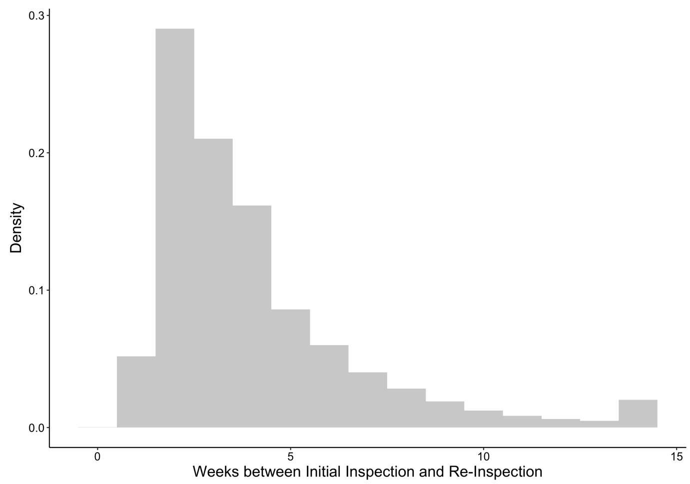
APPENDIX TO

Consumer Reviews and Regulation: Evidence from NYC Restaurants

Chiara Farronato and Georgios Zervas

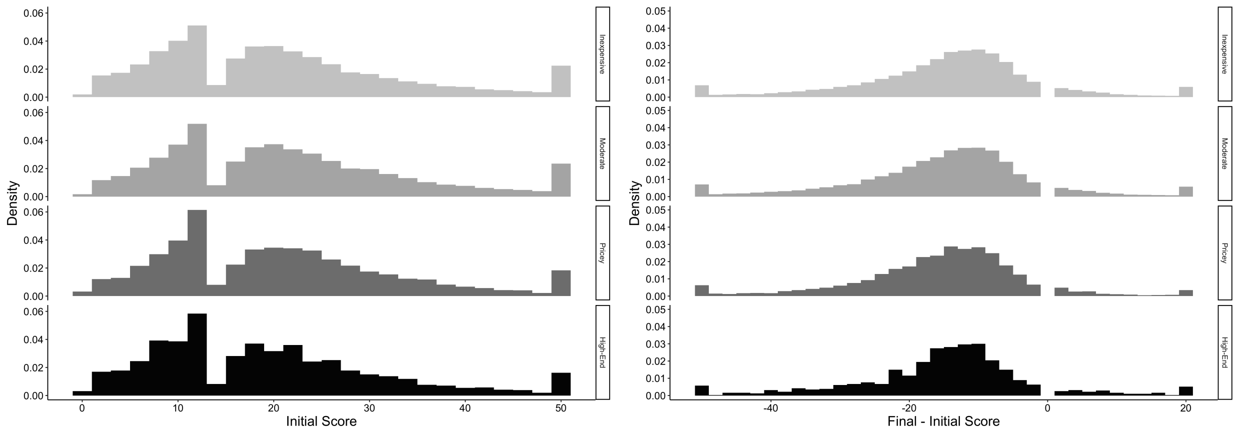
A1 Additional Figures and Tables

Figure A1: Weeks Between Initial Inspection and Re-Inspection



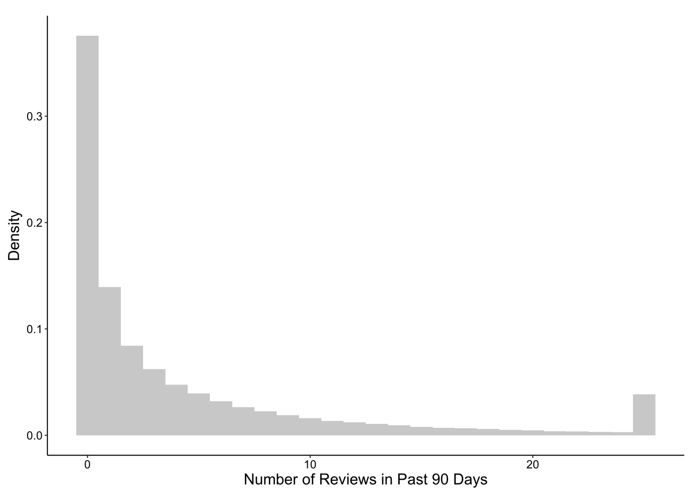
This figure plots the distribution of the time lag in weeks between an initial inspection and a reinspection for the restaurants that received 14 or more points during the initial inspection.

Figure A2: Inspection Outcomes by Price Groups



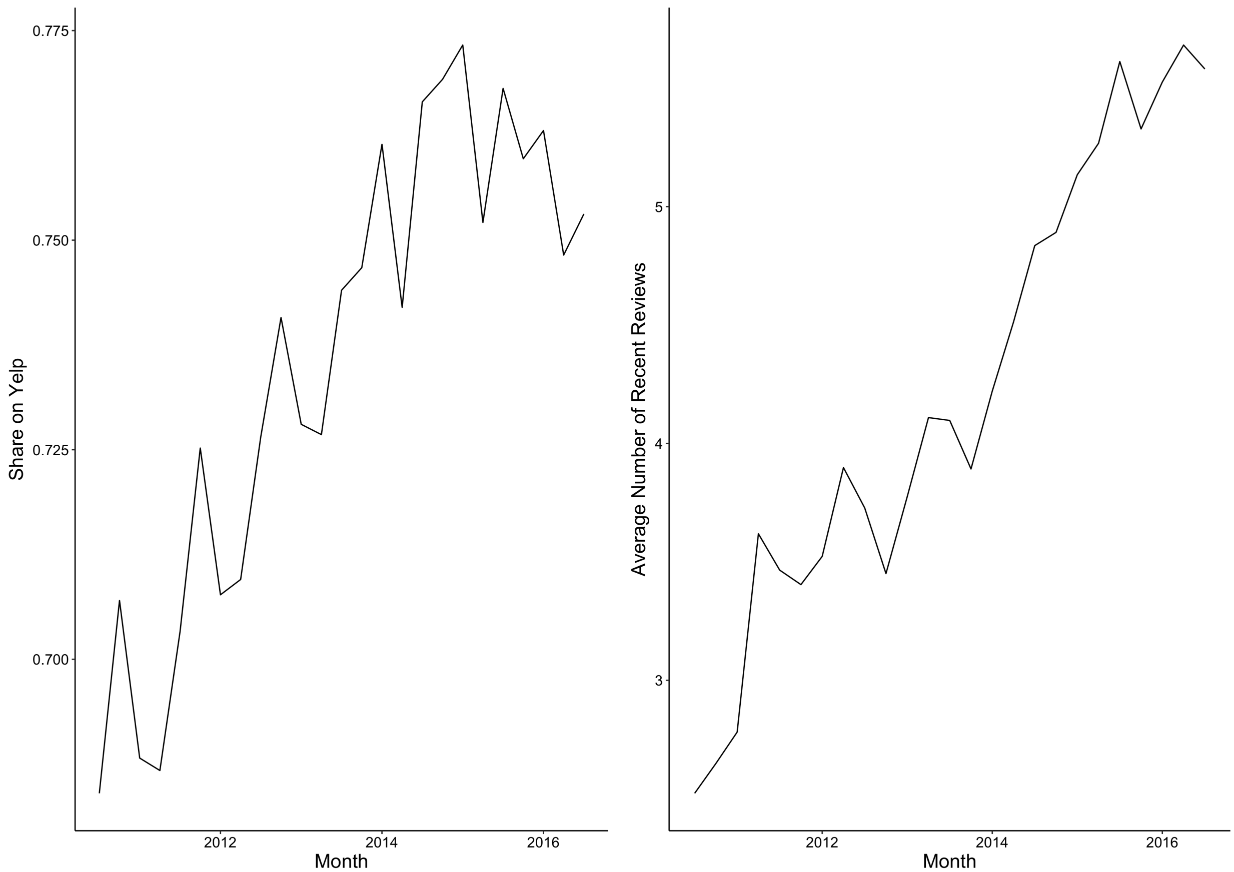
These figures are similar to Figure 3, except that restaurants are divided into four price groups, from "inexpensive" at the top to "high end" at the bottom. For each inspection cycle, the left column shows the distribution of violation scores that restaurants obtain during the initial inspection. For the purpose of these plots, inspection scores are capped at 50. For restaurants that undergo a reinspection, the right column displays the difference between the reinspection score and the initial inspection score (a negative number means that the restaurant improved their hygiene conditions). For the purpose of these plots, the difference in inspection scores is bounded between -20 and 20 (i.e., higher differences in absolute value are capped at +/-20.)

Figure A3: Number of Reviews in Previous 90 Days



This figure plots the distribution of the number of reviews that a restaurant on Yelp obtains in the 90 days preceding an initial inspection. The median number of reviews received prior to an initial inspection is 1, while the mean is 5. For the purpose of this plot, the number of reviews is capped at 25.

Figure A4: Review Frequency over Time



The left figure plots the share of initial inspections during a given quarter whose restaurant is listed on Yelp. The right figure plots how the average number of reviews that a restaurant on Yelp obtains in the 90 days preceding an initial inspection changes over time. Initial inspections, which are the unit of observation, are aggregated at the quarterly level in the time plots.

Figure A5: Comparing Prediction Accuracy Across Classifiers

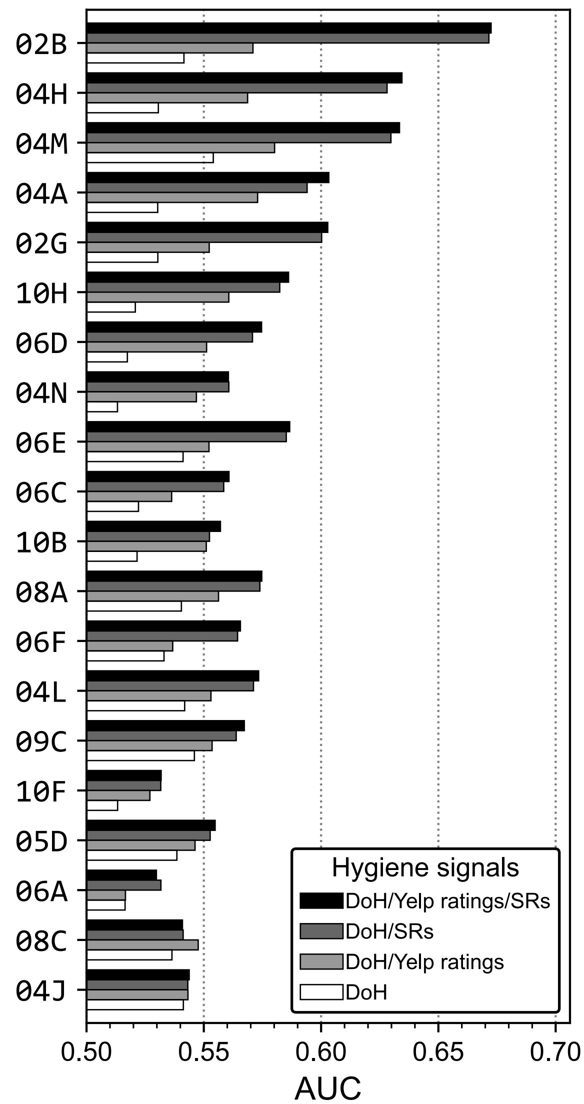
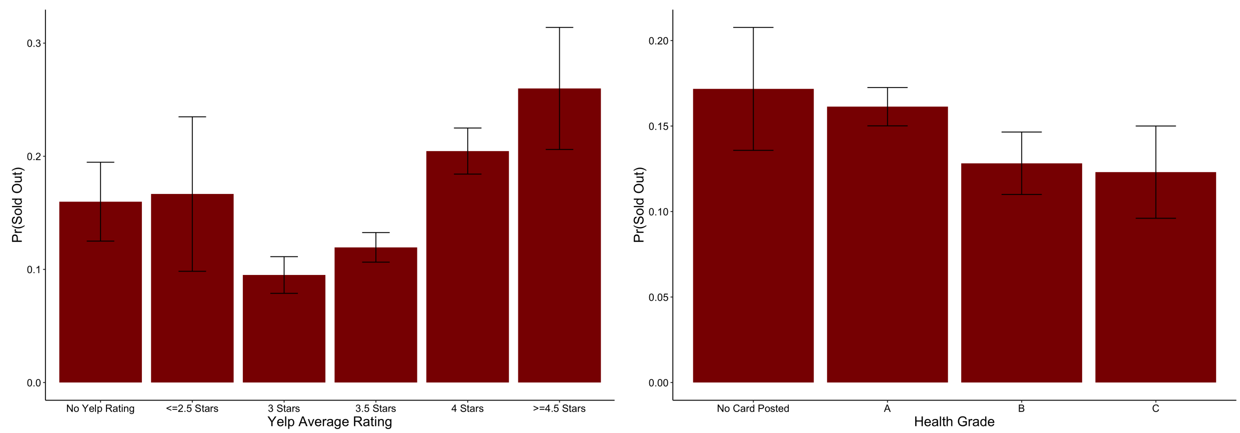


Figure A6: AUC by violation code for all four classifiers.

This figure plots the area under the curve (AUC) of the prediction of the 20 most frequent violation codes, separately for four classifiers: the baseline classifier (white) and the review-augmented classifiers (black), both of which are in the main paper, a classifier that uses letter grades and average star ratings (light grey), and a classifier that uses letter grades and the sufficient reduction projections (dark grey).

Figure A7: Quality Signals and Sold Out Probability



The two bar charts show the average probability of being sold out between 6:30pm and 7:30pm on a given night as a function of Yelp stars (left panel) and hygiene letter grades (right panel). An observation is a restaurant-night for which we have OpenTable availability. Vertical bars denote 95% confidence intervals, computed using standard errors clustered at the restaurant level. 8% of restaurant-nights have no Yelp rating, 2% have 2.5 stars or less, 12% have 3 stars, 39% have 3.5 stars, 34% have 4 stars, and 5% have 4.5 stars or more. For the right panel, 87% of restaurant-nights have a A-grade posted at the door, 10% have a B, 2% have a C, and 1% have a grade pending, or no grade.

Table A1: Yelp Hygiene Signal and Sold Out Probability – Robustness

	Sold Out on OpenTable				
	(1)	(2)	(3)	(4)	(5)
After Review	−0.003	−0.003	−0.004	−0.003	−0.003
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Bad Yelp Hygiene Signal	0.001	0.090	0.003	0.018	0.008
	(0.007)	(0.058)	(0.008)	(0.011)	(0.011)
Bad Yelp Hygiene Signal*After Review	−0.004	−0.004	−0.004	−0.005	−0.003
	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)
Day of Week FE	Yes	Yes	Yes	Yes	Yes
Restaurant-Review FE	Yes	Yes	Yes	Yes	Yes
Specifications	Baseline	Different Control	No Days Around Event	Worst 10% Signal	Worst 40% Signal
Observations	3,430,377	1,366,924	2,877,038	3,430,377	3,430,377
Adjusted R^2	0.510	0.514	0.507	0.510	0.510

Robustness checks to the difference-in-differences specification presented in column 2 of Table 6. Column 1 reproduces the baseline results. The other columns each change one parameter at a time. Column 2 only uses the 1-2-3 star reviews with the 25% best hygiene signal as control group. Column 3 removes 5 days surrounding the review submission from the observations (from 2 days prior to 2 days following the submission of the focal review). Column 4 defines the treated group as restaurants receiving a hygiene signal from Yelp reviews among the top 10% worst signals. Column 5 defines the treated group as restaurants receiving a hygiene signal from Yelp reviews among the top 40% worst signals.

Table A2: Yelp Hygiene Signal and Sold Out Probability – Robustness

Violation Codes	Interaction Coefficient
02B	-0.0018
	(0.0012)
02B, 04H	-0.0014
	(0.0014)
02B, 04H, 04M	-0.0025
	(0.0013)
02B, 04H, 04M, 04A	-0.0037
	(0.0013)
02B, 04H, 04M, 04A, 02G	-0.0043
	(0.0012)
02B, 04H, 04M, 04A, 02G, 10H	-0.005
	(0.0013)
02B, 04H, 04M, 04A, 02G, 10H, 06D	-0.0045
	(0.0012)
02B, 04H, 04M, 04A, 02G, 10H, 06D, 04N	-0.0044
	(0.0011)
02B, 04H, 04M, 04A, 02G, 10H, 06D, 04N, 06E	-0.005
	(0.0012)
02B, 04H, 04M, 04A, 02G, 10H, 06D, 04N, 06E, 06C	-0.0042
	(0.0012)
02B, 04H, 04M, 04A, 02G, 10H, 06D, 04N, 06E, 06C, 10B	-0.0042
	(0.0011)
02B, 04H, 04M, 04A, 02G, 10H, 06D, 04N, 06E, 06C, 10B, 08A	-0.0042
	(0.0012)
02B, 04H, 04M, 04A, 02G, 10H, 06D, 04N, 06E, 06C, 10B, 08A, 06F	-0.0043
	(0.0012)
02B, 04H, 04M, 04A, 02G, 10H, 06D, 04N, 06E, 06C, 10B, 08A, 06F, 04L	-0.0045
	(0.0011)
02B, 04H, 04M, 04A, 02G, 10H, 06D, 04N, 06E, 06C, 10B, 08A, 06F, 04L, 09C	-0.0041
	(0.0012)
02B, 04H, 04M, 04A, 02G, 10H, 06D, 04N, 06E, 06C, 10B, 08A, 06F, 04L, 09C, 10F	-0.0038
	(0.0012)
02B, 04H, 04M, 04A, 02G, 10H, 06D, 04N, 06E, 06C, 10B, 08A, 06F, 04L, 09C, 10F, 05D	-0.0039
	(0.0012)
02B, 04H, 04M, 04A, 02G, 10H, 06D, 04N, 06E, 06C, 10B, 08A, 06F, 04L, 09C, 10F, 05D, 06A	-0.0038
	(0.0012)
02B, 04H, 04M, 04A, 02G, 10H, 06D, 04N, 06E, 06C, 10B, 08A, 06F, 04L, 09C, 10F, 05D, 06A, 08C	-0.0039
	(0.0013)
02B, 04H, 04M, 04A, 02G, 10H, 06D, 04N, 06E, 06C, 10B, 08A, 06F, 04L, 09C, 10F, 05D, 06A, 08C, 04J	-0.0043
	(0.0012)

Robustness checks to the difference-in-differences specification presented in column 2 of Table 6. Each row displays the difference-in-differences coefficient from a different regression. The first row only uses the sufficient reduction from violation code 02B (Hot food item not held at or above $140^o F$), which is the violation code for which Yelp is most informative, to define the event of a bad hygiene signal. Subsequent rows add the sufficient reduction of violation codes for which Yelp is progressively less informative. The eighth row (with violation codes 02B, 02G, 04A, 04H, 04M, 04N, 06D, 10H) reproduces the difference-in-differences coefficient estimate from column 2 of Table 6.

Table A3: Yelp Signal and Restaurants' Hygiene Compliance – Robustness

Violation Codes	IV Coefficient
02B **V.** 04H, 04M, 04A, 02G, 10H, 06D, 04N, 06E, 06C, 10B, 08A, 06F, 04L, 09C, 10F, 05D, 06A, 08C, 04J	-0.049 (0.0052)
02B, 04H **V.** 04M, 04A, 02G, 10H, 06D, 04N, 06E, 06C, 10B, 08A, 06F, 04L, 09C, 10F, 05D, 06A, 08C, 04J	-0.0072 (0.0031)
02B, 04H, 04M **V.** 04A, 02G, 10H, 06D, 04N, 06E, 06C, 10B, 08A, 06F, 04L, 09C, 10F, 05D, 06A, 08C, 04J	-0.0059 (0.0024)
02B, 04H, 04M, 04A **V.** 02G, 10H, 06D, 04N, 06E, 06C, 10B, 08A, 06F, 04L, 09C, 10F, 05D, 06A, 08C, 04J	-0.0178 (0.0021)
02B, 04H, 04M, 04A, 02G **V.** 10H, 06D, 04N, 06E, 06C, 10B, 08A, 06F, 04L, 09C, 10F, 05D, 06A, 08C, 04J	-0.0109 (0.0021)
02B, 04H, 04M, 04A, 02G, 10H **V.** 06D, 04N, 06E, 06C, 10B, 08A, 06F, 04L, 09C, 10F, 05D, 06A, 08C, 04J	-0.0127 (0.002)
02B, 04H, 04M, 04A, 02G, 10H, 06D **V.** 04N, 06E, 06C, 10B, 08A, 06F, 04L, 09C, 10F, 05D, 06A, 08C, 04J	0.0066 (0.0019)
02B, 04H, 04M, 04A, 02G, 10H, 06D, 04N **V.** 06E, 06C, 10B, 08A, 06F, 04L, 09C, 10F, 05D, 06A, 08C, 04J	0.0109 (0.0019)
02B, 04H, 04M, 04A, 02G, 10H, 06D, 04N, 06E **V.** 06C, 10B, 08A, 06F, 04L, 09C, 10F, 05D, 06A, 08C, 04J	0.0093 (0.0018)
02B, 04H, 04M, 04A, 02G, 10H, 06D, 04N, 06E, 06C **V.** 10B, 08A, 06F, 04L, 09C, 10F, 05D, 06A, 08C, 04J	0.0063 (0.0018)
02B, 04H, 04M, 04A, 02G, 10H, 06D, 04N, 06E, 06C, 10B **V.** 08A, 06F, 04L, 09C, 10F, 05D, 06A, 08C, 04J	0.0071 (0.0019)
02B, 04H, 04M, 04A, 02G, 10H, 06D, 04N, 06E, 06C, 10B, 08A **V.** 06F, 04L, 09C, 10F, 05D, 06A, 08C, 04J	0.0095 (0.0018)
02B, 04H, 04M, 04A, 02G, 10H, 06D, 04N, 06E, 06C, 10B, 08A, 06F **V.** 04L, 09C, 10F, 05D, 06A, 08C, 04J	0.0088 (0.0018)
02B, 04H, 04M, 04A, 02G, 10H, 06D, 04N, 06E, 06C, 10B, 08A, 06F, 04L **V.** 09C, 10F, 05D, 06A, 08C, 04J	0.0065 (0.002)
02B, 04H, 04M, 04A, 02G, 10H, 06D, 04N, 06E, 06C, 10B, 08A, 06F, 04L, 09C **V.** 10F, 05D, 06A, 08C, 04J	0.0058 (0.0021)
02B, 04H, 04M, 04A, 02G, 10H, 06D, 04N, 06E, 06C, 10B, 08A, 06F, 04L, 09C, 10F **V.** 05D, 06A, 08C, 04J	0.0185 (0.0018)
02B, 04H, 04M, 04A, 02G, 10H, 06D, 04N, 06E, 06C, 10B, 08A, 06F, 04L, 09C, 10F, 05D **V.** 06A, 08C, 04J	0.0115 (0.0019)
02B, 04H, 04M, 04A, 02G, 10H, 06D, 04N, 06E, 06C, 10B, 08A, 06F, 04L, 09C, 10F, 05D, 06A **V.** 08C, 04J	0.0104 (0.0021)
02B, 04H, 04M, 04A, 02G, 10H, 06D, 04N, 06E, 06C, 10B, 08A, 06F, 04L, 09C, 10F, 05D, 06A, 08C **V.** 04J	0.0197 (0.003)

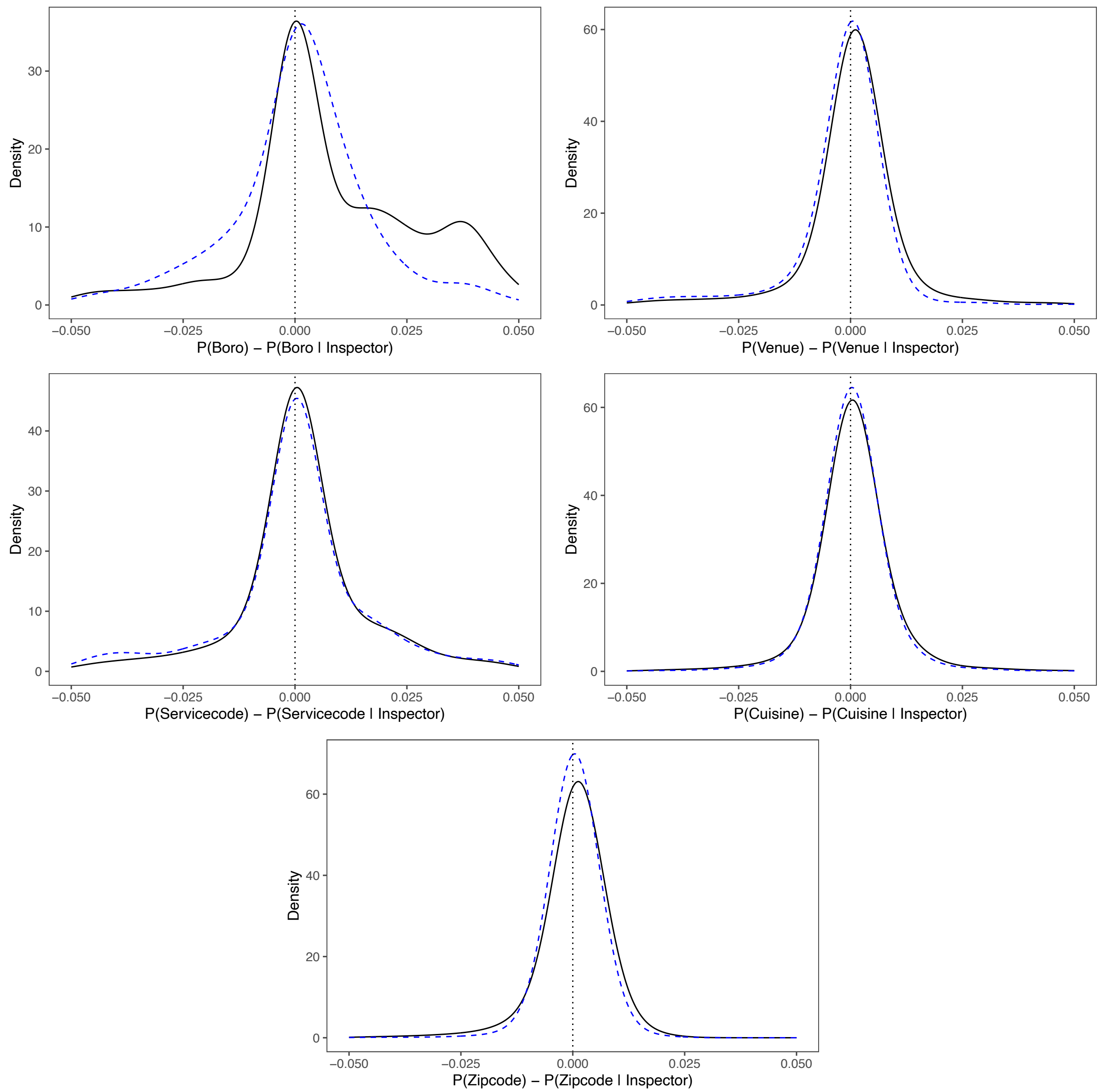
The table displays results of Equation 10 from multiple regressions. Each row considers an increasingly larger set of violation codes for which Yelp is informative. In the first row, the dummy $yelp\_informative_v$ *is 1 only for the violation code for which Yelp is most informative (04M: Live roaches present in facility's food and/or non-food areas). The subsequent rows progressively turn the dummy* $yelp\_informative_v$ *to 1 for additional violation codes, ranked according to the prediction accuracy of Yelp reviews. Row 8 reproduces the results from the main paper (fourth column in Panels A and B in Table 7).*

A2 How are inspectors and reviewers assigned to restaurants?

The instrumental variable approach described in Section 5.1 relies on random assignment of inspectors and reviewers to restaurants. Here we verify that conditional on observables, we cannot reject the hypothesis that inspectors (respectively, reviewers) are randomly assigned to evaluate restaurants. We describe the procedure for inspectors, but it is analogous for reviewers. We compute two probability distributions. First, we compute the unconditional distribution of a particular restaurant characteristic, denoted $P(X)$. Second, we compute the distribution of X conditional on a particular inspector Z, denoted $P(X|Z)$. We then take the difference $P(X) - P(X|Z)$ across all possible values of X and across all inspectors. We compare the distribution of this difference to $P(X) - P(X|Z')$. The only difference between $P(X|Z)$ and $P(X|Z')$ is that $P(X|Z)$ is based on the actual allocation of inspectors to restaurants, while $P(X|Z')$ is a random permutation.

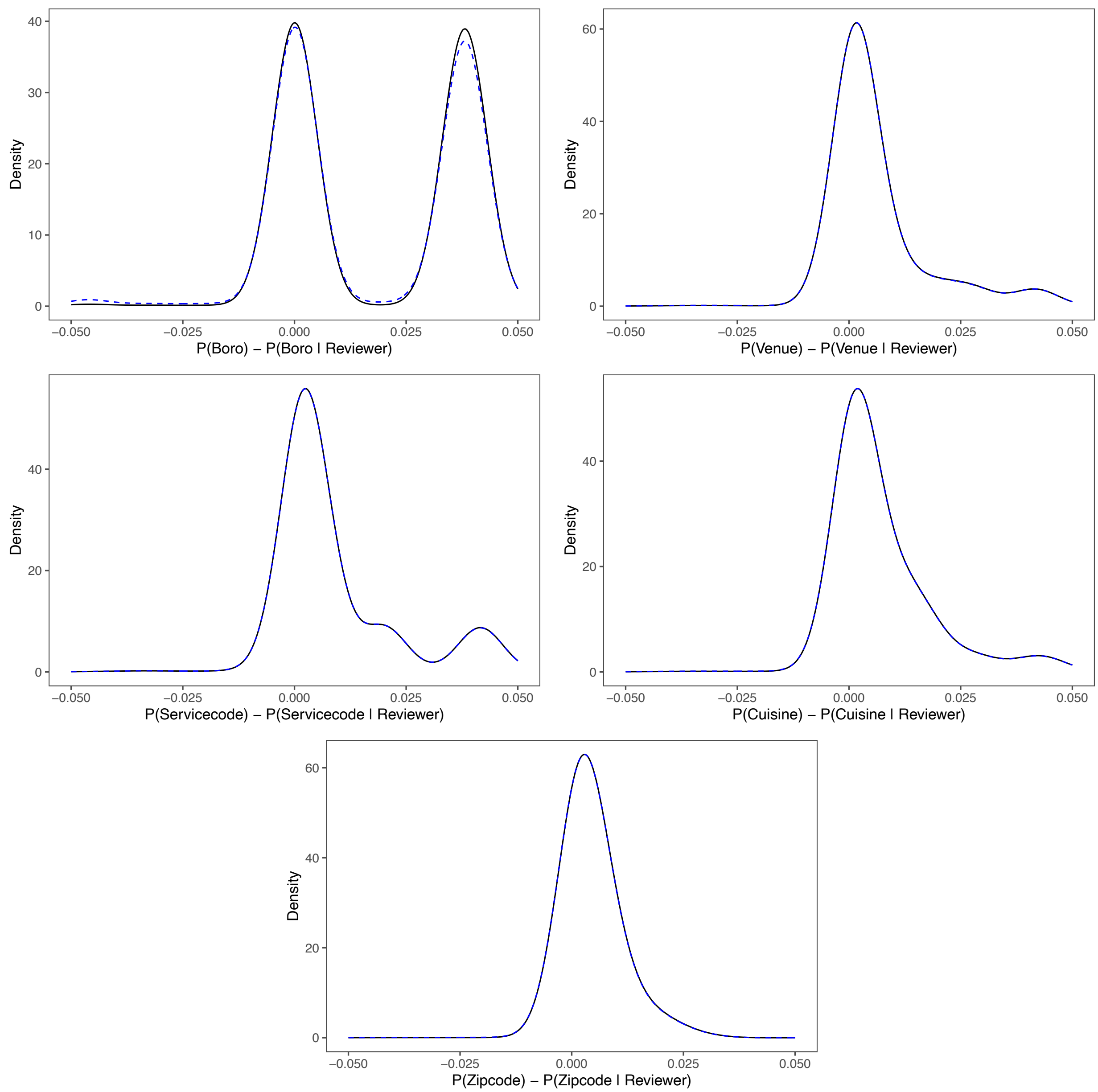
Figure A8 displays the distributions of $P(X)-P(X|Z)$—solid line—and $P(X)-P(X|Z')$—dotted line—across all inspectors and for different observable characteristics. If inspectors were randomly assigned to restaurants conditional on observable X, the dotted and solid density functions could not be distinguished from one another. The figures show that inspectors tend to specialize by geography, inspecting restaurants in one New York City borough more than in other boroughs, and clustering inspections within a few zipcodes. Beyond geography, there does not seem to be specialization of inspectors across other observable restaurant characteristics. As for how reviewers are "assigned" to restaurants, Figure A9 shows that there is no particular pattern for how users choose to submit reviews to restaurants.

Figure A8: Independence of Inspectors and Restaurants' Characteristics.



This figure plots the difference in the distribution of observables unconditional and conditional on inspectors. The solid line plots $P(X) - P(X|Z)$*, where the conditional distribution is a function of the actual allocation of inspectors to restaurants. The dotted line computes the conditional distribution after a random permutation of inspectors to restaurants. The average number of restaurants inspected by each inspector is 807, with a large standard deviation of 822.*

Figure A9: Independence of Reviewers and Restaurants' Characteristics.



This figure plots the difference in the distribution of observables unconditional and conditional on reviewers. The solid line plots $P(X) - P(X|Z)$, where the conditional distribution is a function of the actual allocation of Yelp reviewers to restaurants. The dotted line computes the conditional distribution after a random permutation of reviewers to restaurants. The average number of restaurants reviewed by each Yelp reviewer is 4.3, with a large standard deviation of 14.5.

A3 Review Timing and Ranking of Restaurants on Yelp

We want to verify that restaurants with more recent reviews are ranked higher in Yelp search results compared to restaurants with older reviews. To do this we pulled data from the Yelp API. We submitted the query"Find: Restaurants | Near: New York, NY" on April 8, 2019 at midnight. Yelp places a limit of 1,000 results to be returned, and the order in which they are returned reflects the order shown on the webpage if a user were to perform the same search on Yelp. The restaurants returned in this list are the *ranked* restaurants out of all New York City restaurants. Whether a reastaurant shows up at all in this list, and whether it shows at the top or at the bottom of the search results will be our outcomes of interest.

We also compile the list of all Yelp restaurants in New York by performing a similar query as before, but separately for each zip code.[25] Given the limit to the number of results returned by the Yelp API, a zip code is further disaggregated if the returned results are 1,000. Specifically, if a query for a given zip code returns less than 1,000 restaurants, results are recorded and we move to the next zip code. Otherwise the zip code is split into four quadrants, and we conduct four searches, one for each quadrant using its center and half its diagonal as the search radius. We continue splitting geographies until the results returned are less than 1,000 for each separate search. After dropping duplicates and businesses located outside of New York,[26] we are left with 23,387 restaurants, which constitutes the population of restaurants in New York City that are on Yelp.

For each restaurant, we scrape additional information from Yelp using the URLs that we obtained from the API. This information includes the date of the most recent review, which is our treatment variable of interest, as well as additional controls, such as restaurant category, price, Yelp stars, and total number of reviews.

We run regressions of the following type:

$$y_i = \alpha log(days\_since\_last\_review_i) + \beta X_i + \epsilon_i, \tag{11}$$

where i denotes a restaurant, and y is one of two outcomes: a dummy for whether the

[25]The following website contains the list of zip codes compiled by the New York State Department of Health (NYDOH) used in the grid search: `https://www.health.ny.gov/statistics/cancer/registry/appendix/neighborhoods.htm`.

[26]The search algorithm may pick up restaurants outside of New York City, or in neighboring zip codes, since the search radius is conservatively set to cover larger areas than the quadrant of interest. To determine whether a zip code is located in New York City, we check whether the zip code can be found on the NYDOH list used in the grid search, or whether it is located in New York City according to US Postal Service's ZIP Code Lookup data.

restaurant is *ranked* in the general query for "restaurants in New York City"; and conditional on being ranked, the rank of the restaurant in the search results (where 1 denotes the top result and 999 denotes the last result).[27] We expect that the shorter the time since the last review, the more likely a restaurant is to be ranked, and the lower – i.e. closer to the top of the page – the ranking it obtains. The variable $days\_since\_last\_review_i$ is measured relative to April 7, 2018, the day immediately preceding our data pull. The vector X_i includes number of reviews (in logs) and fixed effects for restaurant category, price grouping ($, $$, $$$, or $$$$), yelp stars, and zip code. To estimate the relationships of interest, we use OLS, logistic, and probit regressions when the outcome is a dummy for being ranked, and OLS, ordered logistic, and ordered probit when the outcome is the exact ranking.

Summary statistics are presented in Table A4 and regression results are presented in Table A5. The results show that businesses with more recent reviews are more likely to be ranked (columns 1-3) and placed higher in search results when ranked (columns 4-6). The effects are sizable. When looking at the probability of being in the top 1,000 results, we discuss the logistic regression estimates given that only 4.3% of New York City restaurants are actually ranked (probit estimates imply similar magnitudes). The estimates from the logistic regression (column 2) suggest that all else held constant, doubling the age of the most recent review is associated with a -0.382 decrease in log odds of being ranked, implying an odds ratio of $exp(-0.382) = 0.682$. This suggests significantly reduced odds of being ranked for businesses with older reviews.

When we focus on the sample of restaurants that are ranked (columns 4-6), the age of the most recent review predicts the actual rank in the search results. All else held constant, OLS results suggest that doubling the age of the most recent review is associated with a *decrease* in rank by 35 positions – e.g. dropping from the top result to the 36^{th} result. Similarly, the odds ratio of $exp(-0.265) = 0.767$ from column 5 confirms reduced odds of a business moving up the rank with an increase in the age of the most recent review.

Table A4: Summary Statistics

	1st Qu.	Median	3rd Qu.	Mean	Std. Dev.	N
Ranked	0	0	0	0.043	0.202	23,385
Days since last review	7	32	185	322.5	754.8	23,383
Days since last review | ranked	0	2	7	5.6	11.7	999

Note: Summary statistics for the Yelp restaurants in New York City. Data were obtained from the Yelp API.

[27] We dropped one observation from the search results because the most recent review date was missing.

Table A5: Review Age and Search Ranking Outcomes

	Ranked 0/1			Rank | Ranked		
	normal	*logistic*	*probit*	*OLS*	*ordered logistic*	*ordered probit*
	(1)	(2)	(3)	(4)	(5)	(6)
reg_log_days_since_last_review_plus1	−0.001	−0.382	−0.218	34.880	−0.265	−0.165
	(0.001)	(0.043)	(0.023)	(9.403)	(0.073)	(0.037)
Observations	23,383	23,383	23,383	999	999	999
Log Likelihood	6,945.889	−1,545.434	−1,550.226			
Akaike Inf. Crit.	−12,829.780	4,152.869	4,162.451			
R^2				0.370		
Adjusted R^2				0.223		

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Note: Regression results of Equation 11. In columns 1-3 the outcome of interest is a dummy for whether a restaurant is in the top 1,000 search results for the query “Find: Restaurants | Near: New York, NY”. In columns 4-6 the outcome is the actual rank in the search results, conditional on being in the top 1,000 results. Controls include number of reviews (in logs) and fixed effects for restaurant category, price grouping ($, $$, $$$, or $$$$), yelp stars, and zip code.