

# **Earning the Grade or Just Window Dressing? The Effects of Information Disclosure on Restaurant Hygiene Quality**

**A Study of New York City Restaurant Hygiene Inspection**

**Jialiang Zhu<sup>† ‡</sup>**

## **Abstract**

This paper explores New York City's new mandatory information disclosure policy of posting hygiene quality grade cards on restaurant windows. I find that the dual inspection process (initial inspection and re-inspection) has been exploited by restaurants. In particular, after getting an A card in the re-inspection, restaurants frequently revert back to the previous bad hygiene conditions by the time of a subsequent initial inspection. As a result, there is no significant overall hygiene improvement in NYC. Furthermore, I use random effects and ordered logit models to find that the reduction in asymmetric information by hygiene letter grades narrows the difference of hygiene score between chain and non-chain restaurants, which was previously due to greater consumer knowledge of chain restaurants' reputations.

*Keywords:* Information Disclosure, Cheating, Chain Reputation, Hygiene Inspection

*JEL codes:* D82, H75, I18, L15, L81, L50

---

<sup>†</sup> Department of Economics, George Washington University, 2115 G St NW, Washington, DC 20052.  
*Email address:* machozhu@gwu.edu

<sup>‡</sup> I am grateful to my advisor Wallace Mullin for his guidance and support. I would also like to thank Arun Malik and Robert Phillips for their comments and suggestions. All remaining errors are my own.

*“So, those [restaurant owners] that don’t want to clean up their kitchen, I know why they’re bitching but I would suggest [you] don’t eat in a restaurant unless they have an A.”*

— Michael Bloomberg

## **1 Introduction**

Darby and Karni (1973) established literature on credence goods, whose quality or characteristics are impossible for consumers to judge, even after consumption. This feature of credence goods leads to asymmetric information between firms and consumers. Due to this feature, firms lack incentives to improve product quality but have strong incentives for opportunistic or even fraudulent behavior. Thus, the equilibrium quality of products tends to be minimal and, in particular, below the quality that is socially optimal.

This paper explores New York City’s new mandatory information disclosure policy of posting hygiene quality grade cards in restaurant windows. Hygiene quality in restaurants, unless egregiously lacking, is often unobservable to a casual diner. Thus, restaurant hygiene quality can be treated as a quasi-credence good, i.e., if hygiene quality is above a certain threshold, it is unobservable and unverifiable; however, if it falls below this threshold, it can be observed by potential consumers. For instance, consumers will immediately question the hygiene in a restaurant if they find a cockroach in the main dish. Health inspectors, on the other hand, are able to enter parts of the restaurant that are inaccessible to patrons, examine cleanliness and conditions where the food is being prepared, and take note of how employees behave while preparing food. Similar to how people need organic products to be labeled, in order to identify them, consumers heavily rely on mandatory information disclosure of restaurant hygiene to help them make purchasing decisions. In light of this, in 2010, the Department of Health and Mental Hygiene (DOHMH) in New York City launched a new letter grade policy on restaurant hygiene inspection, with an intention to improve restaurant hygiene and to give consumers access to information which

would otherwise be unobservable to them.

However, due to this credence-good feature, restaurants still have incentives for opportunistic behavior even when they face this new policy. Therefore, the letter grades on restaurant windows in New York City do not always send consumers a clear signal. For example, a recent news report displayed a large mouse conspicuously perching on a ledge under an A-grade restaurant in New York City.<sup>1</sup> Additionally, some other reports reveal that a number of restaurants with B or C cards try to confuse consumers by adding extra letters by themselves so that a B card sign is modified into a B-E-S-T sign, and similarly, a C card sign turns into a C-A-R-D sign.<sup>2</sup> These strange and confusing signs on restaurant windows prove that restaurants have incentives to cheat and consumers may get confused by these letter grades.

Jin and Phillip (2009) find that reputation incentives effectively motivate restaurants to maintain good hygiene conditions when hygiene information is not available to the public and that chain affiliation provides positive reputation which often allows franchised units to free-ride. For instance, franchise-owned restaurants usually have worse hygiene inspection scores than directly-owned ones. However, they also find that when a new letter grade policy took effect in Los Angeles, it eliminated the hygiene score difference between franchise-owned restaurants and directly-owned ones.

Therefore, the research questions which motivate this paper are: Does the behavior of restaurants change in the face of new regulations which make the information of restaurant hygiene quality more transparent? Does the new policy in New York City actually meet its intended goal of improving restaurant hygiene? And do chain restaurants still have better hygiene than non-chain restaurants due to reputation effect after the policy?

The DOHMH in New York City has been inspecting restaurant hygiene using a scoring system since 2005 for public food safety. In July 2010, the Board of Health in New York City authorized the DOHMH to provide a new grading system with the intention of better informing consumers, and of improving restaurant compliance with Health Code

---

<sup>1</sup> See Heidi Patalano (2015) for the picture.

<sup>2</sup> See Aaron Rutkoff (2010) for pictures.

requirements. Thereby, it might further reduce the burden of foodborne illness in New York City. This new grading system maps the raw hygiene inspection scores—which were not observable to consumers—into A, B, or C grades, and requires restaurants to post the grade cards on their windows for customers’ awareness.

This new policy also supports the restaurant industry by using a dual inspection approach that allows restaurants to improve on bad scores before receiving final grades. According to the new policy, an inspector will visit a restaurant at an unannounced time to conduct the inspection. After the inspection, they will assign a hygiene score to the restaurant along with a letter grade. If the restaurant receives an A level hygiene score, then it will get an A card, the highest level grade, and it has to post it immediately on the outside of its storefront. However, if the restaurant fails to receive the A card, it receives a re-inspection approximately 7 to 30 days later, at which point the final grade is issued.<sup>3</sup>

This paper explores this new hygiene inspection policy in New York City and shows evidence that restaurants have strong incentives for opportunistic behavior even when they face the new mandatory inspection policy. Specifically, because of this second inspection opportunity, restaurants tend to only improve their hygiene conditions in preparation for the re-inspection. Further, once restaurants receive an A card, most of them revert back to bad hygiene qualities.

In order to show that restaurants exhibit this type of opportunistic behavior when faced with mandatory inspection, I use an ordered logit regression model with a panel dataset to find that after the new policy is imposed, the likelihood of receiving an A grade is 76% after the second inspection, while only 47% after the first inspection. Moreover, I find that before this new policy, chain restaurants have better hygiene than non-chain restaurants in terms of hygiene score, but the score differences become smaller only after the re-inspection in the post-policy period due to the reduction of asymmetric information of hygiene reputation by the letter grades. In addition, no evidence emerges from consumer complaints data in New York City to support an overall improvement in restaurant hygiene after this new policy.

---

<sup>3</sup> Restaurants can still refuse to accept the grades and wait for a hearing with the Office of Administrative Trials and Hearings Health Tribunal to contest the score. More details are in the Background Section.

This paper contributes to the literature in several ways. First, it shows empirical evidence that firms have strong incentives for opportunistic behavior in the credence goods market even when they face mandatory regulation. It finds that if regulations have any loopholes—in this case, the re-inspection opportunity—restaurants immediately take advantage of them. Secondly, it finds that the chain reputation effect decreases when letter grades enable consumers to observe hygiene information and equalize consumer learning. Finally, this paper uses an ordered logit model with a panel dataset to demonstrate the opportunistic behavior of restaurants in a more accurate way.

The rest of this paper is organized as follows. In Section 2, I provide a review of the related literature. Section 3 introduces the background of the new inspection policy in New York City. Section 4 provides a description of the data. In Section 5, I present a baseline model of testing the new policy effect. In Section 6, I present the evidence of restaurant cheating and regional differences in restaurant hygiene. In Section 7, I present the framework for testing chain restaurant reputation effect and discuss the results. Section 8 provides a brief conclusion.

## **2 Literature Review**

This section provides a brief review of three important, yet often unconnected strands of literature. First I review several theoretical papers of credence goods. Next, I discuss some empirical papers on the impact of information disclosure on consumer choices and opportunistic behavior of experts. Finally, I review recent work on the impact of restaurant health code inspections on hygiene outcomes.

### **2.1 Credence Goods**

Nelson (1970, 1974) points out that goods can be divided into two categories, search goods, and experience goods, where the former is defined by goods with easily evaluated

characteristics, while the latter are goods whose characteristics are not known until after purchase. In a follow-up study, Darby and Karni (1973) present the idea of credence goods, the quality of which may still be unknown even after consumption. They argue that if the evaluation of credence qualities is costly, government monitoring may be necessary to prevent fraud. Other literature on credence goods largely concentrates on experts' fraudulent behavior. Pitchik and Schotter (1987) use a theoretical model to investigate fraudulent behavior in the expert service market. They find that experts report truthfully in a random manner with a mixed strategy equilibrium in the expert-customer game. Wolinsky (1995) examines markets for informed expert services and argues that despite intense competition, there is still room for fraud. He also emphasizes the role of consumer search for various opinions in disciplining experts. Emons (1997) suggests that if consumers can rationally process *ex ante* information about market conditions, they will be able to infer the seller's incentives and the market mechanism may solve the fraudulent expert problem.

## **2.2 Information Disclosure and Expert Behavior**

Information disclosure methods like advertising and labeling are potential options to narrow the information gap between consumers and sellers. When the Kellogg company first launched a cereal commercial campaign in the 1980s claiming that consumption of fiber could reduce the probability of cancer, it challenged a long-standing ban imposed by the Food and Drug Administration (FDA) on the food market preventing them from claiming that their product "...is adequate or effective in the prevention, cure, mitigation or treatment of any disease or symptom."<sup>4</sup> Since the Nutrition Labeling and Education Act was passed in 1990, food labeling has become mandatory. Ippolito and Mathios (1990) show that Kellogg's commercial about the benefit of consuming fiber boosted the cereal's sale overall and caused more firms to shift to producing high fiber cereal. They also notice that before Kellogg's advertising campaign, publicly available scientific information about

---

<sup>4</sup> Code of Federal Regulation. Washington, D.C: U.S. Government Printing Office, 1980.

fiber had hardly drawn any attention from consumers. They conclude that advertising reduces the cost to consumers of acquiring information, and increases total welfare. Sorensen and Leslie (2010) use a large transaction dataset from Starbucks to find the impact of mandatory calorie posting on consumers' purchase decisions. They find that the average calories per transaction fall by 6% when calorie information is posted. They also find that if Starbucks and Dunkin Donuts are located in close proximity to each other, people tend to visit Starbucks more, because Dunkin Donuts has higher calorie tabs on its products.

There is also a modest empirical literature that shows that a considerable amount of fraudulent behavior exists in markets for expert services, due to asymmetries of information. Hubbard (1998) explores the incentives of experts' in the automotive repair services industry. He finds that inspection failure rates are more than twice as high when inspections are completed by state officials, relative to when they are conducted by private firms. Moreover, the probability of inspection failures is slightly lower when conducted by firms with close geographic competitors. Wambach and Sulzle (2005) examine the impact of variations in insurance on the amount of fraud in a physician-patient relationship. They find that a higher coinsurance rate may lead to less fraud in the market and a lower probability of patients' searching for second opinions. Levitt and Syverson (2008) find that real estate agents tend to invest more time and effort to secure higher prices on their own properties, relative to their customers'.

### **2.3 Hygiene: Policies and Impacts**

Jin and Phillip (2009) examine the changing restaurant hygiene inspection policy in Los Angeles County. They find that reputation incentives effectively motivate restaurants to maintain good quality hygiene and that chain affiliation provides reputation which allows franchised units to free-ride. For instance, before the new mandatory grade cards policy was launched when the inspection score was not available to the public, franchised restaurant units usually had worse hygiene quality scores than units under the same chain which were

directly owned by the company. The new policy enables customers to observe the hygiene quality of restaurants in Los Angeles and leads to a significant improvement in restaurant hygiene quality in the city. It also eliminates the score difference between franchise-owned restaurants and directly-owned ones. In another paper, Jin and Leslie (2003) show evidence that the mandatory grade card policy in Los Angeles County actually reduces the overall foodborne illness rate in Los Angeles hospitals. Ho (2012) questions Jin and Leslie's findings on foodborne illness and uses hygiene inspection data from San Diego and New York City to argue that the targeted hygiene information transparency is flawed in its implementation. He concludes that in San Diego, the restaurant hygiene inspection generates score inflation while in New York City, the hygiene inspection generates large score variations. Meanwhile, Wong et al. (2015) suggest that New York City hygiene inspection does improve the hygiene conditions and in a telephone survey, they find about 90% of adults approved of New York City's letter grade program. However, Wong et al. (2015) fail to capture the changing and inconsistent behavior of restaurant owners when they face a new policy.

In this paper, I make a contribution to the literature on the impact of hygiene inspections by investigating a hitherto under-examined health inspection regime change in New York City. In the next section, I give a brief background on this new hygiene inspection policy and how it affects restaurant behavior.

### **3 Background**

The New York City Department of Health and Mental Hygiene (DOHMH) inspects about 24,000 restaurants every year to monitor restaurant compliance with city and state food safety regulations. Restaurants are inspected approximately once every 300 days, and inspections are not announced ahead of time. Restaurants are issued points for each violation of the New York City health code, with more egregious violations earning more points (thus, a restaurant with no violations will receive a score of 0). If the number

of points earned exceeds a pre-defined threshold, then the inspector will perform a re-inspection at an unannounced time, approximately 7 to 30 days later. This schedule allows the restaurant owner to anticipate the re-inspection with some accuracy.

In July 2010, the DOHMH launched a new inspection policy, which made two major changes to the inspection process. First, they created a letter grade which is mapped from the original raw inspection score. Restaurants with a score between 0 and 13 points earn an A letter grade, scores of 14 to 27 receive a B, and those with 28 or more receive a C. Inspectors give restaurants their grade cards when the inspections are finished.

The second major change to the policy involves giving consumers easier access to hygiene inspection information. Inspection results are made available in three different ways: a searchable Web site, a free smartphone app, and a new requirement that restaurants post letter cards on their storefront. While the website and smartphone app give detailed information about the inspection, and allow users to filter restaurants by ZIP code, cuisine type, and grade, the letter cards merely show A, B, or C. However, this third requirement is the most visible and well-known change to the health inspection code.<sup>5</sup>

When the grading initiative was being developed, industry representatives stressed the need to ensure that grades should be assigned fairly. The DOHMH addressed these concerns in several ways. First, the DOHMH provides a variety of training and practice materials for restaurant owners to learn how to earn an A grade. Additionally, they offer an online course on how to maintain their restaurants in a satisfactory condition.<sup>6</sup> A self-inspection worksheet and a guide to condition levels are also available online for restaurant owners to perform a self-examination.<sup>7</sup> Moreover, training sessions are offered for restaurant supervisors to earn a food protection certificate so they can conduct self-monitoring. The DOHMH also suggests that a supervisor should be on duty whenever the restaurant is receiving or preparing food or is open to the public. All of these materials are available in English, Spanish, Chinese, and Korean.

---

<sup>5</sup> Chapter 23 of Title 24 of the Rules of the City of New York (2010)

<sup>6</sup> available at <http://www.nyc.gov/html/doh/html/services/hanyfood-online.shtml>

<sup>7</sup> available at <http://www.nyc.gov/html/doh/html/food/food-service-inspection.shtml>

Additionally, the DOHMH implemented a dual inspection system that allows restaurants to improve before receiving a final grade. If a restaurant receives a score of 13 or less, it will be assigned an A card, which must be posted immediately. Then it will not be inspected again within the approximately 300-day inspection cycle. If the restaurant does not initially receive an A level score i.e., it gets more than 13 points, the inspector records the score but will not give a grade card. The inspector will return at a later date, usually within 30 days, to conduct a re-inspection. The purpose of this re-inspection is to give the restaurant a second opportunity to improve its hygiene condition. After the re-inspection, restaurants are required to post the results on their storefronts right away. With less than or equal to 13 points, the restaurant would receive an A card and the inspection cycle ends. With a B or C, the restaurant will receive a corresponding grade card if it agrees with the result.

A restaurant can choose to reject the inspection result. If so, it gets a “grade pending” card which must still be posted while the restaurant awaits a hearing with the Office of Administrative Trials and Hearings Health Tribunal to contest the score.<sup>8</sup> One additional complication is that if a restaurant receives a grade of C (> 27) in the initial inspection, then regardless of its result in the re-inspection, the inspection cycle for it is shortened, which means this restaurant is inspected more frequently.

There are several additional small differences between the pre-policy and post-policy periods. First, in the post-policy period, restaurants are re-inspected when they do not initially receive an A grade. However, in the pre-policy period, restaurants would receive a re-inspection if their initial inspection score was 28 points or higher (Wong et al. (2015)). Second, the DOHMH has changed the description and reordered some violation codes in the new policy. However, the Department of Health has claimed that the whole inspection process has remained the same.<sup>9</sup> Figure 1 depicts this new inspection process.

---

<sup>8</sup> The hearings may or may not change the original inspection result. The dataset used in this study does not contain the information on how long restaurants wait for their hearings, and whether the hearing changes the result. Therefore, this paper only focuses on the initial inspection and the following re-inspection.

<sup>9</sup> Chapter 23 of Title 24 of the Rules of the City of New York (2010)

## 4 Data

The hygiene score dataset of restaurants in New York City spans 2006 to 2014 and includes information on restaurant name, restaurant location (an approximate address), inspection date, and cuisine type.<sup>10</sup> The hygiene inspection data was merged with historical temperature data from weather.com, which provides daily information at the borough level between 2006 and 2014.

Table 1 presents the average scores for all inspection records in the pre- and post-policy periods. In the pre-policy period, the overall average hygiene score is about 18, while in the post-policy period, the average score improves by about 2 points. Table 2 then breaks the scores into initial inspection and re-inspection categories. We see that in the pre-policy period, the re-inspection result is only better than the initial inspection result by about 1 point, while in the post-policy period when customers are able to observe the grade, the re-inspection result improves by 4 points in comparison with the initial inspection.

It should be noted, that in the pre-policy period, the dataset does not clearly distinguish between an original inspection and a re-inspection. As mentioned in the introduction, the DOHMH has specific guidelines for the inspection cycle, including the timing of the re-inspections. However, there is evidence, both anecdotal (Nordrum (2014)) and found within the dataset, that these rules have not always been followed, both before and after the policy change. In the pre-policy period, restaurants would be re-inspected if they received a score of 28 or greater, and that re-inspection would occur around one month later (Wong et al. (2015)). If restaurants received a score below 28, they would not be inspected again for approximately one year. However, in the dataset, approximately 40% of the observations are restaurants with a score below 28 still getting re-inspected within 60 days after their initial inspection. These are treated as re-inspections in the analysis that follows.

In the post-policy period, the dataset does indicate whether an inspection is an initial inspection or a re-inspection. However, this indicator tends to also be inconsistent with the rules laid out by the DOHMH. Recall that the criterion for being re-inspected in the

---

<sup>10</sup> The dataset can be found in New York City Open Data website: <https://nycopendata.socrata.com/>

post-policy period is having a score above 13. Again, if a restaurant receives a score of 13 and below, it would not be inspected again for approximately one year. However, many restaurants which received a score of 13 or below were re-inspected within 60 days, and these inspections were marked as re-inspections. These unusual inspections have also been documented in news reports.<sup>11</sup> These irregular observations are kept for analysis and treated as re-inspections in the dataset.

Figure 2 shows the score distributions of initial inspection and re-inspection before and after the new policy. The dotted lines in panel (a) and panel (b) indicate a score of 28. The dotted lines in panel (c) and panel (d) indicate scores of 13 and 27, respectively. We can see the corresponding spikes, which are the thresholds that separate grades A, B, and C. The spike is much more salient in the re-inspection in the post-policy period, panel (d). In the pre-policy period, the score distributions between initial inspections and re-inspections have similar shapes, despite a large volume of zeros during re-inspection in the pre-policy period. Nevertheless, there are no conspicuous spikes in the pre-policy period.

In the next section, I present a baseline model to examine this new policy effect; then I check the policy outcomes by investigating consumer complaints before and after the policy.

## 5 Policy Effect Check

### 5.1 Baseline Model

To estimate the impact of this new letter grade policy, I include a binary variable *Policy* to capture the policy effect; I also include another binary variable *Reinspect* to separate the effect of initial inspection and re-inspection. The baseline model is as follows:

$$\begin{aligned}
 Score_{it} = & \alpha + \beta_1 Policy_{it} + \beta_2 Reinspect_{it} + \beta_3 Policy * Reinspect_{it} \\
 & + Temperature_{it} + TempSquare_{it} + Seasonality_{it} + c_i + u_j + v_{it}
 \end{aligned} \tag{1}$$

---

<sup>11</sup> See Eli Epstein (2012) for more detailed examples.

where  $Score_{it}$  is the hygiene inspection score of restaurant  $i$  at time  $t$ ;  $Policy$  is a binary variable, equal to 1 if the observation is in the post-policy period;  $Reinspect$  is a binary variable, equal to 1 if the inspection is a re-inspection, and equal to 0 if it is an initial inspection;  $Policy * Reinspect$  is the interaction term of these two binary variables.  $Temperature$  is the average temperature (degrees Fahrenheit) on the day when the restaurant was inspected. The temperature is measured at the borough level and included in a quadratic specification ( $TempSquare$ ). Since restaurant hygiene condition could be affected by different seasons, I create a categorical variable  $Seasonality$  in each year, to control the seasonal effect.<sup>12</sup>  $c_i$  denotes the unobservable restaurant-specific time invariant term to control restaurant fixed effect.  $u_j$  is the unobservable time invariant chain fixed effect if the restaurant belongs to a chain.  $v_{it}$  is the error term.

In this baseline model, I use both fixed effects and random effects to check if the policy has caused restaurants to improve their hygiene conditions and whether the re-inspection has made a difference. For the random effects, I cluster the standard errors at chain restaurants level because it is possible that restaurants which belong to one chain share unobservable time-invariant errors,  $u_j$ .

Recall that a lower hygiene score represents a better hygiene condition in a restaurant; thus, I expect the coefficient estimate of  $Policy$  to be negative, which means the new letter grade policy drives more restaurants to maintain good hygiene so that they can earn a lower hygiene score. The coefficient estimate of  $Reinspect$  would also be negative if restaurants improve hygiene conditions for the re-inspection as opposed to the initial inspection, since they can more accurately predict when the inspector will come.

I expect the estimated coefficient on  $Temperature$  to be positive, because Food quality may suffer at higher temperatures. Higher temperatures may also lead to other negative externalities.<sup>13</sup>

---

<sup>12</sup> I treat Mar. to May as Spring; Jun. to Aug. as Summer; Sep. to Nov. as Fall; Dec. to Feb. as Winter.

<sup>13</sup> Cohn and Rotton (2000) document that a higher temperature leads to a higher crime rate. Ranson (2014) also finds that temperature has a strong positive effect on criminal behavior.

## 5.2 Results and Discussion

Table 3 presents the baseline model results of equation (1). Column (1) includes restaurant the fixed effects,  $c_i$  and column (2) uses the random effects for  $c_i + u_j + v_{it}$ , with errors clustered within each chain  $j$ . Both specifications control for year and seasonality effects. The fixed effects results show that the estimated coefficient of *Policy* is about -1.6, implying that the new letter grade policy improved the initial inspection score by about 1.6 points. *Reinspect* is significant with an estimated coefficient of -5.4, while the estimated coefficient of the interaction term, *Policy \* Reinspect* is 0.9. Therefore, in the new policy period, restaurants improve their hygiene score during the re-inspection by more than 6 points, versus 5 points in the pre-policy period. The random effects estimation shows consistent results with only slightly different magnitudes. It shows that in the post-policy period, restaurant hygiene improves by almost 5 points during the re-inspection, versus 2 points in the pre-policy period.

*Temperature* is not significant, while the squared term *TempSquare* is significantly positive with an estimated coefficient of 0.03 in both specifications.<sup>14</sup> It means if the inspection is conducted at a temperature of 90 degrees Fahrenheit, the hygiene score is about 2 points higher than the one conducted at 40 degrees.

The fixed effects show that the improvement in terms of hygiene scores after this new policy is small (only 1.6-point improvement for the initial inspections). It is because restaurants expect a re-inspection opportunity following the initial inspection and initial inspection is hard to anticipate so that there is little incentive for restaurants to keep good hygiene conditions for the initial inspections. This result indicates that for the most part of the inspection cycle, restaurants may not keep good hygiene condition. In other words, the overall restaurant hygiene may only improve during the period when restaurants are expecting a re-inspection. Therefore, the overall restaurant hygiene in New York City may not have a significant improvement after this policy.

---

<sup>14</sup> In the regression, I use  $TempSquare/(100)$  to lower the magnitude of the quadratic form of temperature.

### 5.3 Policy Outcomes Check

One way of examining the effectiveness of this letter grade policy is to check whether it has reduced the occurrence of foodborne illnesses in New York City, as this is one of the ultimate goals of the policy (Wong et al. (2015)). The empirical challenge is that it is almost impossible to identify the link between foodborne hospitalizations and restaurants overall hygiene results. To solve this problem, I use 311 call data from New York City to check whether the number of customer complaints has declined because of this new policy.<sup>15</sup> Each food establishment complaint record has a specific restaurant address and a complaint reason. The reasons for complaint encompass everything related to restaurant hygiene, ranging from spotting rodents or insects to catching workers who handle food with bare hands. This data exists from January 2010 until August 2015.

Table 4 displays the absolute number of complaints and the number of restaurants in the sample. The overall complaints increase over the years from 2010 to 2015 as well as the number of restaurants. However, note that the number of restaurants may not truly reflect the real increase of restaurant establishments, because the DOHMH may increase its inspection areas over the years. Figure 3 shows the adjusted distribution of total customer complaints each month from 2010 to 2015. The adjusted number of complaints is equal to the total number of complaints divided by the total number of restaurants in the sample within each year. The figure does not show a significant decline of complaints, instead, it shows that customers complain more in the summer, confirming the baseline regression results, i.e., temperature increases imply hygiene decreases.<sup>16</sup>

Since there is not strong evidence to support an overall improvement in restaurant hygiene in New York City after this new policy, in the following section, I present evidence of restaurant cheating behavior as a reason for why this policy fails to improve restaurant hygiene. Moreover, I use a model to test whether there are regional differences in restaurant

---

<sup>15</sup> New York City's 311 line is a centralized information line for city services. Residents in New York City call 311 for all kinds of complaints, including food establishment complaints.

<sup>16</sup> The relatively smaller number of adjusted complaints in the year of 2012, 2013, and 2014 is because a significant increase in the number of restaurants in the sample.

hygiene scores before and after the policy.

## **6 Restaurants Cheating and Regional Differences**

Since the timing of the re-inspection is much easier to anticipate and prepare for, than the initial inspection, it is possible that restaurants only maintain a good hygiene condition when they are waiting for the re-inspection, but revert to poor hygiene conditions once they receive A cards. Table 5 presents the hygiene inspection records of one typical restaurant. We can clearly see that the restaurant hygiene score jumps back and forth, depending on the inspection type. It shows that every time the restaurant receives an initial inspection, it receives a B or C level score, but after about 30 days, it receives an A. Since keeping a good hygiene condition is costly, and subsequent initial inspections are at least 100 days later, once restaurants receive an A, they can enjoy the reputation benefits for the rest of the inspection cycle. This leads to a moral hazard situation; once a restaurant's reputation is assured by the A grade, it has no incentive to maintain a good hygiene condition if consumers are not able to observe it. In the next subsection, I check the number of restaurants which follow this cheating behavior.

### **6.1 Restaurant Hygiene Score Check**

In order to check how many restaurants are following this pattern, I create a cheating ratio; the cheating ratio for each restaurant is its number of A letter grades earned only in the re-inspection over the total number of A's. For example, the cheating ratio of the restaurant in Table 5 is 100%. This restaurant has only 6 observations; however, all of the observations are pairs of initial inspection and re-inspection, and all of the re-inspection results are A grades. I also calculate the percentage of A's in each restaurant (total number of A's / total number of inspections). Note that if the cheating ratio is zero, it does not mean that the restaurant earns all its A grades in initial inspections, but simply means it

never earns A's in re-inspections.<sup>17</sup>

Figure 4 displays the histograms of these two numbers (percentage of A's and cheating ratio). The white bars represent the cheating ratio, while the green bars represent the percentage of A's. It shows that about 20% of restaurants earn half of their A's in re-inspections (cheating ratio=0.5), while 8% of restaurants earn all of their A's in re-inspections (cheating ratio=1). For these restaurants with the cheating ratio of 1, their A cards are all earned from re-inspections right after their initial, poor inspections. None of them earn an A directly in the initial inspections. The green bars show that about 12% of restaurants earn A's in half of their inspections, while there are a number of restaurants which earn very few A's through their inspections (percentage of A's is less than 0.2).

It is fairly reasonable to suspect that when consumers observe A cards, it is really just an indicator that the restaurant maintained A-level hygiene quality for the 30 days between the initial inspection and the re-inspection. Once restaurants receive an A card, there are no incentives for them to maintain the same level of hygiene quality. This is why we see so many restaurants jumping back and forth between poor scores and very good scores. The intention of giving restaurants a second chance to improve their hygiene is clearly being taken advantage of by most of the restaurants.

In the following subsection, I present a model to test whether this cheating behavior differs across different regions.

## 6.2 Regional Difference Model

I simplify the model of Jin and Leslie (2009) to test whether there are regional differences in hygiene scores due to consumer learning or other unobservable regional factors. I define information  $I_{ij}$  as a measure of information learning ability that consumers have for restaurant  $i$ 's hygiene in region  $j$ . It is a function of the degree of consumers' repeated visits  $r$  for restaurants in region  $j$ ,  $(r_j)$ . I assume that restaurants in region  $j$  face the same

---

<sup>17</sup> For example, if a restaurant earns B grades in all of its re-inspections, the cheating ratio of this restaurant is also zero.

consumer learning ability so that  $I_{ij} = I(r_j)$ .

A restaurant's hygiene decision is based on its marginal revenue of hygiene,  $MR(S_i, I(r_j), k_j)$  and the marginal cost of hygiene,  $MC(S_i, k_j)$ .  $S_i$  is the hygiene score, while  $k_j$  captures all other regional unobservable factors that might affect restaurant hygiene. In the pre-policy period when restaurant hygiene information is not available to consumers, consumer learning may only have a small impact on restaurant hygiene so that it is expected to be the same across regions. However, when hygiene inspection grades are posted on restaurant windows, consumer learning could make a difference across regions.

The function form for restaurant hygiene score  $S_{ij}^*$  is as follows:

$$S_{ij}^*(Policy = 0 \& Reinspect = 0) = a_1 r_{j00} + a_2 k_j \Rightarrow \text{Initial Inspection in Pre-Policy Period}$$

$$S_{ij}^*(Policy = 0 \& Reinspect = 1) = a_1 r_{j01} + a_2 k_j \Rightarrow \text{Re-inspection in Pre-Policy Period}$$

$$S_{ij}^*(Policy = 1 \& Reinspect = 0) = a_1 r_{j10} + a_2 k_j \Rightarrow \text{Initial Inspection in Post-Policy Period}$$

Define  $\alpha_j = a_1 r_{j01} - a_1 r_{j00}$ , i.e., the difference between re-inspections and initial inspections within region  $j$ , and  $\beta_j = a_1 r_{j10} - a_1 r_{j00}$  as the difference between pre-policy and post-policy periods in the initial inspections within region  $j$ . Note one important assumption here is that letter grades affect consumer learning across regions but do not change other regional factors  $k_j$ .  $r$  and  $k$  are not directly observable, but in the empirical analysis,  $a_1 r_j + a_2 k_j$  is the regional fixed effect in region  $j$ . If there is no difference in consumer learning across regions and the re-inspection does not update consumer information in the pre-policy period, then for all regions  $j$ ,  $\alpha_j = \alpha$ . Therefore,  $a_1 r_{j01} - a_1 r_{j00}$  should be constant across all regions ( $a_1 r_{01} - a_1 r_{00}$ ). On the other hand, if there is a difference in the degree of consumer learning across regions in the post-policy period, then  $\beta_j \neq \beta$ . Therefore, the empirical exercise is to test whether  $\alpha_j$  and  $\beta_j$  are constant across regions.

Table 6 presents the results across two regional definitions: 5-digit ZIP code areas and PUMAs.<sup>18</sup> There are 208 ZIP code areas and 55 PUMAs in the sample. The null hypothesis for the first row is that  $\alpha_j = constant$ .  $\alpha_j$  is the estimated coefficient of each

<sup>18</sup> PUMA (Public Use Microdata Area) is a geographic definition defined by US Census Bureau.

region  $j$  interacted with *Reinspect* in the pre-policy period. Both ZIP code level and PUMA level tests do not reject the null hypothesis. The second row displays the regional fixed effect tests in the initial inspections between pre-policy and post-policy periods.  $\beta_j$  is the estimated coefficient of each region  $j$  interacted with *Policy* in initial inspections. Both geographic definitions reject the null hypothesis with 99 percent confidence. The tests confirm that the letter grades reduce the hygiene asymmetric information so that the once unnoticeable regional difference before the policy becomes conspicuous after the policy.

In summary, after the policy, it is re-inspection, not the initial inspection, that incentivizes restaurants to earn the hygiene reputation by getting a better grade. The tests indicate that there are regional differences in reputation incentives. In the following section, I examine whether the hygiene reputation differences as proxied by hygiene scores exist between different types of restaurants—chain and non-chain.

## 7 Chain Reputation Effect

### 7.1 Empirical Strategy

In this section, I present different models to estimate hygiene score differences between chain and non-chain restaurants in different policy periods.

#### 7.1.1 Chain Effect Model

Before the inspection information became publicly available, franchised restaurants tended to have worse inspection outcomes than those directly owned by the company (Jin and Phillip (2009)). Accordingly, I check whether chain restaurants have better hygiene than non-chain restaurants.<sup>19</sup> I create a binary variable, *Chain*, equal to 1 if a restaurant belongs to a chain. Correspondingly, the variable *Chainnum* is the total number of

---

<sup>19</sup> In the dataset, I only observe the name of the chain but cannot distinguish whether a chain restaurant is a franchised unit or directly-owned by the company.

restaurants that belong to the same chain in New York City. The model is:

$$\begin{aligned}
Score_{ijt} = & \alpha + \beta_1 Policy_{it} + \beta_2 Reinspect_{it} + \beta_3 Policy * Reinspect_{it} + \beta_4 Chain_{ijt} \\
& + \beta_5 Policy * Chain_{ijt} + \beta_6 Reinspect * Chain_{ijt} + \beta_7 Policy * Reinspect * Chain_{ijt} \\
& + \beta_8 Chainnum_{ijt} + \beta_9 Policy * Chainnum_{ijt} + Seasonality_{it} + c_i + u_j + \nu_{it}
\end{aligned} \tag{2}$$

where the binary variable *Chain* distinguishes chain and non-chain restaurants. Restaurants that belong to the same chain usually share the same level of reputation. If one restaurant got involved in a food poisoning scandal, other restaurants in the chain will suffer. Thus, the number of restaurant units in each chain, *Chainnum*, captures chain externality effect. Since a lower hygiene score represents better hygiene, I expect the estimated coefficient of *Chain* to be negative and its corresponding coefficient estimate of *Chainnum* to be negative if chain restaurants have better hygiene than non-chain restaurants because chain affiliated restaurants share the reputation as a whole and reputation is positively associated with number of establishments (Jin and Phillip (2009)). The model also includes the binary variables *Policy* and *Reinspect*, and seasonal control.

$c_i$  is the time invariant restaurant fixed effect;  $u_j$  is the unobservable time invariant chain fixed effect. Standard errors are estimated by clustering restaurants at the chain level.  $\nu_{it}$  is the error term.

Note that the chain restaurant variable, *Chain*, is time invariant which will be dropped in the fixed effects estimation. Therefore, I use random effects to estimate the effect of chain. One problem in using random effects is that the unobserved restaurant fixed effects  $c_i$  may be correlated with *Chain*. To solve this problem, I use another model in the following subsection.

### 7.1.2 Robustness Check on Chain Effect

To estimate the overall chain effect, I create a binary variable for each chain. There are 91 chains in the sample; therefore, I have 91 chain dummies, representing each chain

in the model. The model is as follows:

$$\begin{aligned}
Score_{it} = & \alpha + \beta_1 Policy_{it} + \beta_2 Reinspect_{it} + \beta_3 Policy * Reinspect_{it} \\
& + \beta_{cn1} k_{cn1} + \beta_{cn2} k_{cn2...} + \beta_{cn91} k_{cn91} + \beta_5 Policy * (k_{cn1} + k_{cn2...} + k_{cn91}) \\
& + \beta_6 Reinspect * (k_{cn1} + k_{cn2...} + k_{cn91}) + \beta_7 Policy * Reinspect * (k_{cn1} + k_{cn2...} + k_{cn91}) \\
& + Seasonality_{it} + c_i + \nu_{it}
\end{aligned} \tag{3}$$

$k_{cni}$  is the binary variable, equal to 1 if this restaurant belongs to chain  $i$ . The expected overall chain effect is approximated by:

$$\omega = w_1 \beta_{cn1} + w_2 \beta_{cn2...} + w_{91} \beta_{cn91} \tag{4}$$

where  $w_i$  is the weight for chain  $i$ . The weight  $w_i$  is the ratio of the total number of restaurants in chain  $i$ ,  $n_i$ , to the total number of all chain restaurants in New York City,

$$w_i = (n_i) / \left( \sum_{i=1}^{91} n_i \right)$$

All the other variables are the same in equation (2), where *Policy*, and *Reinspect* are binary variables and *Seasonality* is the seasonality control. Standard errors are clustered at the individual restaurant level. The estimates are expected not to differ significantly from the estimated coefficients in equation (3).

### 7.1.3 Ordered Logit Model

The grade card policy offers another opportunity to examine this inspection regime by using the ordered logit model. I follow the method proposed by Donald Hedeker and Robert D. Gibbons (1994) which uses Gauss-Hermite quadrature to numerically integrate over the distribution of random effects. The model is as follows:

$$y_{it}^* = X_{it}\beta + v_i + \epsilon_{it} \tag{5}$$

$$y_{it} = \begin{cases} 1(A) & \text{if } y_{it}^* \leq 13 \\ 2(B) & \text{if } 13 < y_{it}^* \leq 27 \\ 3(C) & \text{if } y_{it}^* \geq 28 \end{cases}$$

$y_{it}$  is the categorical variable, equal to 1, 2, or 3, representing the A, B, or C grade respectively;  $y_{it}^*$  is the real hygiene score;  $X$  are other explanatory variables, including the same binary variables in the previous chain effect model: *Policy*, *Reinspect*, and *Chain* and seasonality controls. Note that even though letter grades did not exist in the pre-policy period, I assign letter grades corresponding to inspection scores for comparability across time periods.

I expect the results to be consistent with equation (3), where if chain restaurants have better hygiene, then the likelihood of earning an A grade is higher for chain restaurants. Moreover, if *Reinspect* is significantly negative, then it means more restaurants are likely to get an A grade in the re-inspection, rather than in the initial inspection.

## 7.2 Chain Reputation: Results and Discussion

In this section, I discuss the empirical results of the chain effect from the previous models.

### 7.2.1 Chain Effect Model

Table 7 presents the regression results from separating chain restaurants from non-chain restaurants. Recall that *Chain* is a binary variable and *Chainnum* is its corresponding variable that equals the total number of restaurants that belong to the same chain. The results in column (1) show that chain restaurants have better hygiene than non-chain restaurants by 5 points in the initial inspection before the letter grade policy. The number of restaurants that belong to the same chain also has an impact on restaurant hygiene. The estimated coefficient is about -0.01, which means if one chain increases its franchised

stores by 100, then on average, the hygiene score in its related restaurants will improve by 1 point. Note in the re-inspection after the letter grade policy, chain restaurants still earn a better hygiene score than non-chain restaurants, but the score difference is narrowed to approximately 2 points. It is likely that the reduction in the asymmetric information narrows the gap between the restaurants that was previously due to greater consumer knowledge of chain reputation.

Column (2) displays the full list of variables with additional interaction terms. However, the F-test results show that these interaction terms are not jointly significant. The F-test results are displayed in Table 8.

### **7.2.2 Robustness Check on Chain Effect**

Table 9 presents the weighted sum of coefficients of chain dummies following equation (4). The estimated chain effect in the initial inspection before the new policy is -6.33, with a standard error of 0.19 and is significant at the 1% level. This result is consistent with (but slightly higher than) the random effects estimate of *Chain* in equation (2). It also shows that in the re-inspection after the letter grade policy, the hygiene score difference is narrowed between chain restaurants and non-chain restaurants by about 3 points. Nevertheless, these results confirm that chain restaurants on average have better hygiene conditions than non-chain restaurants before the policy; however, the hygiene difference between chain and non-chain is smaller in the re-inspection after the policy. This is because displaying letter grades equalizes the consumers' knowledge of restaurant hygiene reputation between chain and non-chain restaurants after the policy. We, therefore, observe only a small hygiene score difference between them.

### **7.2.3 Ordered Logit Model**

Table 10 presents the regression results from the ordered logit model. It shows estimates consistent with the chain effect model. Table 11 presents the estimated percentage

of each grade based on the ordered logit regression results.<sup>20</sup> In the pre-policy period, 40% of restaurants are predicted to receive an A grade after the initial inspection, while 49% are predicted to receive an A after re-inspection. In the post-policy period, 47% of restaurants are predicted to receive an A after the initial inspection, but almost 76% are predicted to receive an A after re-inspection. The improvement is about 30%, significantly higher than in the pre-policy period. Moreover, if we just compare the initial inspection results between the two periods, we can see the policy does not improve the percentage of those predicted to receive an A. This is consistent with the baseline model results that shows that the policy has an effect on how restaurants perform during the re-inspection but not on their performance during the initial inspection. Table 12 breaks up the estimated results by types of restaurant. It shows that chain restaurants have a higher probability of receiving an A grade. Specifically, 62% of chain restaurants are predicted to receive an A in the initial inspection before the policy, while that number is only 38% for non-chain restaurants. Further, before the policy, hygiene inspection scores are not greatly improved after re-inspection for both chain restaurants and non-chain restaurants, which only improve by about 5%, and 10%, respectively. In the post-policy period, however, the situation is very different. Although the percentages in the initial inspection are still very close to those in the pre-policy period (68% of chain and 48% of non-chain restaurants predicted to receive an A) the re-inspection results show a large improvement. 87% of chain restaurants are predicted to receive an A after re-inspection, improving by about 20%, while 75% of non-chain restaurants are predicted to receive an A, improving by 30%.

These results are consistent with results from the chain effect model. After the policy, although chain restaurants still have a slight edge in hygiene quality score, the gap between chain restaurants and non-chain restaurants is narrowed significantly. The hygiene quality improvement only occurs when the restaurants are inspected in the re-inspection, not the initial inspection.

---

<sup>20</sup> In this table, I average the value of *Chain* variable. I separate chain and non-chain effect in Table 12.

## 8 Conclusion

The intention of the Department of Health in New York City in launching this new mandatory grade card policy was presumably to improve hygiene and transparency in city restaurants. Thus, by helping consumers observe restaurant hygiene quality, a characteristic which cannot otherwise be easily observed, they hoped to incentivize restaurants to maintain high standards of hygiene. However, as I have shown in this paper, the re-inspection policy offers restaurants an opportunity to rationally exploit the inspection regime. The data shows that what consumers observe most frequently is the result of the re-inspection, for which restaurants only need to maintain good hygiene standards for approximately one month. Moreover, the letter grade policy gives restaurants an opportunity to take advantage of an A grade reputation, which is frequently no longer representative of their current hygiene quality. This situation is analogous to the moral hazard problem. Once restaurants display the A grade on their windows, there are no incentives for them to maintain good hygiene conditions.

The paper also confirms the chain reputation effect in maintaining good hygiene conditions. The chain reputation effect decreased as a result of the new policy since the letter grades reduced the hygiene asymmetric information; however, this reduction occurred only after re-inspection.

In summary, this new policy does not help improve restaurant hygiene; instead, it merely gives restaurants a cheap and easy way to increase their hygiene reputation. If the department can discard the re-inspection opportunity, increase the frequency of inspection, and update the letter grades accordingly, then the hygiene grade posted on the window of a restaurant would better reflect genuine hygiene conditions and consumers would correspondingly benefit from this information disclosure policy.

## References

- [1] Akerlof, George. 1970. "The market for 'Lemons': Quality Uncertain and the Market Mechanism." *Quarterly Journal of Economics* 84(3): 488-500.
- [2] Bollinger, Bryan, Phillip Leslie, and Alan Sorensen. 2010. "Calorie Posting in Chain Restaurants." *American Economic Journal: Economic Policy* 3(1): 91-128.
- [3] Cohn, Ellen G., and James Rotton. 2000. "Weather, Seasonal Trends and Property Crimes in Minneapolis, 1987-1988. A Moderator-Variable Time-Series Analysis of Routine Activities." *Journal of Environmental Psychology* 20(3): 257-272.
- [4] Darby, Michael R., and Edi Karni. 1973. "Free Competition and the Optimal Amount of Fraud." *Journal of Law and Economics* 16(1): 67-88.
- [5] Duggan, Mark, and Steven D. Levitt. 2002. "Winning Isn't Everything: Corruption in Sumo Wrestling." *American Economic Review* 92(5): 1594-1605.
- [6] Emons, Winand. 1988. "Warranties, Moral Hazard, and the Lemons Problem." *Journal of Economic Theory* 46(1): 16-33.
- [7] Emons, Winand. 1997. "Credence Goods and Fraudulent Experts." *The RAND Journal of Economics* 28(1): 107-119.
- [8] Epstein, Eli. 2012. "No Money in a Dirty Kitchen: The Repercussions of NYC's Restaurant Grading System." *The Atlantic*, July 23. <http://www.theatlantic.com/>.
- [9] Greene, William H. 2012. "Econometric Analysis." 7th Edition, Prentice Hall.
- [10] Hedeker, Donald, and Robert D. Gibbons. 1997. "Application of Random-Effects Pattern-Mixture Models for Missing Data in Longitudinal Studies." *Psychological Methods* 1997(2): 64-78.
- [11] Ho, Daniel E. 2012. "Fudging the Nudge: Information Disclosure and Restaurant Grading." *The Yale Law Journal* 122(3): 522-851.

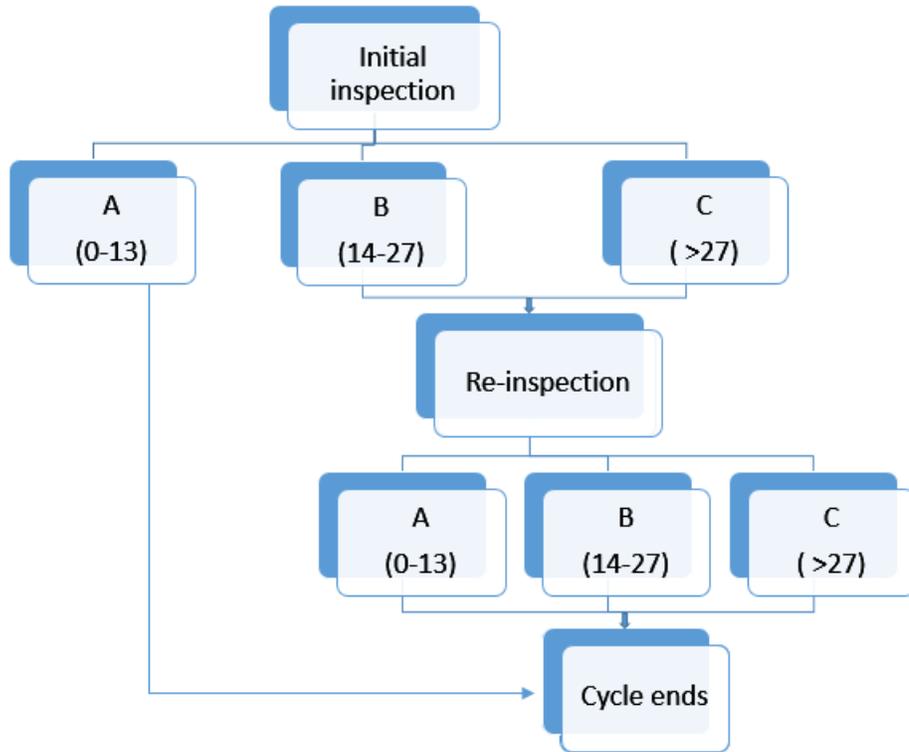
- [12] Hubbard, Thomas N. 1998. "An Empirical Examination of Moral Hazard in the Vehicle Inspection Market." *RAND Journal of Economics* 29(2): 406-426.
- [13] Jacob, Brian A. 2005. "Accountability, Incentive and Behavior: Evidence from School Reform in Chicago." *Journal of Public Economics* 89(5-6): 761-796.
- [14] Ippolito, Pauline M., and Alan D. Mathios. 1990. "Information, Adversiting and Health Choices: A Study of the Cereal Market." *Rand Journal of Economics* 21(3): 459-480.
- [15] Jacob, Brian A., and Steven D. Levitt. 2003. "Rotten Apples: An investigation of the Prevalence and Predictors of Teacher Cheating" *Quarterly Journal of Economics* 118(3): 843-877.
- [16] Jin, Ginger Zhe, and Phillip Leslie. 2003. "The Effects of Information on Product Quality: Evidence from Restaurant Hygiene Grade Cards." *Quarterly Journal of Economics* 118(2): 409-451.
- [17] Jin, Ginger Zhe, and Phillip Leslie. 2005. "The Case in Support of Restaurant Hygiene Grade Cards." *Choice* 20(2).
- [18] Jin, Ginger Zhe, and Phillip Leslie. 2005. "Competition and Disclosure Incentives: An Empirical Study of HMOs." *RAND Journal of Economics* 36(1): 93-112.
- [19] Jin, Ginger Zhe, and Phillip Leslie. 2009. "Reputational Incentive for Restaurant Hygiene." *American Economic Journal: Microeconomics* 1(1): 237-67.
- [20] Jin, Ginger Zhe, Benjamin B. Bederson, Phillip Leslie, Alexander J. Quinn, and Ben Zou. 2016 "Incomplete Disclosure: Evidence of Signaling and Countersignaling." NBER Working Paper No. 22710.
- [21] Jin, Ginger Zhe, and Jungmin Lee. 2013. "Inspection Technology, Detection and Compliance: Evidence from Florida Restaurant Inspections." NBER Working Paper No. 18939.

- [22] Jin, Ginger Zhe, and Jungmin Lee. 2014. "A Tale of Repetition: Lessons from Florida Restaurant Inspections." NBER Working Paper No. 20596.
- [23] Levitt, Steven D., and Chad Syverson. 2008. "Market Distortions When Agents Are Better Informed: The Value of Information in Real Estate Transactions." *The Review of Economics and Statistics* 90(4): 599-611.
- [24] Pitchik, Carolyn, and Andrew Schotter. 1987. "Honesty in a Model of Strategic Information Transmission." *American Economic Review* 77(5): 1032-36.
- [25] Patalano, Heidi. 2015. "This Mouse Doesn't Care About Restaurant's A Grade." *DNAinfo.com*, July 24. <https://www.dnainfo.com/>.
- [26] Nelson, Phillip. 1970. "Information and Consumer Behavior." *Journal of Political Economy* 78(2): 311-329.
- [27] Nelson, Phillip. 1974. "Advertising as Information." *Journal of Political Economy* 82(4): 29-754.
- [28] Nordrum, Amy. 2014. "Q&A: Earning a 'C' on an NYC restaurant inspection." *The Urban Scientist*, May 12. <http://scienceline.org/>.
- [29] Ranson, Matthew. 2014. "Crime, Weather, and Climate Change." *Journal of Environmental Economics and Management* 67(3): 274-302.
- [30] Rob, Rafael, and Tadashi Sekiguchi. 2006. "Reputation and Turnover." *The RAND Journal of Economics* 37(2): 341-361.
- [31] Rutkoff, Aaron. 2010. "Restaurant Makes Best Out of 'B' Grade." *The Wall Street Journal*, September 17. <http://blogs.wsj.com/metropolis/>.
- [32] Sulzle, Kai, and Achim Wambach. 2005. "Insurance in a Market for Credence Goods." *Journal of Risk and Insurance* 72(1): 159-176.

- [33] Taylor, Curtis R. 1995. "The Economics of Breakdowns, Checkups, and Cures." *Journal of Political Economy* 103(1): 53-74.
- [34] McKelvey, Wendy, Wong, Melissa R., and Bailey Matis. 2015. "Letter Grading and Transparency Promote Restaurant Food Safety in New York City." *Journal of Environmental Health* 78(2).
- [35] Wellington, Ben. 2014. "Think NYC Restaurant Grading is Flawed? Here's the Proof." *I Quant NY*, February 17. <http://iquantny.tumblr.com/>.
- [36] Wolinsky, Asher. 1993. "Competition in a Market for Informed Experts' Services." *The RAND Journal of Economics* 24(3): 380-398.
- [37] Wolinsky, Asher. 1995. "Competition in Markets for Credence Goods," *Journal of Institutional and Theoretical Economics* 151(1): 117-131.
- [38] Wong, Melissa R., Wendy McKelvey, Kazuhiko Ito, Corinne Schiff, J. Bryan Jacobson, and Daniel Kass. 2015. "Impact of a Letter-Grade Program on Restaurant Sanitary Conditions and Diner Behavior in New York City" *American Journal of Public Health* 1(20): 1-7.
- [39] Wooldridge, Jeffrey. 2010. "Econometric Analysis of Cross Section and Panel Data." 2nd Edition, The MIT Press.

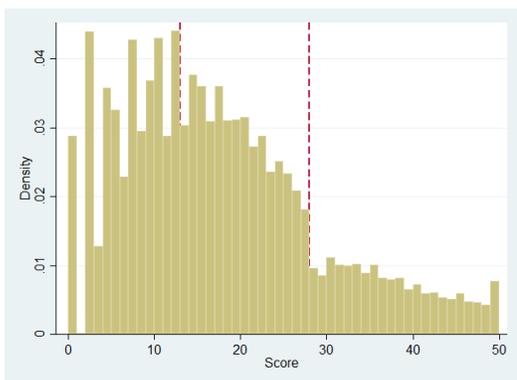
## 9 Figures

Figure 1: Inspection Process

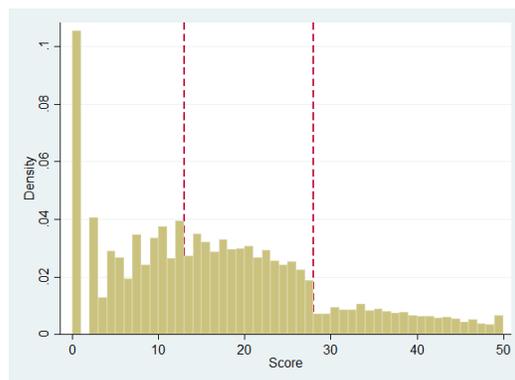


*Note:* This figure shows the inspection process of this new letter grade policy. If a restaurant receives a hygiene score lower or equal to 13 in the initial inspection, it will get an A card. If the restaurant receives a hygiene score higher than 13 in the initial inspection, then it will not get any grade card at the moment, instead, it will have another chance—a re-inspection. No matter how many points this restaurant receives in the re-inspection, it will get a grade card accordingly. However, if this restaurant doesn't agree with the re-inspection result, it can choose to post a grade pending card and has an opportunity to be heard at the office of Administrative Trials and Hearings Health Tribunal

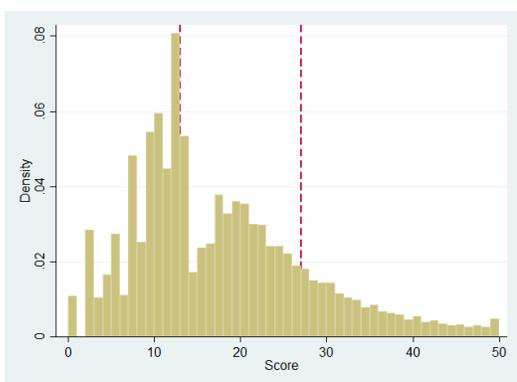
Figure 2: Hygiene Score Distributions



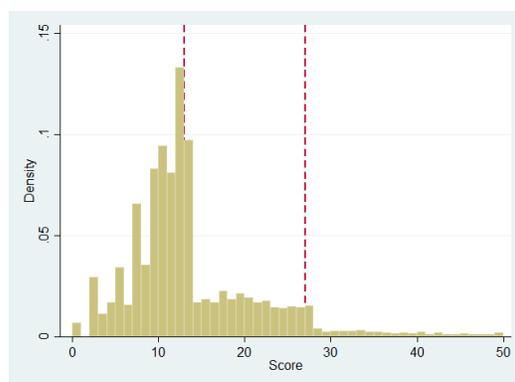
(a) Initial Inspection Score Distribution Before Policy



(b) Re-inspection Score Distribution Before Policy



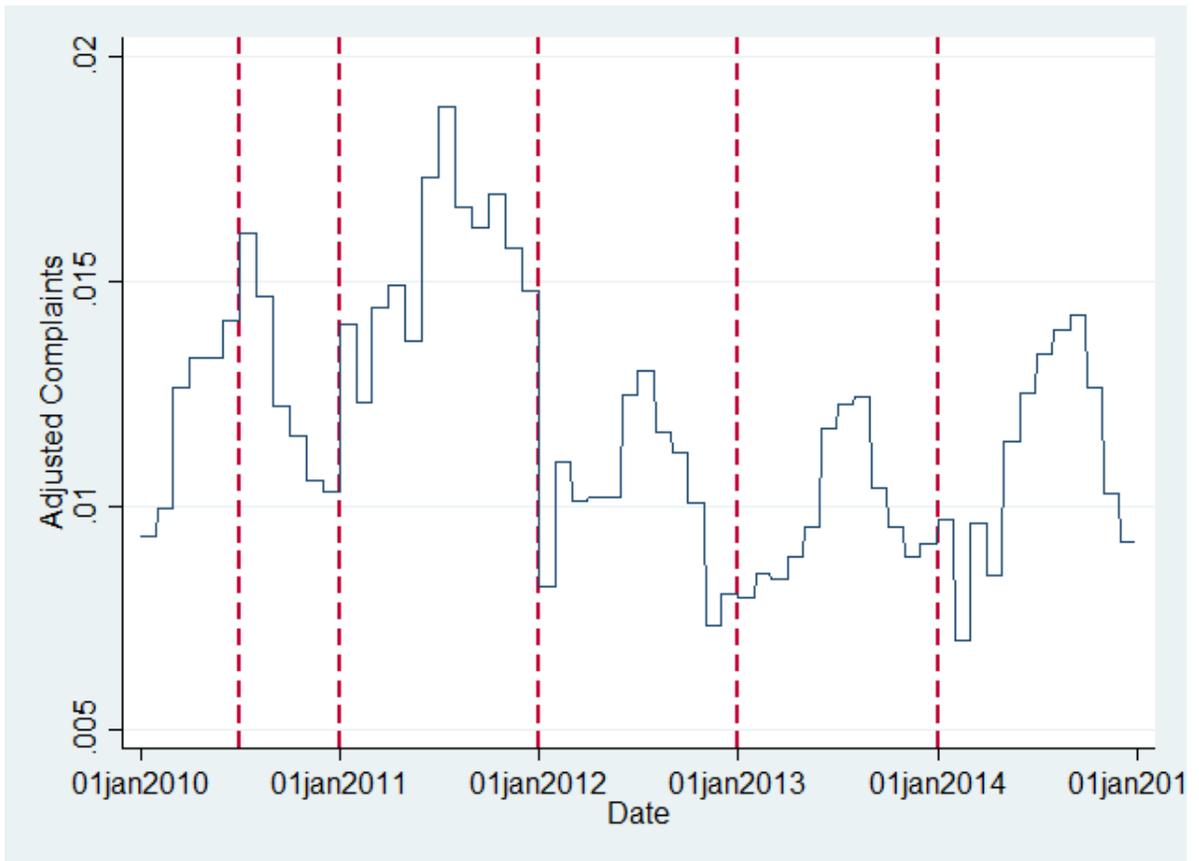
(c) Initial Inspection Score Distribution After Policy



(d) Re-inspection Score Distribution After Policy

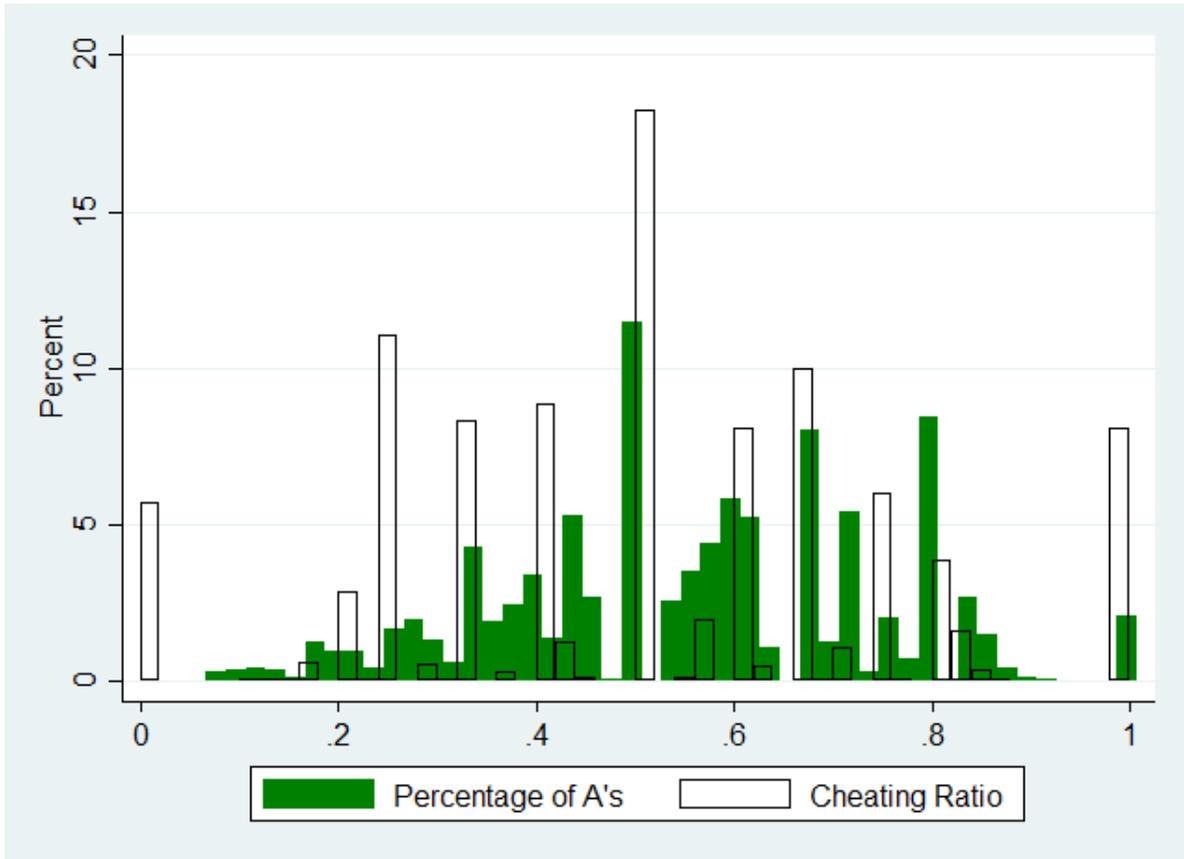
*Note:* This panel of figures shows the distributions of hygiene inspection score before and after the new policy. The dotted lines in panel (a) and panel (b) indicate the threshold score of 28 that trigger a re-inspection before the policy. The dotted lines in panel (c) and panel (d) indicate the threshold scores of 13 (13 or lower is A ) and 27 (27 or lower is B) after the policy, respectively.

Figure 3: Consumer Complaints in New York City



*Note:* The figure shows the adjusted distribution of total customer complaints in each month from Jan. 2010 to Aug. 2015. It is the ratio of the total consumer complaints over the number of restaurants in New York City every month. The first dotted line indicates the day when the new policy is launched. Other dotted lines separate different years.

Figure 4: Histogram of Cheating Ratio



*Note:* The figure shows the histograms of percentage of A's ( $\#$  of A's /  $\#$  of inspections) and the cheating ratio ( $\#$  of A's earned only in the re-inspection /  $\#$  of A's). Green bar represents the percentage of A's and white bar represents the cheating ratio. If cheating ratio is equal to one, it implies that restaurant earns all its A grades in re-inspections.

## 10 Tables

Table 1: Statistics of Inspection Scores Before & After Policy

	<b>Mean</b>	<b>Std. Dev.</b>	<b>Observations</b>	<b>Difference</b>	<b>t</b>
Pre-Policy	19.15	15.68	94425	2.60	45.96***
Post-Policy	16.54	11.82	142054		

*Notes:* This table shows the basic statistics of hygiene inspection score before and after the new letter grade policy. \* $p < 0.1$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$

Table 2: Inspection Scores Summary by Inspection Category

<b>Initial Inspection</b>					
	<b>Mean</b>	<b>Std. Dev.</b>	<b>Observations</b>	<b>Difference</b>	<b>t</b>
Pre-Policy	19.64	15.78	65478	1.32	18.06***
Post-Policy	18.32	12.54	84851		
<b>Re-inspection</b>					
	<b>Mean</b>	<b>Std. Dev.</b>	<b>Observations</b>	<b>Difference</b>	<b>t</b>
Pre-Policy	18.47	15.97	19039	4.42	44.15***
Post-Policy	14.05	9.86	51690		

*Notes:* This table shows the basic statistics of hygiene inspection score by different inspection category before and after the new letter grade policy. \* $p < 0.1$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$

Table 3: Hygiene Score Regression On Policy and Re-inspection

	(1)	(2)
Policy	-1.609* (0.887)	-3.780*** (0.833)
Reinspect	-5.430*** (0.108)	-2.096*** (0.147)
Policy × Reinspect	-0.905*** (0.128)	-2.671*** (0.148)
Temperature	0.012 (0.013)	0.004 (0.013)
TempSquare/(100)	0.030** (0.012)	0.040*** (0.012)
Restaurants Fixed Effects	YES	NO
Observations	220754	220754
$R^2$	0.03	0.03

*Notes:* Dependent variable is hygiene inspection score. *Policy* and *Reinspect* are both binary variables. Variable *Temperature* is the average temperature of the day when the inspection was inspected, using Fahrenheit (°F) unit. *TempSquare* is the square form of temperature. Column (1) uses fixed effects. Column (2) uses random effects with standard errors clustered at chain level. Year and seasonality effect is controlled in both regressions. Standard errors are in parenthesis. \* $p < 0.1$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$

Table 4: Consumer Complaints

	2010	2011	2012	2013	2014	2015
Consumer Complaints	2471	2502	2739	3118	3122	3438
# of Restaurants	10737	13078	15819	18488	19882	.

*Notes:* This table shows the number of consumer complaints from 2010 to 2015 and the number of restaurants in the sample from 2010 to 2014. The number of restaurants in 2015 is missing.

Table 5: Hygiene Score History from One Restaurant

ID	Inspection Date	Score	Interval Days <sup>†</sup>
40280083	2011/3/8	41	.
40280083	2011/4/12	11	35
40280083	2011/8/25	46	135
40280083	2011/9/19	10	25
40280083	2012/1/4	22	107
40280083	2012/1/20	11	16

*Notes:* This table shows the hygiene inspection score records of one restaurant

<sup>†</sup> Interval days are the days between inspections.

Table 6: F-Tests on Regional Fixed Effects

Regional Fixed Effect	ZIP Code Level	PUMA Level
Before Policy: Initial VS. Re-inspection	F Statistic $Prob > chi2$	F Statistics $Prob > chi2$
$\alpha_j = constant$	202.4 0.36	63.0 0.16
Initial Inspection: Before Policy VS. After Policy	F Statistic $Prob > chi2$	F Statistics $Prob > chi2$
$\beta_j = constant$	332 0.00	88.7 0.00

Notes: This table presents the F-tests on the regional fixed effects in two different regional definitions: 5-digit ZIP code level areas and PUMA (Public Use Microdata Area) level areas. There are 208 ZIP code areas and 55 PUMAs in the sample. The null hypothesis for the first row is that  $\alpha_j = constant$ .  $\alpha_j$  is the regional fixed effect difference within the same region  $j$  between initial inspections and re-inspections in the pre-policy period. The F test does not have strong evidence to reject the null hypothesis in both ZIP code level areas and PUMA areas. The null hypothesis for the second row is that  $\beta_j = constant$ .  $\beta_j$  is the regional fixed effect difference in the initial inspection between pre-policy and post-policy period in region  $j$ . The F-test rejects the null hypothesis with 99 percent confidence in both ZIP code level areas and PUMA level areas.

Table 7: Chain Effect on Hygiene Score

	(1)	(2)
Policy	-3.291*** (0.762)	-3.293*** (0.762)
Reinspect	-2.281*** (0.132)	-2.280*** (0.132)
Policy×Reinspect	-2.811*** (0.147)	-2.812*** (0.147)
Chain	-4.964*** (0.563)	-4.667*** (0.624)
Policy×Chain	1.175*** (0.373)	0.889* (0.517)
Reinspect×Chain	1.152*** (0.385)	0.836 (0.539)
Policy×Reinspect×Chain	0.921* (0.476)	0.916 (0.646)
Chainnum	-0.009*** (0.003)	-0.011*** (0.003)
Policy×Chainnum		0.002 (0.002)
Reinspect×Chainnum		0.002 (0.002)
Policy×Reinspect×Chainnum		0.000 (0.003)
Observation	220912	220912
$R^2$	0.05	0.05

*Notes:* Dependent variable is hygiene inspection score. *Policy*, *Reinspect*, and *Chain* are all binary variables. *Chainnum* is the number of restaurant that belong to the same chain. Both regressions use random effects with standard errors clustered at chain levels. Year and seasonality effect is controlled in both regressions. Standard errors are in parenthesis. Column (1) presents results with fewer variables while column (2) displays a full set of variables. An overall significance F-test is conducted and confirms the results in column (1) are valid. \* $p < 0.1$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$

Table 8: F-Test on the Number of Chain Establishments

Joint Hypothesis	F Statistic	<i>Prob &gt; chi2</i>
Policy × Chainnum=0		
Reinspect × Chainnum=0	4.12	0.249
Policy × Reinspect × Chainnum=0		

*Notes:* This F-test is based on the estimated results of equation (2) from Table 7. *Policy* and *Reinspect* are binary variables separating policy periods and inspection types, respectively. *Chainnum* is the number of restaurants that belong to the same chain. The results show that we cannot reject the null hypothesis that these interaction terms are not jointly significant.

Table 9: Chain Effect on Hygiene Score

Overall Chain Effect	$w_1\beta_{cn1} + w_2\beta_{cn2}\dots + w_{91}\beta_{cn91}$
Initial Inspection before Policy	-6.33*** (0.193)
Initial Inspection after Policy	-5.49*** (0.116)
Re-inspection before Policy	-5.84*** (0.376)
Re-inspection after Policy	-3.65*** (0.131)

*Notes:* This table presents the weighted sum of chain effect. The first row shows the weighted chain effect in initial inspection before the policy. The second row displays the weighted chain effect in initial inspection after the policy. The third and fourth row presents the re-inspection before and after the policy, respectively. All of these coefficients are statistically significant at 1% level. Standard errors are in parenthesis.

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

Table 10: Orderd Logit in Panel Model Regression

	(1)
Policy	-0.413*** (0.153)
Reinspect	-0.409*** (0.017)
Policy×Reinspect	-0.964*** (0.021)
Chain	-0.977*** (0.099)
Policy×Chain	-0.005 (0.075)
Reinspect×Chain	0.158*** (0.052)
Policy×Reinspect×Chian	0.026 (0.096)
Observations	220898

*Notes:* Dependent variable is categorical, with value 0, 1, and 2 representing A, B, and C grade. *Policy* and *Reinspect* are both binary variables. The regression uses random ordered logit with standard errors clustered at the chain level. Year and seasonality effect is controlled. Standard errors are in parenthesis.  
 $*p < 0.1$ ,  $**p < 0.05$ ,  $***p < 0.01$

Table 11: Ordered Logit Probability Results Overall

Initial Inspection		Re-inspection	
<b>After Policy</b>	<b>Before Policy</b>	<b>After Policy</b>	<b>Before Policy</b>
<b>Pr(y=A)=46.6%</b>	Pr(y=A)=40.4%	<b>Pr(y=A)=75.9%</b>	Pr(y=A)=48.8%
<b>Pr(y=B)=38.6%</b>	Pr(y=B)=41.7%	<b>Pr(y=B)=19.7%</b>	Pr(y=B)=37.9%
<b>Pr(y=C)=14.8%</b>	Pr(y=C)=17.9%	<b>Pr(y=C)=4.4%</b>	Pr(y=C)=13.3%

*Notes:* This table shows the predicted probability by averaging the chain binary variable for each grade category based on the estimation results from the random ordered logit model in Table 10.

