Impact of a Letter-Grade Program on Restaurant Sanitary Conditions and Diner Behavior in New York City

Melissa R. Wong, MPH, Wendy McKelvey, PhD, Kazuhiko Ito, PhD, Corinne Schiff, JD, J. Bryan Jacobson, MPH, and Daniel Kass, MSPH

Restaurant food safety is increasingly important, with almost half of the US food dollar spent on restaurant food and about one third of caloric intake from foods prepared outside the home. In New York City (NYC), residents eat out nearly 1 billion times each year. Although most diners do not get sick, foodborne pathogens cause millions of preventable illnesses in the United States annually. The exact proportion of restaurant-attributable foodborne illness is unknown, but national surveillance in the United States found that two thirds of reported foodborne outbreaks from 1998 through 2008 occurred in the restaurant or deli setting and consumption of food prepared outside the home has been linked to an increased risk of sporadic foodborne diseases.

Regular inspection of restaurants for food safety is a core function of local health authorities, guided by the US Food and Drug Administration (FDA) Food Code. Although all states have sanitation codes modeled after the FDA Food Code, implementation methods vary by jurisdiction. The NYC Department of Health and Mental Hygiene (hereafter, Health Department) is charged with inspecting restaurants, coffee shops, bars, nightclubs, employees or university cafeterias, bakeries, and fixed-site food stands (hereafter, restaurants). Its inspection program uses a scoring system to measure compliance with the NYC Health Code, which is updated regularly to maintain consistency with the FDA Food Code and the New York State Sanitary Code. Restaurants are entitled to an impartial review of inspection results by an administrative tribunal, which can improve an assigned score and reduce associated monetary fines.

Before letter grading, the Health Department aimed to inspect restaurants at least once per year and imposed monetary fines for violations cited at inspections. Inspection results were available on the Health Department Web site. However, financial disincentives and the Web site posting were insufficient to drive improvements across the industry, with most restaurants cited for multiple public health hazards. Mean inspection scores and restaurant sanitary conditions were stagnant (D. Kass, email communication, February 2009).

In an effort to improve restaurant food safety and increase transparency of inspection information, the Health Department launched its letter-grade program on July 27, 2010. The program uses public disclosure of inspection scores in the form of letter grades at point of decision-making; a more finely tuned, risk-based inspection schedule; and financial incentives to encourage high food-safety standards. It began after an 18-month planning process that included a public announcement of the intent to begin letter grading; meetings with restaurant industry representatives, food safety experts, and regulators from a jurisdiction with a restaurant sanitary grade program; promulgation of 2 regulations subject to notice and comment; and training and education for restaurateurs. The process was covered by the media, and by July 2010, restaurateurs were aware of the program and anticipating the launch.

We evaluated the impact of the restaurant letter-grade program by assessing (1) hygiene and food-safety practices as characterized by inspection outcomes before and after program implementation and (2) public response to the program measured by 2 population-based telephone surveys.

METHODS

The NYC restaurant inspection program has been using a point system to score inspections since 2005. Presence and severity of violations contribute to an inspection score. Under the grading program, an inspection score of 0 to 13 points is in the A-range; 14 to 27 points is in the B-range; and 28 or more points is in the C-range. Restaurants scoring 0 to 13 points on the first inspection of their inspection cycle (initial inspection) are issued an A grade. Restaurants not earning an A grade on initial
inspection receive a full reinspection no less than 7 days later. The grade card is issued based on the reinspection score. The initial inspection and any reinspection together are an “inspection cycle.” Upon completion of an inspection cycle, there is an interval before the next cycle. Restaurants earning an A grade on initial inspection of a cycle are inspected in 11 to 13 months. Restaurants scoring 28 or more points on either initial or reinspection of a cycle have a 3- to 5-month interval. The remaining restaurants scoring 14 to 27 points on either initial inspection or reinspection of a cycle have a 5- to 7-month interval.

Before the grading program was launched, the Health Department aimed to conduct at least 1 inspection in all restaurants annually. Restaurants scoring 28 or more points received a follow-up compliance inspection about 1 month later. A score of 28 points or higher could result in a restaurant being placed on a twice-yearly inspection schedule. Administrative violations (e.g., expired permit) were included in the scoring system before implementation of letter grading, but they are not included under the grading program.

Health Department inspectors cite violations with standardized forms on handheld computers. They also collect data on restaurant descriptors such as cuisine, service method to customer (e.g., wait service, counter service), venue description (e.g., restaurant, bar), and chain status (15 or more national outlets). Inspectors are trained in the classroom and under an experienced inspector in the field before they are allowed to work independently.

Data Analysis

We analyzed prejudicadation inspection scores and points for violations cited on initial or reinspections conducted between July 27, 2007, and July 26, 2013. We subtracted administrative violation points from pregrading inspection scores to make pregrading scores more comparable with postgrading scores.

We calculated measures that used “most recent initial inspection” among restaurants in business as of July 27 in each year. “Most recent initial inspection” is used in crude analyses to depict a restaurant’s usual sanitary conditions closest to the specified period end date. We consider initial inspections of a cycle the best indicator of usual sanitary conditions because they occur at the longest interval after the previous inspection and they are unannounced to operators. Crude metrics were percentage of restaurants scoring in the A-, B-, or C-range; percentage scoring 40 points or higher (85th percentile score on initial inspection in the program’s first year); median inspection score; and average points for specific violations or violation groups. Average violation points characterize both presence and severity of violations over time.

We assessed performance on reinspection of a cycle by calculating percentage of restaurants scoring in the A-range on reinspection among those with B-range or C-range initial inspection scores. We tracked the percentage of restaurants with A, B, or C grades on a cycle that went on to earn an A grade on their next cycle.

We modeled the probability of scoring 0 to 13 points (A-range score) across all initial inspections in all 43,448 restaurants by fitting a binomial regression model that included 5 indicators of time: 13 to 36 months before grading (reference), 0 to 12 months before grading, 0 to 12 months after grading, 13 to 24 months after grading, and 25 to 36 months after grading. We fit restaurant random intercepts to account for repeated observations and variation across individual restaurants. We used indicator variables to adjust for potential confounding by season of inspection (January—March, April—June, July—September, October—December), because pest and holding temperature–related violations increase during the warmest season and the distribution of inspection date varied over time. We did not think chain status was a potential confounder because the distribution before and after grading remained constant, but we included it to estimate the probability that a chain restaurant scored 0 to 13 points relative to a nonchain. We also ran the fully adjusted model for the subset of restaurants with inspections in the first and last year of the study (n = 7059) to evaluate whether improvement differed among the most stable restaurants.

To assess whether an excess or deficit in the frequency of inspection scores around grade cut-offs could have biased our results, we estimated the underlying (unbiased) smooth frequency distribution of scores by fitting a generalized additive model with penalized splines and used the smoothed distribution to estimate the “bias-corrected” percentage of A-range scores in the postgrading period. The percentage of A-range scores across initial inspections in the postgrading period dropped only slightly from 30.7% to 27.4% upon correction. We therefore deemed it unnecessary to correct for potential bias resulting from an excess or deficit of scores around grade cut-offs.

We conducted analyses in SQL Management Studio 2008 R2 (Microsoft, Redmond, WA), SAS version 9.2 (SAS Institute, Cary, NC), and R version 3.0.1 (R Project, Vienna, Austria).

Public Perception Surveys

The Health Department worked with Baruch College Survey Research (BCSR) to conduct 2 English/Spanish bilingual telephone surveys in July 2011 and February 2012 to assess public perceptions of the grading program. Landline samples on a random-digit-dial design and respondents were selected randomly within the household; cell phones were randomly selected from a mobile number database for NYC county telephone numbers. Respondents were screened for NYC residency and age of 18 years or older.

In July 2011 and January 2012, 502 and 511 adults completed surveys, respectively. Based on the American Association for Public Opinion Research (AAPOR) standard definitions,14 response rates were 26% and 22%, and cooperation rates were 60% and 51%, respectively. AAPOR response rates incorporate estimates of the proportion of respondents of unknown eligibility that might have been eligible. Data were weighted to the US Census 2009 American Community Survey to ensure the samples represented the age, gender, race, Hispanic origin, and borough distribution of NYC adults. Confidence intervals (CIs) for proportions were calculated with SAS version 9.3 (SAS Institute, Cary, NC).

RESULTS

Approximately 24,000 restaurants operate in NYC on any given day. A total of 43,892 restaurants were in business at some point between July 2007 and July 2013, and 46% (20,005) of those were in operation at some point both before and after grading. During the 3 years before grading, 31,226 restaurants...
After we controlled for chain status, season of inspection, and correlation within restaurants, the probability of attaining an A-range score on an unannounced initial inspection among all restaurants increased 26% (success ratio [SR] = 1.26; 95% CI = 1.22, 1.31) by the 2-year mark (Table 2). The SR increased at the 3-year mark to 1.35 (95% CI = 1.31, 1.40). Compared with the warmest season (July–September), the other seasons exhibited higher SRs, with the highest (SR = 1.30; 95% CI = 1.26, 1.35) in the coldest season (January–March). The SRs for the subset of restaurants in business during the whole period (data not shown in Table 2) were slightly higher—1.32 (95% CI = 1.25, 1.40) and 1.41 (95% CI = 1.33, 1.49) for the 2- and 3-year mark, respectively. Chain restaurants showed a high SR (95% CI = 3.31, 3.61) and the subset of restaurants operating during the whole period (SR = 3.79; 95% CI = 3.54, 4.07).


<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Operating Between July 27, 2007, and July 26, 2010 (n = 31 226), No. (%)</th>
<th>Operating Between July 27, 2010, and July 26, 2013 (n = 32 700), No. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Borough</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bronx</td>
<td>3 267 (10)</td>
<td>3 222 (10)</td>
</tr>
<tr>
<td>Brooklyn</td>
<td>7 538 (24)</td>
<td>8 047 (25)</td>
</tr>
<tr>
<td>Manhattan</td>
<td>11 828 (38)</td>
<td>12 584 (38)</td>
</tr>
<tr>
<td>Queens</td>
<td>7 307 (23)</td>
<td>7 552 (23)</td>
</tr>
<tr>
<td>Staten Island</td>
<td>1 286 (4)</td>
<td>1 295 (4)</td>
</tr>
<tr>
<td>Chain restaurants</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>3 393 (11)</td>
<td>3 627 (11)</td>
</tr>
<tr>
<td>No</td>
<td>27 833 (89)</td>
<td>29 073 (89)</td>
</tr>
<tr>
<td>Restaurant typea</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wait service restaurant or diner</td>
<td>8 589 (31)</td>
<td>10 564 (33)</td>
</tr>
<tr>
<td>Quick-service establishment with take-out or limited seating</td>
<td>12 706 (46)</td>
<td>14 443 (45)</td>
</tr>
<tr>
<td>Baked goods, ice cream, or cafe only</td>
<td>3 539 (13)</td>
<td>4 050 (13)</td>
</tr>
<tr>
<td>Bar or wine bar</td>
<td>1 206 (4)</td>
<td>1 289 (4)</td>
</tr>
<tr>
<td>Cafeteria and banquet-style service or deli buffet</td>
<td>992 (4)</td>
<td>1 236 (4)</td>
</tr>
<tr>
<td>Food service at attraction</td>
<td>534 (2)</td>
<td>662 (2)</td>
</tr>
<tr>
<td>Missing</td>
<td>3 660</td>
<td>456</td>
</tr>
</tbody>
</table>


aPercentage excludes missing values.

Certain critical food safety violations contributed fewer average points in July 2013 compared with the 2 years before grading (Table 3). In July 2012, the average points given to all restaurants declined substantially for evidence of any type of vermin (rats, mice, flies, or roaches), inadequate hand-washing facilities, and no food safety–certified supervisor on-site. Points given for improper storage or use of equipment or utensil and inadequate food worker hygiene also declined to a lesser extent. These overall point reductions were maintained in July 2013. Meanwhile, average points increased for improperly maintained food contact surfaces, and the points given for inadequate protection of food from contamination, cross contamination, and holding food at improper temperatures increased slightly (Table 3). Although average points for temperature and cross-contamination violations increased slightly, average severity of cited violations decreased (data not shown).

We observed inverse trends for C-range scores on recent initial inspection over time. The proportion of C-range scores decreased from 29% as of July 2008 and 27% as of July 2010 to 22% as of July 2013 (Figure 1). The percentage of extreme C-range (≥ 40 points)–scoring restaurants dropped from 14% in the year before grading to 13% in July 2011, 7% in July 2012, and 9% in July 2013, and the 80th percentile decreased from 36 points in July 2008 to 30 points in July 2013.

Three years after grading, more restaurants corrected unsanitary conditions observed on initial inspection of most recent inspection cycle. In July 2013, 45% of restaurants requiring reinspection earned A grades upon reinspection, up from 34% in July 2011. Likewise, there has been a decrease in the proportion of poorly performing restaurants that do not improve on reinspection (28+ point scores on both initial and reinspection). The proportion of restaurants that scored poorly on both initial and reinspection dropped from 28% as of July 2009 to 22% as of July 2013.

When we tracked performance from inspection cycle to inspection cycle, we found that 80% and 79% of A-grade restaurants maintained their A grade on their next cycle at 2 and 3 years after grading, respectively. Among B-grade restaurants, 53% and 54%...
improved to an A grade on the next cycle as of the 2- and 3-year mark, respectively.

Public Perception Surveys

Results from 2 independent telephone surveys suggested that New Yorkers dine out frequently and support and use letter grades to help them decide where to eat. Among NYC adults, 67% (95% CI = 63%, 71%) and 68% (95% CI = 63%, 72%) reported eating meals from a restaurant, deli, coffee shop, or bar at least once per week at the 1-year and 18-month mark, respectively. At the 1-year mark, 90% (95% CI = 87%, 93%) approved of the program and 71% (95% CI = 66%, 74%) had seen a grade card in restaurant windows. At 18 months, support remained at 91% (95% CI = 88%, 94%) and 81% (95% CI = 77%, 84%) had seen grade cards. Among those who had seen grade cards, 88% (95% CI = 85%, 92%) considered them in their dining decisions at the 1-year and 18-month mark.

Results suggested that grades reassure diners about food safety; 76% (95% CI = 71%, 80%) felt more confident in a restaurant’s food safety when an A grade was posted. An estimated 70% (95% CI = 66%, 74%) expressed concern about getting sick from eating from restaurants, delis, and coffee shops, with 38% (95% CI = 34%, 43%) being very concerned. A majority of 88% (95% CI = 85%, 91%) supported more frequent inspections for restaurants that do not earn an A grade.

DISCUSSION

The NYC Health Department launched the restaurant letter-grading program to motivate restaurants to improve food safety, inform the public about inspection results, and reduce illness associated with dining out. The program introduced multiple changes to the enforcement landscape, including the mandatory posting of letter grades summarizing sanitary inspection scores, a fine-tuned risk-based inspection schedule, and a revised policy on financial penalties. Survey results suggest that New Yorkers approve of the program and use it when making dining decisions. Our restaurant hygiene analysis suggests that the program provided an effective incentive for operators to comply with regulations and improve practices. We also found that there is an incentive to maintain hygiene practices, with the majority of A-grade restaurants earning A grades on their next inspection cycle.

Our ultimate goal is to reduce foodborne illness, but evaluating the impact of 1 program on such a multifactorial outcome is challenging. Past foodborne illness studies have noted that case finding suffers from underreporting and potential misclassification. Among cases that are identified, it can be difficult to know if exposures occurred in a restaurant. Certain hygiene and food-safety conditions monitored in restaurants are known risk factors or environmental antecedents for foodborne illness outbreaks, so we think measurement of sanitary conditions alone serves as a good proxy for public health risks.
Improvement in hygiene conditions appeared to be driven by certain categories of violations. Having a certified kitchen manager on site is important because it has been associated with fewer critical violations on inspection and identified as an important factor for preventing foodborne outbreaks. Decreases in violations for inadequate hand-washing facilities and worker hygiene and improper storage or use of equipment or utensils are also likely to decrease risk for foodborne illness. Decreases in presence and severity of vermin violations contributed in large part to improvements in inspection scores, but vermin violations remain the largest average contributors to inspection score on initial inspection, suggesting a need for more restaurant operator education on this topic. The increase in average violation points related to food contact surface maintenance was likely an artifact related to a tendency for inspectors to cite this violation under a “miscellaneous” section before grading.

Although overall inspection performance improved in the second and third year of grading, A-range scores (0–13 points) decreased slightly in the first year of grading compared with the year before. We believe this decrease reflects the method in which the program was rolled out. The first restaurants inspected under the grading program were those that scored poorly under pregrading program rules. These poorer-performing restaurants were overrepresented during year 1. We call attention to the strong association between chain restaurant status and A-range score on initial inspection. This finding is consistent with other studies that reported better sanitary conditions (i.e., fewer critical violations) in chain restaurants compared with non-chains. It is instructive to consider the mechanisms used by chains to ensure food safety, such as use of standardized procedures, specialized equipment, and additional worker training and internal mock inspections, when conducting educational outreach among nonchains.

New York City is not alone in requiring public disclosure of restaurant inspection results at the point of decision-making. This type of disclosure program is becoming more common in North America at the state, county, and local level and several jurisdictions have published program evaluation findings. Similar to our results, the Toronto and Los Angeles evaluations found their disclosure programs were used by consumers and led to improved restaurant sanitary practices. Leslie found that mandatory posting of grade cards in Los Angeles County improved inspection scores after they controlled for restaurant characteristics. Similar to our findings, Toronto Public Health found overwhelming program approval by diners and that diners felt safer making purchases with their program. Both of these evaluations were also able to detect decreases in foodborne illness after program implementation.

A previous study of the NYC restaurant grading program analyzed a public-use restaurant inspection data set and concluded that the program was not associated with an improvement in scores. However, the analysis included only 17 complete months of inspection data after grading. We identified improvements in sanitary conditions only after the 2-year mark, which may partially explain the inconsistency in results. The previous analysis also did not account for overrepresentation of poorer-performing restaurants resulting from more frequent inspection for poorer performers after grading. By contrast, our regression analysis addressed oversampling by including random intercepts for individual restaurants.

**Limitations**

This study has certain limitations. We compared inspection performance across time among inspected restaurants. In our earliest period (July 2007–July 2008), about 25% of restaurants were uninspected because of reduced staffing and other inspectional priorities. Because initial inspection assignment before grading was random, we believe inspections during this period were not biased toward poorer-performing restaurants. Use of inspection scores over time may have also been problematic. Subtracting administrative violation points from pregrading inspection scores to make them comparable with grading scores may have underestimated inspection scores pregrading, because the scoring system did not always include points from every violation to calculate inspection score. The impact would be an underestimate of the success of the program. We were unable to find an adequate comparison group (e.g., nongraded jurisdiction) because of jurisdictional differences in food-safety regulations and inspection scoring systems, but we used time and within-restaurant analysis as controls to isolate the impact of the program over time.

---

**TABLE 2—Estimated Success in Scoring in the A-Range on Initial Inspection in Restaurants:**

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Inspections, No.</th>
<th>Model I: SR (95% CI)</th>
<th>Model II: SR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time period</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13–36 mo before grading</td>
<td>42 016</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>0–12 mo before grading</td>
<td>26 200</td>
<td>1.05 (1.01, 1.09)</td>
<td>1.05 (1.01, 1.09)</td>
</tr>
<tr>
<td>0–12 mo after grading</td>
<td>32 594</td>
<td>0.86 (0.83, 0.89)</td>
<td>0.87 (0.84, 0.90)</td>
</tr>
<tr>
<td>13–24 mo after grading</td>
<td>38 339</td>
<td>1.24 (1.20, 1.29)</td>
<td>1.26 (1.22, 1.31)</td>
</tr>
<tr>
<td>25–36 mo after grading</td>
<td>32 918</td>
<td>1.33 (1.29, 1.38)</td>
<td>1.35 (1.31, 1.40)</td>
</tr>
<tr>
<td>Season</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>July–September (Ref)</td>
<td>36 598</td>
<td>. . .</td>
<td>1.00</td>
</tr>
<tr>
<td>October–December</td>
<td>41 697</td>
<td>1.20 (1.16, 1.24)</td>
<td>1.26 (1.22, 1.35)</td>
</tr>
<tr>
<td>January–March</td>
<td>45 825</td>
<td>. . .</td>
<td>1.30 (1.26, 1.35)</td>
</tr>
<tr>
<td>April–June</td>
<td>47 947</td>
<td>. . .</td>
<td>1.20 (1.16, 1.24)</td>
</tr>
<tr>
<td>Chain restaurant</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No (Ref)</td>
<td>151 374</td>
<td>1.00</td>
<td>. . .</td>
</tr>
<tr>
<td>Yes</td>
<td>20 693</td>
<td>3.46 (3.31, 3.61)</td>
<td></td>
</tr>
</tbody>
</table>


*Model includes random intercepts for unique restaurants.

*Model includes random intercepts for unique restaurants and adjusts for restaurant characteristics. Similar to our studies before grading.

Finally, the NYC restaurant grading program involved multiple changes to the enforcement landscape—more nuanced risk-based inspection frequency, greater exposure of restaurants to the risk of fines, grade posting, improvements to online resources, and additional training opportunities. We cannot tease out which factors contributed most to improving hygiene or grades.

Conclusions

The results from our analysis indicate that the NYC restaurant letter-grading program exhibited a positive impact on restaurant hygiene, food-safety practices, and public awareness, suggesting that the program is an effective tool for improving food safety. Our analysis also identified violation areas that can be targeted for improvement in future program operations.

About the Authors

Melissa R. Wong, Wendy McKeel, Kazuiko Ito, and J. Bryan Jacobson are with the Bureau of Environmental Surveillance and Policy, New York City Department of Health and Mental Hygiene, New York, NY. Corinne Schiff and Daniel Kass are with the Division of Environmental Health, New York City Department of Health and Mental Hygiene, New York.

Correspondence should be sent to Melissa R. Wong, MPH, Bureau of Environmental Surveillance and Policy, New York City Department of Health and Mental Hygiene, 125 Worth St, 3rd Floor, RN 34E, New York, NY 10013 (e-mail: maowong2@health.nyc.gov). Reprints can be ordered at http://www.aph.org by clicking the “Reprints” link. This article was accepted October 17, 2014.

Contributors

M. R. Wong contributed to program evaluation design, performed analysis, and drafted the article. W. McKeel contributed to program evaluation design and assisted with drafting the article. K. Ito and J. B. Jacobson conducted the analysis and assisted with drafting the article. C. Schiff and D. Kass conceptualized the program and assisted with drafting the article. All authors helped to interpret findings and review drafts of the article.

Acknowledgments

This work was supported by federal funds from the Centers for Disease Control and Prevention (grants 5U01EH000692 and 1U38EH000939). Note. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

Human Participant Protection

The New York City Department of Health and Mental Hygiene institutional review board determined that the program evaluation protocol was not human participant research in accordance with 45 CFR Part 46.

References


