

# Can Monitoring and Transparency Make Food Safer, and at What Cost?

Jason Huang

June 4, 2017

## Abstract

I estimate the impact of food inspections on restaurant cleanliness by estimating how inspection results affect subsequent inspection results and 311 complaint calls. To deal with the endogeneity of the inspection results, I exploit the random assignments of inspectors and construct inspector-specific measures of stringency as an instrumental variable. I find that this variable is highly predictive of not only the total scores but also the specific violations that are cited, despite the random assignment process that results in similar restaurant characteristics across inspectors. I find that, on the margin, more citations leads to better subsequent inspections. I also find that the establishments not only address the areas for which they received citations but also improve in other areas, suggesting complementarity across multiple dimensions of cleanliness. Furthermore, I find that consumers also respond to these improved sanitation conditions by finding lower probabilities of receiving 311 complaint calls inspection results that do not contribute to public information have little effect on call frequencies. But inspection results that contribute to lower letter grades lead to a significant increase in calls. Taken together, my findings suggest that the violation citations have incentivize restaurants to improve, but these efforts are not noticed by customers. Instead, consumers react mostly to the letter grades, which may at times not fully reflect the underlying qualities. Current findings suggest that showing a numerical score instead of or concurrently with the letter grades can be more informative. This paper also has important implications for other cities as they plan to design or modify their methods of displaying their food inspection results.

*JEL: D02, D81, D82, K32, L51*

# 1 Introduction

Food hygiene at restaurants is a great public health concern. The CDC estimates that over 3 million incidents of food-borne illness occur in the US annually, with 60% of the cases resulting from dining in restaurants<sup>1</sup>. Unlike food quality or service, which consumers can easily observe, food sanitation remains opaque to consumers since most diners cannot venture back to the kitchens and monitor the food preparation process. Local health departments have long conducted regular food inspections to ensure that restaurants meet certain hygiene standards. In the last two decades, however, major cities have started make these inspection results publicly available. In the case of New York City (NYC), starting from July 2010, the Department of Health and Mental Hygiene (DOHMH) has required restaurants to post letter grades on their entrance windows. During each inspection, the inspector calculates a score based on the number of violations and their severities. The numerical score is then converted to a letter grade.

This paper estimates the causal impact of NYC's food inspection results on restaurant compliances and customer perceptions, measured by frequencies of 311 complaints calls. Inspection results can potentially influence restaurant behaviors through for two reasons. First one is monetary. The scores determine the letter grades restaurants post on their entrance windows. These grades can influence consumer demand and affect revenues. The second one is behavior. restaurants may improve by learning from previous mistakes through the citations. Restaurants may commit violations not because they are profit maximizing and purposefully cutting enough corners to equate the marginal savings with the expected marginal cost of penalties from DOHMH. Instead, restaurants may be simply making mistakes because of lack of knowledge or experience.<sup>2</sup> Given that cleanliness is a multi-dimensional task, I examine how the distribution of citations impact the distributions of efforts across these dimensions. Finally, I test whether consumers behaviors respond to changes in restaurant hygiene efforts by measuring the changes in the frequencies of 311 complaint calls. The inspection results may affect consumer behaviors through two channels. In the first channel, worse results incentivize restaurants to make substantive improvements, which lead to decreases in calls. In the second channel, worse results lead to lower grades, causing consumers to view the establishment more negatively and more likely to complain.

The initial results find that, on the margin, restaurants respond to poor inspection results by improving their cleanliness, as measured by improved scores from subsequent inspections. Furthermore, I find that the establishments improve by not only focusing on the areas for which they received citations but on almost every aspect. In regards to the complaint calls, I find that inspection results that do not contribute to letter grades have minimal effect. However, for inspection results that contribute to the posted letter grade, failing to earn the top grade leads to more than

---

<sup>1</sup>Surveillance for Food borne Disease Outbreaks United States: 2013 Annual Report. Available: <https://www.cdc.gov/foodsafety/pdfs/foodborne-disease-outbreaks-annual-report-2013-508c.pdf>

<sup>2</sup>A potential reason is that scores are directly associated with monetary penalties. Each violation is associated with a fine, and the amount of fine depends on the severity cited by the inspector. However, the amounts of these fines are very small, ranging from 200 to 1000 per violation.

200% increase in the number of complain calls relative to the average of . The results suggest that the cleanliness of restaurants improve in response to citations, but along dimensions that are not easily visible to consumers. Instead, consumers react mostly to the letter grades, which may at times not fully reflect the underlying sanitation qualities. A simple model to explain this result is that a customer calls to complain whenever one's perceived experience at a restaurant drops below a certain threshold. *Ceteris paribus*, a lower letter worsens the perceived experience, making the complaint call more likely.

An empirical challenge to studying the impact inspection is that the variable of interest, the inspection results, is endogenous. It includes both the effects of monitoring and compliance. For example, suppose we find that restaurants that had high numbers of citations tend to have high number of citations in the future. Do more citations lead to worse compliance or do restaurants that did poorly in the past continue to do poorly in future? This paper uses the random assignment of food inspectors to inspections for restaurants in New York City as an instrumental variable to overcome the endogeneity of inspection results. As different inspectors have different levels of stringencies, the results that restaurants receive are partly determined by chance.<sup>3</sup>

A growing literature has focused on the role of transparency to improve food sanitation and reduce food-borne illness. The first type is certification, or what [Ho \(2012\)](#) calls targeted transparency. Letter grade system, first introduced in Los Angeles and later adopted by New York City is one example of certification. [Jin and Leslie \(2003\)](#) and [Simon et al. \(2005\)](#) study the introduction of the grading system in Los Angeles in 1998 and find strong evidence for a decrease in hospitalization for foodborne illnesses in jurisdictions after the enactments of the grading systems. [Jin and Leslie \(2003\)](#) argue that the reduction is driven by people cooking more at home and eating out less. [Wong et al. \(2015\)](#) and [Meltzer et al. \(2015\)](#) study New York City's food inspection program. They use inspection scores to measure restaurant cleanliness. After finding that improved sanitation scores and lower violation fines followed the reform, they conclude that the reform improved sanitation practices. These papers look at impact of the overall inspection program while I study the effect of individual inspections. [Filion and Powell \(2009\)](#) document the growing popularities of these grading programs across the globe. The second type of transparency arise from customer reviews. [Kang et al. \(2013\)](#) use Yelp reviews to predict health code violations, and they argue that inspection selection can be optimized based on information from social media. However, sending food inspectors to restaurants for inspections remains as the most direct tool of enforcement, and understanding the impact of these inspections on compliance remains important. Hence, I see this study about monitoring and detection as a complement to instead of a substitute of studies about the role of transparency.

This paper also relates to a broad literature that studies the connection between government monitoring and compliance. Researchers have explored this issue in other areas such as taxation

---

<sup>3</sup>Other areas of research that have used this random assignments of decision makers to recover causal impacts of include the impact of incarceration on recidivism ([Bhuller et al., 2016](#)), the effect of disability insurance on labor supply ([Maestas et al., 2013](#)), and the impact of foster care system on child outcomes ([Doyle, 2007](#)).

(Feinstein (1991), Kleven et al. (2010), Gemmell and Ratto (2012)), workplace safety (Levine and Johnson, 2012), nuclear power plants (Feinstein, 1989), and environmental protection (Duflo et al., 2016). Jin and Lee (2014a) and Jin and Lee (2014b) also look at the interaction between food inspectors and restaurants. But they focus on the inspectors' behaviors and detection technology, while this paper asks a related but different question of how restaurants respond to these inspections results.

The rest of the paper proceed as follows. Section 2 discusses the New York City's food inspection program. Section 3 describes the food inspection and 311 data. Section 4 discusses the construction of the instrumental variable and describes various validity tests. Section 5 presents the main empirical strategy and shows the results. Section 6 ends this draft with a discussion of the policy implications and the steps going forward.

## 2 New York City Food Inspection and Grading System

New York City's DOHMH oversees the food safety program. It sends out food inspectors to restaurants to conduct inspections. Each restaurant receives at least one inspection a year. Starting in 2005, the city implemented a numerical scoring system. It outlined ninety-eight violations, broken into eight categories: food temperature, food source (general and critical), food protection, facility design, personal hygiene, vermin/garbage, and facility maintenance. Each violation is associated with a range of scores, and the inspector assigns the numerical value that reflects the severity of the violation. For example, a violation 10F for not non-food contact surface improperly constructed is worth from 2 to 5 points, depending on how many of these surfaces the inspector finds<sup>4</sup>.

Starting in 2010, the sum of the scores were converted to letter grades, A, B, C, based on the following cutoffs. Score between 0 to 13 earns an A, between 14 to 27 earns a B, and 28 or above earns a C. In addition, the city required the restaurants to post these letter grades on the entrance so that they are clearly visible to the customers. The grading policy also changed the inspection cycles. It introduced dual inspection cycles, which involved the a random initial inspection and a possible re-inspection within a month if the initial inspection results in a letter grade lower than an A. During the time in between the initial inspection and re-inspection, restaurants can either post their assigned letter grade or a sign saying "Grade Pending." The score from re-inspection is then used to determine the letter grade the restaurant posts until the next initial inspection, and the previous initial inspection result has no effect. Inspectors to both cycles are randomly assigned so rarely do the same inspector inspect a restaurant twice in a row. Finally, the time between an initial inspection and the subsequent initial inspection depends on the inspection results from the original inspection. If the initial inspection score is lower than 14, the subsequent inspection occurs

---

<sup>4</sup>See Food Service Establishment Inspection Procedures Handbook:  
<http://www1.nyc.gov/assets/doh/downloads/pdf/about/healthcode/health-code-chapter23.pdf>.

in roughly a year. If the score is between 14 and 27, the median time until the next inspection is 201 days. If the score is greater than 27, the median length until the next inspection is 176 days.

DOHMH may temporarily close a restaurant for two reasons. The first reason is that the establishment poses a public health hazard that cannot be addressed by the end of the inspections. Such hazard includes issues such hot food not cooked to high enough temperature or cold food not stored at low enough temperature, presence of unpasteurized milk, or harmful noxious gas detected. The second reason is that restaurants receive scores above 28 for three consecutive inspections. If closed, DOHMH immediately posts a closure sign on the window. To reopen, the establishment must submit a written statement to DOHMH that describes the steps it has taken to correct the violations. Then re-opening inspection occurs. The median time between closure and re-opening inspection is 4 days, with 95% of the cases occurring within 24 days. The establishment can re-open if it passes the re-opening inspection, but it receives its next initial inspection within three months. Repeat violations may prompt DOHMH to shutdown an establishment for longer or permanently.

All inspections are conducted with a handheld computers in which the inspectors record the violations. The sum of the individual scores exceeding certain threshold triggers either re-inspections or closure of the restaurant. The average inspection takes one hour and forty-five minutes.

### 3 Data

#### 3.1 Food Inspection Data

The main data source is the universe of food inspection results conducted by New York City's Department of Health and Mental Hygiene (DOHMH) from July 2007 to September of 2016. Each observation in the data corresponds to a violation cited during each inspection. It has information on the date of the inspection, the type of inspection (whether initial inspection or re-inspection), anonymized ID of the inspector, the total score from all the violations, date of adjudication if it occurred, and the modified score after adjudication. It also has the name, street address, zipcode, cuisine type, service type, and venue type of the establishment.

This study focuses mainly on the inspections conducted after the enactment of the grading system, which occurred on July 2010. While sources say the implementation was swift, I use all inspections that occurred after October 1st of 2010 to give 3 month buffer to account potential lags.<sup>5</sup> In the period following the grading system, on which this paper focuses, I see over 337,000 inspections, which include both initial inspections and re-inspections, across over 44,000 establishments. Figure 1 shows the distributions of inspection types across the inspection types. We see that over 60% of initial inspections earn scores higher than 13, resulting in re-inspections. Of the re-inspections,

---

<sup>5</sup>Ho (2012) says that the actual law was enforced on July 15, 2010.

14.2% earn Cs, 37.7% earn Bs, and the remaining earn As. But almost 90% of re-inspections that do not earn As go to adjudication, with a bit over half getting modified scores that result in improved grades and over 80% resulting in lower modified scores.

Table 2 presents summary statistics for both the restaurant samples and inspection samples. By comparing the two samples, we can see whether inspections tend to skew towards certain types of restaurants. The statistics appear fairly comparable across the two samples, with restaurants that have counter service being inspected more often and those with takeout service inspected less often.

### 3.2 311 Calls

New York's 311 is a special telephone number designated for non-emergency requests to municipal services. The phone line also served as a way to report complaints, such as sighting of rodents or food poisoning, against retail food establishments to DOHMH. The data contain the date of the call, the types of complaints, and the address of the incident.

The 311 calls are matched to the food inspection data based on the address using fuzzy string matching. The appendix provide more detail on the matching process.

### 3.3 Descriptive Analysis: Bunching Around Thresholds

As documented in other economic literatures, the discontinuous mapping from integral scores to discrete letter grades creates jumps in incentives and leads to bunching in the score distributions<sup>6</sup>. More specifically, as shown in the left plot of figure 2, the distribution of scores from initial inspections clearly show bunching around the cutoff between an A and a B at 13. We do not see any bunching above 13 since all those inspections re-inspections follow and those scores do not contribute to the final and those restaurants often opt to post Grade Pending in the meantime. In the right plot in figure 2, we see bunching around both 13 and 27 in the scores from re-inspections since those scores contribute to the determination of the letter grades.

Strategic behaviors of both the inspectors and the restaurants can contribute to the observed bunching in the score distribution. Some potential reasons for inspectors to bunch during the first inspection include saving extra trips for re-inspections or the administrative costs of adjudications. But given that the same inspector does not inspect conduct the re-inspection, and the adjudication is handled by another department, this explanation is unlikely. The bunching may be driven by restaurant behaviors. The inspection worksheet is publicly available, so the restaurants may that are around the threshold between grades may exert extra effort to improve. A final reason, which

---

<sup>6</sup>See Gouriou and Roys (2014) for discontinuity of firm sizes caused by size-based regulations.

is a behavioral explanation, may be that inspectors experience awkwardness handing out a score that is a point or two what is needed to earn a better grade.

## 4 Stringencies of Inspectors as Instrumental Variable

According to staff of DOHMH, whenever a restaurant needs to be inspected, whether for an initial inspection or a re-inspection, a inspector is randomly assigned to inspect that establishment. Inspectors do not specialize in any particular geographies or types of restaurants, so every restaurant at any point in time should be equally likely to receive any of the inspectors. In the next subsections, I discuss how I construct the instrumental variable and show its predictive power on the endogenous variable.

### 4.0.1 Inclusion Restriction

The inclusion restriction of a instrumental variable requires it to have high predictive power on the endogenous variable. In this case, we need our inspector stringency to be highly statistically correlated with the scores that the inspector assigns to the restaurants. I calculate the instrumental variable for each inspection by computing the average score that the assigned inspector gives to all the other restaurants that the inspector has inspected for that inspection type (initial or re-inspection). Specifically, I define the instrument between restaurant  $i$  and inspector  $j$  of inspection type  $p$  as

$$Z_{ijp} = \frac{1}{n_{jp} - I_{ijp}} \left( \sum_{k:tp \neq i:tp} SCORE_{k:tp} \right)$$

where  $n_{jp}$  is the total number of type  $p$  inspections conducted by inspector  $j$ ,  $I_{ijp}$  is the total number of type  $p$  inspections inspector  $j$  has conducted at restaurant  $i$ , and  $SCORE_{k:tp}$  is the score inspector  $j$  gave to restaurant  $k$  on date  $t$  for an inspection type  $p$ .<sup>7</sup>

Figure 3 plots the distribution of the inspector average scores. The histogram shows a considerable amount of heterogeneity in terms of inspectors' levels of stringencies<sup>8</sup>. Figure 3 also shows the fitted value of locally weighted regression of the inspection scores on the inspectors' leave-out average scores<sup>9</sup>. The inspection scores and the inspector average scores follow a monotonic relationship, closely tracing along the 45° line. The plot suggests that being inspected by a more stringent inspector leads to a higher score.

<sup>7</sup>The types of inspections can be either initial inspection or re-inspection.

<sup>8</sup>See Feinstein (1989), Feinstein (1991), and Macher et al. (2011) for additional documentation of inspector heterogeneity pertaining to IRS examiners, nuclear power plant inspectors, FDA regulators, respectively.

<sup>9</sup>I use a bandwidth of 0.5.

We set up the first stage of our two-stage least square analysis here.

$$score_{it} = \gamma Z_{it} + \delta_i + \tau_t + \epsilon_{it} \quad (1)$$

Table 3 shows the first stage result of (1). I show the results of both with and without restaurant fixed effects. We see that the point estimates are around 1, which suggests that an inspection done by an inspector with a 1 point higher average score increases the expected inspection score by 1 point. The estimates are also highly statistically significant, suggesting that the inspector stringency is not a weak instrument.

#### 4.1 Assessing the Randomness of Inspector Assignment

While we cannot directly test for the randomness of the inspectors' assignment process, we can test for some empirical patterns that random assignment implies.

The first implication of random assignment is that the inspector propensity should not be correlated with observables of restaurants to which one is assigned. Hence, we test whether restaurant observable characteristics have predictive power of the stringency levels of the inspectors assigned to them. If the assignments are random, the covariates should all be small and statistically insignificant. For example, a coffee shop in Brooklyn should have the same chance of getting a strict inspector than a Michelin five star restaurant in Manhattan. Table 4 shows the regression results of inspection scores and inspector leave-out average score on various restaurant characteristics and lag terms. We see that all the covariates are highly statistically significant for the score regression. In the second column, which the depend variable is the inspector propensity but includes inspections conducted by inspectors who have at least 50 inspections, most of the covariates have no statistical significance and the point estimates are very small. There are few covariates, such as Pizza/Italian cuisine, or whether the establishment offers buffet or cafeteria service, that are statistically significant at 90% or higher level. However, their point estimates are fractions of those in the score regression. Calculating inspector propensity using only 50 inspections may cause some covariates to be significant by chance. The third column restrict the sample to inspections done by inspectors with at least 650 inspections. We see that the covariates that were significant in the second columns are no longer significant.

A second implication of the randomness is that the inspections each inspector conduct should not be geographically clustered, and each geographic area should not be inspected by only a few inspectors. To test this claim, I calculate the Herfindahl-Hirschman index (HHI) of the zipcodes of the restaurants each inspector has inspected and the HHI of the inspectors that have inspected each zipcode. The results are presented in figure 4. The red dotted lines represent the HHI of zipcodes and inspectors under the perfect random case<sup>10</sup>. The takeaway from the figure is that, while no

---

<sup>10</sup>Because inspections are not uniformly distributed, the baseline is calculated based on the sample distributions of inspectors across zipcodes



zipcode or inspector exhibit a perfectly random distribution, I find no evidence that inspectors tend to focus on a small geographic area or that certain zipcodes get inspected by only a handful of inspectors.

## 5 Empirical Analysis

### 5.1 Impact of Inspection Scores on Subsequent Scores

We start with analyzing the impact of scores from initial inspections on subsequent initial inspection results. To measure the effect, I estimate the following equation.

$$Score_{i,t^{next}} = \beta Score_{it} + \delta_i + \tau_t + \tau_{t^{next}} + \varepsilon_{it}, \quad (2)$$

where  $Score_{i,t}$  is the score earned by restaurant  $i$  on date  $t$ ,  $\delta_i$  and  $\tau_t$  are the restaurant and date fixed effects, respectively, and  $t^{next}$  is the date of the subsequent initial inspection. The error term  $\varepsilon_{it}$  is two-way clustered at the inspector and zipcode levels.

The term  $Score_{it}$  may be endogenous since it may be correlated with unobservable restaurant specific time varying variables, such as the management or the overall business conditions of the establishments. Hence, we instrument for it using equation (1).

Table 5 shows the regression results. The first table shows the result using the entire sample. The second and third tables show the results broken down by The first two columns show the results of the OLS regressions, followed by the results from the IV regression. The fact that OLS yields positive coefficients while the IV yields negative coefficients suggest that sanitation qualities persists across time. A point higher in the current initial inspection is associated with 0.13 point higher score in the subsequent initial inspection. However, a point higher inspection score leads to an 0.23 point improvement in the subsequent inspection.

In the case that the restaurant's initial inspection is above 13 and experiences

$$Score_{i,t^{next}} = \beta_1 Score_{it} + \beta_2 Score_{it^{re-inspect}} + \delta_i + \tau_t + \tau_{t^{re-inspect}} + \tau_{t^{next}} + \varepsilon_{it}, \quad (3)$$

### 5.2 Impact of Inspection Results on Getting an A and Temporary Closure

In addition to subsequent scores as measurement for compliance, I consider two more outcomes during the next initial inspection: the event of the restaurant receiving an A and the event of getting temporarily closed by DOHMH. While previous subsection has explored the impact of inspection scores on subsequent inspection, it assumes a linear model. Given the bunching of scores around the 13-14 threshold, we do not expect the treatment effect to be constant across different score

ranges. The outcome of obtaining an A during initial inspection is also of interest for DOHMH, since a restaurant's failing to do so triggers an re-inspection that imposes

Table 6 reports the results. Given that the standard deviation of scores are around 14, the result suggests that one standard deviation increase in inspection scores increases the chance of getting an A during next initial inspection by about 3%, which is almost a 10% increase relative to the average 36%. And one standard deviation increase in score decreases the chance of getting closed

### 5.3 Multi-task

So far, I have used the total score as an one dimensional summary of the inspection results. However, cleanliness is truly a multi-dimensional task<sup>11</sup>. There are 82 violation codes, grouped into eight categories: food temperature, food source, food protection, facility design, personal hygiene, vermin, and facility maintenance. Figure 5 provides a visual breakdown of the frequencies by which these groups are cited. I now measure the result of an inspection using an eight-dimensional binary vector. Each element of the vector represents one of the eight violation groups and encodes whether that inspection had at least one citation within that group.

I test how restaurants' effort allocation responds to the citation profile using the following regression.

$$Pr(Cited_{igt^{next}}) = \sum_{g' \in \mathcal{G}} \beta_{gg'} Cited_{igt} + \delta X_i + \tau_t + \varepsilon_{igt}, \quad (4)$$

where  $Cited_{igt}$  is an indicator for whether restaurant  $i$  receives a citation of violation in group  $g$  on date  $t$ ,  $\tau_t$  is the date fixed effect and  $X_i$  is the restaurant controls. As before,  $Cited$  is an endogenous, so need to instrument it using inspector propensity. The coefficient  $\beta_{gg'}$  measures how a restaurant's effort in group  $g$  responds to a citation in group  $g'$ . To instrument eight endogenous variables, I construct each inspector's tendency to cite at least one violation in group  $g$  by using a leave-out strategy similar to the one done in section 4.0.1. More specifically, I construct the instrument of inspector  $j$ 's tendency to cite group  $g$  at restaurant  $i$  as

$$Z_{ijg} = \frac{1}{n_j - n_{ij}} \left( \sum_{k|jgt \neq ijgt} Cited_{ijgt} \right),$$

where  $n_j$  is the number of inspections made by inspector  $j$  and  $n_{ij}$  is the number of inspections inspector  $i$  has made at restaurant  $j$ . The first stage is a system of equations defined as

$$Pr(Cited_{ijgt}) = \sum_{g' \in \mathcal{G}} \theta_{gg'} Z_{ijg'} + \varepsilon_{ijgt}, \quad (5)$$

---

<sup>11</sup>See Lu (2012) for literature on multi-tasking.

where  $\mathcal{G}$  is the set of violation groups. The coefficients  $\theta_{gg'}$  measures the predictive power of an inspector's tendency to cite a violation in group  $g'$  on the citation of a violation in group  $g$ . We should expect  $\theta_{gg}$  for all the groups to be statistically significant, since

The result of the (5) is presented in table 7. As expected, almost all the  $\theta_{gg'}$ 's are highly statistically significant and close to 1. Furthermore, most of the off-diagonal coefficients are either positive or negative but statistically insignificant, suggesting that receiving an inspector who is more stringent on one dimension does not decrease the chance of being cited in another dimension. There are some coefficients, boxed in blue, that show the

Table 9 presents the results of (4) instrumenting with (5). We see that the diagonal coefficients are all negative, and six of the eight groups are statistically significant at 90% level. Food protection seems to be the most sensitive.

The signs of the diagonals shed insight on the production function of restaurant cleanliness. We see that many off-diagonal elements are negative and statistically significant, which suggests that a citation in one dimension leads to an improvement in another dimension. This finding makes sense in light of the fact that various dimensions may be complementary. One example of such complementarity is that a citation in facility maintenance reduces the probability of a subsequent facility maintenance citation by 3%, it also reduces the probabilities of subsequent citations in food protection, food temperature, and vermin/garbage by 3.0%, 3.7%, and 4.37%, respectively. A citation in facility maintenance may lead the restaurant to improve along this dimension, which might include fixing a broken thermometer (violation 10E) that improves the food temperature dimension or better maintaining contact surface (violation 10F) that improves food protection dimension and makes the kitchen more vermin proof.

## 5.4 Impact of Inspection Results on 311 Complaint Calls

The ultimate goal of any food inspection programs is to reduce food-borne illness, but directly studying food poisoning is difficult. This outcome variable suffers from under-reporting, and more severe cases that result in hospitalization, the origin of the illness cannot be traced back to a specific establishment. Food inspection results serve as a convenient proxy for restaurant cleanliness. A couple studies have found evidence of association between inspection results and the likelihood of food-borne illness outbreaks.<sup>12</sup> In this section, I use 311 complaint calls of restaurants to supplement the findings in my previous sections.

To not bias the analysis against restaurants with no address or address that are difficult to match, I exclude all establishments that share the same address with at least another establishment. This process excludes many concession stands or fast food establishments in major hubs like JFK airport or Grand Central Station. This leaves me with a sample of 40,531 establishments that had

---

<sup>12</sup>See Irwin et al. (1989) and Petran et al. (2012a) and Petran et al. (2012b).

at least one inspection after October 2010. The average number of complaint calls between initial inspections is 0.14.

I transform the inspection level data to a restaurant-month level data where each observation represents a month for each restaurant between the first and last recorded inspection. For each month, I calculate the inspection score as the score from the most recent inspection. For months in which inspections occur and scores change, I calculate the proportion of the month after the inspection. For example, if an inspection occurs March 15 2011, then for the year-month cell of 3-2011, I calculate the weight as  $(31 - 15/31) = \frac{16}{31}$ .

I estimate the following equation.

$$called_{it} = \beta Score_{it} + \theta Days\_to\_next_{it} + \delta_i + \tau_t + \varepsilon, \quad (6)$$

where  $called_{it}$  a binary variable that equals one if DOHMH received at least one complaint call for restaurant  $i$  in month  $t$ . For the IV regression, I again instrument the score using equation (1).

Table 10 presents the result of (6) for both initial inspections and re-inspections. Compared to the point estimate of score from re-inspection regression, the point estimates from the initial inspection regression is much smaller and statistically insignificant. The point estimates of both the OLS and IV regressions of re-inspections are positive and statistically significant. One possible explanation for this counter-intuitive result is that higher score contributes to worse grades. A lower grade lowers consumers' threshold of tolerance, making them more likely to complain.

To investigate the role the letter grades, I estimate the following non-parametric regression of complaint frequencies on scores and on modified scores, i.e

There are two types of scores: the scores from the cycle inspections and the scores determining the letter grade scores. [TODO: More explanation here]

$$called_{it} = \sum_s \beta_s \mathbf{1}_{\{Score_{it}=s\}} + \delta_i + \tau_t + \varepsilon_{it}$$

$$called_{it} = \sum_s \beta_s \mathbf{1}_{\{Visible\_Score_{it}=s\}} + \delta_i + \tau_t + \varepsilon_{it}$$

Figure ?? show the  $\beta_i$ 's for the above two regressions, along with the histogram of modified scores. As expected, we see a noticeable jump in the average number of complaints between modified score of 13 and 14, but we do not see such jump between score of 13 and 14. We also see a noticeable but noisier jump between 27 and 28. These discontinuities are not present in the coefficients of the non-parametric regression on scores. Furthermore, the point estimates for the visible scores

Because of the bunching we see in the modified scores, the jumps of the coefficients are suggestive but not conclusive evidence to establish a causal link between the letter grades and the frequencies of complaint calls. [TODO: fix writing] Inspector assignment instrument too noisy

$$called_{it} = \sum_{l=-5}^5 \sum_{g \in \{A,B,C,P\}} \beta_{gl} \mathbf{1}_{\{grade_{it+l}=g\}} + \delta_i + \tau_t + \varepsilon_{it}.$$

Figure ?? shows the coefficients.

## 6 Discussion and Steps Going Forward

An effective targeted transparency program requires the information made public to be simple for the general public to understand. However, by simplifying a multi-dimensional and continuous result into a discrete letter grade, NYC's food inspection program has designed a system that can lead to widely different consumer responses to restaurants with comparable underlying qualities. The result that lower grades invite more complaints, even when the underlying qualities are arguably comparable, raises questions regarding the best way to display the inspection results. One alternative to the displaying only the letter grade is to display the numerical score instead of or concurrently with a letter grade.

I plan to add an additional section discussion the potential pitfalls of the instrumental variable. One concern is the exclusion restriction, which requires that the instrument affects the outcome only through the endogenous variable. This condition fails if more stringent inspectors also happen to be more knowledgeable about cleanliness practice, and they affect the restaurants' subsequent inspection results not only through the scores they give but the knowledge they impart. Another potential problem is that inspectors adjust their grading based on the identities of the previous inspectors to an establishment. A second concern is the monotonicity of the instrument.

Future steps also include additional specifications and robustness checks. For example, instead of using raw counts, the regressions in subsection 5.4 can use the number of complaints divided by the number of days in the relevant windows. I also plan to tests whether the effects are different for different types of complaints.

## References

- Bhuller, M., Dahl, G. B., Løken, K. V., and Mogstad, M. (2016). Incarceration, Recidivism and Employment . *Mimeo*.
- Doyle, J. (2007). Child Protection and Child Outcomes: Measuring the Effects of Foster Care. *American Economic Review*, 97(5):1583–1610.
- Duflo, E., , Greenstone, M., Pande, R., and Ryan, N. (2016). The Value of Regulatory Discretion: Estimates from Environmental Inspections in India. *NBER Working Paper Series*.
- Feinstein, J. (1989). The Safety Regulation of U.S. Nuclear Power Plants: Violations, Inspections, and Abnormal Occurrences. *Journal of Political Economy*, 97(1):115 – 154.
- Feinstein, J. (1991). An Econometric Analysis of Income Tax Evasion and Its Detection. *The RAND Journal of Economics*, 22(1).
- Filion, K. and Powell, D. A. (2009). The Use of Restaurant Inspection Disclosure Systems as a Means of Communicating Food Safety Information. *Journal of Food Service*, 20(6):287–297.
- Gemmell, N. and Ratto, M. (2012). Behavioral Responses to Taxpayer Audits: Evidence from Random Taxpayer Inquiries. *National Tax Journal*, 65(1):33–58.
- Gourio, F. and Roys, N. (2014). Size-dependent regulations, firm size distribution, and reallocation. *Quantitative Economics*, 5:377416.
- Ho, D. H. (2012). Fudging the Nudge: Information Disclosure and Restaurant Grading. *The Yale Law Journal*, 112(3):522 – 851.
- Irwin, K., Ballard, J., Grendon, J., and Kobayashi, J. (1989). Results of Routine Restaurant Inspections Can Predict Outbreaks of Foodborne Illness: The Seattle-King County Experience. *American Journal of Public Health*, 79(5):586–590.
- Jin, G. Z. and Lee, J. (2014a). A Tale of Repetition: Lessons from Florida Restaurant Inspections. *Mimeo*.
- Jin, G. Z. and Lee, J. (2014b). Inspection Technology, Detection, and Compliance: Evidence from Florida Restaurant Inspections. *The RAND Journal of Economics*, 45(4):885917.
- Jin, G. Z. and Leslie, P. (2003). The Effect of Information on Product Quality: Evidence from Restaurant Hygiene Grade Cards. *The Quarterly Journal of Economics*, 118(2):409–451.
- Kang, J. S., Kuznetsova, P., Choi, Y., and Luca, M. (2013). Where Not to Eat? Improving Public Policy by Predicting Hygiene Inspections Using Online Reviews. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*.
- Kleven, H. J., Knudsen, M. B., Kreiner, C. T., Pedersen, S., and Saez, E. (2010). Unwilling or Un-

- able to Cheat? Evidence from a Randomized Tax Audit Experiment in Denmark. *Econometrica*, 79(2):651–692.
- Levine, D. I. . M. W. T. and Johnson, M. S. (2012). Randomized Government Safety Inspections Reduce Worker Injuries with No Detectable Job Loss. *Science*, 336(6038):907–911.
- Lu, S. F. (2012). Multitasking, Information Disclosure, and Product Quality: Evidence from Nursing Homes. 21(3):673705.
- Macher, J. T., Mayo, J. W., and Nickerson, J. A. (2011). Regulator Heterogeneity and Information Asymmetry. *The Journal of Law and Economics*, 54:25 – 54.
- Maestas, N., Mullen, K. J., and Strand, A. (2013). Does Disability Insurance Receipt Discourage Work? Using Examiner Assignment to Estimate Causal Effects of SSDI Receipt. *American Economics Review*, 103(5):1797–1829.
- Meltzer, R., Calabrese, T., Silver, D., and Mijanovich, T. (2015). Is Public Grading Worth the Costs? An Evaluation of New York Citys Restaurant Grades Policy. *Mimeo*.
- Petran, R. L., White, B. W., and Hedberg, C. W. (2012a). Health department inspection criteria more likely to be associated with outbreak restaurants in minnesota. *Journal of Food Protection*, 75(11):2007–2015.
- Petran, R. L., White, B. W., and Hedberg, C. W. (2012b). Using a theoretical predictive tool for the analysis of recent health department inspections at outbreak restaurants and relation of this information to foodborne illness likelihood. *Journal of Food Protection*, 75(11):2016–2027.
- Simon, P. A., Leslie, P., and Run, G. (2005). Impact of Restaurant Hygiene Grade Cards on Foodborn-Disease Hospitalization. *Journal of Environmental Health*, 67(7):32– 36.
- Wong, M. R., McKelvey, W., Ito, K., and Jacobson, J. B. (2015). Impact of a letter-grade program on restaurant sanitary conditions and diner behavior in New York City. *American Journal of Public Health*, 105(3):81–87.

## 7 Appendix

### 7.1 Monotonicity

The instrument needs only the conditional independence and exclusion restriction to recover the causal effect of inspection results if the effect is constant across establishments. With heterogeneous effects, however, we need monotonicity condition. In this setting, the condition states that the score that an establishment gets is strictly increasing in the stringency of the inspector. This is a condition that we cannot directly test since it requires the scores of counterfactual establishments inspected by inspectors of different stringency. This condition can be in question when the treatment depends on multi-dimensions and inspectors differ not only in overall stringency but across those dimensions. For example, suppose that we have a stringent and lenient inspector, but the lenient inspector is much more picky about keeping cold food under the proper temperature than the stringent inspector. Then a sushi restaurant that serves a lot of raw fish can conceivably receive a worse score when it gets the lenient inspector.

To appease such concern, I follow the strategy used in [Bhuller et al. \(2016\)](#) and conduct two tests. First, I test whether the first stage holds for various sub-samples based on overall cleanliness, boroughs, cuisines, and service. The overall cleanliness is calculated as the predicted values from regressing scores on restaurant fixed effect, inspection date fixed effects and inspection type fixed effects. I split the samples into four quartiles based on the predicted values. Second, I reconstruct the inspector stringency by excluding the sub-sample of interest. For example, in the case of inspections done in Manhattan, I recalculate the inspectors' propensities based on their inspections outside of Manhattan. Monotonicity condition implies that inspectors who are strict for certain establishments are also strict at others.

Table 1 reports the results for the two tests, organized by cleanliness quartiles, boroughs, cuisine, and service. Each subtable shows the regressions of the original leave-out stringency levels and the re-calculated stringency based on inspections outside of the subsamples. All the estimates are large and statistically significant, suggesting that inspectors' tendencies remain consistent across different types of establishments.



Table 1: Monotonicity Test

## (a) Cleanliness Quartiles

Baseline-Sample					Inverse-Sample				
VARIABLES	(1) 1st Quartile	(2) 2nd Quartile	(3) 3rd Quartile	(4) 4th Quartile	VARIABLES	(1) 1st Quartile	(2) 2nd Quartile	(3) 3rd Quartile	(4) 4th Quartile
Estimate	0.407*** (0.0270)	0.681*** (0.0240)	0.852*** (0.0213)	1.231*** (0.0256)	Estimate	0.369*** (0.0340)	0.612*** (0.0247)	0.779*** (0.0254)	1.609*** (0.0674)
Observations	80,172	81,817	81,901	86,433	Observations	75,188	81,328	81,481	85,851
Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1					Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1				

## (b) Boroughs

Baseline-Sample						Inverse-Sample					
VARIABLES	(1) Manhattan	(2) Bronx	(3) Brooklyn	(4) Queens	(5) Staten-Is.	VARIABLES	(1) Manhattan	(2) Bronx	(3) Brooklyn	(4) Queens	(5) Staten-Is.
Estimate	1.005*** (0.0156)	1.082*** (0.0275)	0.985*** (0.0198)	0.969*** (0.0206)	0.944*** (0.0286)	Estimate	0.990*** (0.0269)	1.082*** (0.0344)	0.954*** (0.0238)	0.934*** (0.0248)	0.938*** (0.0308)
Observations	131,900	31,010	79,664	77,186	10,507	Observations	128,977	30,334	76,973	74,933	10,468
Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1						Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1					

## (c) Cuisines

Baseline-Sample						Inverse-Sample					
VARIABLES	(1) American	(2) Pizza/Italian	(3) Chinese	(4) Coffee	(5) Japanese	VARIABLES	(1) American	(2) Pizza/Italian	(3) Chinese	(4) Coffee	(5) Japanese
Estimate	0.957*** (0.0241)	0.994*** (0.0180)	1.086*** (0.0402)	0.796*** (0.0354)	1.096*** (0.0262)	Estimate	0.929*** (0.0298)	0.990*** (0.0199)	1.070*** (0.0449)	0.782*** (0.0363)	1.098*** (0.0273)
Observations	75,330	38,432	38,785	13,303	10,915	Observations	74,954	38,312	38,424	13,292	10,910
Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1						Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1					

## (d) Service

Baseline-Sample				Inverse-Sample			
VARIABLES	(1) Counter Service	(2) Takeout Service	(3) Wait Service	VARIABLES	(1) Counter Service	(2) Takeout Service	(3) Wait Service
Estimate	1.055*** (0.0177)	0.894*** (0.0140)	1.127*** (0.0178)	Estimate	1.064*** (0.0269)	0.810*** (0.0181)	1.160*** (0.0236)
Observations	127,981	127,365	73,381	Observations	126,011	124,867	72,645
Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1				Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1			

Standard errors are two-way clustered at inspector and zipcode level. To reduce noise, only inspections conducted by inspectors who have done at least 50 inspections remain in the regressions.

## 7.2 Matching 311 calls to Food Inspection Data

I use fuzzystring matching to link 311 complaint calls with food inspection data based on the address. I perform steps to standardize the address

- Add suffix to numbered streets, i.e. change "1 Street" to "1st Street"

[TODO: to be completed]



Table 2: Summary Statistics of Restaurants and Inspections

	Restaurant Sample		Inspection Samples	
	mean	sd	mean	sd
chain	0.11	0.31	0.10	0.31
fast_food	0.02	0.14	0.02	0.13
bar	0.06	0.23	0.06	0.24
Buffet Service	0.01	0.12	0.02	0.13
Cater Service	0.00	0.07	0.00	0.06
Counter Service	0.34	0.47	0.39	0.49
Take-out Service	0.44	0.50	0.39	0.49
Wait Service	0.18	0.39	0.22	0.42
Cafeteria Service	0.01	0.10	0.01	0.10

This table presents the summary statistics on the sample of inspections after October 1 2010. The restaurant sample contains one observation per restaurant and the inspection sample contains one observation per inspection.

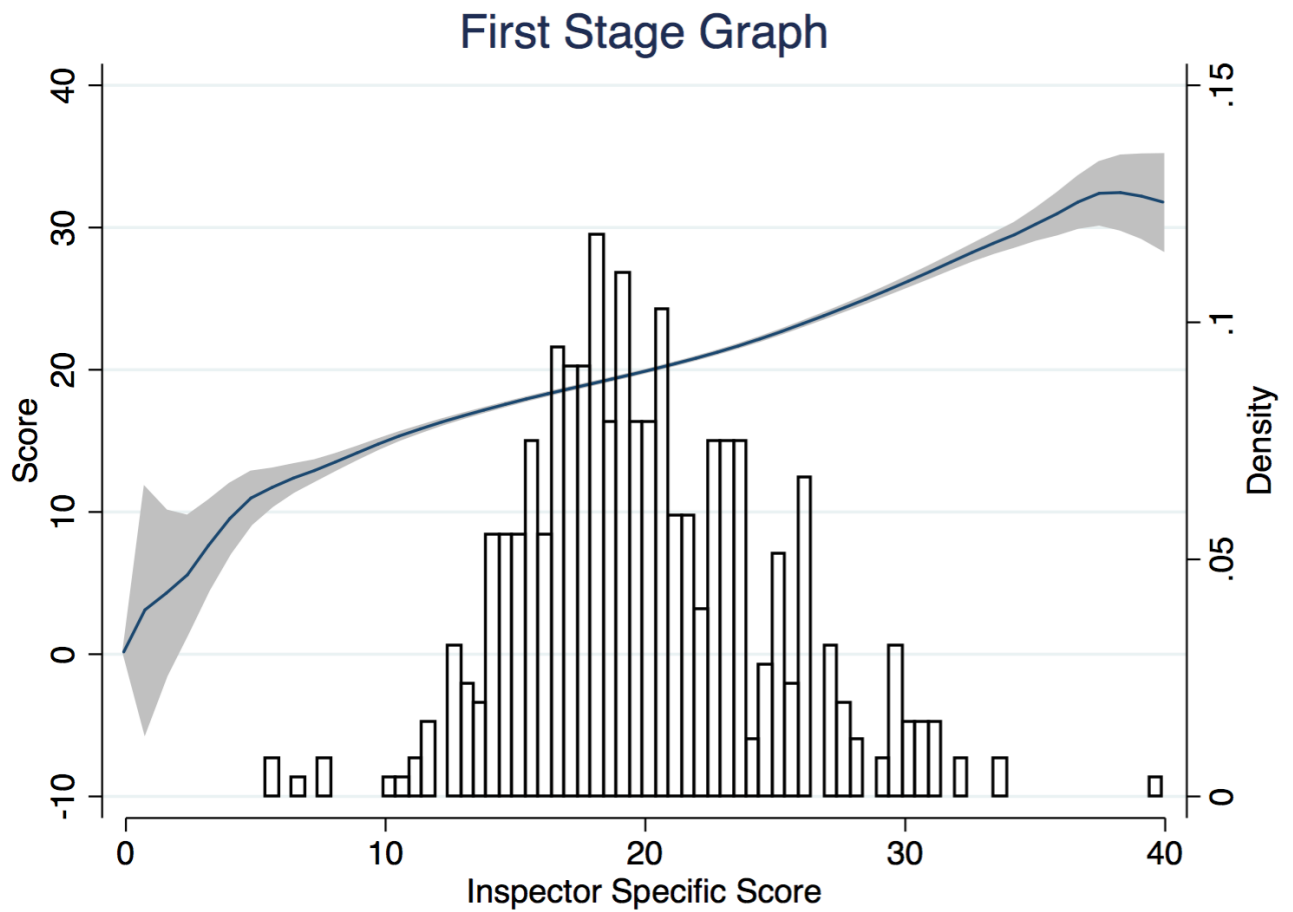


Figure 3: The plot is generated based on 156,323 initial inspections conducted by inspectors who had done at least 50 initial inspections. The expected inspection score on the left y-axis is plotted against the leave-out average inspector scores on the x-axis. The right axis shows the density of inspector average scores.

Table 3: First Stage Regression of Inspection Score on Inspector Stringency

(a) Full			
VARIABLES	(1) Score	(2) Score	(3) Score
Z	0.961*** (0.00948)	0.997*** (0.00983)	1.135*** (0.0162)
Observations	330,469	330,466	325,681
R-squared	0.149	0.201	0.414
Restaurant Controls	NO	YES	NO
Restaurant FE	NO	NO	YES
F Statistics	10271	10270	4908
Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1			
(b) Initial Inspections			
VARIABLES	(1) Score	(2) Score	(3) Score
Z	0.945*** (0.0125)	0.952*** (0.0131)	0.989*** (0.0147)
Observations	200,623	200,620	192,356
R-squared	0.141	0.208	0.475
Restaurant Controls	NO	YES	NO
Restaurant FE	NO	NO	YES
F Statistics	5685	5264	4537
Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1			
(c) Re-inspections			
VARIABLES	(1) Score	(2) Score	(3) Score
Z	0.948*** (0.0111)	0.959*** (0.0111)	0.997*** (0.0150)
Observations	123,226	123,220	113,385
R-squared	0.125	0.162	0.460
Restaurant Controls	NO	YES	NO
Restaurant FE	NO	NO	YES
F Statistics	7320	7528	4431
Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1			

First column consists of all inspections after 10/1/2010. The sample for the second column is reduced to inspections with non-empty zipcode, chain indicator, cuisine type, venue type, and service type. Standard errors are two-way clustered at the inspector and zipcode level.

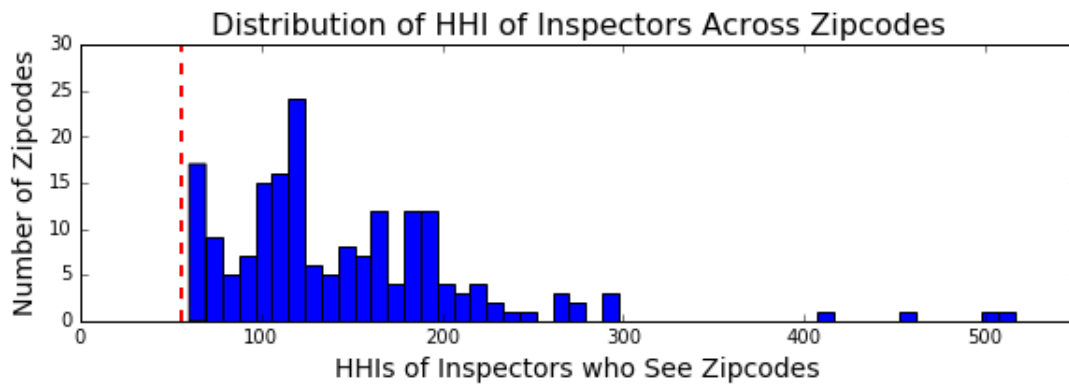
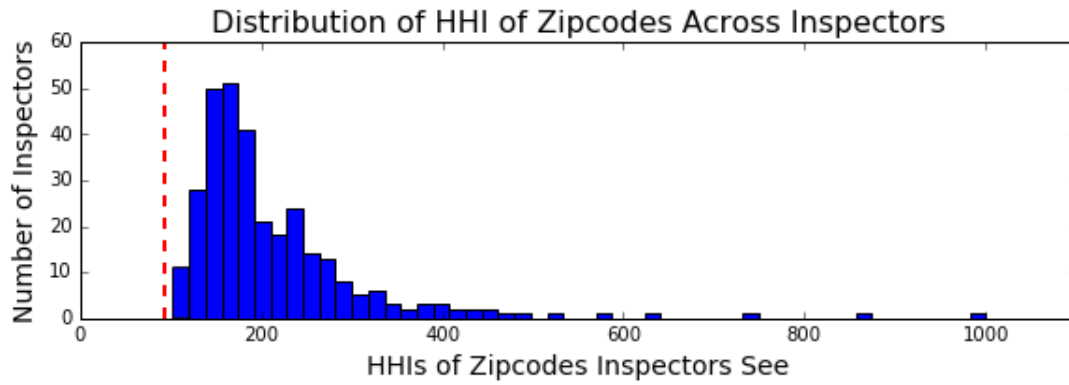


Figure 4: Distribution of HHI

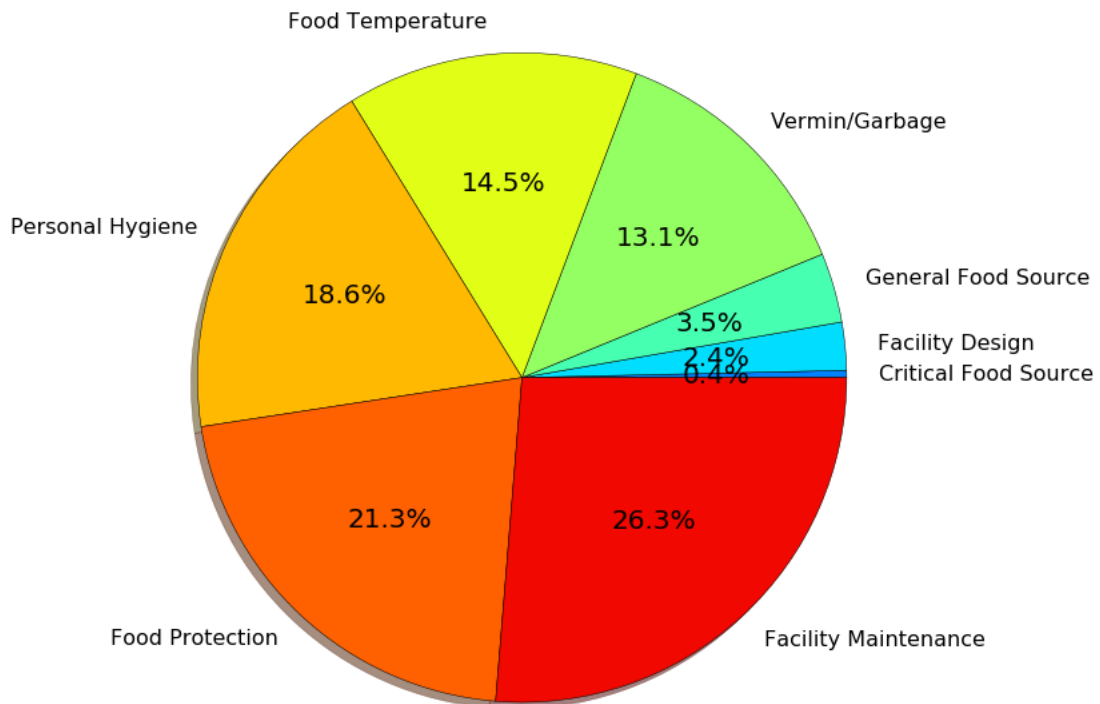


Figure 5: Breakdown of Frequencies of Violation Groups

Table 4

VARIABLES	Score		Inspector Stringency		Inspector Stringency	
VARIABLES	(> 50 Inspections)	se	(> 50 Inspections)	se	(> 650 Inspections)	se
last score	0.218***	(0.00746)	-0.000440	(0.00199)	-0.00193	(0.00205)
last grade = B	1.430***	(0.131)	0.0178	(0.0628)	0.0837	(0.0683)
last grade = C	1.495***	(0.206)	-0.0494	(0.0764)	0.0240	(0.0807)
last inspector propensity	-0.298***	(0.0104)	-0.00363	(0.00491)	-0.00529	(0.00551)
chain	-3.644***	(0.179)	-0.0274	(0.0644)	-0.0300	(0.0734)
Sea Food	0.301	(0.412)	-0.0687	(0.129)	-0.0648	(0.136)
Chinese	1.588***	(0.307)	-0.0952*	(0.0510)	-0.0511	(0.0546)
Pizza/Italian	0.305**	(0.118)	-0.0893***	(0.0342)	-0.0614	(0.0384)
Coffee/Tea	-2.219***	(0.162)	-0.0421	(0.0889)	-0.0514	(0.0966)
Latin	1.534***	(0.303)	-0.0714	(0.0556)	-0.0390	(0.0583)
Spanish	1.560***	(0.261)	0.0125	(0.0615)	-0.0435	(0.0691)
Caribbean	1.542***	(0.288)	-0.0637	(0.0608)	-0.0619	(0.0544)
Sandwich	0.661**	(0.297)	-0.00916	(0.0472)	0.0180	(0.0488)
Concession Stands	-4.379***	(0.722)	0.295	(0.197)	0.209	(0.211)
Fast Food Restaurant-Food Court	0.570***	(0.187)	0.0394	(0.0881)	-0.0211	(0.0868)
Restaurant	1.453***	(0.158)	0.0449	(0.0662)	0.00343	(0.0641)
Buffet Service	2.564***	(0.365)	-0.182*	(0.107)	-0.196	(0.118)
Cater Service	-1.897***	(0.609)	-0.200	(0.190)	-0.193	(0.219)
Counter Service	-0.635***	(0.185)	0.0273	(0.0613)	0.0123	(0.0684)
Take-out Service	-1.410***	(0.185)	-0.153	(0.110)	-0.173	(0.123)
Wait Service	1.205***	(0.166)	0.0418	(0.0481)	0.0404	(0.0525)
Cafeteria Service	-3.541***	(0.500)	-0.280*	(0.156)	-0.200	(0.168)
Observations	299,174		299,174		244,099	
F Statistics	101.6		2.423		2.249	

Robust standard errors in parentheses

\*\*\* p&lt;0.01, \*\* p&lt;0.05, \* p&lt;0.1

Table shows the result of regressing Inspector Leniency on Restaurant characteristics and previous inspection results. Most variables that are highly predictive of restaurant scores are not predictive of inspector propensity, with the exception of the indicator for whether the establishment is a bar or offers wait service. The regressions include date, zipcode, inspection type fixed effects, and the errors are two-way clustered at the inspector and restaurant level.

Table 5: Impact of Inspection Score on Subsequent Inspection Result

(a) Full Sample				
VARIABLES	(1) OLS	(2) OLS	(3) IV	(4) IV
Score	0.182*** (0.00763)	0.134*** (0.00654)	-0.262*** (0.0330)	-0.233*** (0.0253)
Observations	156,323	156,317	156,323	156,317
Restaurant Controls	No	Yes	No	Yes
Restaurant FE	No	No	No	No
Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1				
(b) Score > 13				
VARIABLES	(1) OLS	(2) OLS	(3) IV	(4) IV
Score	0.155*** (0.00726)	0.129*** (0.00680)	-0.283*** (0.0363)	-0.252*** (0.0314)
next score	0.145*** (0.00617)	0.112*** (0.00564)	-0.130*** (0.0119)	-0.126*** (0.0114)
Observations	101,385	101,377	101,382	101,374
Restaurant Controls	No	Yes	No	Yes
Restaurant FE	No	No	No	No
Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1				
(c) Score ≤ 13				
VARIABLES	(1) OLS	(2) OLS	(3) IV	(4) IV
Score	0.613*** (0.0229)	0.398*** (0.0209)	-3.041*** (0.532)	-1.902*** (0.256)
Observations	51,451	51,442	51,451	51,442
Restaurant Controls	No	Yes	No	Yes
Restaurant FE	No	No	No	No
Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1				

Standard errors are two-way clustered at zipcode and inspector levels.



Table 6: Impact of Inspection Scores on Temporary Closure and Obtaining Grade A

VARIABLES	(1) Score (OLS)	(2) Score (IV)	(3) Closure (OLS)	(4) Closure (IV)	(5) Grade A (OLS)	(6) Grade A (IV)
Score	-0.139*** (0.00492)	-0.242*** (0.0162)	-0.000565*** (7.05e-05)	-0.00101*** (0.000185)	-8.49e-05 (0.000127)	0.00222*** (0.000556)
Observations	149,831	149,831	134,365	138,674	149,831	149,831
Inspection Date FE	YES	YES	YES	YES	YES	YES
Restaurant FE	YES	YES	YES	YES	YES	YES
dependent mean	21.08	21.08	0.0154	0.0162	0.372	0.372

Robust standard errors in parentheses

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

Standard errors are two-way clustered at zipcode and inspector levels.

Table 7

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Facility Maintenance	Food Protection	Personal Hygiene	Food Temperature	Vermin/Garbage	Gen. Food Source	Facility Design	Crit. Food Source
Facility Maintenance	1.012*** (0.0193)	-0.0212 (0.0202)	0.0411*** (0.0143)	0.00688 (0.0151)	0.00619 (0.0213)	0.000455 (0.0117)	-0.0111 (0.0103)	0.00262 (0.00166)
Food Protection	-0.0597 (0.0637)	0.940*** (0.0557)	0.0543 (0.0494)	0.234*** (0.0600)	0.0642 (0.0629)	0.0196 (0.0386)	-0.0107 (0.0295)	0.00770 (0.00536)
Personal Hygiene	0.0581*** (0.0170)	0.0394** (0.0187)	0.965*** (0.0146)	0.0898*** (0.0198)	0.0651*** (0.0195)	0.0439*** (0.0133)	0.0326*** (0.00927)	0.00183 (0.00211)
Food Temperature	0.0310 (0.0202)	0.113*** (0.0240)	0.0177 (0.0167)	0.906*** (0.0222)	0.0999*** (0.0246)	0.0369*** (0.0134)	0.0442*** (0.00993)	0.00192 (0.00156)
Vermin/Garbage	0.0449 (0.0540)	0.0280 (0.0465)	-0.0909** (0.0442)	-0.178*** (0.0564)	0.908*** (0.0520)	-0.0261 (0.0346)	0.0158 (0.0253)	-0.00511 (0.00504)
Gen. Food Source	-0.0560* (0.0317)	-0.0629* (0.0353)	0.00496 (0.0286)	-0.0256 (0.0281)	-0.0760** (0.0350)	0.947*** (0.0241)	-0.0262 (0.0159)	-0.000915 (0.00359)
Facility Design	-0.184*** (0.0631)	-0.337*** (0.0824)	0.0223 (0.0515)	0.00730 (0.0731)	-0.400*** (0.0739)	-0.168*** (0.0408)	0.791*** (0.0457)	-0.00582 (0.00844)
Crit. Food Source	-0.130 (0.204)	-0.132 (0.237)	-0.199 (0.151)	-0.111 (0.213)	0.0138 (0.229)	-0.265 (0.173)	-0.0209 (0.114)	0.891*** (0.0334)
Observations	337,615	337,615	337,615	337,615	337,615	337,615	337,615	337,615
Time FE	YES	YES	YES	YES	YES	YES	YES	YES

Robust standard errors in parentheses

\*\*\* p&lt;0.01, \*\* p&lt;0.05, \* p&lt;0.1

Standard errors are two-way clustered at zipcode and inspector levels

Table 8

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Facility Maintenance	Food Protection	Personal Hygiene	Food Temperature	Vermis/Garbage	Gen. Food Source	Facility Design	Crit. Food Source
Facility Maintenance	<b>-0.236***</b> (0.00961)	0.00733 (0.0119)	0.000184 (0.0107)	-0.00506 (0.00944)	-0.000922 (0.00766)	-0.00159 (0.00380)	0.000981 (0.00295)	-0.000509 (0.00140)
Food Protection	<b>-0.0776**</b> (0.0378)	<b>-0.177***</b> (0.0306)	-0.00731 (0.0410)	-0.0142 (0.0253)	<b>-0.0397*</b> (0.0227)	-0.0100 (0.0113)	0.00372 (0.0108)	<b>0.00780*</b> (0.00400)
Personal Hygiene	0.0222 (0.0156)	-0.00484 (0.0103)	<b>-0.189***</b> (0.0148)	-0.00771 (0.00957)	-0.000551 (0.00892)	0.00458 (0.00489)	0.00129 (0.00320)	0.00217 (0.00173)
Food Temperature	0.0148 (0.0216)	-0.0261 (0.0163)	0.0133 (0.0162)	<b>-0.214***</b> (0.0134)	-0.00923 (0.0128)	0.00613 (0.00722)	-0.00622 (0.00573)	-0.00153 (0.00243)
Vermis/Garbage	<b>0.132***</b> (0.0439)	<b>-0.0784*</b> (0.0404)	0.00134 (0.0544)	0.0446 (0.0318)	<b>-0.174***</b> (0.0283)	0.0128 (0.0159)	-0.0105 (0.0138)	-0.00126 (0.00404)
Gen. Food Source	-0.0114 (0.0291)	-0.00139 (0.0348)	-0.0456 (0.0374)	-0.0103 (0.0334)	<b>-0.0395*</b> (0.0202)	<b>-0.191***</b> (0.0199)	-0.0147 (0.00933)	-0.000895 (0.00395)
Facility Design	0.0265 (0.0787)	<b>-0.195**</b> (0.0925)	-0.00742 (0.0776)	0.0168 (0.0551)	-0.0235 (0.0612)	0.00421 (0.0296)	<b>-0.199***</b> (0.0239)	<b>0.0169*</b> (0.00977)
Crit. Food Source	0.274 (0.248)	-0.284 (0.207)	0.279 (0.244)	0.125 (0.195)	-0.0346 (0.142)	-0.0233 (0.0848)	-0.0643 (0.0682)	<b>-0.185***</b> (0.0298)
Observations	149,831	149,829	149,831	149,829	149,829	149,831	149,829	149,831
Dependent mean	0.989	0.802	0.842	0.645	0.503	0.125	0.0694	0.0126

Robust standard errors in parentheses

\*\*\* p&lt;0.01, \*\* p&lt;0.05, \* p&lt;0.1

Standard errors are two-way clustered at zipcode and inspector levels.

Table 9

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Facility Maintenance	Food Protection	Personal Hygiene	Food Temperature	Vermin/Garbage	Gen. Food Source	Facility Design	Crit. Food Source
Facility Maintenance	-0.0299** (0.0134)	-0.0304** (0.0142)	-0.0131 (0.0148)	-0.0372** (0.0148)	-0.0437** (0.0184)	-0.00911 (0.00866)	-0.00103 (0.00569)	0.00330 (0.00264)
Food Protection	0.0134 (0.0398)	-0.106** (0.0487)	-0.0282 (0.0466)	-0.251*** (0.0456)	-0.155*** (0.0567)	-0.0612*** (0.0231)	0.000699 (0.0259)	0.0166* (0.00853)
Personal Hygiene	0.0136 (0.0160)	-0.0767*** (0.0177)	-0.0583*** (0.0190)	-0.106*** (0.0185)	-0.0804*** (0.0210)	-0.0219** (0.00967)	-0.00694 (0.00697)	0.00373 (0.00372)
Food Temperature	-0.0215 (0.0185)	-0.122*** (0.0198)	-0.0202 (0.0163)	-0.0869*** (0.0200)	-0.108*** (0.0222)	-0.00905 (0.0131)	-0.0246*** (0.00894)	-0.00258 (0.00437)
Vermin/Garbage	-0.00984 (0.0292)	-0.0749** (0.0307)	-0.0152 (0.0367)	0.0950*** (0.0337)	-0.0675* (0.0388)	0.0262 (0.0168)	-0.0221 (0.0203)	-0.00139 (0.00550)
Gen. Food Source	-0.0120 (0.0173)	-0.0154 (0.0236)	-0.0410* (0.0214)	-0.0359 (0.0235)	-0.0459* (0.0262)	-0.0227 (0.0147)	-0.0222** (0.00881)	0.000948 (0.00399)
Facility Design	-0.0241 (0.0365)	-0.137*** (0.0473)	-0.0732 (0.0505)	-0.0378 (0.0452)	-0.152*** (0.0526)	0.0154 (0.0248)	-0.0247 (0.0236)	0.0205** (0.00838)
Crit. Food Source	-0.0391 (0.136)	-0.258** (0.131)	0.102 (0.125)	-0.114 (0.150)	-0.0939 (0.136)	0.00645 (0.0698)	0.00234 (0.0703)	-0.0492* (0.0258)
Observations	156,317	156,316	156,316	156,317	156,317	156,316	156,316	156,316

Robust standard errors in parentheses

\*\*\* p&lt;0.01, \*\* p&lt;0.05, \* p&lt;0.1

Standard errors are two-way clustered at zipcode and inspector levels

Table 10

VARIABLES	(1) Initial Inspections	(2) Re-inspections (OLS)	(3) Re-inspections (OLS)	(4) Re-inspections (IV)
Score	0.000219 (0.000208)	0.00143*** (0.000236)		0.00160*** (0.000532)
Days Until Next	0.000380*** (3.78e-05)	0.000434*** (4.76e-05)	0.000125*** (2.04e-05)	0.000182*** (3.86e-05)
Modified Score			0.00222*** (0.000250)	
Observations	136,276	83,361	83,361	83,361

Robust standard errors in parentheses

\*\*\* p&lt;0.01, \*\* p&lt;0.05, \* p&lt;0.1

Sample include inspections conducted by inspectors who have conducted at least 50 inspections of the same type. Standard errors are two-way clustered at zipcode and inspector levels.