THE NEW SCHOOL

A game of Cat and Mouse:

Making NYC Restaurants More Consistent in Performance

(Final Project)

Kiril Traykov

NURP 5027: Quantitative Research Design

Professor Richard Hendra

July 20, 2016

At the end of our summer course in “Quantitative Research Design”, some of us may go and have a celebratory dinner in NYC! It is perhaps one of those special occasions when the price of a skirt steak or a beautifully glazed squid would feel justified after a packed summer study of experimental designs. While restaurants in NYC come in every cuisine and form, they are all united by the regulations set by the Department of Health and Mental Hygiene for sanitation and food handling. So while we revel in our skirt steak aroma when it gets served to us, we probably think little if we should have any concerns beyond how well it has been cooked. And why should we – all restaurants in NYC receive a grade of A, B, or a C by the Department of Mental and Mental Hygiene and the one we are at right now has a grade of A!

However, as it turns out, a grade of A is not a reliable indicator – as this paper will show, almost 90% of restaurants earn a grade of A and while they comply at one time with all the regulations for food safety, they wildly abandon them for a while until they have to be re-graded. This creates a moral hazard problem in which NYC diners may be on the losing end.

I. Definition of the Problem

The Department of Health and Mental Hygiene conducts more than 21,000 restaurant inspections in NYC every year[1]. The distribution by borough shows that Manhattan dominates the restaurant scene with more than 8,500 establishments:

Fig.1: Restaurants by borough in 2015:

Borough	Total
MANHATTAN	8,572
BROOKLYN	5,122
QUEENS	4,829
BRONX	2,016
STATEN ISLAND	732
NYC	**21,271**

Although there is no hard-written rule about how often a restaurant may get inspected, the Department strives to conduct an inspection once every 6 months, which is called a “Cycle Inspection”. Data from Full Year 2015[1] reveals that 51% of restaurants (or the majority) indeed got inspected at least twice a year, 34% got inspected at least three times and 20% - at least four times. Some restaurants even got inspected 9 times, but the reasons why they were chosen are unclear. A cursory scan of their scores after the inspection also did not reveal any distinguishing clues.

Such Cycle Inspections constitute 84% of all inspections. The other 16% are for pre-permit inspections, administrative inspections, smoke-free air inspections, trans-fat inspections, calorie-posting inspections, and inter-agency inspections. The focus of this study will be on the Cycle Inspections since they measure the food safety practices that we are concerned about.

[1] All data in this study has been downloaded from the “NYC Open Data” portal available at https://nycopendata.socrata.com/Health/DOHMH-New-York-City-Restaurant-Inspection-Results/xx67-kt59

When a Cycle Inspection occurs, a restaurant is graded against a set of 65 possible violations in two categories – critical and non-critical. Inspectors give penalty points for each violation and the more penalty points a restaurant receives, the worse its safety practices are. Although there are general guidelines how to award penalty points, inspectors have a range they can use based on severity.

Figure 2: Violations per category:

# of Critical violations per category:

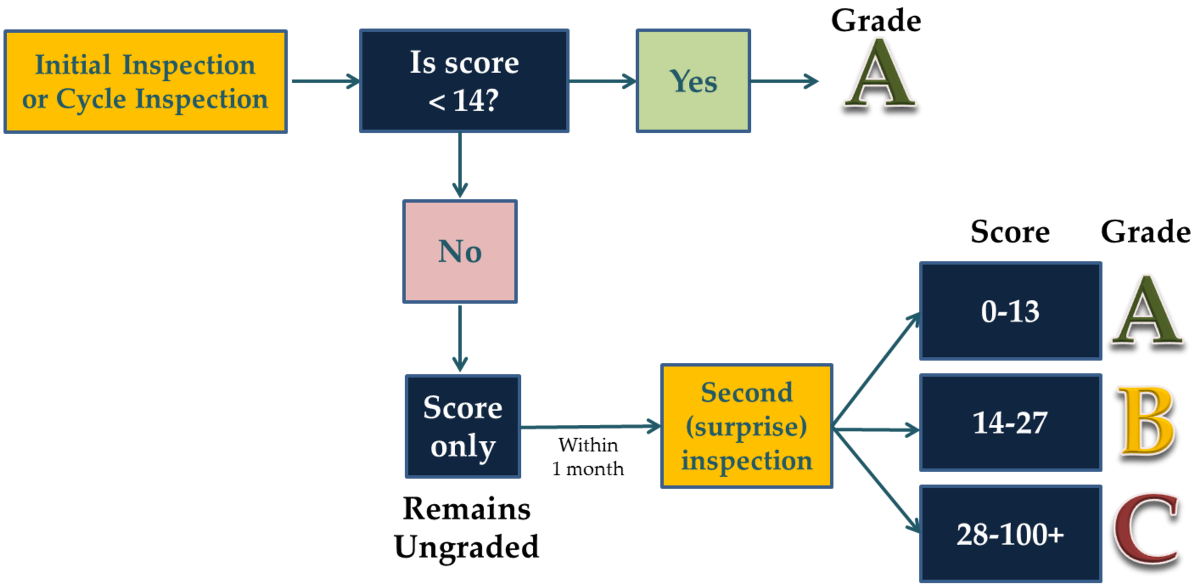
15	Food contamination during prep
10	Food temperature problem
9	Inadequate sanitization
7	Milk/Water/Eggs/Fish/Shellfish
7	Facility not suitable for food handling
1	Obstructing inspector

# of General violations per category:

10	Poor Facility Maintenance
3	Dented Cans/Thawing/Contact Surface
3	Pesticide & Garbage

The penalty points that a restaurant accrues are called a "score" and the process for how a restaurant earns a grade of A, B, or C is as follows:

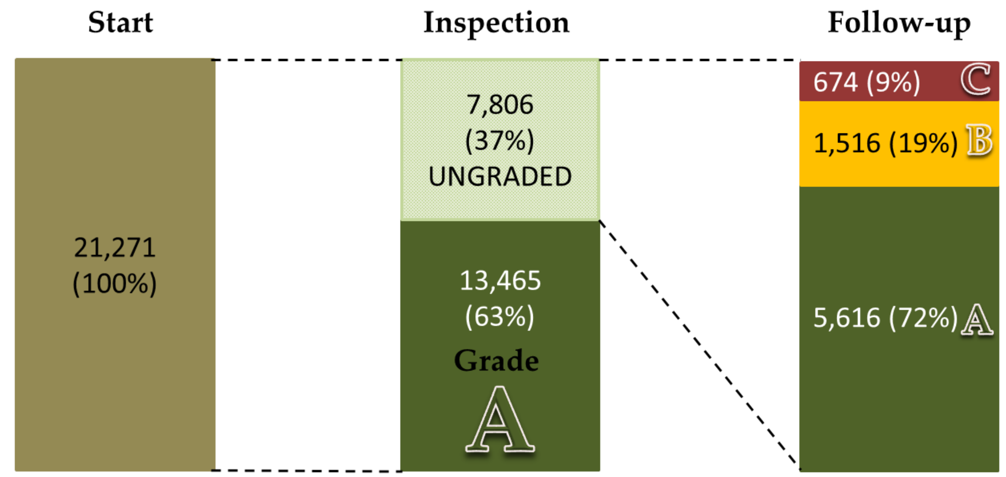
Figure 3: Restaurant Grading Process:



A distinct (and crucial) element in this process is that restaurants are given two tries to earn an A. Thus, even if the first time they are not successful, they will have a second inspection usually within a month when they can redeem their faults from the first inspection.

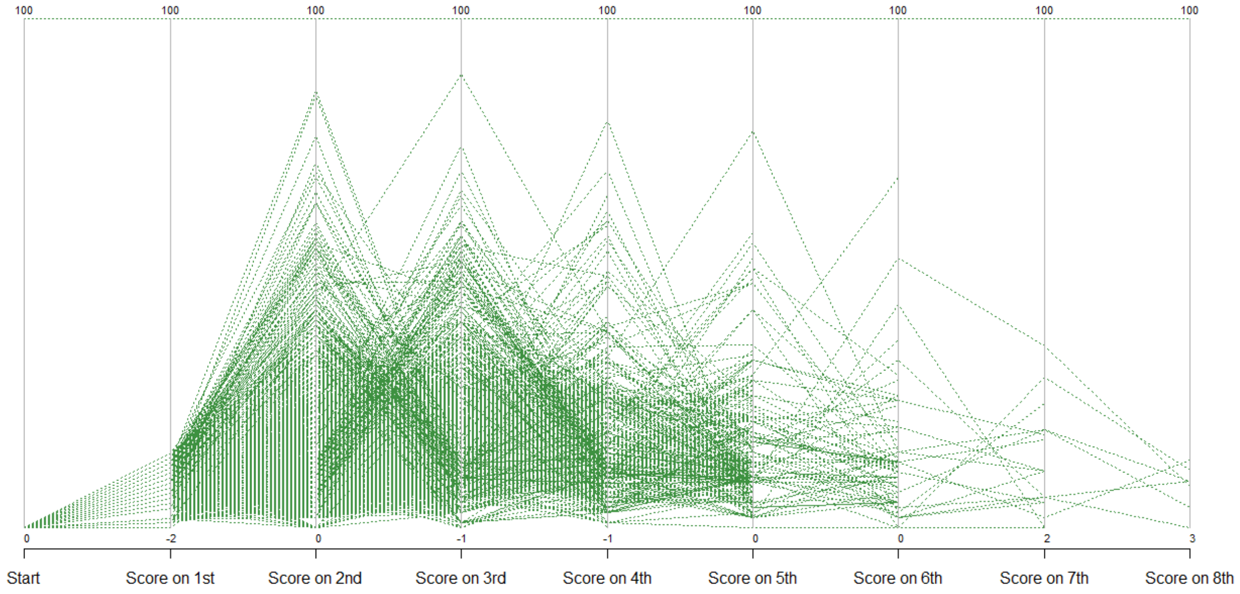
An analysis of the first two calendar inspections in 2015 shows that 63% of restaurants achieve an A on the first inspection by earning a score of less than 14. During the follow-up inspection, when restaurants anticipate that they will be checked, an additional 72% of the remaining restaurants earn an A. So overall, close to 90% of all restaurants eventually receive an A grade, leaving grades B and C as rare.

Figure 4: Passing rates for restaurants in 2015 (first calendar inspection)



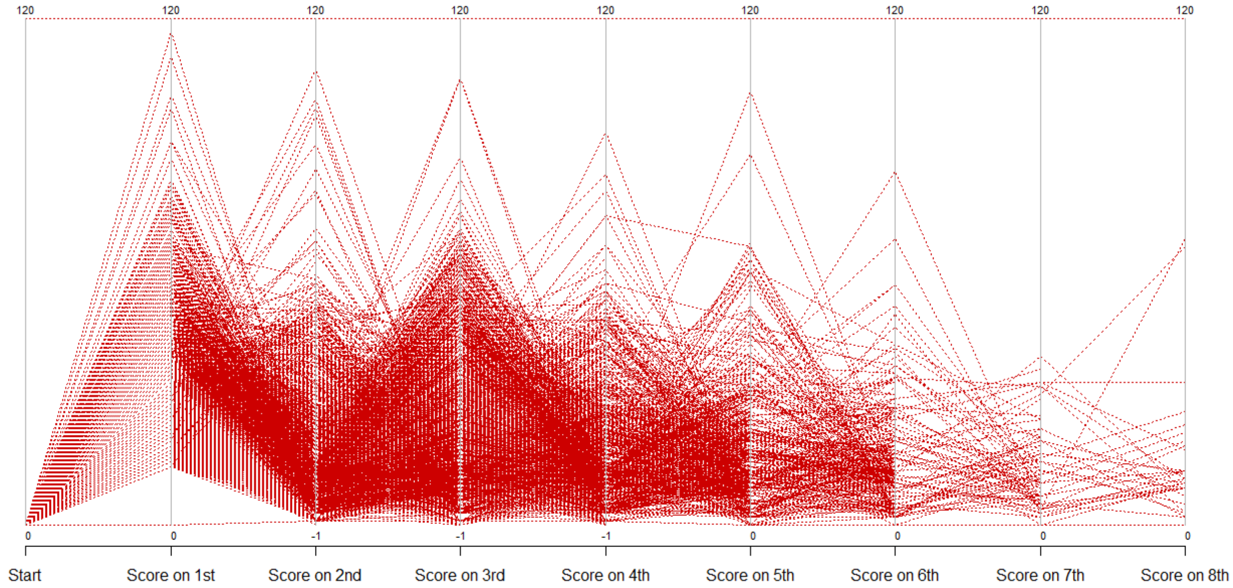
Given this high quality of performance in which almost every restaurant earns an A, one may ask where the problem then may be. The problem is disguised in the lenient decision to give restaurants two opportunities to achieve an A. This only becomes apparent when we graph restaurants' performance for the entire 2015 year. What such an analysis reveals is a concerning yo-yo pattern of scores in which restaurants, which start with an A (13,464 as indicated above), become negligent of their practices until the next inspection warns them that they may lose their A rating. In light of a pending score higher than 13, restaurants take the necessary improvements and appear flawless during the follow-up inspection.

Figure 5: Yearly performance of restaurants which initially earn an A



The same behavior is visible in restaurants whose initial calendar score is higher than 13 (7,806 as indicated above):

Figure 6: Yearly performance of restaurants which initially earn a score higher than 14



Thus, the overall problem is lack of consistency in upholding the sanitation and food safety regulations. The situation almost resembles a game of “cat and mouse” in which restaurants try to circumvent the inspectors by having two chances to earn an A after which they regress back to unacceptable scores of higher than 14, *while* keeping their A grade. For the customers, this means that one can never be sure at which moment of time restaurants truly practice A-grade sanitation standards and the quality of experience is left to the chance of timing of visit.

II. Policy Change and Hypothesis

The erratic behavior of NYC restaurants calls for a re-examination of the existing food safety trainings in place. Currently, the Department of Health and Mental Hygiene requires every restaurant supervisor to pass a “Food Protection” exam after a 15-hour course offered online or in-person. The course is offered in English, Spanish, and Chinese, and if taken in-person, a Korean version is available as well. The exam can only be completed in-person.

Thus, if we do not wish to change the existing “two opportunities” system for restaurant evaluation (which will likely trigger restaurant owners’ outcry and negative publicity), we can look to strengthen the preparedness of restaurant managers to meet and maintain the evaluation criteria. This could be accomplished by a revamp of the “Food Protection” course into a more comprehensive and criteria-focused “Restaurant Management” course.

The rationale for proposing an augmentation to the existing course is because it does not prepare restaurant managers for the entire set of evaluation criteria. It only focuses on food safety, while inspectors check the whole premise, including facilities management, cleanliness, and equipment.

Thus, the curriculum may need a lift to more comprehensive restaurant management content. The design of such a new course could be left to further debate but some options are to:

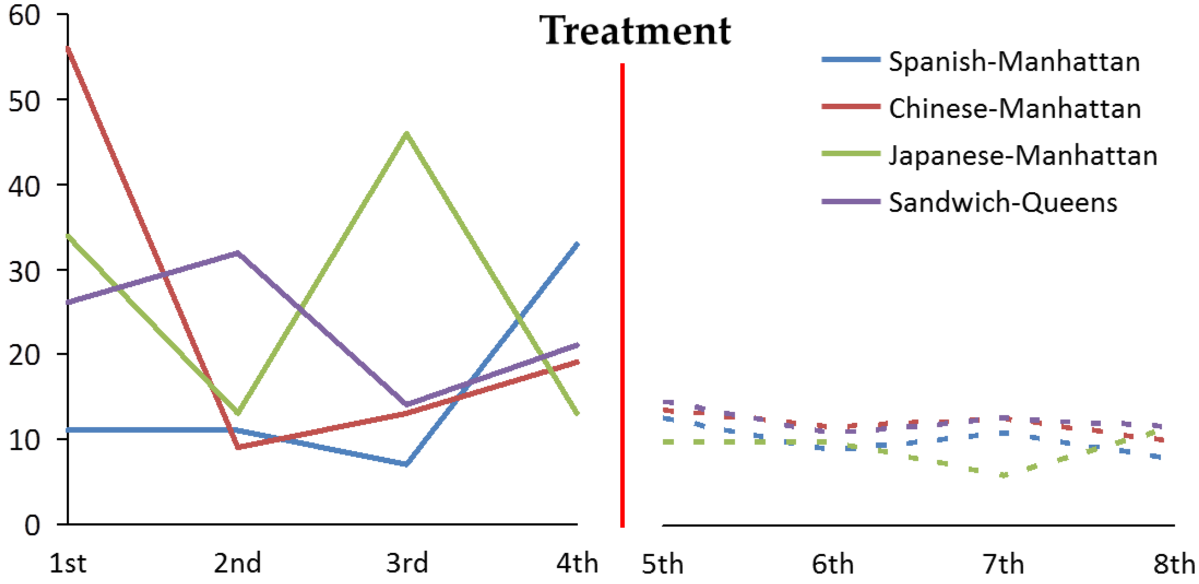
- Offer the course in flexible modules and increase the total number of contact hours to accommodate new content (4 consecutive Fridays for 6 hours each day, for example)
- Require restaurant owners/managers to do a self-evaluation and list corrective actions suitable for their own establishment
- Have an actual inspection as a Final Exam

The hypothesis of the author is that after a curriculum change like that:

Main Hypothesis: Restaurants would exhibit improved inspection performance if their managers are more knowledgeable about proper holistic restaurant management practices. The improvement will come in the form of less variability from inspection to inspection, ideally with scores below 14.

The expected effect / improvement (which will be detailed below in the Analytical Framework section) would result in both a change in slope (reduction in value to less than 1.0, ideally 0.2 and below) and an intercept shift to lower ranges. This is illustrated in the example below:

Figure 7: Expected effect for 4 randomly selected restaurants in NYC (2015 data)



III. Analytical Framework and Data Analysis

In choosing a design framework to evaluate the effect of the treatment, the new Restaurant Management program, one is intuitively drawn to first think whether the Regression Discontinuity (RD) design could apply because of the presence of a defined cut-off (score = 14).

On first look, such an approach would meet the two required assumptions for an RD study:

a. the probability of treatment is discontinuous at the cut off as restaurants would clearly be either to the left or the right of the cut-off point and
b. no alternative interpretations (called causal confounds) exist to determine which restaurants fall on either side of the cut-off.

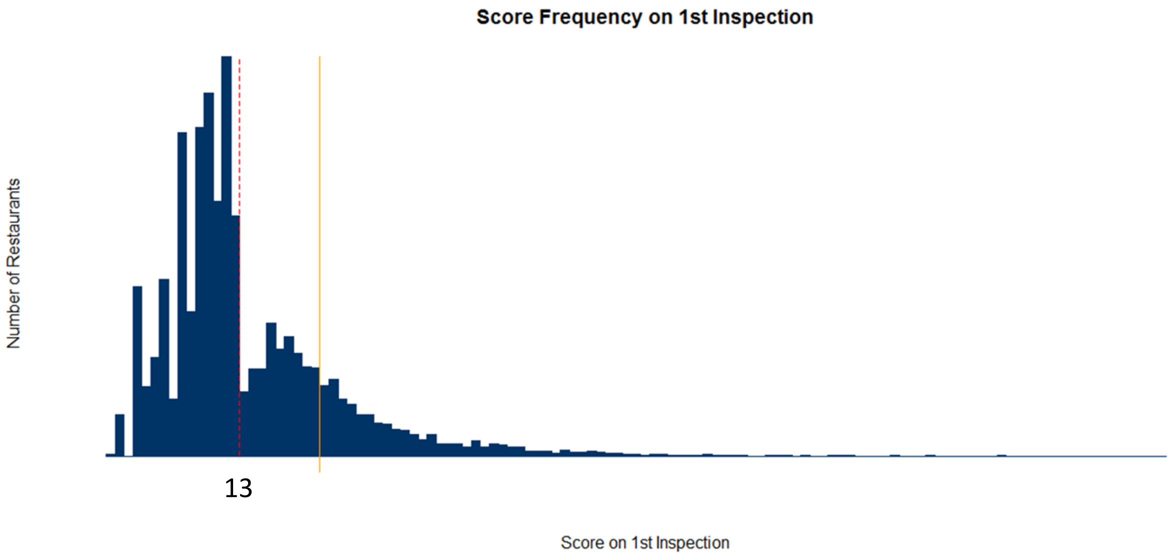
To implement the RD design framework, we could pick a single inspection (e.g. the first calendar one in 2015) as the pre-treatment assignment variable and then plot it against the second inspection results after treatment has occurred. The treatment group in this case would consist of restaurants whose score on the pre-treatment assignment variable is higher than 14. This scenario will ensure no cross-overs (and hence a sharp RD design!) since only restaurants with scores higher than 14 would be in treatment.

However, it should be noted that the treatment group would already have a very strong incentive to improve even without the treatment so effect sizes may be overestimated since the control group will become negligent and commit more violations, while the treatment group would decrease theirs. Thus, there is a force behind the study that causes the desired behavior to occur even without treatment, which may render overoptimistic results.

Another practical problem would be the timeline to conduct treatment because the second inspection for the treatment group will usually occur within a month of the first one. This means that the Restaurant Management course will need to be administered in less than a month! This would be an extremely aggressive and unrealistic deadline.

To add to the complications above, on foundational level, there is strong evidence that there is manipulation in the assignment variable. When we graph the density of scores that restaurants received in the first inspection in 2015, we see a clear drop of density after the cut-off.

Figure 8: Histogram of scores on the first calendar inspection in 2015



From the histogram above, it is clearly visible that inspectors do not give a score of 14 or 15 very often. The hypothesis of the author for this behavior is that inspectors do not wish to hand a score which is just above the A-grade cut-off because they may face remonstration and disgruntlement from restaurant managers/owners, who will be frustrated that their restaurant did not pass for a single point or two. Since inspectors have a range of points for each violation based on severity, they may choose to either keep a restaurant closely under 14 or give a score of 17 or above. Such “manipulation” endangers the otherwise sharp RD design and in conclusion, the author concludes that an RD design is, regrettably, not a fitting analytical framework in this context.

After the RD consideration, we could look into the “golden standard” of experimental design, random assignment (RA), as a possible analytical framework to evaluate the effect of the new Restaurant Management course.

Although we have good starting conditions to conduct an RA experiment – access to baseline data, balanced characteristics, and statistical power – this approach falls victim to its own data richness because the pool of restaurants is more than 21,000 and a real-world cost-benefit analysis would eviscerate any plan to implement it. Since we would have to randomly assign all restaurants in NYC to either a control or treatment group and keep the ratio between them at least 60:40, respectively, this would mean that we need to conduct treatment for at least 8,508 restaurants. This will be very challenging and costly, especially for a program which has no proven effects yet and is just an enhancement of an already existing one. Hypothetically, if this were to still occur, we could also face a high number of “no shows” since it would be difficult to put efforts to individually stay in touch with all restaurants assigned to treatment. This would severely jeopardize our Intention-To-Treat (ITT) sought-after effect, which will become a Treatment on Treated (TOT) effect and would decrease the validity of the study.

In thinking how to remedy the situation and still use RA as a solution, the author thought of two potential variations:

1) Limit the study to restaurants from one borough only or
2) Limit the study to one type of cuisine only

If we choose the first option, we could take Bronx, for example, with its 2,016 restaurants (Figure 1) and form a treatment group with around 800 restaurants. If we go with the second approach, we could pick one of 85 available cuisines (e.g. Italian, Japanese, Sandwich Place, Coffee Shop) and have a more manageable number of control establishments.

Although these RA approaches could still work, the main problem with this design would be external validity. The results that we get from either approach cannot be held as true for the city-wide network of restaurants because of the inherent characteristics within each borough or cuisine. In the case of the borough, we could not assume that conditions for restaurants in Bronx are the same as they are in Manhattan. Everything from square feet to budget to building condition may be inherently different than the same variables in Manhattan. Similarly, food in one cuisine, Italian, for example, could require much different treatment and handling than food in another cuisine, Vietnamese, for example. Thus, results cannot be held valid for all of NYC.

Because of these limitations, we would have to cross the RA design as potential solution in our case.

At this point, we can evaluate what we have in order to guide our pursuit of non-experimental design in the right direction. Our assets are a rich data set and multiple historic observations. This leads us to examine if propensity scores and Interrupted Time Series (ITS) could be the right choice.

Even though random assignment was not suitable in our context, *random selection*, on the other hand may be the place to start. To gain statistical power, we could randomly select 200 from our pool of 21,271 restaurants and then assign a control group to them through propensity scores and measure their post-treatment performance using ITS. The mechanics of this approach could be the winning ones.

In terms of propensity scores, we would have a pool of 21,071 restaurants to find matches for our 200 chosen ones. To ensure we find as close a match as possible, we would approximate restaurants based on 3 co-variates: i) cuisine, ii) borough (or distance if we have a geo-mapping tool), and iii) price range. The first two already exist in our baseline data, while the third one could be available through a third-party API, such as Yelp, for example. All three co-variates are exogenous to the treatment, readily observable, and within the same geographical borders (NYC) as the randomly chosen treatment group, which would satisfy the conditions to go with propensity scores as a holistic approach.

One additional co-variate, which could be useful is “seating capacity” (to understand the size of the restaurant), but data for it, unfortunately, does not exist anywhere.

To further reduce any bias, we would use the matching with replacement method in which one and the same restaurant can be used as a control proxy for two or more restaurants in treatment if it happens to best approximate both restaurants. The author’s expectation, though, is that such an event will be rare and could only occur for rare cuisines if such restaurants get randomly chosen (e.g. Hawaiian, Portuguese, or Nuts Shop, for example).

The propensity score estimation in our study will be a one-to-one match on cuisine and price range and a similarity match on geo-proximity (or nearest-neighbor matching - the closer, the better). We could use a logistic model to estimate the propensity score and discover how different/similar the treatment and control groups are, but we have to note that there are no other narrower co-variates in our data set, which could be used to optimize the propensity score.

Next, we would use the co-variates to check for balance by looking at the distribution of restaurants by borough, cuisine type, and price range. We would also do a T-test or Chi-square test to check if characteristics vary at a statistically significant level. However, because of the one-to-one strength of our matching method, the author expects that such tests will not reveal any disparities. The measurement error is also expected to be small.

Overall, our degrees of freedom will be 600 (200 cases * 3 co-variates), which will give us very good statistical power.

Once we have a control group, we could then turn to ITS as the co-method to evaluate the new Restaurant Management course. From historical data from the Department of Health and Mental Hygiene, we have 4 pre-test time points for each restaurant, which is sufficient to apply this analytical framework and measure against 4 post-treatment outcomes. We hope that the post-treatment measurements will result in both flattening of the slope (which would mean that restaurants have a smoother, more stable performance) and a shift to a lower intercept (which would mean that the violations have fallen in number).

To estimate the mean and the slope for the randomly chosen 200 restaurants better, we could also take all inspection observations that exist on record in the last two years and extend the pre-test series to more than 4. Under this scenario, some restaurants will only have their 4 observations, but others may end up with as many as 10 or more. This will allow us to assess variation in linear change for each restaurant. To the knowledge of the author, there aren't any historical events in the last 2 years in NYC, which could have influenced the scores of the restaurants so history is not a threat in our design.

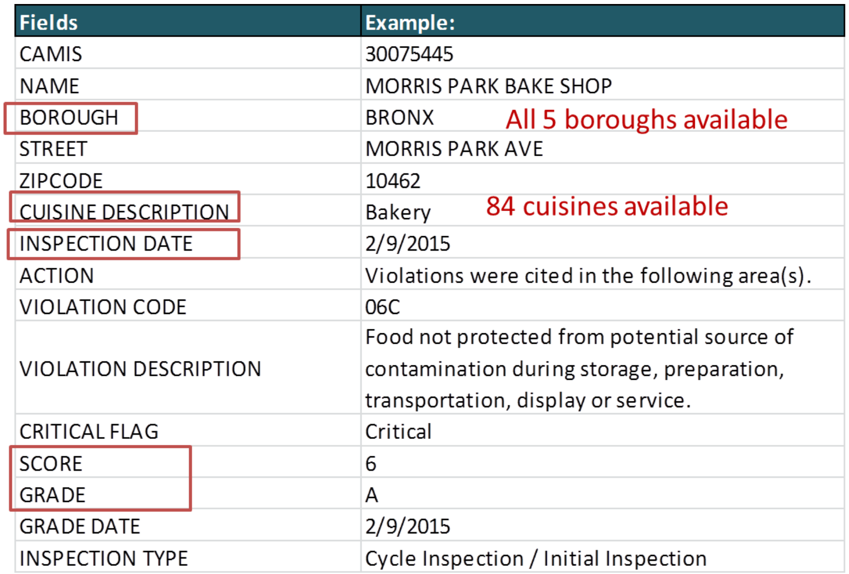
All in all, the propensity scores will give us comparison time series, while ITS will measure the pre-treatment and post-treatment changes for the randomly selected 200 restaurants. Combined like that, we have found our analytical framework to assess the efficacy of the new Restaurant Management course.

IV. Data Collection

Data collection will largely remain intact from what currently exists as a process. The post-treatment data will be collected through normal Cycle Inspections with one important modification: the inspections for our 200 randomly chosen restaurants will be attempted to be done on the same day. This is because the managers of those 200 restaurants will already know that they are part of this new program assessment and during the Restaurant Management course may have exchanged business contacts or struck friendships. Thus, they could possibly warn each other that an inspection is happening if inspections are phased over a period of time. This will create information asymmetry and lead to bias in the results as some restaurants may be more prepared than others. Thus, to avoid all threats of bias, inspections will be attempted to be done in a cohort style.

Once inspection has been performed, the data sheet that will be used for our study design will be in the following format:

Figure 9: Example of a data sheet after inspection



Fields	Example:
CAMIS	30075445
NAME	MORRIS PARK BAKE SHOP
BOROUGH	BRONX
STREET	MORRIS PARK AVE
ZIPCODE	10462
CUISINE DESCRIPTION	Bakery
INSPECTION DATE	2/9/2015
ACTION	Violations were cited in the following area(s).
VIOLATION CODE	06C
VIOLATION DESCRIPTION	Food not protected from potential source of contamination during storage, preparation, transportation, display or service.
CRITICAL FLAG	Critical
SCORE	6
GRADE	A
GRADE DATE	2/9/2015
INSPECTION TYPE	Cycle Inspection / Initial Inspection

All 5 boroughs available

84 cuisines available

V. Learnings from Other Studies and Literature Review

Similar attempts to assess the effect of improving food safety procedures at food service establishments have been encouraging. Most of the existing literature focuses on training the actual food service workers, not only their managers/supervisors, which is the biggest difference between other published studies and our own. In the ideal case scenario, all food service workers would get training in NYC, but the author recognizes that such a case is unrealistic as that would mean training a workforce easily outnumbering 150,000 workers with high rates of turnover. This is why requiring at least the managers of restaurants to improve their understanding of best practices and management skills will be the realistic policy to move forward with and make an impact.

An important consideration that the literature review brought up is the question of what happens when a restaurant changes its manager. In our own study design those questions would be if we should require every new manager to attend the Restaurant Management course and how would we be able to track when changes like that occur. Potentially, manager changes will also have an effect on the inspection score, so knowing and addressing such changes should be part of our study design. At present, the most realistic policy would be to require restaurants to send two trainees to our Restaurant Management course so that even if the manager leaves, there is yet another person at the restaurant who would ensure business continuity and carry on the best practices. Then, in due time, the new manager could get trained as well.

In a 2008 assessment of “Food safety training and foodservice employees’ knowledge and behavior”[2], a group of authors found that overall knowledge and compliance with standards of behavior improved significantly between pre- and post-treatment measurements. The authors start with the revelation that 59% of foodborne illnesses are traced to restaurant operations. They took a random sample of 31 restaurants in 3 Midwestern states to conduct a 4-hour ServSafe® training on food safety (cross contamination, poor personal hygiene, and time/temperature abuse) and observe the results. A total of 402 employees participated, but the design of the study is somewhat questionable. The authors took an initial group of 242 employees to conduct a pre-test and then administered the training to a different set of 160 employees and measured their responses. The only binding factor to preserve a dose of internal validity is that all workers were from the same set of restaurants so because of common kitchen practices and behaviors, the authors believed that each worker was substitutable with others from their restaurant and could be used interchangeably. Their study, however, is positive in its findings that a training course can improve knowledge *and* behavior.

In another study conducted in Korea[3], a similar training program on food safety was offered to 12 restaurants in two groups: a treatment one of 7 restaurants, and a control of the remaining 5. What is peculiar in the findings is that the authors found that although knowledge increased, food safety behavior did not show any improvement after the training! This means that there may have been a lack of leadership to push for the full behavioral adoption of the acquired food safety knowledge, which means that our intention to focus on restaurant managers who can channel this behavior may be the

[2] Roberts, Kevin R.; Barrett, Betsy B.; Howells, Amber D.; Shanklin, Carol W.; Pilling, Valerie K.; Brannon, Laura A., “Food safety training and foodservice employees’ knowledge and behavior”, *Food Protection and Trends* (journal), vol. 28, issue 4, pp.252-260., 2008.

[3] Park, Sung-Hee; Kwak, Tong-Kyung; Chang, Hye-Ja, “Evaluation of the food safety training for food handlers in restaurant operations”, *Nutrition Research and Practice*, pp 58-68, February 2010 edition.

right step to ensure that there is an observable change in their scores. The study from Korea shows that it may be more important to start the training from top to bottom than the other way around.

Then, in another study[4] focused specifically on the training effect for food service managers (most relevant for our study design), the authors examined 1034 food inspection reports from a 12-month period (2005-06) from a county in Ohio and compared food hygiene violations between food service facilities (including hospital, school, workplace, and daycare cafeterias) with certified and non-certified food managers. The total number of food service establishments examined was 605. Their findings state that restaurants with trained and certified food managers had significantly fewer critical food safety violations but more non-critical violations than restaurants without certified personnel. The value of having certified personnel was only observed in independent restaurants and those with few branches, which is primarily the case of the NYC dining scene. This study shows why a holistic management training is needed since the non-critical violations span facilities practices other than food safety, which still add up to the penalty points incurred by NYC restaurants.

Thus, based on other empirical work from other scholars, we can expect that a Restaurant Management course designed for restaurant managers in NYC will result in a more balanced performance by the NYC restaurants around the acceptable A-grade levels of 14 or below.

[4] Kassa, H.; Silverman, K Baraudi; “Effect of a manager training and certification program on food safety and hygiene in food service operations”, *Environmental Health Insights*, issue 4, pp13-20, May 2010