance—and of the fiscal impact on the public sector. Restaurants with better grades are less likely to close, receive fewer fines, have greater sales, and remit more taxes than they would have with worse grades. The financial rewards of a higher grade (increased revenues) are far greater than the impacts on fines, suggesting that providing consumers with increased salient information can give governments a powerful regulatory tool.

The effects on economic activity are both statistically significant and substantively meaningful. The impact of earning an A rather than a B grade is a 3–5 percentage point decrease in the probability of restaurant closure within a year, compared to a base rate of 12 percent. Conversely, the impact of earning a C rather than a B is a 2–5 percentage point increase in the probability of restaurant closure. Similarly, an A grade increases restaurant sales by $80–$120 a day compared to a B (about $30,000–$40,000 a year). In contrast, C grades decrease sales by $110–$140 a day (about $40,000–$50,000 a year). The impacts on economic activity create strong incentives for improved food safety compliance. In addition, restaurant grades (both posted grades and provisional grades) improve substantially during the observed period, suggesting that restaurants indeed respond to the economic incentives.

Together, our results indicate that grading mechanisms, as a means of making formerly obscured information immediately accessible, are effective at changing both consumers' and restaurant operators' behaviors. Patrons appear to migrate toward the establishments with better grades, as evinced by the grade-induced increase in sales revenues. As a result, the returns to an A grade are much larger than the size of traditional fines for food safety compliance violations, creating greater financial incentives for food safety compliance than traditional forms of regulation. Restaurants appear to improve their food safety compliance (decreasing fines and increasing the share of A grades), presumably in pursuit of maintaining and/or increasing their patronage. Thus, policies that bring consumers into this coproductive incentive structure may increase compliant behavior without using a punitive regulatory approach, as suggested by Bovaird (2007). Further, these results indicate that this regulatory tool leads to better outcomes—reduction in number and severity of violations, likely a result of improved restaurant hygiene—through self-policing rather than top-down deterrence efforts (Potoski and Prakash 2004).

There are, however, trade-offs to this regulation. Restaurants that do not (or cannot) improve compliance (i.e., reduce the number or severity of violations) and achieve higher grades suffer in terms of revenues and the likelihood of survival. The upside, presumably, is welfare (and fiscal) gains derived from reduced foodborne illness, depending on the effect of improved inspection scores on foodborne illness. Existing evidence is limited but promising. Firestone and Hedberg (2018), for example, found that *Salmonella* infections declined in New York City relative to the rest of the state following the implementation of the letter grades law. Moreover, Wong et al. (2015) found that restaurant grades are quite popular, suggesting that information-based regulation might be more politically feasible than issuing fines that create commensurate financial incentives for compliance (tens of thousands of dollars annually).

In addition, the effects on public revenues are statistically and substantively significant. While earning an A leads to a large decrease in fines assessed on the restaurant, there is little evidence of such an effect from earning a B over a C (by program design). The impact of A grades on fine revenues (a reduction), however, is offset by the impact of A grades on sales taxes. As increase sales taxes by a magnitude commensurate with the changes in sales ($4–$6 a day; $1,500–$2,000 a year) as compared to B grades. Similarly, C grades decrease sales taxes ($1,800–$2,000 a year). Thus, there is a substitution of tax revenue raised from A restaurants, replacing forgone fine revenue. Earning an A may lead to substitution of a more popular, less punitive source of public revenues (sales taxes), for an unpopular, punitive one (fines).

Combining estimates of impacts on sales taxes with the change in the citywide grade distribution (greater share of A grades) would suggest a $350–$500 annual increase in restaurant sales taxes, but only if restaurant sales are not zero-sum.[37] Conversely, observed mean fines per operating restaurant fall between $360 and $650 annually after program rollout (figure 4), largely resulting from the change in citywide grade distribution (more A grades) and reductions in violations. At first glance, this suggests that the rise in mean sales taxes offsets declines in mean fine revenue. However, Meltzer et al. (2019) found little effect on overall restaurant sales, so caution is required here; if impacts of grades are zero-sum then public grading may lead to a net decrease in total public revenues. Future work should focus on the long-term effects of public restaurant grading in order to assess implications for total public revenues once high rates of compliance (A grades) are sustained for multiple years.

Still, the composition of New York City's revenue sources changes, perhaps an intended consequence of public grading. Restaurants earning A grades are more likely to stay open, pay fewer fines, and remit more sales taxes (because of increased sales) than restaurants earning Bs. Over time, the city becomes increasingly reliant on these establishments for revenues and decreasingly reliant on restaurants with worse food safety practices. This decreases the long-term viability of food safety fine revenue, as more restaurants earn A grades and fewer pay fines; in turn, the city becomes more reliant on sales taxes. To be clear, there is no double dividend here—getting more fine revenue and improving food safety compliance are trade-offs. Fines are intended to incentivize compliance, not serve as a form of taxation, and so the substitution of these public revenue sources might indicate better public administration practices. Still, sources of revenues may affect availability of resources for use, so changes in the level and source of public revenues are notable for budget management, as well. More broadly, these findings suggest that municipal governments weighing the efficacy of restaurant grades should consider the impact on both total public revenues and the sources of public revenues, rather than just potential health gains from increased compliance.

While this article adds to the evidence about a particular restaurant hygiene regime, future comparative research should examine whether different systems result in desired outcomes. Wide variation exists between the restaurant inspection regulations between locations. Perhaps these differences can be exploited to determine what type of system is most effective in producing desired outcomes and what types result in the smallest fiscal and economic disruptions. For example, does one system reduce foodborne illnesses more than others, or are different systems of equal value to public health? Further, while this research has focused on the restaurants, future work might address how consumers view and assess grades using direct surveys. Finally, our work presents average effects across a large and diverse city and restaurant mix, but it is possible that the effect depends on the characteristics of the restaurant, among other things. Unfortunately, we were unable to explore this heterogeneity in the effects because limitations from grouping sales data, required by privacy rules. To be concrete, sales in high-revenue restaurants may be less responsive to grades because they have an established brand or substantial marketing. That said, consumers of low-revenue restaurants might be less sensitive to grades and the impact on these restaurants could be smaller. Future research should explore heterogeneity of these effects, especially where individual sales data are available.

In sum, we find that integrating consumers into the regulatory framework for food safety compliance provided new and substantial incentives for restaurants to improve their practices. Moreover, restaurants responded quickly to these incentives with improved food safety compliance—a reasonable response given the size of grades' impacts on sales revenues relative to the size of fines. In arenas where standards of practice, such as food safety, are commonly established, policy makers may be able to incorporate public information efforts into their regulatory approach with positive implications for consumers, compliant service providers, and the public sector.

Acknowledgments and Disclaimers

The results and opinions shown are entirely of the authors, but this research would not have been possible without the contributions of the New York City Department of Health and Mental Hygiene and Department of Finance and the Robert Wood Johnson Foundation. Special thanks are due to Meryle Weinstein, Dan Kass, Wendy McKelvey, J. Bryan Jacobson, Melissa Wong, Corinne Schiff, Karen Schlain, and Steven Robinson.

Notes

1. A similar logic motivates letter grading of schools (Rockoff and Turner 2010; Winters and Cowen 2012) and hospitals and enthusiasm for using grades for subway stations, transportation services, and street vendors, among others.
2. The New York City Department of Health and Mental Hygiene (2012, 2016) provides details on the relationship between violations, violation severity, and violation points assessed.
3. Restaurants are also closed temporarily if they pose a large public safety risk. This article does not explore temporary closures.
4. After two inspections (initial and reinspection), restaurants may challenge inspection violations at a third-party tribunal. Silver, Rothbart, and Bae (forthcoming) estimate the impact of grades and public grading on adjudications in New York City.
5. Some research has explored the fairness of grading policies. Ho (2012), for example, found that prior scores predicted less than 2 percent of future grades and interpreted this as program inconsistency. Note, however, that the public grading law was explicitly intended to encourage restaurants to take actions to improve (change) their grades in future inspections. The "inconsistency" may instead indicate the success of the program.
6. Descriptive statistics from the pre-grading period, two and half years before public grading, are shown to provide context for the policy and magnitudes of the effects.
7. Meltzer et al. (2019) provide more conclusive evidence on the impact of the public grading policy on inspection scores.
8. While a majority of restaurants earn A grades, those that get very high inspection scores raise mean inspection scores above the threshold for an A. The mean final inspection score is much higher than the median final inspection score because inspection scores have a floor (0 violation points) but no ceiling.
9. In robustness checks, closures within 390 days of inspection are used.

10. Meltzer et al. (2019) found that increased inspection frequency drove the initial increase in fines, while fines *per inspection* fell dramatically every quarter of grading.
11. A full list of cuisine, service, and venue types is available from the authors upon request.
12. For example, the DOHMH inspects hotel restaurants. Sales tax returns, however, aggregate all sales for the hotel and do not separate them by business activities.
13. The sample includes NAICS codes beginning with 722, as well as 445299, 445291, and 445120, to flag primary food and beverage establishments; the same set of codes recommended by the DOF is used in Meltzer et al. (2019) to identify entities likely to be graded.
14. In this article, "sales taxes" are sales taxes collected by restaurants and owed to New York City. 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 New York City. The state collects the entire sales tax and remits the city's portion in the following month. Restaurants with sales of $300,000 or less in the previous quarter may remit sales taxes quarterly, while restaurants with sales of more than $300,000 remit monthly.
15. Meltzer et al. (2019) found that increases in sales are significant but little evidence that the grading policy caused it.
16. As noted later, revenue data are reported by groups of establishments to preserve the privacy of individual establishments' sales data.
17. The building class variable was constructed using the RPAD variable AV-BLDGCL (see appendix A). In all, 44 percent of restaurants operate in mixed use retail buildings, 34 percent in retail commercial, and less than 10 percent operate in each of the other four building classes.
18. A small number of groups have 11 rather than 10 restaurants in order to include all restaurants.
19. Thus, the 5,145 restaurants operating in all 20 quarters were randomly assigned to 509 groups of 10 and 5 groups of 11; the 149 restaurants operating in all but the last quarter were grouped in 5 groups of 10 and 9 groups of 11; the 244 operating in all but the first were grouped in 20 groups of 10 and four groups of 11. Restaurants were assigned sequentially, until all were assigned to groups, homogeneous in quarters of operation, and no observation provides information on fewer than 10 establishments.
20. The policy was implemented in the middle of the second sales tax quarter of 2011. The analytic sample includes data observed from the third quarter of 2011 to the third quarter of 2013.
21. Results are available from the authors upon request.
22. Estimates of the impact on the percentage of changes in average daily sales and sales taxes using log(sales) and log(sales taxes) as the outcomes are qualitatively similar to those presented and are available upon request from the authors.
23. Note that if a restaurant in group *g* earns an A for an entire quarter, the value of A_{gq} is 0.1 higher than if that restaurant earns a B for the entire quarter. The impact on sales would be divided evenly among the whole group, but it should be inflated by a magnitude of 10 if own restaurant grades only affect own restaurant sales. Thus, one should divide coefficients by 10 to reflect the impact of a single restaurant's grade change on mean grades, but then multiply by 10 to reflect that impacts on group sales result from only a single restaurant's grade change.
24. Provisional grades measured at the beginning of the quarter best reflect grades posted throughout the quarter because of the inspection cycle. For example, a restaurant can post "Grade Pending" after earning an provisional C, but the C is often not posted until late in the quarter (if at all) because of adjudication timing.
25. Inspection scores may differ across inspectors because of inspector stringency. Differences in inspector stringency cannot be tested, because inspector IDs were not made available for privacy concerns. Because of the random assignment of inspectors, however, this only increases noise in estimates rather than biasing them.
26. Grades are assigned mechanically based on post-adjudication scores, so this study at first appears to be an opportunity to use a regression discontinuity design. Underlying scores, however, are not smoothly distributed across grade cut points, which is a key test for the appropriateness of a regression discontinuity (figure 6, panel A). This empirical result undermines the consistency of regression discontinuity estimates. Yet adjudicated scores are likely to be unsmooth without any restaurant manipulation at inspection because of at least three features of program design: (1) restaurants may vary in their ability or desire to win at adjudication (Silver, Rothbart, and Bae, forthcoming), which is one reason analyses focus on final inspection scores before restaurants have the opportunity to prepare arguments, hire experts to represent them, and hearing timing is known; (2) only restaurants that fail to achieve an A on initial inspection are reinspected, mechanically leading to an uneven score distribution near the A grade threshold because of multiple pulls from an underlying score distribution for non-A restaurants only (this unsmooth distribution is unrelated to gaming or ability); (3) scores are a sum of violation points, which accrue in units other than 1, leading to an integer problem. Taken in sum, programmatic details explain a large portion of the lack of smoothness in the score distribution. Initial inspection scores are more smoothly distributed than after adjudication (panel B of figure 6) and the integer problem existed two years before the grading policy (panel C of figure 6). Thus, preferred model estimates rely on restaurant fixed effects and linear controls for underlying inspection scores, showing regression discontinuity results as robustness checks in the appendix.
27. Appendix E presents further evidence that successful restaurant gaming at the point of inspection is unlikely.
28. Fines and closure models use provisional grades earned during final inspection. Sales and tax models use provisional grades at the beginning of the quarter.
29. Results are also insensitive to the window for closure. Estimates using 390 days following final inspection and preferred model specifications (table 4, columns 1 and 4) are similar (3.2 percentage point decrease for an A and a 1.6 percentage point increase for a C) and not statistically distinguishable. Appendix B shows local linear estimates using the optimal bandwidth, minimizing mean squared error (Imbens and Kalyanaraman 2012). Results are similar. The optimal bandwidth for determining the effect of an A is 1.293 and of a C is 1.939. Thus, Wald regressions use 1- or 2-point bandwidths, to reflect violation integers.
30. Regression tables are available from the authors upon request.
31. Local linear estimates use an optimal bandwidth (Imbens and Kalyanaraman 2012). The optimal bandwidth for determining the effect of an A is 3.774 and of a C is 10.35.
32. Appendix F shows results using measures of grades at alternative times in the quarter.
33. Restaurants may post a B or "Grade Pending" during the "Grade Pending" period, but the decision is unobserved in the data.
34. Appendix F shows results using measures of grades at alternative times in the quarter.
35. Appendix G shows results using measures of grades at alternative times in the quarter.
36. Appendix H shows results using measures of grades at alternative times in the quarter.
37. A 10 percentage point increase in the A (C) grade share, for example, is associated with a $150–$220 increase ($180–$220 decrease) in sales taxes per restaurant. The share of provisional A (provisional C) grades rises (falls) by about 20 (2) percentage points annually after rollout.

References

Bovaird, Tony. 2007. Beyond Engagement and Participation: User and Community Coproduction of Public Services. *Public Administration Review* 67(5): 846–60.

City of New York, Office of the Press Secretary. Mayor Bloomberg, Deputy Mayor Gibbs and Health Commissioner Farley Announce Decrease in Foodborne

Illness and Increase in Restaurant Revenue Since Letter Grading Began. March 6. http://www1.nyc.gov/office-of-the-mayor/news/076-12/mayor-bloomberg-deputy-mayor-gibbs-health-commissioner-farley-decrease-foodborne [accessed July 10, 2019].

Dranove, David, and Ginger Zhe Jin. 2010. Quality Disclosure and Certification: Theory and Practice. *Journal of Economic Literature* 48(4): 935–63.

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): 287–97.

Firestone, Melanie J., and Craig W. Hedberg. 2018. Restaurant Inspection Letter Grades and Salmonella Infections, New York, New York, USA. *Emerging Infectious Diseases* 24(12): 2164–68.

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

Imbens, Guido, and Karthik Kalyanaraman. 2012. Optimal Bandwidth Choice for the Regression Discontinuity Estimator. *Review of Economic Studies* 79(3): 933–59.

Jin, Ginger Zhe, and Phillip Leslie. 2003. The Effect of Information on Product Quality: Evidence from Restaurant Hygiene Grade Cards. *Quarterly Journal of Economics* 118(2): 409–51.

———. 2009. Reputational Incentives for Restaurant Hygiene. *American Economic Journal: Microeconomics* 1(1): 237–67.

Jones, Timothy F., Boris I. Pavlin, Bonnie J. LaFleur, L. Amanda Ingram, and William Schaffner. 2004. Restaurant Inspection Scores and Foodborne Disease. *Emerging Infectious Diseases* 10(4): 688–92.

King, Hal. 2013. *Food Safety Management: Implementing a Food Safety Program in a Food Retail Business*. New York: Springer Science & Business Media.

Meltzer, Rachel, Michah W. Rothbart., Amy E. Schwartz, Thad Calabrese, Diana Silver, Todor Mijanovich, and Meryle Weinstein. 2019. What Are the Financial Implications of Public Quality Disclosure? Evidence from New York City's Restaurant Food Safety Grading Policy. *Public Finance Review* 47(1): 170–201.

New York City Department of Health and Mental Hygiene. 2012. How We Score and Grade. https://www1.nyc.gov/assets/doh/downloads/pdf/rii/how-we-score-grade.pdf [accessed July 10, 2019].

———. 2016. Self-Inspection Worksheet for Food Service Establishments. https://www1.nyc.gov/assets/doh/downloads/pdf/rii/self-inspection-worksheet.pdf [accessed July 10, 2019].

Potoski, Matthew, and Aseem Prakash. 2004. The Regulation Dilemma: Cooperation and Conflict in Environmental Governance. *Public Administration Review* 64(2): 152–63.

Rockoff, Jonah E., and Lesley J. Turner. 2010. Short-Run Impacts of Accountability on School Quality. *American Economic Journal: Economic Policy* 2(4): 119–47.

Saul, Michael H. 2012. Grading Eatery Grading. *Wall Street Journal,* March 7.

Silver, Diana, Michah W. Rothbart, and Jean Bae. 2019. *Do the Stakes in Food Safety Inspections Change Restaurant Outcomes in Court?* Unpublished manuscript.

Simon, Paul A., Phillip Leslie, Grace Run, Ginger Zhe 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(7): 32–36.

Thaler, Richard H., and Cass R. Sunstein. 2008. *Nudge: Improving Decisions about Health, Wealth, and Happiness*. New Haven, CT: Yale University Press.

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(3): 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(3): e81–e87.

Supporting Information

A supplementary appendix may be found in the online version of this article at http://onlinelibrary.wiley.com/doi/10.1111/puar.13091/full.