Bipartisan Policy Center

EVIDENCE
WORKS

CASES WHERE EVIDENCE MEANINGFULLY INFORMED POLICY

EDITORS

Nick Hart
Fellow
Bipartisan Policy Center

Meron Yohannes
Research Analyst
Bipartisan Policy Center

ACKNOWLEDGMENTS

DISCLAIMER

RECOMMENDED CITATION

Nick Hart and Meron Yohannes (eds.) *Evidence Works: Cases Where Evidence Meaningfully Informed Policy.* (Washington, D.C.: Bipartisan Policy Center, 2019).

13. MAKING STREET-LEVEL BUREAUCRACY WORK

Safer Food in Seattle and King County

Daniel Ho is the William Benjamin Scott and Luna M. Scott Professor of Law, Professor (by courtesy) of Political Science, and Senior Fellow at the Institute for Economic Policy Research at Stanford University. He served as a co-project lead for the Food Safety Inspection for King County.

Becky Elias was the manager of the food program in the public health department in Seattle, King County. She also served as a co-project lead.

Every year foodborne pathogens make 15 percent of the American public ill and lead to 128,000 hospitalizations.[1] The Government Accountability Office classifies the U.S. food safety system as "high risk" due to "inconsistent oversight, ineffective coordination, and inefficient use of resources."[2] Food safety inspectors play a critical role in protecting the public and are housed at the Food Safety and Inspection Service, the Food and Drug Administration, and local health departments. While there has been a popular push for greater consumer disclosure of inspection results, there is also widespread concern that enforcement efforts are inconsistent across and within departments.

Improving the effectiveness of food safety enforcement is critical for protecting public health, including efforts to ensure permitted businesses comply with food safety regulations. This challenge of quality and consistency is also not unique to food safety. Agencies such as the Patent and Trademark Office, the Nuclear Regulatory Commission, the Social Security Administration, and the Board of Veterans Appeals all struggle with similar challenges of quality and consistency in applying, enforcing, and administering the law. Quality-improvement initiatives that work therefore have implications for many other areas of governance characterized by the decentralized administration of complex bodies of law.[3,4]

ISSUE BACKGROUND

In 2001, King County was the first county in Washington State to make food safety inspections available to the public on the internet, responding to demands for increased transparency. Twelve years later, county residents organized a petition with 2,000 signatures, calling for an improved food safety rating system that would include still-greater

transparency, this time at the point of purchase. With increased public and media scrutiny, the county took steps to require restaurants to publicly display a summary rating from food inspectors so that consumers can make more informed dining decisions.

Stakeholders offered a series of recommendations during public meetings about the King County food safety system:

1. The system should use more than a single inspection to determine ratings. This was perceived as fairer and more informative than existing systems, overcoming conventional critiques of inspections as providing only a single snapshot in time.[5,6]

2. Ratings should convey relative performance. King County only permits businesses to be open when they meet a minimum safety standard; stakeholders perceived the rating should convey how well a business performs beyond meeting the minimum standards and how businesses fare relative to one another.

3. The rating system should consider inspection consistency. Stakeholders in the community, including restaurant operators and inspection staff, focused on consistency in light of perceptions and evidence of variability in inspection style and stringency, both in King County and other jurisdictions.

EVIDENCE AVAILABILITY

When King County's placarding efforts began, there was limited evidence to inform regulatory design and implementation. One study examined grading systems in 17 large jurisdictions, finding substantial variability in the design of such systems. Collecting inspection data from 10 jurisdictions, the study identified (a) substantial inconsistencies in how the same establishment was scored over time, and (b) substantial evidence of grade inflation after the adoption of grading.[7]

Some observational studies found health benefits comparing jurisdictions before and after adoption of restaurant grading.[8,9,10] The leading study of Los Angeles was confounded by the state's largest salmonella outbreak in Southern California occurring before Los Angeles implemented restaurant grading.[11] There is also mixed evidence about whether food safety inspection scores are correlated with foodborne illness outbreaks.[12,13,14]

With limited evidence available, King County was the first jurisdiction in the state to implement restaurant grading with a rigorous evaluation plan in mind.

Developing a stronger evidence base

In 2014, the Public Health department in Seattle and King County engaged what is now the Regulation, Evaluation, and Governance Lab (RegLab) at Stanford University to support the evaluation plan. RegLab is a research laboratory whose mission is to promote evidence-based regulatory policy and administration through rigorous demonstration projects using

data science and technology. RegLab and King County formed a collaborative agreement to develop and improve the evidence base for food safety enforcement. King County's food program is responsible for health inspections of the county's 11,000 permitted establishments. Three projects emerged out of this collaboration.

Evaluation of peer-review pilot program

Building on information from existing academic literature, the county and RegLab designed a pilot program for health inspectors to engage in peer review process for conducting inspections. Existing research suggested reviewing each other could improve the accuracy and consistency of inspections.[15] Using an experimental evaluation design, researchers randomly assigned half the health inspectors into the four-month trial program. In the program, inspectors spent one day each week visiting establishments with peers, independently scoring health code violations, and comparing differences. Surprisingly, inspectors cited major code items differently in 60 percent of inspections, even when observing identical conditions and agreeing on food risks. For a food safety rating system, these citation differences could affect restaurant inspection scores and summary ratings. As a result of the variations in citations, the county also developed standards for high-prevalence and inconsistently scored infractions and trained inspectors to improve their overall consistency. The evaluation results concluded that the peer-review program caused (a) a 17 to 19 percent increase in violations detected, and (b) an improvement in inter-inspector consistency based on independent inspections.[16]

Design of rating system

While peer review improved the consistency of inspections, substantial inter-inspector variability remained. To further improve consistency as part of the rating design, the partners used results from the peer-review evaluation and retrospective observational data to inform the rating methodology.[17] First, inspectors were more likely to consistently cite businesses with "critical" violations—those more directly linked to foodborne illness—during peer-review inspections. Second, analysis of observational data suggested that repeat violations did not substantially predict future performance. Third, to determine the number of inspections of a business before assigning a grade, the partners analyzed how well historical information predicted performance. While each additional inspection provided better predictions, the gains were more limited after four inspections.

Policy evaluation of rating system

RegLab developed an experimental evaluation to assess the health impacts of the new rating system as part of its implementation. Public Health agreed to roll out the grading system in phases, which both facilitated operational implementation and the evaluation.[18] In addition, due to concerns about how inspectors would issue citations as a result of having a grading system in place, an independent team of contractor inspectors, who were trained in part by the Food and Drug Administration, observed risk factors in establishments without delivering any inspection report to operators. This allowed for direct observation for

AS A RESULT OF BROAD ENTHUSIASM FOR THE PEER-REVIEW PROGRAM, THE FOOD PROGRAM MANAGEMENT TEAM INSTITUTED PEER REVIEW AS PART OF AN ONGOING QUALITY-ASSURANCE PROGRAM … PEER REVIEW AND OBSERVATIONAL EVIDENCE ALSO INFORMED THE DEPARTMENT'S GRADING-SYSTEM DESIGN.

risk-factor prevalence in food-handling practices, independent of the rating system. Public Health also designed a qualitative assessment through community-based participatory research to receive feedback from stakeholders.

Although early in the evaluation process, no appreciable effects were identified from the rating system on foodborne illnesses, foodborne hospitalizations, or risk factors after eighteen months into the roll-out. The rating system may have increased public engagement in the form of submission of complaints in the short run. On the other hand, inspectors decreased citation of critical violations in a way not reflected by noncritical violations of the risk-factor study, suggesting that the rating system has a direct deleterious effect on inspector stringency.

EVIDENCE USE

As a result of broad enthusiasm for the peer-review program, the food program management team instituted peer review as part of an ongoing quality-assurance program for all food safety inspectors on a monthly or bimonthly basis. In other words, the control group from the evaluation was brought into the peer-review process, and management continues to use reports from peer-review inspections to target training materials.

Peer review and observational evidence also informed the department's grading-system design. Researchers calculated ratings based on the relative performance of an establishment of average critical violations over the past four routine inspections. Researchers placed no special weight on repeat violations given lack of predictive power. They assessed relative performance within a ZIP code area in recognition of the fact that most dining choices are local, which also had the collateral benefit of adjusting for inspector differences, as inspectors are assigned principally based on ZIP code. This choice proved contentious to elected officials, but there are few alternatives that would provide meaningful distinctions between restaurants and account for inter-inspector and regional variability.

Finally, how the county will use the experimental evaluation remains to be seen. Due to political constraints, such as the popular enthusiasm for and the King County executive support of restaurant grading, it is unlikely that the county will abandon grading. The management team is considering additional quality-assurance and performance-management efforts to ensure accurate signposting and to mitigate the impact grading systems have on inspectors, causing them to be more reluctant in writing violations. The trial results may be the most beneficial for jurisdictions facing the choice of whether to adopt a grading system, as it provides the first rigorous evidence of the effects of restaurant grading.

LESSONS

- *Clear benefits to government-academic collaborations exist.* Academic-agency partnerships are a critical way to develop the evidence base when agency research resources may be lacking, and researchers can craft evaluations to address broader scholarly questions. Both the peer-review trial and the evaluation of the grading system were first of a kind in this particular context.

- *Rigorous evaluations can inform policy and operations.* Policy implications are strongest from the experimental evaluation of peer review, with implications across many domains of regulatory enforcement and adjudication. The evaluation of the rating system at this point, however, does not sustain popular enthusiasm for restaurant grading as a way to lower foodborne illness, as there was no evidence of any health benefits.[19] If anything, the direct effect on citation behavior, which leads to well-documented forms of grade inflation, is something that public health agencies will need to take seriously in training and managing frontline staff when implementing rating systems. The citation effects, however, also corroborate the design of the rating system to mute "citation behavior" (using relative rankings within each ZIP code area).

- *Practical implementation barriers and public expectations must be acknowledged.* Although the rating system was grounded in evidence, the complexity of the design also posed practical challenges. RegLab developed open-source software to implement the system, but it was nonetheless challenging to integrate into information-technology systems. In addition, the complexity of the rating system, most notably the ZIP code adjustment, may be difficult to explain to the lay public, causing challenges for public messaging and outreach. The underlying difficulty stems from a public interest for a rating system based on inspections never intended for that purpose, a challenge not unique to food safety inspections. Ultimately, there may be considerable misunderstandings of what can and cannot be expected of food safety inspections.[20]

- *Decision-makers want positive results*. Substantial political pressures and the desire for immediate results created tension around the grading evaluation. Once a grading system is implemented, the champions of such a system desire studies to generate positive effects. A chief virtue of the grading experiment is a pre-commitment to an analysis plan (preventing "specification searching" to reach desired results), but how that evidence is used in prospective operations remains an open question.[21]

- *Continued engagement for building evidence is key.* The case study illustrates the need for active and ongoing stakeholder relations through the system design, implementation, and evaluation to support the integrity of the project throughout. Notwithstanding the challenges mentioned above, the case study illustrates the tremendous scientific and public-policy payoffs to academic-agency collaborations like the Stanford RegLab-Public Health model. ◾

1   U.S. Centers for Disease Control and Prevention. "Estimates of Foodborne Illness in the United States: Findings." Last reviewed November 5, 2018.
    Available at: https://www.cdc.gov/foodborneburden/2011-foodborne-estimates.html.

2   U.S. Government Accountability Office. *High Risk Series: An Update*. GAO 15-290, 2015, 78.

3   Daniel Ho. "Does Peer Review Work. An Experiment of Experimentalism." *Stanford Law Review*, 69(1): 1-119, 2017.

4   Daniel Ho and Sam Sherman. "Managing Street-Level Arbitrariness." *Annual Review of Law and Social Science*, 13: 251-272, 2017.

5   Richard Boehnke and Catherine Graham. *International Survey on Public Posting of Restaurant Inspection Reports, and/or Grade Card Posting Schemes Based upon Health Inspections*. RHB Consulting & Associates, August 2000.

6   Chris Wiant. "Scores, Grades, and Communicating about Food Safety." *Journal of Environmental Health*, 61(9): 37-38, 1999.

7   Daniel Ho. "Fudging the Nudge: Information Disclosure and Restaurant Grading." *The Yale Law Journal*, 122(3): 522-851, 2012.

8   Ginger Zhe Jin and Phillip Leslie. "The Effect of Information on Product Quality: Evidence from Restaurant Hygiene Grade Cards" *The Quarterly Journal of Economics*, 118(2): 409-51, 2003.

9   T. Serapiglia, et al. "Association of Food Premises Inspection and Disclosure Program with Retail-Acquired Foodborne Illness and Operator Noncompliance in Toronto." *Journal of Environmental Health*, 70(1): 54-59, 2007.

10 Melissa Wong, et al. "Impact of a Letter-Grade Program on Restaurant Sanitary Conditions and Diner Behavior in New York City." *American Journal of Public Health*, 105(3): 81-87, 2015.

11 Daniel Ho, Zoe Ashwood, and Cassandra Handan-Nader. "New Evidence on Information Disclosure through Restaurant Hygiene Grading." *American Economic Journal: Economic Policy*, Forthcoming 2019.
    Available at: https://www.aeaweb.org/articles?id=10.1257/pol.20180230.

12 M.A. Cruz,, D. J. Katz, and J. A. Suarez. "An Assessment of the Ability of Routine Restaurant Inspections to Predict Food-Borne Outbreaks in Miami-Dade County, Florida." *American Journal of Public Health*, 91(5): 821-823, 2001.

13 K. Irwin, et al. "Results of Routine Restaurant Inspections Can Predict Outbreaks of Foodborne Illness: the Seattle-King County Experience." *American Journal of Public Health*, 79(5): 586-590, 1989.

14 T.F. Jones, et al. "Restaurant Inspection Scores and Foodborne Disease." *Emerging Infectious Diseases*, 10(4): 688-692, 2004.

15 Kathleen Noonan, Charles Sabel, and William Simon. "Legal Accountability in the Service-Based Welfare State." *Law & Social Inquiry*, 34(3): 523-568, 2009.

16 Ho. "Does Peer Review Work."

17 Daniel Ho, Zoe Ashwood, and Becky Elias. "Improving the Reliability of Food Safety Disclosure: A Quantile Adjusted Restaurant Grading System for Seattle-King County." Forthcoming, 2019.

18 Cassandra Handan-Nader, Daniel Ho, and Becky Elias. *Feasible Policy Evaluation by Design: A Randomized Synthetic Stepped-Wedge Trial in King County*. Stanford Institute for Economic Policy Research, Working Paper 18-040, 2018.

19 Archon Fung, Mary Graham, and David Weil Fung. *Full Disclosure* (New York: Cambridge University Press, 2007).

20 Lauren Dundes. "Scores and Grades: A Sampling of How College Students and Food Safety Professionals Interpret Restaurant Inspection Results." *Journal of Environmental Health*, 64(5): 14-22, 2001.

21 Edward E. Leamer. "Let's Take the Con Out of Econometrics." *The American Economic Review*, 73(1): 31-43, 1983.