

**UC Irvine**  
**UC Irvine Law Review**

**Title**

Equity in the Bureaucracy

**Permalink**

<https://escholarship.org/uc/item/9mr0c58p>

**Journal**

UC Irvine Law Review , 7(2)

**ISSN**

2327-4514

**Author**

Ho, Daniel E.

**Publication Date**

2017-06-01

# Equity in the Bureaucracy

Daniel E. Ho\*

*Many governments have proposed “equity initiatives.” Seattle-King County’s Equity and Social Justice initiative, for instance, calls for applying an “equity lens” to policy analysis and for “all county employees to advance equity through their daily work.” How should such initiatives be understood and implemented in bureaucratic decisions? This Essay argues that equity dimensions of regulatory enforcement are pervasive, if unrecognized. Based on collaborative work with the food-safety program of the largest health department in Washington State (Public Health—Seattle and King County), I use large-scale inspection microdata, merged with census and social media information, to illustrate three such dimensions. First, as an empirical matter, equity considerations pervade the exercise of discretion by the many frontline employees who implement an agency’s policies. When it comes to enforcement—including equity considerations—an agency is a “they,” not an “it.” Second, I demonstrate that recent proposals for food-safety agencies to use big data to target inspections would have considerable distributive implications, shifting public enforcement resources away from minority and immigrant areas and importing the digital divide into regulatory enforcement. Third, information disclosure in the form of restaurant letter grading presents complex equity trade-offs. Conventional letter grading ignores differences in inspection stringency, resulting in dramatic geographic differences solely due to the identity of the inspector. I demonstrate how to design a more accurate and consistent grading system, but I also show that this system can magnify differences across cuisines. These case studies demonstrate both the profound*

---

\* William Benjamin Scott and Luna M. Scott Professor of Law, Stanford Law School; Senior Fellow, Stanford Institute for Economic Policy Research; Address: 559 Nathan Abbott Way, Stanford, CA 94305; Tel: 650-723-9560; Fax: 650-725-0253; Email: dho@law.stanford.edu. Thanks to Zoe Ashwood, Adrianna Boghozian, and Sam Sherman for terrific research assistance; to Becky Elias, manager of the Food Program in King County, for many insights, suggestions, and conversations about food equity; to participants at the 2016 “Establishing Equity in our Food System” symposium at UC Irvine School of Law; and to Damarys Espinoza, Stephen Lee, Benjamin van Rooij, and Laura Trice for helpful comments. No county provided any funding for this research.

*challenge and opportunity presented by data-driven approaches in giving content to “equity initiatives” in the bureaucracy.*

Introduction .....	402
I. Food Safety and Equity on the Frontline.....	407
A. Background.....	407
B. Inspection Data.....	413
C. Peer Review .....	417
II. Targeting.....	419
A. Promises .....	420
B. Challenges.....	423
III. Disclosure.....	433
A. Unadjusted Grading .....	434
B. Equitable Grading.....	437
IV. Implications.....	445
A. Peer Review and Data-Driven Training .....	445
B. Information Reporting and Information Disclosure.....	447
C. Language Access.....	447
Conclusion.....	449

## INTRODUCTION

On a cloudy Saturday in August, University of Washington attorney James Buder and elementary school teacher Deanna Buder took their four-year-old daughter, known to most as Scout, to a farmers market east of Seattle. They shared a plate of pork carnitas tacos from Chilangos, a food truck and catering business selling at seven local farmers markets under the direction of Mexico City–born chef Oscar Mendez. Several days later, Scout developed severe stomach cramps, fever, and diarrhea. The symptoms worsened, and two weeks later, the Buders rushed Scout to the Seattle Children’s Hospital, where she was diagnosed with *E. coli*.<sup>1</sup> Scout experienced kidney failure and was placed in the intensive care unit (ICU), receiving blood transfusions and dialysis. Said Deanna, “We couldn’t pinpoint it because we hadn’t eaten beef, and that’s what you think of when you think of *E. coli*.”<sup>2</sup>

When Scout was hospitalized, the health department (Public Health–Seattle and King County) investigated. Interviews with thirteen individuals infected with

---

1. See Complaint for Personal Injury and Damages at 5, *Buder v. Mendez Bros.*, No. 15-2-25355-7 (Wash. Super. Ct. Oct. 15, 2015); Tom Corrigan, *County Still on Track of E. Coli Outbreaks Linked to Local Farmers’ Markets*, ISSAQUAH PRESS, Sept. 10, 2015, at 1.

2. Tina Patel, *6 Sickened by E. Coli Linked to Local Food Vendor at Farmers Markets*, Q13 FOX NEWS (Sept. 1, 2015, 9:58 PM), <http://q13fox.com/2015/09/01/health-officials-say-e-coli-outbreak-linked-to-vendor-at-farmers-market> [<https://perma.cc/6ZPW-BMD5>].

*E. coli* revealed they had all eaten at Chilangos.<sup>3</sup> The department issued a cease and desist order to Chilangos and shut down its shared “commissary kitchen,” used by eleven other vendors to prepare food.<sup>4</sup> As the news broke, the local CBS broadcast reported the story with the headline, “*E. coli* cases linked to food sold at farmers markets,” along with pictures shared over social media of Scout in the ICU.<sup>5</sup> Commentary speculating about the outbreak was relentless. Some blamed it on culture: “This is not a race issue, but a cultural one. In some cultures, to see a fly or cockroach in the kitchen is ‘no big deal’ and food handling safety is not a top priority.”<sup>6</sup> Others assailed local food: “This is the problem with farmers market food. It can never get traced back, all evidence is long gone and no one is ever held accountable.”<sup>7</sup> And the Buders themselves denounced not just Chilangos, but also county officials for lax enforcement and maintenance of food-safety standards.<sup>8</sup> While suspicion initially fell on meat, further investigation focused on cilantro, but the ultimate source could not be identified. County officials worked with vendors to provide additional food-safety training to all affected staff, and Chilangos reopened roughly a month later.<sup>9</sup>

Fortunately, after three weeks of hospitalization, Scout recovered. Represented by Marler Clark, a leading foodborne illness litigation firm that sued Jack in the Box in 1993, the Buders have filed suit against Chilangos, asserting strict product liability and negligence claims.<sup>10</sup>

Scout’s tragic story highlights profound issues in food-safety regulation and enforcement in a highly fragmented and decentralized system. Each year, some 48 million Americans become sick from foodborne illness, 128,000 are hospitalized, and 3000 die.<sup>11</sup> Particularly vulnerable populations include children, the elderly, pregnant women, and immunocompromised patients.<sup>12</sup> Roughly 60% of

---

3. See Lindsay Bosslet, *Public Health Investigates E. Coli Outbreak*, PUB. HEALTH INSIDER (Sept. 1, 2015), <http://publichealthinsider.com/2015/09/01/public-health-investigates-e-coli/> [<https://perma.cc/X75W-DXL2>].

4. *Id.*

5. *Health Investigators: E. Coli Cases Linked to Food Sold at Farmers Markets*, KIRO 7 (Sept. 1, 2015, 3:54 PM), <http://www.kiro7.com/news/e-coli-case-being-investigated-king-county/27086005> [<https://perma.cc/R87N-LTPD>].

6. Kristin Morgan, Comment to *Update: 9 Confirmed, 1 Probable Case in Seattle E. Coli Outbreak*, FOOD SAFETY NEWS (Sept. 4, 2015) [hereinafter *9 Confirmed in Seattle E. Coli Outbreak*], <http://www.foodsafetynews.com/2015/09/6-e-coli-cases-linked-to-mexican-food-sold-at-washington-farmers-markets/> [<https://perma.cc/39YE-K9EZ>].

7. Megan, Comment to *9 Confirmed in Seattle E. Coli Outbreak*, *supra* note 6.

8. See Patel, *supra* note 2; Cathy Siegner, *9 Confirmed in Seattle E. Coli Outbreak*, *supra* note 6.

9. Bosslet, *supra* note 3.

10. See Complaint for Personal Injury and Damages, *supra* note 1.

11. See Ctrs. for Disease Control and Prevention, *Estimates of Foodborne Illness in the United States*, U.S. DEP’T HEALTH & HUMAN SERVS., <http://www.cdc.gov/foodborneburden> [<https://perma.cc/E7M8-E45D>] (last updated Aug. 19, 2016).

12. U.S. Food & Drug Admin., *Foodborne Illness: Especially Dangerous for the Vulnerable*, U.S. DEP’T HEALTH & HUMAN SERVS., FOOD & DRUG ADMIN. [FDA], <http://www.fda.gov/ForConsumers/ConsumerUpdates/ucm354783.htm> [<https://perma.cc/QUH7-GK9U>] (last updated Sept. 20, 2016).



documented outbreaks came from restaurants,<sup>13</sup> and securing regulatory compliance with these establishments is the purview of states, counties, and municipalities.

Many of these jurisdictions have proposed so-called “equity initiatives,” calling for agencies to take action to reduce inequities and to promote social justice.<sup>14</sup> In King County, all executive agencies are charged to apply an “equity lens” to policy initiatives and to “establish systems to engage and empower all county employees to advance equity through their daily work.”<sup>15</sup> Yet what does equity mean in the bureaucratic context? When and how should line-level food-safety officials “advance equity” in the inspection process? How can administrative agencies apply an “equity lens” in implementing the health code to preserve cultural differences associated with food while protecting the public at large? How should food-safety officials accurately and equitably disclose information about food-safety risks? And what distributive implications does media or social media have on the deployment of public resources?

This Essay argues, based on inspection data from the largest health department in Washington State (Public Health—Seattle and King County), that

---

13. CTRS. FOR DISEASE CONTROL & PREVENTION, *New CDC Data on Foodborne Disease Outbreaks*, U.S. DEP’T HEALTH & HUMAN SERVS., <http://www.cdc.gov/features/foodborne-diseases-data> [https://perma.cc/JU9H-9BES] (last updated May 27, 2015).

14. Of the twenty largest counties in the United States, eight counties (apart from King County) have established initiatives to promote equity and/or social justice. *See, e.g.*, News Release, Miami-Dade Cty, Board of County Commissioners Passes Legislation Requiring Social Equity Statements on All New Ordinances, (Sept. 1, 2015), <http://www.miamidade.gov/district08/releases/2015-09-01-social-equity-statements-on-all-new-ordinances.asp> (requiring social equity statements for ordinances); *County Equity Oversight Panel*, BD. OF SUPERVISORS, CTY. OF L.A., <http://bos.lacounty.gov/About-Us/Executive-Office-of-the-Board/County-Equity-Oversight-Panel> [https://perma.cc/TD2Y-YBM9] (last visited Mar. 3, 2016) (describing equity oversight panel); *Chronic Disease and Health Equity (CDHE) Unit*, SAN DIEGO HEALTH & HUMAN SERVS. AGENCY, [http://www.sandiegocounty.gov/content/sdc/hhsa/programs/phs/chronic\\_disease\\_health\\_disparities.html](http://www.sandiegocounty.gov/content/sdc/hhsa/programs/phs/chronic_disease_health_disparities.html) [https://perma.cc/CE2A-WBGK] (last visited Mar. 3, 2016) (discussing responsibility of unit to promote “wellness and prevention of illness, disability, and premature death from chronic disease by increasing health equity”); *Center for Health Equity*, N.Y. DEP’T HEALTH & MENTAL HYGIENE, <http://www1.nyc.gov/site/doh/health/neighborhood-health/center-for-health-equity.page> [https://perma.cc/44MJ-XRJI] (last visited Mar. 3, 2016) (discussing mission to “strengthen and amplify the Health Department’s work to eliminate health inequities, which are rooted in historical and contemporary injustices and discrimination, including racism”); *Health Equity*, HARRIS CTY. PUB. HEALTH, <http://publichealth.harriscountytexas.gov/Resources/Health-Data/Health-Equity> [https://perma.cc/C9U6-WFGB] (last visited Mar. 3, 2016) (discussing initiative to strengthen the “health equity capacity of [its] workforce” and to promote policies with a “health equity lens”); *Human Relations Commission*, CTY. OF SANTA CLARA, <https://www.sccgov.org/sites/ohr/human%20relations%20commission/Pages/default.aspx> [https://perma.cc/8HB]-THKH] (last visited Mar. 3, 2016) (discussing Human Relations Commission’s responsibility to “advise the Board of Supervisors on issues that affect the human and civil rights of all County residents and advocate for and take positive action to eliminate prejudice and discrimination based on race, religion, national origin, cultural background, sex, sexual orientation, age, disability or other factors”).

15. KING CTY. EXEC., ACO 9-2 (AEO), ADVANCING EQUITY AND SOCIAL JUSTICE THROUGH DEVELOPMENT OF A STRATEGIC INNOVATION PRIORITY PLAN AND EXECUTIVE DEPARTMENT ACTIONS (2014).

equity dimensions of regulatory enforcement are pervasive, if unrecognized. Part I demonstrates that individual inspectors exhibit wide variability in enforcement styles, including on equity grounds. Many qualitative reports suggest that jurisdictions with large immigrant populations encounter difficulties in food-safety enforcement in “ethnic” establishments due to linguistic, cultural, and food preparation differences.<sup>16</sup> Consistent with these reports, I show empirically that differences in individual inspection styles can be broken down into two distinct dimensions. First, inspectors vary generally in their stringency of code enforcement. Second, inspectors vary—independent of general stringency—in their stringency specific to ethnic establishments, specifically Asian restaurants in King County. These two dimensions combine to generate the most inspection uncertainty for Asian establishments across different inspectors. I then provide suggestive evidence from a randomized controlled trial showing that peer review has the potential to reduce differences in how Asian establishments are inspected.<sup>17</sup>

The Essay then considers two policies often advanced by New Governance approaches<sup>18</sup>: first, improving regulatory performance by risk targeting of enforcement resources, and second, disclosure based on letter grading of establishments.<sup>19</sup> Although these approaches are seductive and intuitively appealing, I show that their distributive implications can be complex and their health consequences nonobvious.

Part II argues that agencies should be cautious about targeting, the idea that enforcement resources should be increasingly targeted toward high-risk establishments. A particularly novel and widely reported idea is to use social media, such as Yelp reviews or Twitter feeds, to target inspections. One widely cited

---

16. See, e.g., *Chinatown Neighborhood Ass’n v. Harris*, 794 F.3d 1136 (9th Cir. 2015) (affirming dismissal of challenge to California’s “Shark Fin Law”); Kimberly J. Harris et al., *Food Safety Inspections Results: A Comparison of Ethnic-Operated Restaurants to Non-Ethnic-Operated Restaurants*, 46 INT’L J. OF HOSPITALITY MGMT. 190 (2015); Jee Hye Lee et al., *Popular Ethnic Foods in the United States: A Historical and Safety Perspective*, 13 COMPREHENSIVE REV. FOOD SCI. & FOOD SAFETY 2 (2014); Alison Rudder, *Food Safety and the Risk Assessment of Ethnic Minority Food Retail Business*, 17 FOOD CONTROL 189 (2006); Stephanie K. Baer, *Peking Duck Is So Important to Chinese Culture It Got a Health Code Exception: Here’s Why*, SAN GABRIEL VALLEY TRIB. (Aug. 22, 2015), <http://www.sgvtribune.com/health/20150822/peking-duck-is-so-important-to-chinese-culture-it-got-a-health-code-exception-heres-why> [<https://perma.cc/SLK8-8PMW>]; Eveline Chao, *The Roast Duck Bureaucracy*, OPEN CITY (Mar. 11, 2014), <http://opencitymag.com/the-roast-duck-bureaucracy/> [<https://perma.cc/JVR7-XFCM>]; Lillian G. Po et al., *Food Safety Education for Ethnic Audiences*, FOOD SAFETY (June/July 2011), <http://www.foodsafetymagazine.com/magazine-archive1/junejuly-2011/food-safety-education-for-ethnic-audiences/> [<https://perma.cc/G3A9-F6FM>].

17. The full results of the peer review randomized controlled trial are reported in Daniel E. Ho, *An Experiment of Experimentalism: Does Peer Review Work?*, 69 STAN. L. REV. 1 (2017).

18. Although varied, “New Governance” theories depart from traditional command-and-control models in favor of decentralized, collaborative, and performance-based solutions. Such approaches, for instance, favor information disclosure over bans and decentralized learning over hierarchical control. See Gráinne de Búrca & Joanne Scott, *Introduction: New Governance, Law and Constitutionalism*, in LAW AND NEW GOVERNANCE IN THE EU AND THE US 1, 2–3 (Gráinne de Búrca & Joanne Scott eds., 2006).

19. *Id.*

algorithm, for instance, uses descriptions like “Mexican” and “Vietnamese” that would pose serious questions under antidiscrimination law.<sup>20</sup> Moreover, I show empirically that because social media use is highly skewed along demographic lines, such targeting would shift enforcement resources away from areas that are more heavily populated by minorities, immigrants, and the less educated. These findings inform longstanding concerns about the digital governance divide and equity implications of regulatory use of big data.<sup>21</sup>

Part III examines the distributive implications of restaurant grading, which has been hailed as the poster child for information disclosure.<sup>22</sup> By converting numerical inspection scores into letter grades of A, B, and C, and posting these in entryways, many argue that regulators can nudge consumers to act as if fully informed and to incentivize operators to improve safety practices. But grading, as conventionally practiced, fails to take into account the core challenge of vast differences in inspection styles among frontline inspectors.<sup>23</sup> I present collaborative work with King County to design a more accurate, equitable, and meaningful grading system. Adjusting for inter-inspector differences across neighborhoods can improve accuracy and consistency and mitigate differences solely due to the identity of the inspector. On the other hand, such adjustments magnify differences between Asian and non-Asian restaurants, showing that reducing inequity along one dimension (neighborhoods) can heighten concerns along another.

Part IV discusses how equity considerations might better be addressed. Our research suggests that peer review and data-driven training can help resolve difficult equity questions with greater care and consistency. While data-driven efforts to increase the reporting rate are promising, the more important efforts involve training frontline inspectors and educating food staff in principles of food science, which can be aided by performance data, as well as peer review, but cannot be sidestepped by algorithm or grade. King County has taken bold and commendable steps in hiring multilingual staff and providing materials in multiple languages from which other counties should learn, but there remains considerable room for

---

20. See WASH. REV. CODE § 49.60.400 (2013).

21. See FED. TRADE COMM’N, BIG DATA: A TOOL FOR INCLUSION OR EXCLUSION? UNDERSTANDING THE ISSUES (2016), <https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf> [https://perma.cc/UJN5-H7QC]; John Clayton Thomas & Gregory Streib, *The New Face of Government, Citizen-Initiated Contacts in the Era of E-Government*, 13 J. PUB. ADMIN. RES. & THEORY 83, 98–99 (2003) (finding that users of government websites are “better educated, more likely to be white, wealthier, and younger than Internet users in general”); Kay Lehman Scholzman et al., *Weapon of the Strong? Participatory Inequality and the Internet*, 8 PERSP. ON POL. 487, 489–490, 503 (2010) (finding strong associations between socioeconomic status and most measures of Internet-based political engagement); Michael J. Magro, *A Review of Social Media Use in E-Government*, 2 ADMIN. SCI. 148, 155 (2012) (“The digital divide is a major barrier to e-participation.”).

22. See ARCHON FUNG, MARY GRAHAM & DAVID WEIL, FULL DISCLOSURE: THE PERILS AND PROMISE OF TRANSPARENCY 82–83 (2007); Omri Ben-Shahar & Carl E. Schneider, *The Failure of Mandated Disclosure*, 159 U. PA. L. REV. 647, 743–48 (2011).

23. See Daniel E. Ho, *Fudging the Nudge: Information Disclosure and Restaurant Grading*, 122 YALE L.J. 574, 592 (2012).

improvement regarding language access. The Essay concludes by noting that quantitative research can help to concretize otherwise diffuse notions of “equity” and that rigorous evaluation of equity initiatives can promote more effective deployment of public resources.

### I. FOOD SAFETY AND EQUITY ON THE FRONTLINE

The frontline of the food-safety system consists of local health inspectors who exhibit considerable variability in experience, expertise, and discretion. While much of administrative law and regulatory theory considers agencies to be monolithic, this Part makes one core point: equity dimensions of regulatory enforcement play out through the discretionary decisions of numerous frontline employees. In this sense, an agency is a “they,” not an “it.”<sup>24</sup>

Section I.A provides background on King County and the contentious question of how food-safety inspections engage with Asian establishments. Some of the most disputed elements are the role of culture, language, and the engagement of immigrant communities with the law. Section I.B analyzes nearly 60,000 inspections from 2007–2014, matched with establishment information from Yelp, to empirically demonstrate that inspectors vary along two critical dimensions: general stringency and stringency specific to Asian establishments. Section I.C then turns to the information from the peer review to show that while such differences still materialize under a peer review intervention, they are less pronounced. The intervention suggests that peer review can help a department grapple with tough issues surrounding how to inspect Asian establishments.

If an agency is a “they,” peer review promotes acting more like an “it.”

#### A. Background

As depicted in Figure 1, King County is a fast-growing area with a substantial immigrant population, due in part to the county’s historical openness to refugees.<sup>25</sup> From 1990 to 2011, the county’s population grew by 29%, with nearly 60% of the population growth coming from immigration, roughly half from Southeast Asia.<sup>26</sup> Asians comprise the largest minority group in the county, constituting roughly 16% of the population, with roughly 48% of those foreign-born.<sup>27</sup> Of individuals

---

24. With apologies to Kenneth A. Shepsle, *Congress is a “They,” Not an “It”: Legislative Intent as an Oxymoron*, 12 INT’L REV. L. & ECON. 239 (1992).

25. Elizabeth Rhodes, *New Immigrants Changing Look of Life in Seattle*, SEATTLE TIMES (Oct. 10, 1993), <http://community.seattletimes.nwsourc.com/archive/?date=19931010&slug=1725371> [https://perma.cc/46XP-TLT6].

26. CHANDLER FELT, KING CTY. COUNCIL, KING COUNTY’S CHANGING DEMOGRAPHICS: A VIEW OF OUR INCREASING DIVERSITY, 11–12 (2013), <http://www.kingcounty.gov/~media/exec/PSB/documents/AGR/KingCountyDemographics2012.ashx?la=en> [https://perma.cc/F3MM-C6T6].

27. See U.S. Census Bureau, *State & County QuickFacts: King County, Washington*, U.S. DEP’T COMM., <http://www.census.gov/quickfacts/table/PST045215/53033,00> [https://perma.cc/GL7E-CWW8] (last visited Feb. 26, 2016).

speaking an Asian language at home, 43% have limited English proficiency.<sup>28</sup> The restaurant sector is, if anything, even more diverse. By one estimate, roughly 20% of the 11,500 permitted food establishments serve Asian cuisines,<sup>29</sup> 35% of restaurant workers were born abroad, and 29% of food preparation workers have limited English proficiency.<sup>30</sup> As the *Seattle Times* reported, immigrants “who lack English end up working within their ethnic communities—in restaurants, small groceries and the like.”<sup>31</sup>

---

28. U.S. Census Bureau, *American FactFinder: King County, Washington*, U.S. DEP'T COMM., <http://factfinder.census.gov/> [<https://perma.cc/HDV7-C69J>] (last visited Feb. 26, 2016) (search “Advanced Search” for “King County, Washington;” then follow ID “S0505,” then “Selected Characteristics of the Foreign-Born Population by Region of Birth: Asia, 2010–02014 American Community Survey 5-Year Estimates” hyperlink).

29. This is based on all establishments that could be matched from Yelp to King County establishment data. *See infra* note 38.

30. 35% (+/- 1.5%) and 29% (+/- 1.4%) statistics calculated using 2010–02014 American Community Survey (ACS) 5-year Public Use Microdata Samples (PUMS) data. PUMS data obtained from U.S. Census Bureau, *American FactFinder: ACS Public Use Microdata Samples*, U.S. DEP'T COM., <http://factfinder.census.gov/> (search “Advanced Search” for “Public Use Microdata Sample;” select “2010-2014 ACS 5-year Public Use Microdata Samples (PUMS) – CSV format,” and download “Washington Population Records”) (last visited July 11, 2016). Estimates were calculated by selecting Public Use Microdata Areas (PUMAs) in King County, accounting for code differences between 2011 and 2012. We used Standardized Occupation Codes within the data to identify entries corresponding to restaurant workers: we defined a restaurant worker as a chef, head cook, “first-line supervisor of food preparation and serving worker,” cook, “combined food preparation and serving worker (including fast food),” waiter, waitress, dishwasher, host (restaurant, lounge, and coffee shop), or hostess (restaurant, lounge, and coffee shop). We defined “food preparation worker” as all of the above, with the exception of host, hostess, waiter, waitress, and dishwasher. We used the “Nativity” field in the PUMS data to identify people born overseas, and we used the “Ability to speak English” field to identify those who speak English less than very well. In order to calculate estimates and standard errors, we used the methodology outlines in the “ReadMe” accompanying the PUMS data. *See* U.S. CENSUS BUREAU, AMERICAN COMMUNITY SURVEY 2010–2014 ACS 5-YEAR PUMS FILES *README* (Jan. 21, 2016), [https://www2.census.gov/programs-surveys/acs/tech\\_docs/pums/ACS2010\\_2014\\_PUMS\\_README.pdf](https://www2.census.gov/programs-surveys/acs/tech_docs/pums/ACS2010_2014_PUMS_README.pdf) [<https://perma.cc/NY3J-VVP3>]. We used “Person’s Weight” and “Person’s Weight replicate” fields to calculate the number of restaurant workers satisfying the criteria expressed above, and to calculate the total number of restaurant workers in King County.

31. Rhodes, *supra* note 25.

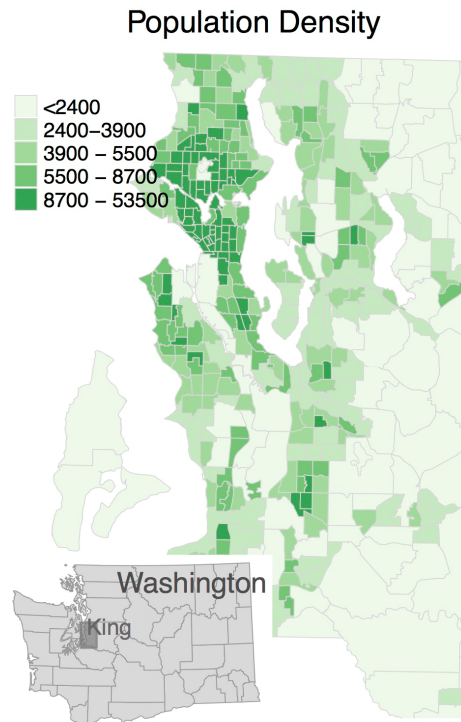


Figure 1: King County Population Density (2015). The bottom left insert plots counties in Washington State, with the gray box indicating the area of King County displayed in the main panel. King County displays census tracts, with darker shading corresponding to higher population density (broken into quintiles).

Restaurant inspections are conducted by the food-safety program in the Public Health Department (Seattle and King County). The program has fifty-five staff members, including thirty-four frontline inspectors assigned to inspect specific areas.<sup>32</sup> These areas are based primarily on ZIP codes and are switched in an “area rotation” once every few years. The typical full-service restaurant receives two routine inspections and one educational inspection per year.<sup>33</sup> Fifty violations are classified either as critical (red) or noncritical (blue) violations.<sup>34</sup> In a routine inspection, an inspector determines

32. See PUB. HEALTH – SEATTLE & KING CTY., KING COUNTY FOOD PROTECTION PROGRAM REVIEW: FINAL REPORT 1 (2014); *Common Questions About Food Safety*, PUB. HEALTH—SEATTLE & KING CTY, <http://www.kingcounty.gov/healthservices/health/ehs/foodsafety/foodfaq.aspx> [https://perma.cc/B5NQ-8GFU] (last updated Mar. 18, 2013).

33. Conventional full-service restaurants are typically classified as “risk III” establishments. Risk I and II establishments have more limited food preparation and are therefore subject to fewer inspections per year. See *Permanent Food Risk Based Inspection Program*, PUB. HEALTH—SEATTLE & KING CTY, <http://www.kingcounty.gov/depts/health/environmental-health/food-safety/food-business-permit/risk-levels-permit-classifications.aspx> [https://perma.cc/KX7U-25AQ] (last visited Nov. 4, 2016).

34. See *Food Establishment Inspection Report Form*, PUB. HEALTH—SEATTLE & KING CTY., <http://www.kingcounty.gov/healthservices/health/ehs/foodsafety/inspections/~media/health/>

whether to cite each of these violations, and each violation is assigned a fixed number of points, from two to twenty-five points.<sup>35</sup> County policies provide that thirty-five red points require a return visit within thirty days, and ninety red points result in immediate closure.<sup>36</sup> A food-safety training class is required of all workers.<sup>37</sup>

The food-safety program also conducts foodborne illness investigations based largely on reports of infections.<sup>38</sup> The county maintains a dedicated phone line staffed by a public health nurse to conduct a thorough intake process, recording all sources for meals, symptoms, and related facts.<sup>39</sup> As is well known in food science, public perceptions of foodborne illness are often misguided, and this intake process helps to focus departmental resources on well-founded claims. For instance, callers often present with symptoms that are inconsistent with claims about a particular establishment: e.g., eating at a restaurant and getting sick two hours later. During King County's Chipotle E. coli outbreak in 2014, the intake nurse received some 150 calls, but of these calls, fewer than fifteen presented with symptoms consistent with E. coli.<sup>40</sup>

For years, inspectors have struggled with how to accurately and consistently implement the health code, with tensions particularly acute with respect to Asian establishments. As part of King County's Equity and Social Justice initiative, the food-safety program engaged in substantial hiring efforts to diversify the cultural and linguistic backgrounds of inspectors.<sup>41</sup> Roughly 22% of the staff speaks a foreign language, covering Mandarin, Cantonese, Vietnamese, Korean, Cambodian, Tagalog, Amharic, Japanese, and Punjabi.<sup>42</sup> Some inspectors have tried on a limited basis to use translation services. For instance, one inspector noted on the inspection report that he "[r]eviewed [the] inspection report with translator, daughter of owner, on the telephone." Another reported "[f]ac[ing] difficulty communicating with the operator. Will do an educational visit with Vietnamese interpreter to go over proper food safety process." Food Worker Manuals are available in English, Cambodian, Chinese, Korean, Punjabi, Russian, Spanish, Tagalog, Thai, and Vietnamese.<sup>43</sup>

---

publichealth/documents/foodsafety/inspectionform.ashx [https://perma.cc/985E-U7MV] (last visited Sept. 10, 2016).

35. Two violations are disaggregated based on severity, so our analysis at the violation-level examines fifty-two violations. *See id.*

36. *See Common Questions About Food Safety, supra* note 32.

37. *Id.*

38. *See Foodborne Illness Complaints*, PUB. HEALTH—SEATTLE & KING CTY., <http://www.kingcounty.gov/healthservices/health/communicable/diseases/FoodBorneIllness.aspx> [https://perma.cc/32RE-EZZY] (last visited Sept. 8, 2016).

39. *See How to Report Possible Foodborne Illness*, PUB. HEALTH—SEATTLE & KING CTY., <http://www.kingcounty.gov/healthservices/health/ehs/foodsafety/inspections/illness.aspx> [https://perma.cc/3BVL-M4RV] (last visited Sept. 10, 2016).

40. Telephone Interview with Becky Elias, Pub. Health—Seattle and King Cty. (Feb. 7, 2016).

41. *Id.*

42. *Id.*

43. *See Food Worker Manual*, WASH. ST. DEP'T OF HEALTH, <http://www.doh.wa.gov/CommunityandEnvironment/Food/FoodWorkerandIndustry/FoodWorkerManual> [https://perma.cc/BY6A-V838] (last visited September 26, 2016).

As a sign of staff tension, a local television news broadcast in 2013 reported on allegations that the department was “go[ing] easy” on Asian restaurants.<sup>44</sup> One inspector appeared as an anonymous whistleblower, accusing the department of “turning a blind eye” to ethnic restaurants.<sup>45</sup> In 2006, another sued for wrongful termination purportedly after “refusing to go easy on ethnic restaurants.”<sup>46</sup> The suit was settled for \$125,000.<sup>47</sup> Another former inspector claimed, “If we close down too many Asian restaurants, then it’s going to start looking like we are singling them out and discriminating against them.”<sup>48</sup> At the same time, the news report also noted that nearly 80% of restaurants with the highest number of health violations serve Asian food.<sup>49</sup> The Director of Public Health’s response: “All restaurants, ethnic restaurants and other restaurants, have to follow the same high standard.”<sup>50</sup> Others believe that rigid code application can systematically disadvantage certain ethnic cuisines with distinct food preparation techniques. As one commentator wrote, “Restaurant regulation is a complicated issue, especially when concerning restaurants run by immigrants to the country, people who were raised eating food prepared and treated differently.”<sup>51</sup>

Tensions in bureaucratic equity are by no means unique to King County. In New York, major misunderstandings between restaurant inspectors and restaurateurs, for instance, have surrounded the application of the code to Chinese roast duck,<sup>52</sup> Korean kimchi,<sup>53</sup> and Japanese sushi rice.<sup>54</sup>

Because the department does not itself retain cuisine information, we merged information from Yelp, based on a record linkage algorithm using name similarity,

---

44. *Whistleblower: Health Inspectors Turning Blind Eye to Ethnic Restaurants* (KIRO 7 television broadcast May 6, 2013), <http://barfblog.com/2013/05/do-health-inspectors-turn-blind-eye-to-ethnic-restaurants/> [https://perma.cc/68AU-GTUA].

45. *Id.*

46. *Id.*

47. *Id.*

48. *Id.*

49. *Id.*

50. *Id.*

51. *Editorial: Restaurant Scrutiny Good, But Don’t Jump to Conclusions*, NORTHWEST ASIAN WEEKLY., June 1–7, 2013, <http://www.nwasianweekly.com/2013/05/editorial-restaurant-scrutiny-good-but-dont-jump-to-conclusions/> [https://perma.cc/PN9K-WL6P].

52. See Chao, *supra* note 16 (describing history of citations of roast ducks when one food science writer opined, “[t]he fact of the matter is that they cook the crap out of it, the skin is dry, they baste it when it’s up on the thing so there’s very little water activity, and the stuff underneath has been killed pretty good”).

53. Sumathi Reddy, *Restaurants Sour on Rules Over Kimchi*, WALL ST. J. (Dec. 20, 2011), <http://www.wsj.com/articles/SB10001424052970204879004577108632273070376> [https://web.archive.org/web/20170105204944/http://www.wsj.com/articles/SB10001424052970204879004577108632273070376] (describing enforcement of cold holding temperatures against kimchi, despite acidity level, and quoting one restaurateur as saying, “[the inspectors] don’t understand the kimchi”).

54. Jill Colvin, *Restaurant Owners Skewer Health Department Grading System*, DNAINFO (Mar. 7, 2012, 7:29 PM), <https://www.dnainfo.com/new-york/20120307/manhattan/restaurant-owners-skewer-health-department-grading-system> [https://web.archive.org/web/20170105205040/https://www.dnainfo.com/new-york/20120307/manhattan/restaurant-owners-skewer-health-department-grading-system] (reporting that the owner of Sushi Yasuda “received three different answers from three different inspectors about rules overseeing sushi rice”).



phone number, and geocoded addresses.<sup>55</sup> Figure 2 provides descriptive statistics of the relative performance across cuisines types. Each dot represents the average number of critical (red) violation points, with the dashed line representing the county average for the first routine inspection in 2014. It becomes immediately apparent that Asian cuisines (e.g., Indian, Japanese, Chinese, Thai, and Vietnamese) fare worse than Western cuisines. There is no evidence, on the other hand, for the cultural claim made in the context of the Chilangos outbreak. Establishments serving Mexican food perform the same as the average King County restaurant.

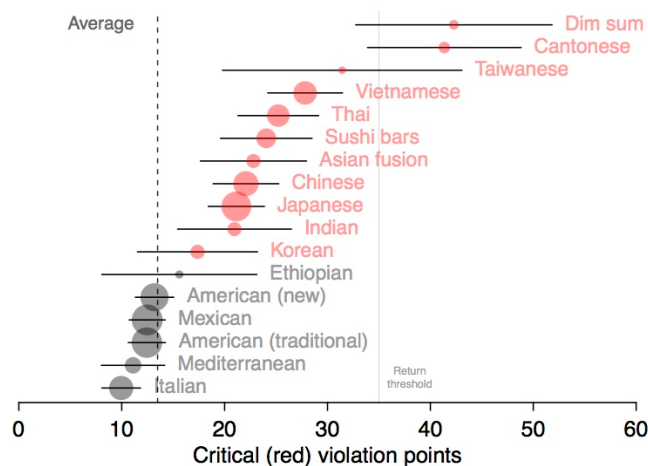


Figure 2: Average number of critical (red) violation points by cuisine type. Each dot represents a (nonexclusive) cuisine classification according to Yelp, weighted by the number of establishments. Horizontal lines indicate 95% confidence intervals. The dashed vertical line represents the average number of red violations. These statistics are calculated for the first routine inspection for an establishment in the 2014 calendar year. Only cuisine types with at least twenty establishments are presented.

These descriptive statistics might of course stem from underlying differences in food-safety practices across establishment types, so to understand how individual inspectors grapple with these cuisine differences, we turn to more comprehensive data.

55. We collected information for 12,025 establishments from Yelp. This list was intentionally overinclusive, including, for instance, drug stores, gas stations, and wineries. To cast the search as widely as possible, we utilized multiple search strings (e.g., “Food,” “Restaurants,” “General Food”) across all cities and neighborhoods in King County. In order to merge the Yelp establishment information with King County’s data, we regularized names and addresses in Yelp and King County data and geocoded addresses to augment and validate county GIS information, using ArcGIS and Google’s API. Our matching algorithm matches exactly based on ZIP code, phone number, and similarity of addresses and establishment names. Out of 11,568 establishments to be inspected in 2014, we identify matches for 53% of establishments, a higher rate than that of Jun Seok Kang et al., *Where Not to Eat? Improving Public Policy by Predicting Hygiene Inspections Using Online Reviews*, in PROCEEDINGS OF THE 2013 CONFERENCE ON EMPIRICAL METHODS OF NAT. LANGUAGE PROCESSING (EMNLP) 1443 (2013).

### B. Inspection Data

We study data from 58,594 routine inspections conducted by thirty-four frontline inspectors at 4930 high-risk restaurants for which cuisine classifications could be matched from Yelp from 2007–2014.

We hypothesize that equity considerations can affect the exercise of discretion in two ways. First, the primary effect may be on the general stringency level of inspections. Consider an analogous university setting with no enforced grade curve. Equity considerations may lead some instructors to grade leniently across the board. The observable implication is that inspectors most concerned with equity considerations might inspect all establishments more leniently. Second, due to cultural and linguistic factors, inspectors may differ in stringency specifically with regard to Asian establishments.<sup>56</sup> For instance, in the case of an Asian establishment with a large number of potential violations, one inspector might cite only the most critical violations, given the potential language difficulties of conveying remedial measures. Another inspector, however, might decide simply to write up all of the violations, on the assumption that only a return visit or closure will generate compliance. In psychometric terms, we might refer to this propensity by some inspectors who are otherwise similarly stringent to inspect Asian establishments differently as “differential item functioning” (DIF).<sup>57</sup>

The left panel of Figure 3 examines to what extent the raw difference for Asian establishments exists across inspectors. Each gray dot represents the scores awarded by an individual inspector to non-Asian establishments on the  $x$ -axis and those awarded to Asian establishments on the  $y$ -axis. The difference is stark. Nearly every inspector assigns worse inspection scores to Asian establishments, strongly suggesting that the performance difference reflects underlying food-safety practices. The inspector differences, however, are not homogeneous. Some inspectors fall in the lower left corner and are more lenient overall, thereby also generating smaller absolute differences between Asian and non-Asian establishments. Amongst stricter inspectors, scoring at least ten points, there are dramatic differences in the relative performance of Asian to non-Asian establishments, suggesting the presence of DIF.

The middle panel of Figure 3 provides some intuition of such DIF in practice for one inspector. The  $x$ -axis presents the average score that peers assigned to an establishment and the  $y$ -axis presents how that inspector scored the same establishment. Each dot represents an establishment, weighted by the number of peers who have inspected it, with red indicating an Asian establishment. This panel shows two trends.

56. Although there are of course distinct Asian cuisines and communities, as well as deviations between cuisine and ethnicity of operators, to make the analysis tractable, we use the term “Asian establishment” to refer to establishments denoted by Yelp to serve Asian cuisines. These are primarily comprised of categories displayed in red in Figure 2, but also include smaller categories of cuisine (i.e., Asian Fusion, Hot Pot, Indonesian, Malaysian, Mongolian, Ramen, Shanghaiese, Szechuan, Himalayan Nepalese, Laotian, Pakistani, and Singaporean).

57. In the educational testing context, DIF refers to the fact that some test items function differently for test takers of equal ability. Here, Asian establishments may function differently for inspectors of otherwise equal stringency. *See generally* DIFFERENTIAL ITEM FUNCTIONING (Paul W. Holland & Howard Wainer eds., 1993).

First, the inspector generally appears to be more stringent than peers, assigning an average of 13.8 red points compared to 11.5 red points by peers, even though these are for the same establishments ( $p$ -value = 0.01). Second, stringency appears to uniquely affect Asian establishments: the fitted red line is uniformly higher for Asian establishments, compared to the gray line for non-Asian establishments. (If there were no inspector differences at all, the lines would fall on the forty-five-degree line.) Note that this is not simply an artifact of Asian establishments generally faring worse: if that were the case, the red dots would simply be shifted to the right and fitted lines would not diverge.

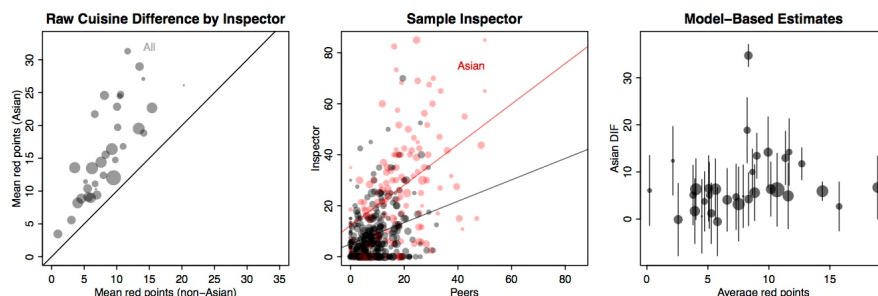


Figure 3: The left panel plots inspector average red points for non-Asian establishments on  $x$ -axis and Asian establishments on  $y$ -axis, with dots weighted by the number of Asian establishments. The middle panel presents how one inspector's scores compared to inspections of the same establishments by peers. Asian establishments receive higher violation scores, holding constant how other inspectors inspected those establishments. The right panel presents parameter estimates for stringency on the  $x$ -axis and Asian DIF on the  $y$ -axis, showing that inspectors appear distinct on these two dimensions. For establishments with identical inspection histories, Asian DIF represents the additional points that the specific inspector would assign to an Asian establishment versus a non-Asian establishment. The vertical lines are 95% confidence intervals on the DIF parameter.

To characterize this more formally, we fit the following regression:

$$E(Y_{ijt}) = \gamma_i + \phi_j + \lambda_i(A_j) + \alpha_t$$

where  $Y_{ijt}$  represents the number of red points assigned by inspector  $i$  to establishment  $j$  in year  $t$ . The first term on the right-hand side  $\gamma$  is an inspector fixed effect representing the general stringency of each inspector. The second term  $\phi$  is an establishment fixed effect, which controls for time-invariant, establishment-specific differences in food-safety practices. This term is primarily identified off of area rotations (and vacations, which lead some inspectors to conduct inspections outside of their home area) and accounts as well for general food-safety differences across Asian and non-Asian establishments. Because we control for establishment-specific attributes, the third term represents the DIF estimate for each inspector ( $\lambda$ ), where  $A_j$  is an indicator for whether establishment  $j$  serves Asian food. The last term  $\alpha$  is a year fixed effect, to account for general changes over years. We estimate the above equation with ordinary least squares, with the least stringent inspector as the baseline category, clustering standard errors at the employee level.

The right panel of Figure 3 presents estimates of general stringency on the  $x$ -axis and Asian DIF on the  $y$ -axis. Each dot represents the relative stringency and Asian DIF of an inspector.<sup>58</sup> These estimates suggest that there are indeed two distinct dimensions that characterize frontline inspectors. Some inspectors, for instance, are lenient generally, but relatively strict on Asian establishments. One inspector, who speaks Cantonese and Mandarin, falls in the midrange of general stringency, scoring roughly eight points on average for non-Asian establishments, but assigns thirty-five additional points to Asian establishments. Some inspectors in the bottom right part of the panel are generally tough but relatively lenient toward Asian establishments compared to their peers. The vertical (95% confidence) intervals are statistically distinct across different inspectors, suggesting that Asian DIF is unlikely due to chance alone. (We can reject the null hypotheses that (a) inspector effects jointly, or (b) DIF effects jointly are equal to zero.<sup>59</sup>)

It is important to note that there is no obvious “right answer” to code enforcement based on what we observe. Inspectors who inspect Asian establishments more stringently, for instance, could be biased against such establishments. But it is equally possible that these inspectors exhibit no bias, and that inspectors who inspect Asian establishments more leniently exhibit favoritism toward such establishments. Nor is it obvious a priori whether more stringent enforcement against Asian establishments is beneficial toward minority communities. Enforcement actions can negatively affect a business (e.g., via a requirement to install a new refrigerator), but can also protect its patrons. Absent more information (e.g., on customers, employment), one cannot determine whether a higher level of code enforcement would be beneficial to the Asian community per se.

---

58. These coefficient estimates are all relative to one baseline inspector, who, for ease of interpretation, is lenient in both dimensions.

59. The model explains roughly 32% of the variation in inspection outcomes based on  $R^2$ .

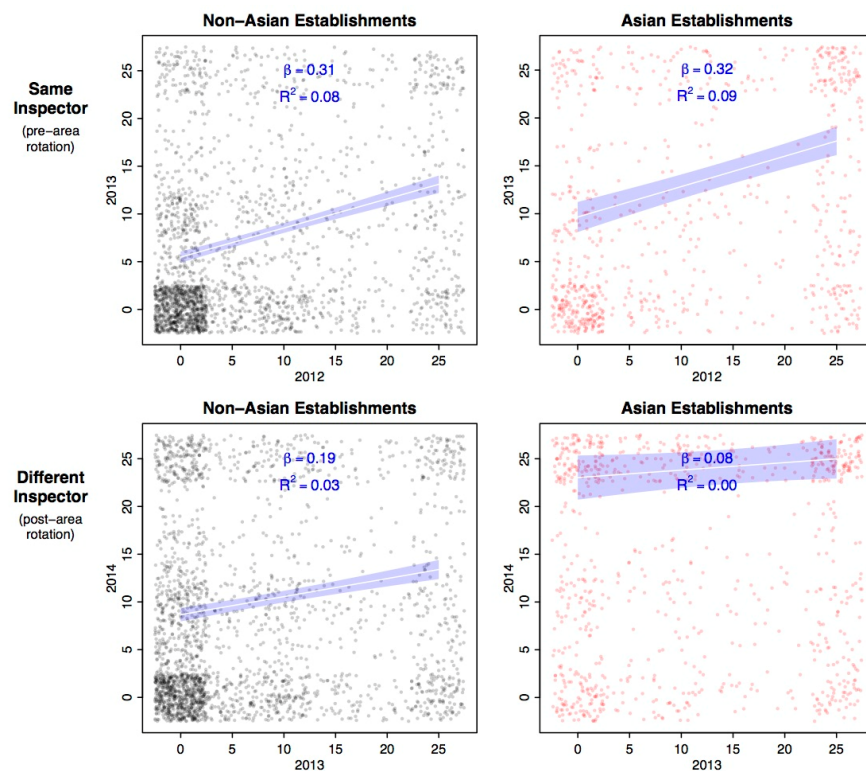


Figure 4: Consistency of inspections across inspection cycles before an area rotation (top panels) and after an area rotation (bottom panels). Each panel depicts the number of red points for an establishment (censored at twenty-five red points) across years.

Figure 4 shows how these two dimensions interact in practice. The top two panels present how the same establishment performs over two calendar years in routine inspections conducted by the same inspector (prior to an area rotation). Each dot represents an establishment (randomly jittered for visibility) and the blue bands present linear fits, with 95% confidence bands, for the correlation over cycles. The slope is positive, indicating that a higher score in 2012 is associated, albeit weakly, with a higher score in 2013. The degree of association is captured by the regression coefficient  $\beta$ , which indicates that a ten-point increase in 2012 is associated with a three-point increase in 2013. The association is essentially identical for Asian establishments presented in the top right panel. Roughly 8–9% ( $R^2$ ) of the variation in 2013 is explained by prior performance.

The bottom panels display what happens after an area rotation occurs in 2014 and a new inspector is assigned to these establishments. Comparing the strength of association after an area rotation provides us a sense of how differences in individual inspection styles affect the predictability of scores. For non-Asian establishments in the bottom left panel, the association decreases but is still statistically significantly positive—a ten-point increase in 2013 is associated with a roughly two-point increase

in 2014. The bottom right panel shows that this attenuation is far more severe for Asian establishments, so much so that there is no longer a statistically significant correlation across inspection cycles. Because differences in general stringency and Asian DIF interact in the bottom right panel, an area rotation leads to the greatest uncertainty for Asian establishments.<sup>60</sup>

This observational evidence corroborates the qualitative evidence that inspectors operationalize the food code and equity concerns quite differently, with unique consequences for Asian establishments.

### C. Peer Review

In 2015, we collaboratively designed a randomized controlled trial of peer review within the health department. For sixteen weeks, we randomly assigned half of the inspection staff to engage in joint inspections one day a week, during which they observed identical conditions on the ground, but independently scored health code violations for the establishment. Our general results are reported elsewhere,<sup>61</sup> but we focus here on how peer review might inform equity considerations, such as linguistic, cultural, and socioeconomic differences in establishments.

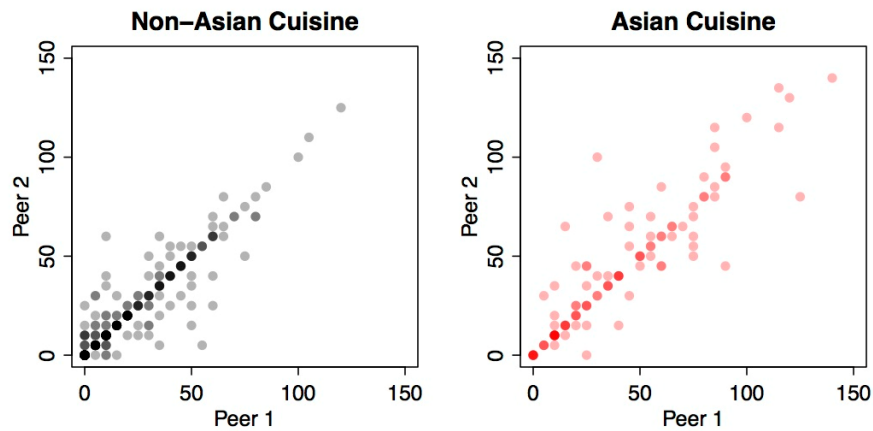


Figure 5: Correlation across pairs of peers during peer review inspections. Asian establishments generally score on the higher end of the range and as a result generate more opportunities for disagreement. Regressing Peer 2 scores against Peer 1 scores, an Asian establishment indicator, and an Asian interaction with Peer 1, we cannot reject the null hypothesis that the correlation between peers is identical across cuisine type ( $R^2 = 0.85$ ).

60. We found no evidence that stringency and DIF are correlated with the language skills of the inspector.

61. See Becky Elías & Daniel E. Ho, *Government Under Review: Could Peer Review for Public Servants Make the Law More Consistent?*, BOS. REV. (June 27, 2016), <https://bostonreview.net/editors-picks-us-books-ideas/daniel-e-ho-becky-elias-government-under-review> [h <https://perma.cc/QK9C-NFAD>]; Ho, *supra* note 17.

Figure 5 presents evidence of the correlation across peers visiting the same establishment. We see that the correlation is quite high for both non-Asian and Asian cuisines ( $R^2 = 0.85$ ), even though inspectors disagreed about specific health code citations nearly 60% of the time.<sup>62</sup> While Asian establishments receive higher scores, we cannot reject the null hypothesis that the correlation between peers is the same for either set of establishments. This suggests that Asian DIF is moderated in peer review. This result stands in contrast to the result in independent inspections, which appeared to exhibit greater inter-inspector inconsistency for Asian establishments (compare the bottom right panel of Figure 4). These results together suggest that peer review may be particularly useful for improving the consistency of inspections for high-scoring Asian establishments.

To understand how peer review might alleviate Asian DIF, we can examine comments submitted in a weekly survey of the peer review team, some of which mention the particular language difficulties encountered during joint visits. For instance, inspectors typically select comments on violations based on a drop-down field in a computer tablet. One inspector noted that while her “peer wanted to use the canned comments[,] I wanted to use my own comments because the people were English as a second language folks [and] I could use easier words for them to understand.”<sup>63</sup> Another language difficulty encountered was between the manager and staff of an establishment. One inspector reported, “The operator also asked me to provide her with Spanish language because her staff [was] mostly Hispanic food workers. I emailed the operator the food worker’s manual [in] Spanish, Thai and in English the same day.”<sup>64</sup> One supervisor who had not been in the field for some time noted, “[N]ot understanding a word that was said gave me a greater appreciation of ESL [English as a Second Language] difficulties in the field.”<sup>65</sup> Another indicated that peer review highlighted differences in enforcement, particularly surrounding language: “I learned that some of us will write the violation as they see it and some of us are interested in preventing foodborne illness by making sure the operator[s] understand food safety. The understanding part may require the inspector to spend more time educating the food workers using necessary tools (interpreter services, food worker’s manual in their own language, video) to achieve the goal.”<sup>66</sup>

These accounts from peer inspections demonstrate not only how differences in addressing language barriers can generate inconsistency in independent inspections, but also how peer review can transmit such skills and ultimately increase consistency in handling such issues.

\* \* \* \*

---

62. In 60% of nearly 400 peer inspections, inspectors disagreed on citing one or more code items.

63. Daniel E. Ho, Peer Review Comments Findings (2011) [hereinafter Peer Review Comments Findings] (some results from this source presented in Ho, *supra* note 17).

64. *Id.*

65. Ho, *supra* note 17, at 72 (citation omitted).

66. Peer Review Comments Findings, *supra* note 63.

To sum up, Part I has shown that equity considerations pervade the exercise of frontline discretion, a point largely overlooked by regulatory theories and administrative law doctrines that often treat an agency as monolithic. In this context, the challenge lies less in lack of awareness of equity dimensions, but rather in the sharp differences across inspectors regarding how such considerations are implemented on the ground. How should inspectors best overcome language barriers? To what extent is a more educational approach warranted when there may be cultural differences in food preparation? How can inspectors best apply food science to a wider range of food preparation techniques that can surface in non-Western cuisines? Answering such questions alone is challenging. Peer review, in the words of King County’s equity initiative, constitutes a promising “system[] to engage and empower [frontline inspectors] to advance equity through their daily work.”<sup>67</sup>

Parts II and III will show that equity dimensions also unexpectedly affect two popular regulatory strategies. We begin with attempts to “target” public resources based on big data.

## II. TARGETING

In 2012, the city of Boston developed an iPhone app called “Street Bump” that used the iPhone Accelerometer to detect street bumps for city repair. At first glance, it seemed ingenious. Rather than relying on spotty self-reports, Street Bump could efficiently trigger service requests, allocating resources where needed most. The first version, however, had very high false positive rates (“phantom potholes”<sup>68</sup>). And because iPhone users were not a random sample of the Boston population, the app appeared to deploy far more city resources to more affluent areas of Boston.<sup>69</sup> While such distributive implications can be addressed *ex post*,<sup>70</sup> the Street Bump example is instructive. Equity considerations should not be afterthoughts in the deployment of public resources. In evaluating current and proposed systems, policymakers should proactively identify, assess, and mitigate potential equity problems in the design of policies.

---

67. KING CTY. EXEC., *supra* note 15, at 3.

68. Phil Simon, *Potholes and Big Data: Crowdsourcing Our Way to Better Government*, WIRED, <http://www.wired.com/insights/2014/03/potholes-big-data-crowdsourcing-way-better-government/> [https://perma.cc/Z7U2-Q7SH] (last visited Feb. 27, 2016); see Civile, *Street Bump—Useful But Not Ready for Prime Time Engineering*, PUB. WORKS GRP. BLOG (Feb. 10, 2013), <http://www.publicworksgroup.com/blog/2013/02/street-bump-review/> [https://perma.cc/CA9E-56KM].

69. See Edith Ramirez, Chairwoman, Fed. Trade Comm’n, Opening Remarks at Big Data: A Tool for Inclusion or Exclusion?, at 13 (Sept. 15, 2014), [https://www.ftc.gov/system/files/documents/public\\_events/313371/bigdata-transcript-9\\_15\\_14.pdf](https://www.ftc.gov/system/files/documents/public_events/313371/bigdata-transcript-9_15_14.pdf) [https://perma.cc/V97U-WMYB]; FED. TRADE COMM’N, *supra* note 21, at 27–28.

70. See Ramirez, *supra* note 69, at 13 (noting that as a partial remedy, the City of Boston sent its own inspectors out with Street Bump apps to cover the city more equitably).



### A. Promises

Using data-driven techniques to target discretionary enforcement resources has been proposed, discussed, and implemented in many regulatory contexts, including criminal enforcement,<sup>71</sup> occupational safety and health inspections,<sup>72</sup> environmental inspections,<sup>73</sup> mining safety inspections,<sup>74</sup> and consumer safety.<sup>75</sup> Outside the enforcement policy context, prediction algorithms have been shown to help reduce Medicare costs by avoiding futile joint replacements,<sup>76</sup> to make bail decisions,<sup>77</sup> to target youth mentoring interventions,<sup>78</sup> to forecast domestic violence calls,<sup>79</sup> and to allocate community supervision of convicted criminals based on forecasts of the likelihood of murder.<sup>80</sup> Prior targeting approaches have used the agency's own data, but a newer proposal is to use "big data" from social media to target inspection resources. Such approaches have particular appeal in the area of food safety because of limited enforcement resources and the well-known difficulties of reporting, detecting, and tracing sources of foodborne illness.<sup>81</sup>

71. See ANTHONY A. BRAGA & DAVID L. WEISBURD, *POLICING PROBLEM PLACES: CRIME HOT SPOTS AND EFFECTIVE PREVENTION* (2010); PRISCILLA HUNT, JESSICA SAUNDERS & JOHN S. HOLLYWOOD, RAND CORP., *EVALUATION OF THE SHREVEPORT PREDICTIVE POLICING EXPERIMENT* (2014) (finding no statistically significant differences between treatment and control districts); G.O. Mohler et al., *Randomized Controlled Field Trials of Predictive Policing*, 110 J. AM. STAT. ASS'N 1399 (2015) (finding that model-based predictions outperformed analyst predictions).

72. See Sidney A. Shapiro & Randy S. Rabinowitz, *Punishment Versus Cooperation in Regulatory Enforcement: A Case Study of OSHA*, 49 ADMIN. L. REV. 713, 745 (1997).

73. See Eric Helland, *The Enforcement of Pollution Control Laws: Inspections, Violations, and Self-Reporting*, 80 REV. ECON. & STAT. 141 (1998).

74. See Alison D. Morantz, *Mining Mining Data: Bringing Empirical Analysis to Bear on the Regulation of Safety and Health in U.S. Mining*, 111 W. VA. L. REV. 45, 61 (2008) (discussing theory and history of targeted enforcement and arguing for application to mining safety).

75. See Richard A. Berk, *Forecasting Consumer Safety Violations and Violators*, in *IMPORT SAFETY: REGULATORY GOVERNANCE IN THE GLOBAL ECONOMY* 131 (Cary Coglianese et al. eds., 2009).

76. See Jon Kleinberg et al., *Prediction Policy Problems*, 105 AM. ECON. REV. 491, 493 (2015) (illustrating how mortality forecasts can save Medicare costs on joint replacements).

77. See LAURA & JOHN ARNOLD FOUND., *DEVELOPING A NATIONAL MODEL FOR PRETRIAL RISK ASSESSMENT* (2013), [http://www.arnoldfoundation.org/wp-content/uploads/2014/02/LJAF-research-summary\\_PSA-Court\\_4\\_1.pdf](http://www.arnoldfoundation.org/wp-content/uploads/2014/02/LJAF-research-summary_PSA-Court_4_1.pdf) [<https://perma.cc/VP7W-94WN>]; Shaila Dewan, *Judges Replacing Conjecture with Formula for Bail*, N.Y. TIMES, June 28, 2015, at A18.

78. See Dana Chandler, Steven D. Levitt & John A. List, *Predicting and Preventing Shootings Among At-Risk Youth*, 101 AM. ECON. REV. 288, 288 (2011).

79. See Richard A. Berk, Yan He & Susan B. Sorenson, *Developing a Practical Forecasting Screener for Domestic Violence Incidents*, 29 EVALUATION REV. 358, 360 (2005).

80. See Geoffrey C. Barnes et al., *Low-Intensity Community Supervision for Low-Risk Offenders: A Randomized, Controlled Trial*, 6 J. EXPERIMENTAL CRIMINOLOGY 159, 186 (2010); Richard Berk et al., *Forecasting Murder Within a Population of Probationers and Parolees: A High Stakes Application of Statistical Learning*, 172 J. ROYAL STAT. SOC'Y 191, 208 (2009).

81. According to the CDC, "[s]urveillance statistics [of foodborne illness] reflect a fraction of cases that occur in the community. Underdiagnosis and underreporting of foodborne illnesses present challenges for surveillance and the detection of outbreaks." Ctrs. for Disease Control and Prevention, *Foodborne Illness Surveillance Systems*, U.S. DEP'T HEALTH & HUMAN SERVS. (2011), <http://>

I here review some of the leading work in the food-safety context.

Chicago's Cross-Agency Data Project. The city of Chicago conducted a pilot project that merged multiple municipal databases to forecast which establishments were likely to have critical violations.<sup>82</sup> The city tested the algorithm for a two-month period in 2014, comparing violations detected in field visits to predicted violations had inspections counterfactually been prioritized by the predictive algorithm. Predictors included variables internal to the food-safety program, such as the inspector assigned, the type of facility, prior critical violations, and length of time since operation and last inspection, but also included information outside the food program such as nearby garbage and sanitation complaints and burglaries. The retrospective study concluded, without reporting statistical significance, that critical violations would, on average, be found 7.5 days earlier with the forecasting exercise.<sup>83</sup>

Sadilek Twitter Targeting. Adam Sadilek and coauthors studied how Twitter data could be used to allocate enforcement resources.<sup>84</sup> They downloaded all tweets in Las Vegas for a three-month period, associated these tweets with restaurants within fifty meters, and collected tweets by these users for the following five days.<sup>85</sup> They used Amazon's Mechanical Turk to recruit humans to answer for a sample of tweets the question: "Do you think the author of this tweet has an upset stomach today?"<sup>86</sup> Using this hand-coded sample as the training dataset, they applied machine-learning techniques (support vector machines) to classify all tweets for upset stomachs.<sup>87</sup> Top predictive terms, for instance, included the words "stomach" and "stomachache."<sup>88</sup> They then used these predictions to schedule health inspections with restaurants.<sup>89</sup> Working together with the Southern Nevada Health District, they sent inspectors to pairs of establishments, one being identified via Twitter and the other used as a control.<sup>90</sup> Inspectors were not informed which establishment was targeted by the algorithm.<sup>91</sup> For a sample of 142 total inspections, they found that Twitter-identified

---

[www.cdc.gov/foodborneburden/PDFs/FACTSHEET\\_G\\_SURVEILLANCE.pdf](http://www.cdc.gov/foodborneburden/PDFs/FACTSHEET_G_SURVEILLANCE.pdf) [<https://perma.cc/G3YV-E3LU>].

82. See *Food Inspection Forecasting: Optimizing Inspections with Analytics*, CITY OF CHI. [hereinafter *Food Inspection Forecasting*], <http://chicago.github.io/food-inspections-evaluation/> [<https://perma.cc/KEP3-LBNQ>] (last visited Feb. 26, 2016).

83. See Julian Spector, *Predictive Policing Comes to Restaurants*, ATLANTIC (Jan. 7, 2016), <http://www.theatlantic.com/technology/archive/2016/01/predictive-policing-food-poisoning/423126/> [<https://perma.cc/P4ZA-8QFA>].

84. See Adam Sadilek et al., *Deploying nEmesis: Preventing Foodborne Illness by Data Mining Social Media*, PROCEEDINGS OF THE 28TH CONF. ON INNOVATIVE APPLICATIONS OF ARTIFICIAL INTELLIGENCE 3982 (2016), <http://www.aaai.org/ocs/index.php/IAAI/IAAI16/paper/view/11823/12317> [<https://perma.cc/B8M7-2PW8>].

85. *Id.* at 3983.

86. *Id.* at 3985.

87. *Id.* at 3986.

88. *Id.* at 3983 fig.2.

89. *Id.* at 3983.

90. *Id.*

91. *Id.* at 3985.

establishments had nine violation points, compared to six points for control establishments.<sup>92</sup>

Chicago Twitter Reporting. Chicago deployed Twitter primarily to increase the reporting rate of foodborne illnesses in the city.<sup>93</sup> In 2013, the city set up a website (Foodborne Chicago) for online reporting of foodborne illnesses.<sup>94</sup> From March 2013 to January 2014, the city collected 2241 tweets mentioning the term “food poisoning,” which were then reviewed by staff for signs consistent with foodborne illness.<sup>95</sup> Two hundred seventy of these tweets were deemed genuine reports, resulting in a Twitter reply to encourage the individual to report the incident via the website.<sup>96</sup> A total of 193 individuals reported the incident via the web form, resulting in 133 additional inspections.<sup>97</sup> During the same period, Chicago conducted 1941 complaint-based inspections; the rates of detection of critical violations (roughly 92%) were comparable across these two samples.<sup>98</sup>

New York Yelp Targeting. New York City used Yelp review information to increase the reporting rate analogously to Chicago’s use of Twitter. Based on 294,000 Yelp reviews, the city’s Department of Health and Mental Hygiene retrospectively searched for food poisoning terms (“sick”, “diarrhea”, “vomit”, and “food poisoning”), yielding 893 reviews.<sup>99</sup> Epidemiologists manually read the reviews for signs consistent with a foodborne illness outbreak, namely whether the incubation period exceeded ten hours and whether multiple individuals appeared to be affected.<sup>100</sup> This manual review yielded 129 individuals, whom the department attempted to contact via Yelp.<sup>101</sup> Of these, 102 did not consent to be interviewed, but of the twenty-seven who were interviewed, three outbreak cases were identified, though there was no laboratory confirmation due to the passage of time.<sup>102</sup>

Kang Yelp Targeting. Kang et al. scraped and merged data from Yelp with King County health inspection data from 2006–2013.<sup>103</sup> The dataset covered 1756 restaurants and 152,000 reviews.<sup>104</sup> Applying machine-learning techniques (support vector

---

92. *Id.* at 3987.

93. See Jenine K. Harris et al., *Health Department Use of Social Media to Identify Foodborne Illness—Chicago, Illinois, 2013–2014*, 63 MORBIDITY & MORTALITY WEEKLY REP. 681 (2014); *About, FOODBORNE CHI.* [hereinafter *Foodborne Chicago*], <https://www.foodbornechicago.org/pages/about> [<https://perma.cc/ZJ2Z-Q8AD>] (last visited Feb. 26, 2016).

94. Harris et al., *supra* note 93, at 681.

95. *Id.* at 682.

96. *Id.*

97. *Id.*

98. *Id.*

99. See Cassandra Harrison et al., *Using Online Reviews by Restaurant Patrons to Identify Unreported Cases of Foodborne Illness—New York City, 2012–2013*, 63 MORBIDITY & MORTALITY WEEKLY REP. 441, 441 (2014), <http://www.cdc.gov/mmwr/pdf/wk/mm6320.pdf> [<https://perma.cc/5HQ5-XRLA>].

100. *Id.*

101. *Id.* at 444.

102. *Id.* at 443–44.

103. See Kang et al., *supra* note 55, at 1444.

104. *Id.*

machines), they identified review terms and cuisines associated with inspection scores.<sup>105</sup> For instance, a review mentioning the term “friendly” was associated with lower violation scores,<sup>106</sup> while the terms “ramen”, “pho”, or “gross” were associated with higher violation scores.<sup>107</sup> Similarly, German and European cuisines were associated with lower violation scores,<sup>108</sup> and Vietnamese, Thai, Mexican, Japanese, Indian, and Chinese cuisines were associated with higher violation scores.<sup>109</sup> Using inspection history alone resulted in a classification accuracy rate of 72%, compared to 81% when using review information.<sup>110</sup>

### B. Challenges

While promising, naively adopting the above approaches for food-safety inspections poses profound challenges. This subsection discusses these challenges in antidiscrimination law, statistical validity, public misperceptions, unarticulated implicit normative commitments, and opportunity costs.

1. Antidiscrimination Law. Some algorithmic approaches are likely to run afoul of antidiscrimination law. Kang’s proposal, for instance, to target inspection resources away from European establishments toward Vietnamese, Thai, Mexican, Japanese, Indian, and Chinese establishments may well violate equal protection, Title VI, and/or state law.<sup>111</sup> Especially because the algorithm adds only marginal value on top of simply using the inspection history alone, using such suspect characteristics is likely a non-starter for a public agency.

These questions tie into a broader, and increasingly important, question about whether state agencies are allowed to take into account race, national origin, or gender in algorithmic decisions,<sup>112</sup> a debate that has been most developed in the criminal justice

105. *Id.* at 1446.

106. *Id.* at 1446 tbl.3.

107. *Id.* at 1446 tbl.2.

108. *Id.* at 1446 tbl.3.

109. *Id.* at 1446 tbl.2.

110. *See id.* at 1445–47 (determining accuracy by dichotomizing violation scores across a range of thresholds and calculating the ability of predictors to classify establishments correctly).

111. *See* 42 U.S.C. § 2000(d) (2012) (“No person in the United States shall, on the ground of race, color, or national origin, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any program or activity receiving Federal financial assistance.”); WASH. REV. CODE § 49.60.030(1) (2009) (“The right to be free from discrimination because of race, creed, color, national origin, sex, honorably discharged veteran or military status, sexual orientation, or the presence of any sensory, mental, or physical disability or the use of a trained dog guide or service animal by a person with a disability is recognized as and declared to be a civil right.”); *Castaneda v. Partida*, 430 U.S. 482 (1977); *Blair v. Wash. State Univ.*, 108 Wash. 2d 558 (1987) (finding state Law Against Discrimination to apply to public university’s athletics scholarship program); 45 C.F.R. § 80.3 (2015); Dow Constantine, *King County Title VI Policy Statement* (May 28, 2010), <http://kingcounty.gov/depts/civil-rights/title-nine.aspx> [https://perma.cc/9YTS-SDU8] (“King County further assures every effort will be made to ensure nondiscrimination in all of its programs and activities, whether those programs and activities are federally funded or not.”).

112. In the gender context, see, for example, *Ariz. Governing Comm. for Tax Deferred Annuity & Deferred Comp. Plans v. Norris*, 463 U.S. 1073 (1983) (finding that payment of lower retirement benefits to women based on actuarial statistics violated Title VII of the Civil Rights Act); *City of Los*

context.<sup>113</sup> For instance, the model by Berk et al. includes race and gender as predictors in a forecasting model of murder.<sup>114</sup> They argue that formalizing the predictive power of race would allow a polity to make an informed determination about the trade-off between reduced forecasting accuracy and racial neutrality.<sup>115</sup> In other contexts, algorithms might reduce bias compared to an alternative decision making system. The weight placed on race explicitly in a parole algorithm, for instance, may in fact be *lower* than the weight placed on race implicitly via human clinical assessment.<sup>116</sup> In the inspection context, however, the vast majority of inspections are routine inspections with fixed annual frequencies across all establishments. Relative to that baseline, an algorithm that prioritizes Asian establishments for routine inspections cannot possibly reduce national origin bias.<sup>117</sup>

Even setting aside the highly questionable use of cuisine categories, the use of reviews by private individuals in an unstructured setting can raise questions under antidiscrimination laws. Kang's algorithm, for instance, uses highly subjective language like "friendly" as a predictor of risk.<sup>118</sup> One study has shown that online reviews exhibit signs of bias, describing minority neighborhoods in reviews as "dark," "dangerous," and "sketchy."<sup>119</sup> Predictors that are themselves affected by racial bias—even if not racial classifications per se—may pose problems.<sup>120</sup> Of course, determining which predictors are subject to racial bias is a difficult determination: for instance, if reviewers were more

Angeles, Dep't of Water & Power v. Manhart, 435 U.S. 702 (1978) (finding that a municipal agency requirement that women make larger contributions to pension funds based on actuarial statistics violated Title VII of the Civil Rights Act).

113. See, e.g., Jordan M. Hyatt & Richard A. Berk, *Machine Learning Forecasts of Risk to Inform Sentencing Decisions*, 27 FED. SENT'G REP. 222, 227 (2015) (arguing that we do not know whether racial bias is more or less pronounced with actuarial vs. clinical judgments in the sentencing context); Anna Maria Barry-Jester et al., *Should Prison Sentences Be Based on Crimes that Haven't Been Committed Yet?*, FIVETHIRTYEIGHT.COM (Aug. 4, 2015, 11:42 AM), <http://fivethirtyeight.com/features/prison-reform-risk-assessment/> [<https://perma.cc/PA9C-JZPC>].

114. See Berk et al., *supra* note 80, fig. 1.

115. *Id.* at 201.

116. Cf. Ian Ayres & Sydney Foster, *Don't Tell, Don't Ask: Narrow Tailoring After Grutter and Gratz*, 85 TEX. L. REV. 517, 563 (2007) (arguing that the extent of the racial preference was empirically greater in a discretionary as opposed to quantified preference system).

117. A more nuanced claim could be made for return inspections, as discretion enters both the likelihood of scoring thirty-five or more points and when to return to the establishment. Proposals for targeting algorithms, however, have focused only on routine inspections and return inspections constitute a small fraction of the full inspection workload.

118. Kang et al., *supra* note 55, at tbl.3.

119. Sharon Zukin et al., *The Omnivore's Neighborhood? Online Restaurant Reviews, Race, and Gentrification*, J. CONSUMER CULTURE 1 (2015) (comparing reviews from a white-gentrifying and a black-gentrifying neighborhood in Brooklyn, New York, and finding that the former is more likely to be associated with Europe and the latter with loaded descriptions of the neighborhood).

120. The Arnold Foundation's bail algorithm, for instance, uses prior traffic violations, which could itself be a function of racial bias. See MARIE VANNOSTRAND & CHRISTOPHER T. LOWENKAMP, ARNOLD FOUND., ASSESSING PRETRIAL RISK WITHOUT A DEFENDANT INTERVIEW 12 tbl.4 (2013). See also Bernard E. Harcourt, *Risk as a Proxy for Race*, 27 FED. SENT'G REP. 237 (2015) ("[R]isk today has collapsed into prior criminal history, and prior criminal history has become a proxy for race.").

likely to call Asian restaurants “gross,” conditional on the same inspection score, this result may reflect (anti-Asian) bias from consumers or (pro-Asian) bias from inspectors.

2. Statistical Validity. While interesting as proofs-of-concept, big data techniques tend to gloss over conventional principles of statistical inference and research design, which threaten the validity of wide-scale deployment. We highlight here two common challenges with causal inference and representativeness.

(a) Causal Inference. Enthusiasm over big data often glosses over principles of causal inference. Big claims are made, but little attention is paid in predictive analytics to evaluating the actual impact of the potential intervention. The Chicago Cross-Agency Data Project estimated that critical violations could be detected 7.5 days earlier, without reporting statistical significance.<sup>121</sup> The intervention simultaneously introduced a web reporting form with Twitter outreach, and, as a result, it is impossible to disentangle the beneficial effects of the Twitter outreach from the effects of a web reporting form.

Sadilek’s Las Vegas study assesses the potential impact of targeting by providing a matched comparison on Twitter-targeted inspections versus control inspections.<sup>122</sup> But even here, inferences are limited. The paired comparisons include only 142 inspections.<sup>123</sup> The number of tweets could simply reflect the number of customers at (or close to) the establishment, regardless of safety practices. The matched comparison might hence be confounded by the popularity of an establishment. The more food is served, the easier it may be for an inspector to cite a violation, and the more patrons an establishment has, the higher the absolute risk even if per capita risk is the same, so the differences in point scores may be an artifact of differences in popularity. The ultimate insight from the Las Vegas study might then be that inspection resources should be targeted toward popular restaurants (with Tweets providing one measure of popularity), but state agencies hardly need social media to develop a risk-grading system based on the frequency of visits or revenue.<sup>124</sup>

(b) Representativeness. Public health agencies exist to protect the public at large, but social media data are not representative of the population as a whole. Consider Kang’s study, which was able to match only 1756 restaurants from Yelp to King County’s health inspection database.<sup>125</sup> Seattle, the city that Kang focused on, in fact

---

121. See *Food Inspection Forecasting*, *supra* note 82.

122. See Sadilek et al., *supra* note 84.

123. *Id.* at 3983.

124. *Id.* (incorporating seating levels and state income revenue data in a more comprehensive fashion to achieve the same aim).

125. Jun Seok Kang and his colleagues imply that the health department “has only limited resources to dispatch inspectors, leaving out a large number of restaurants with unknown hygiene grades” and infer that “[m]ore than 50% of the restaurants listed under Yelp did not have inspection records, implying the limited coverage of inspections.” Kang et al., *supra* note 55, at 1443–44. This reflects some basic misunderstandings about the inspection system. First, Yelp’s restaurant categories do not directly match the jurisdiction of King County’s food program. Nearly all permitted establishments are in fact visited by an inspector during the year. In 2014, for instance, 99% of risk III establishments received at least one visit. More importantly, matching Yelp entries to inspection data is not straightforward, as the names can differ considerably across the two datasets. For instance, the restaurant named “Von’s Gustobistro” in Yelp is named “1000 Spirits” in King County’s inspection

had over 7789 permitted establishments during that time period.<sup>126</sup> Are Yelp reviews representative? Sadilek's study uses exclusively individuals who tweeted within fifty meters of a restaurant.<sup>127</sup> How representative are those establishments of the population of Las Vegas restaurants? Table 2 in the Appendix provides some basic descriptive statistics about how comparable Yelp and Twitter users are to the nation. Unsurprisingly, social media users are younger, more educated, and wealthier.<sup>128</sup>

Figure 6 plots the distribution of the number of reviews across establishments in King County. The distribution is highly skewed, with a small number of establishments drawing the vast majority of Yelp reviews. Out of some 6000 matched establishments, the top 100 establishments account for over 25% of all reviews submitted. The modal establishment receives no reviews at all.

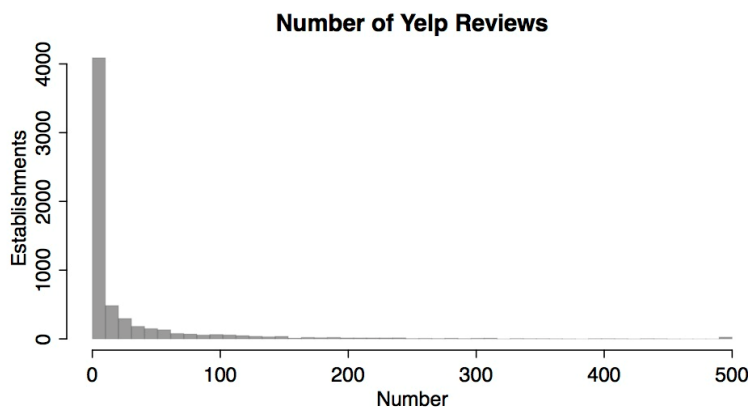


Figure 6: Distribution of number of Yelp reviews across establishments in King County. Very few establishments receive a high frequency of reviews.

To study how the Yelp targeting algorithm would affect the deployment of inspection resources in King County, we merge census information from the American Community Survey with our Yelp and inspection data. Figure 7 plots the proportion of Asians and median income levels across census tracts. From Figure 1, population density is highest in the Seattle area. Asians tend to live in the southern and eastern portions of

---

database. Our match is almost surely correct, given that the website refers to “Von’s Gustobistro 1000 Spirits.” See *100 Years, 1,000 Spirits*, VON’S GUSTOBISTRO, <http://www.vons1000spirits.com/> [<https://perma.cc/6N2N-CJ53>] (last visited Feb. 29, 2016).

126. This figure is based on all establishments subject to an inspection during Kang’s observation period from 2006–2013.

127. Sadilek et al., *supra* note 84, at 3984.

128. See Alan Mislove et al., *Understanding the Demographics of Twitter Users*, in PROCEEDINGS OF THE FIFTH INTERNATIONAL AAAI CONFERENCE ON BLOGS AND SOCIAL MEDIA 554, 557 (AAI Press 2011), <http://www.ccs.neu.edu/home/amislove/publications/Twitter-ICWSM.pdf> [<https://perma.cc/69FP-VXG2>] (“Twitter users significantly overrepresent the densely populat[ed] regions of the U.S., are predominantly male, and represent a highly non-random sample of the overall race/ethnicity distribution.”).

the county, with significant geographic clustering. The northern and eastern portions of the county tend to be higher income.

Figure 8 compares the restaurant distribution to the presence and depth of reviews on Yelp. The left panel plots the number of permitted establishments subject to inspections in King County. The middle panel plots the number of restaurants that could be matched from Yelp against King County data. Out of 11,500 permitted establishments, only around 6000 could be reliably matched. High schools, cafeterias, and nursing homes, for instance, are generally not reviewed on Yelp, but are subject to health inspections. The right panel displays the penetration by the number of Yelp reviews per establishment identified on Yelp. We observe a dramatic shift toward central and north Seattle to the detriment of the rest of King county. Targeting inspections based on such reviews would hence shift enforcement resources considerably across the county, undercutting the premise of joint city-county health agency.



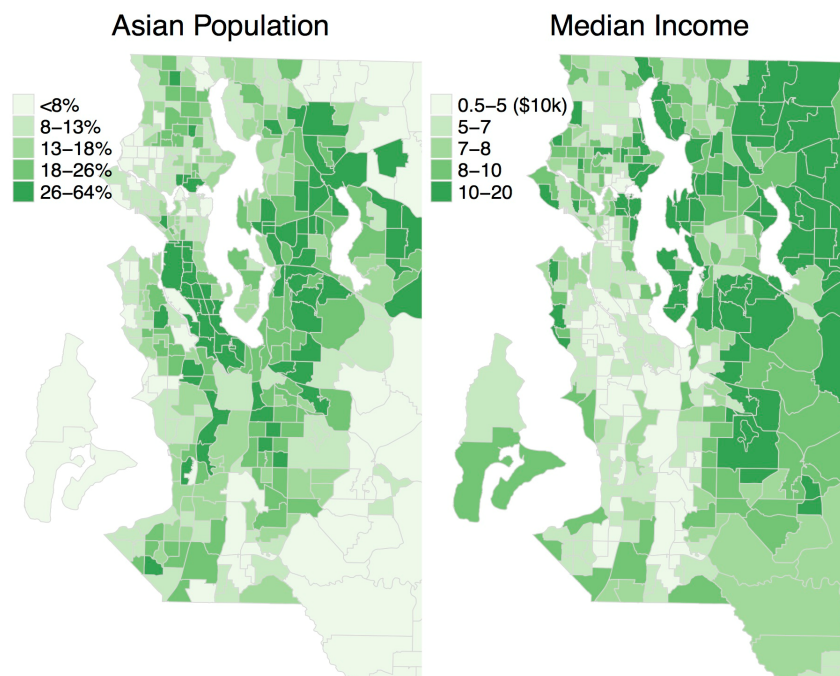


Figure 7: King County Demographics. The left panel plots the percentage of the population that identifies as Asian and the right panel the median household income by census tract.

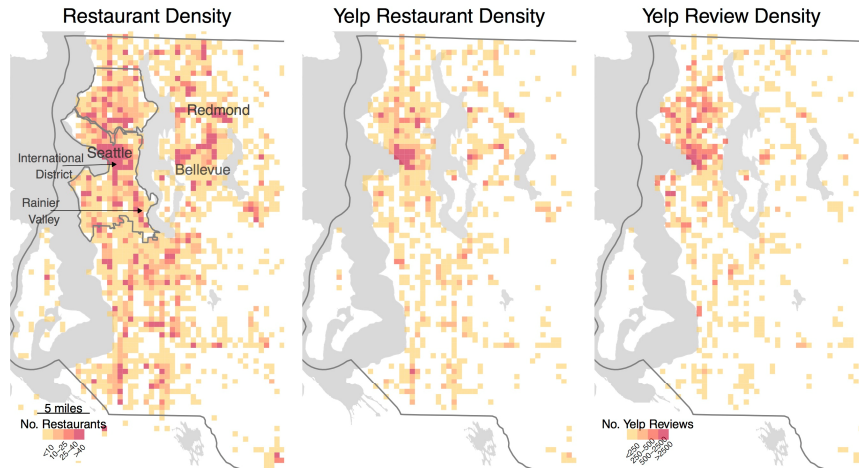


Figure 8: Yelp Penetration in King County. The panels compare all permitted establishments on the left panel to establishments matched on Yelp in the middle panel to the distribution of reviews (conditional on a match) in the right panel. This figure shows that targeting based on Yelp reviews would overwhelmingly focus inspection resources on central and north Seattle to the detriment of the rest of King County.

What equity implications would this shift have? Figure 9 shows how the penetration of Yelp reviews (on the  $y$ -axis) varies with demographic characteristics. Yelp reviews are far less prevalent in areas with high Asian, Hispanic, and limited English-speaking populations, and far more prevalent in areas that are disproportionately white.

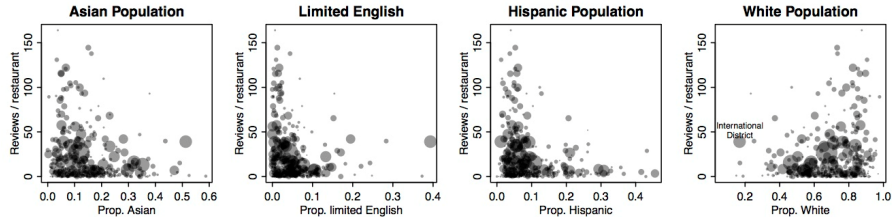


Figure 9: Correlates of Yelp review penetration. Each panel presents demographic attributes from the American Community Survey, with dots representing census tracts, weighted by the number of establishments. The  $y$ -axis represents the number of reviews per restaurant. The International District is an area with a high number of ethnic restaurants that is frequently visited by tourists.

To formally test this, we conduct simple regression tests (presented in the Appendix, Table 3). The Table confirms that Yelp presence and number of reviews are highly (and statistically significantly) correlated with demographic covariates. The presence of an establishment on Yelp and the number of reviews, for instance, are each negatively associated with the proportion of the population that is Asian, foreign born, and has only a high school education. A 10% increase in the Asian population is associated with a decrease in five to twelve reviews per establishment. These results

corroborate extant evidence that governmental use of online platforms can exacerbate the digital divide.<sup>129</sup>

3. Food Science vs. Food Perceptions. Targeting proposals cannot easily overcome the fundamental difficulties with detecting and investigating foodborne illnesses.<sup>130</sup> Outbreaks with unknown food or etiology comprise the majority of reported outbreaks.<sup>131</sup> Symptoms often manifest themselves after two days, but a common public misconception is that illness is caused by the last place one ate.<sup>132</sup>

Relying on freeform, online speculation may exacerbate these problems. Sadilek's study, for instance, exclusively identifies whether a tweet indicated a stomachache (regardless of whether foodborne at all) five days after posting a tweet within fifty meters of an establishment.<sup>133</sup> But the incubation period for common foodborne illnesses (e.g., campylobacteriosis or listeriosis) can be up to several days.<sup>134</sup> By failing to verify whether reporting is consistent with foodborne illness, the algorithm skips over a critical step of the intake process, thus making the prediction subject to a high degree of error.

New York's Yelp study is exemplary in its attempt to seriously vet the basis for complaints. But the effort of reviewing nearly 900 reviews and attempting to interview individuals resulted in only three suspected foodborne illnesses. The absence of any lab confirmation or identification of a pathogen made this information much less useful from a public health perspective. As seen in Figure 6, reviews are very sparse over time for the vast majority of establishments. This sparseness makes it nearly impossible for

---

129. Of course, while social media skew in one direction, administrative data might skew in another. Garbage complaints and burglary complaints, as used by Chicago, for instance, might actually result in greater deployment of resources toward poorer neighborhoods. None of this is to say that the addition of such information cannot be helpful, but government agencies must be mindful of the trade-offs involved. In other contexts, weighting and post-stratification may be a plausible way to adjust for sampling bias. YouGov, for instance, conducts post-stratification to estimate nationally representative parameters. See *Panel Methodology*, YOU GOV UK, <https://yougov.co.uk/about/panel-methodology/> [<https://perma.cc/NE29-XQVC>] (last visited Feb. 29, 2016). Here, however, the absence of *any* reviews for large numbers of restaurants makes such an adjustment infeasible.

130. See Clare Leschin-Hoar, *Outbreaks of Foodborne Illness Are Becoming Harder to Detect*, SCI. AM. (Nov. 26, 2012), <http://www.scientificamerican.com/article/food-poisoning-outbreaks-become-harder-to-detect/> [<https://perma.cc/UY3T-4SCF>] (noting how the increase in non-culture tests prevents DNA fingerprinting of pathogens and source tracing).

131. See Caroline Smith DeWaal, *Underreporting of Foodborne Illness: Strategies to Increase Awareness and Diagnosis*, CTR. FOR SCI. IN THE PUB. INT. (2006), [http://www.fsis.usda.gov/wps/wcm/connect/815f89ea-088a-4ba8-ad91-038a17248b62/Slides\\_092706\\_CSmithDeWaal.pdf?MOD=AJPERES&CACHEID=48d200fc-8ebb-4ab9-86d8-90f10b6a33d5](http://www.fsis.usda.gov/wps/wcm/connect/815f89ea-088a-4ba8-ad91-038a17248b62/Slides_092706_CSmithDeWaal.pdf?MOD=AJPERES&CACHEID=48d200fc-8ebb-4ab9-86d8-90f10b6a33d5) [<https://perma.cc/862Y-W44Z>].

132. See *Foodborne Disease: Frequently Asked Questions*, VA. DEP'T OF HEALTH, <http://www.vdh.virginia.gov/EnvironmentalHealth/FOOD/FoodSafety/FAQ/> [<https://perma.cc/L8NT-D4ZD>] (last visited Feb. 29, 2016) ("A common misconception is that gastrointestinal illness was caused by the last food item that was eaten before symptoms started.").

133. See Sadilek et al., *supra* note 84, at 3984.

134. *Foodborne Illnesses: What You Need to Know*, U.S. DEP'T OF HEALTH & HUMAN SERVS., FDA, <http://www.fda.gov/Food/ResourcesForYou/Consumers/ucm103263.htm> [<https://perma.cc/TNN6-29XW>] (last visited Mar. 4, 2016).

such information to be reliably used as a real-time monitoring device of foodborne illness.

Overcoming inaccurate public health perceptions is a common challenge for food safety, but social media reporting systems may exacerbate the distortion from such misperceptions.

4. Normative Desirability. The notion that algorithmically driven governance can sidestep normative and equity concerns is illusory. As we know from Part I, inspectors differ dramatically in stringency. Both Kang's algorithm and the Chicago Cross-Agency Data Project naively rely on inspection history to forecast violations.<sup>135</sup> This means that the algorithm may simply be predicting whether a tough inspector is inspecting the area. Perversely, the algorithm may hence shift enforcement resources *precisely into the wrong area*, namely one where inspectors are already stringently inspecting establishments.<sup>136</sup> Algorithmic allocation may hence be entirely circular: a tough inspector cites more violations, and therefore we should inspect more stringently. But the real public health problem and challenge of institutional design may be exactly the reverse. The greater concern may be about violations that are *not* scored because of the lenience of inspectors.

Similarly, it is not obvious whether information from social media should serve as a complement or as a substitute for enforcement resources. Implicitly, targeting algorithms assume the former. But if social media already disclose a suspected risk (publicly shaming an establishment into taking remedial measures), perhaps inspection resources should be deployed precisely where social media disclosures fail to function as a complement to conventional regulation. Concretely, if social media provide more information about food safety in central and northern Seattle, perhaps we should target inspection where there is an underprovision of food-safety information (i.e., the rest of King County). Understanding the direct effects of social media on establishment and consumer practices is a critical step in ascertaining whether, as a normative matter, such information should act as a complement or substitute for regulatory enforcement.

5. Opportunity Costs. The costs to large-scale predictive data analytics can be substantial. First, the direct engineering time can be nontrivial. Sadilek's Twitter project consumed six months of computer engineering to build the system and interface.<sup>137</sup> New York's Yelp system "required substantial resources; in addition to programming expertise, staff members were needed to read reviews, send emails, interview reviewers, and perform follow-up inspections."<sup>138</sup> Some twenty individuals were credited for developing the foodborne illness reporting system in Chicago.<sup>139</sup> Higher quality systems

---

135. Chicago, for instance, includes the identity of the inspector and Kang's algorithm includes ZIP codes that form the basis of inspector area assignments. *See* Harris et al., *supra* note 93, at 683; Kang et al., *supra* note 55, tbl.1.

136. Unlike in other contexts, objective outcome data are hard to come by, and laboratory-confirmed foodborne illnesses are rare events that cannot easily be used to train an algorithm.

137. *See* Sadilek et al., *supra* note 84.

138. Harrison et al., *supra* note 99, at 445.

139. *See Foodborne Chicago, supra* note 93.

(Chicago's reporting system and New York's Yelp system) required substantial human resources of food-safety professionals for their implementation.

Second, if deployed on a large scale, such systems introduce perverse incentives. Competitors might have the incentive to strategically tweet or mention food poisoning in reviews.<sup>140</sup> And restaurants could monitor Twitter and Yelp to anticipate when to expect an inspector, thereby undercutting the randomness of current, unannounced routine inspections.

While costs (and the digital divide) might decrease over time,<sup>141</sup> the core needs of health departments—and government agencies generally—may be met not so much by advanced machine learning algorithms as by basic information infrastructure internal to the agency. Information technology infrastructure is generally recognized to be weak in government agencies.<sup>142</sup> As a McKinsey report found, “Although many cities, counties, states, and agencies have chief information or chief technology officers, they often lack data expertise lower in the organization.”<sup>143</sup> King County, for instance, adopted computer tablets for food inspections, but failed to develop an adequate means for supervisors to review the data input by inspectors, making supervision and management more difficult. The opportunity cost of a public health nurse reading Yelp reviews may be time spent debriefing investigation results with frontline staff, a task that might ultimately produce better long-term consequences. After assessing Web 2.0 initiatives

140. Strategic sabotage by competitors via social media and the Internet has been well documented. *See, e.g.*, Mike Blumenthal, *Fake Reviews Starting to Get Mainstream Media Attention*, UNDERSTANDING GOOGLE MY BUS. & LOCAL SEARCH (Mar. 2, 2011), <http://blumenthals.com/blog/2011/03/02/fake-reviews-starting-to-get-mainstream-media-attention/> [<https://perma.cc/6QHE-NP8R>]; Samuel Gibbs, *Restaurant Owner Sues Google Over Maps Listing ‘Sabotage,’* THE GUARDIAN (July 8, 2014), <http://www.theguardian.com/technology/2014/jul/08/restaurant-owner-sues-google-over-maps-listing-sabotage> [<https://perma.cc/VV8H-CXBS>]; Susanna Kim, *Business Owners Suspect Sabotage Tactics from Competitors*, ABC NEWS (May 25, 2011), <http://abcnews.go.com/Business/business-owners-suspect-sabotage-tactics-competitors/story?id=13674139> [<https://perma.cc/QR26-2A76>]; Kevin Poulsen, *How Google Map Hackers Can Destroy a Business At Will*, WIRED (July 7, 2014), <http://www.wired.com/2014/07/hacking-google-maps/> [<https://perma.cc/T73X-RSYN>] (“Beneath its slick interface and crystal clear GPS-enabled vision of the world, Google Maps roils with local rivalries, score-setting, and deception.”).

141. For instance, the argument may be (a) by sharing code for nEmesis, the development costs should reduce over time for other counties adopting the system, and (b) demographic disparities on Twitter will decline over time, making the sample increasingly representative. There are reasons to doubt this. First, the code still has to be adopted to the jurisdiction at hand, and local health departments have very limited computer programming capacity. Second, users shift rapidly across social media platforms (Friendster, MySpace, Facebook, Twitter, Yik Yak), making it likely that investments into these tools will have short lifespans.

142. *See* U.S. GOV'T ACCOUNTABILITY OFFICE, GAO-15-290, HIGH-RISK SERIES: AN UPDATE 37 (2015); U.S. GOV'T ACCOUNTABILITY OFFICE, GAO-12-346, INFORMATION TECHNOLOGY: FDA NEEDS TO FULLY IMPLEMENT KEY MANAGEMENT PRACTICES TO LESSEN MODERNIZATION RISKS 14 (2012); Stuart Shapiro & Cary Coghlanese, *First Generation E-Rulemaking: An Assessment of Regulatory Agency Websites*, in PROCEEDINGS OF THE 8TH ANNUAL INTERNATIONAL CONFERENCE ON DIGITAL GOVERNMENT RESEARCH: BRIDGING DISCIPLINES AND DOMAINS 19–25 (2007) (finding wide variation in the quality of government agency websites).

143. Michael Chui, Diana Farrell & Kate Jackson, *How Government Can Promote Open Data and Help Unleash Over \$3 Trillion in Economic Value*, INNOVATION IN LOCAL GOVERNMENT: OPEN DATA AND INFORMATION TECHNOLOGY 4, 15 (McKinsey & Co. 2014).

by Los Angeles County, Raoul Freeman and Peter Loo concluded, “While these technologies promise enhanced user experiences and civic participation, their implementation must be considered with policy and organizational implications. Often, additional resources are required to ensure that these technologies are effectively implemented and that their benefits are fully realized.”<sup>144</sup>

While it may grab headlines and be politically salable to invest in highly technical projects, these analytical resources might be better deployed to help the department manage, learn, and evaluate based on existing data.<sup>145</sup>

\* \* \* \*

None of this is to say that this kind of local experimentation is not worth it. It absolutely is, and we applaud health departments for their entrepreneurialism. But given the concerns identified above, agencies should not let such attempts crowd out first-order concerns with public health enforcement. After all, Boston’s remedy for the “rich bias” of Street Bump 1.0<sup>146</sup>—namely, to send city *inspectors* out with the iPhone app—itsself required an effective and equitable inspection system.

### III. DISCLOSURE

Can disclosure improve food safety? For years, scholars have conceived of restaurant letter grading as the model for effectively engaging in information disclosure. By providing a simplified signal at the time relevant to consumer decision-making, scholars posit that letter grading overcomes the deficiencies of information overload. David Weil and coauthors, for instance, conclude from a synthetic review of a wide range of information disclosure regimes that letter grading is one of the few “highly effective” forms of disclosure.<sup>147</sup> Grade disclosures and perceptions of food safety certainly seem to affect consumer decisions.<sup>148</sup> In earlier work, however, I showed that the frontline differences in inspection styles undercut the informational basis for grading.<sup>149</sup> In New York, scores from one routine, unannounced inspection have virtually no predictive power over scores down the road.<sup>150</sup> The long-standing public

---

144. Raoul J. Freeman & Peter Loo, *Web 2.0 and E-Government at the Municipal Level*, in PROCEEDINGS OF THE IEEE 2009 WORLD CONGRESS ON PRIVACY, SECURITY, AND TRUST AND THE MANAGEMENT OF E-BUSINESS 70 (2009).

145. As another illustration, Stanford Law School has been engaged in dozens of policy practicums where a government agency or nonprofit organization is the client, with Stanford faculty and students providing research resources (often quantitatively driven) to help the organization learn from data.

146. See EXEC. OFFICE OF THE PRESIDENT, BIG DATA: SEIZING OPPORTUNITIES, PRESERVING VALUES 51–52 (2014), [https://www.whitehouse.gov/sites/default/files/docs/big\\_data\\_privacy\\_report\\_may\\_1\\_2014.pdf](https://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf) [<https://perma.cc/V8CC-XBX6>].

147. See FUNG ET AL., *supra* note 22, at 82–83.

148. See Jee Hye Lee, Johye Hwang & Azlin Mustapha, *Popular Ethnic Foods in the United States: A Historical and Safety Perspective*, 13 COMPREHENSIVE REV. FOOD SCI. & FOOD SAFETY 2, 11 (2014); Ginger Zhe Jin & Phillip Leslie, *The Effect of Information on Product Quality: Evidence from Restaurant Hygiene Grade Cards*, 118 Q.J. ECON. 409 (2003).

149. See Ho, *supra* note 17.

150. *Id.* at 4–45.

health critique of letter grading—which originally stemmed from the New Deal—is that restaurant inspections are merely a “snapshot-in-time,” with inspectors, food-safety practices, and conditions in continual flux.<sup>151</sup>

### A. Unadjusted Grading

Nearly all jurisdictions engaging in letter grading determine the grade on the basis of a single routine inspection.<sup>152</sup> The left panel of Figure 10 displays King County’s score distribution based on one inspection cycle. Over 50% of establishments receive zero red points, so a conventional grading system by construction could not draw distinctions among the top 50% of establishments. The distribution is also very skewed, with some establishments receiving very high point totals. The right panel displays the distribution across an area rotation, with each cell shaded based on the density of establishments. For instance, the lower left corner represents roughly 31% of establishments that receive zero red points in both time periods. Across area rotations and inspection cycles, however, the correlation is quite low. Using a simple least squares fit, a ten-point increase is associated with a roughly 2.4-point increase in the subsequent cycle ( $R^2 = 0.04$ ).

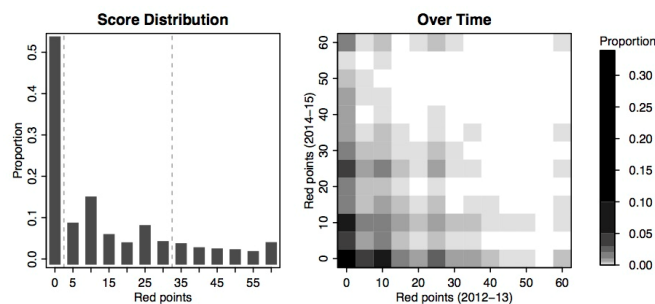


Figure 10: Distribution of points for one inspection cycle in the left panel and across area rotations in the right panel. In the right panel, each cell is shaded by the density of establishments falling in that point combination, with the legend indicating the proportion of observations. Points are censored at sixty for visibility. Using only a single inspection cycle, over 50% of establishments receive zero red points. Findings are comparable in other years. While there is a statistically significant correlation ( $\beta = 0.24$ ,  $SE = 0.02$  from least squares regression), unexplained variation is quite high ( $R^2 = 0.04$ ).

A superficially appealing grading system might assign A’s to restaurants scoring zero red points, B’s to restaurants scoring between zero to thirty-five red points, and C’s to any establishment scoring thirty-five or more red points (and hence subject to a return visit). We call this the “unadjusted” grading scheme, which closely approximates

151. *Id.* at 50.

152. Many grading systems do allow for reinspections to change a grade, but because these inspections are anticipated (typically within thirty days), they are unlikely to reflect a genuine assessment of food-safety practices. By default, the principal grading basis is a routine inspection in nearly all jurisdictions.

how grading is practiced in other jurisdictions. Based on that system, at the end of 2015, roughly 50% of establishments would have earned A's, 40% B's, and 10% C's. Figure 11 displays the geography of grades assigned on this system. The grades would suggest that there is substantial variability across regions in King County. At the end of 2015, for instance, Redmond would receive nearly 20% C's. The International District, which most inspectors would point to as a high-risk area, would appear to be middle-of-the-road.

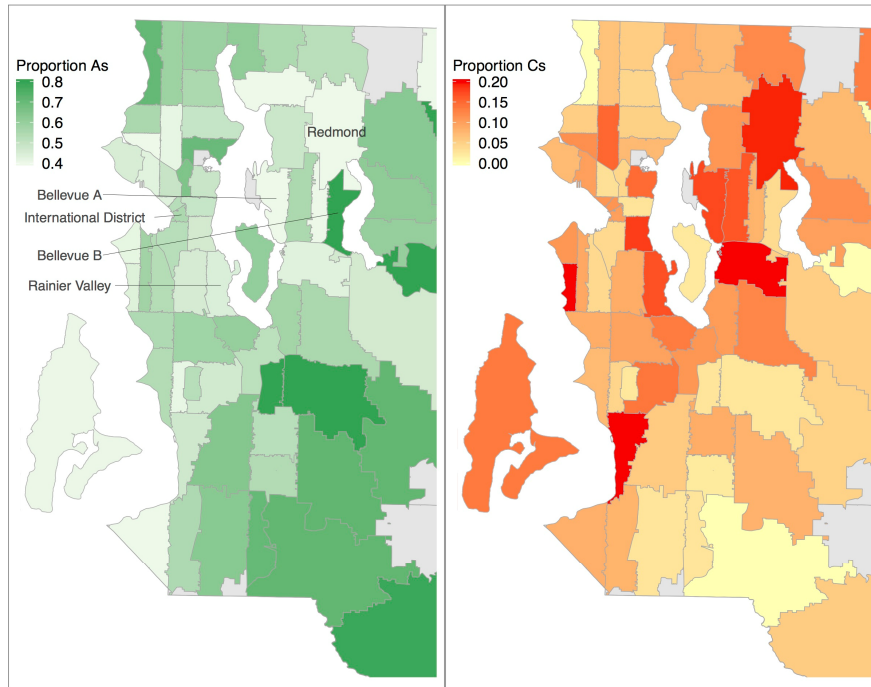


Figure 11: Geography of naïve grading system applied to King County at the end of 2015. Each establishment is assigned a grade based on a single routine inspection, with zero red points corresponding to an A grade, red points between zero and thirty-five points corresponding to a B grade, and thirty-five or more points corresponding to a C grade. The left panel displays the proportion of A establishments in a ZIP code, with darker shading corresponding to more A's. The right panel plots the proportion of C's in a ZIP code, with darker shades of red indicating more C's.

But these geographic differences are unlikely to accurately represent genuine geographic differences in food risk. For instance, Figure 12 shows that inspectors largely anchor around the same mean before and after an area rotation. To understand the substantive impact of such inspector differences on grading, we conducted a simulation exercise, estimating a model that accounts for establishment, month, and inspector fixed effects.<sup>153</sup> Based on this model, we can counterfactually predict how the assigned grades

153. See Ho, *supra* note 17.



might change if a single inspector had conducted all inspections across the county. Moving from the tenth to the ninetieth percentile inspector, a lenient inspector would assign 55% of establishment A's, while a tough inspector would assign 2.5% A's. Because inspectors are assigned principally by ZIP code, these geographic differences are far more likely to represent differences in inspection styles than underlying food safety.

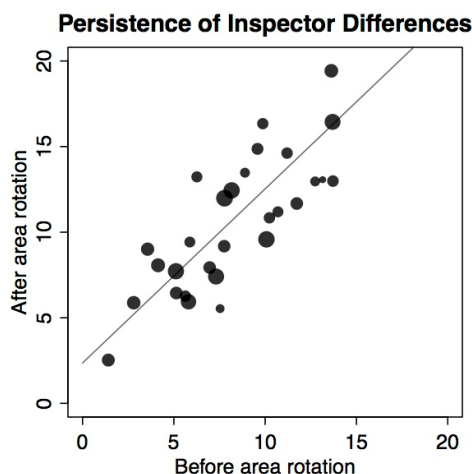


Figure 12: Persistence of inspector differences across area rotations. Each dot represents an inspector who actively inspected both before and after the area rotation in January 2014, with dots weighted by the minimum number of inspections in either the pre or post period. Based on a simple linear fit, a one-point increase in the pre-area rotation mean is associated with a one-point increase after the post-area rotation mean ( $R^2 = 0.65$ ).

In addition, grading cutoffs themselves can change inspection behavior. The difference between no red points and some red points is one that is already subject to considerable variability across inspectors. The additional pressure for operators to want to earn an A can lead some inspectors to use their discretion to not cite a violation. Indeed, this is what appears to have happened in the vast majority of grading jurisdictions. In San Diego, for instance, there is a sharp discontinuity around the ninety-point threshold for an A grade, with 99.9% of establishments earning A's.<sup>154</sup> As one San Diego inspector noted, "Some inspectors will give out a B for an eighty-nine. I usually warn somebody at that point. It's a judgment call."<sup>155</sup>

Similarly, while the thirty-five-point threshold for a C grade appears consistent with the food code (and accounts for the fact that some inspectors do not tally all violations once it becomes clear that a return visit is required), some inspectors would like to avoid the additional workload of a return visit. By introducing additional tension

154. See Ho, *supra* note 23, at 611.

155. Ed Sylvester, *Making Sure Your Eating Places Are A-OK: Inspectors Rate S.D. Restaurants*, L.A. TIMES, May 25, 1980, at A1, A3.

with an operator over the line between a B and C grade, the thirty-five-point threshold may further disincentivize inspectors from engaging in return visits that are arguably the most important from a public health perspective.

In short, unadjusted grading presents serious flaws.

### B. Equitable Grading

In collaboration with King County, we sought to design a better grading system. The principal criteria were that a grading system should (a) draw meaningful distinctions in terms of food risk among restaurants deemed safe enough to open by the county (i.e., with fewer than ninety red points), and (b) be easily explained to the public and to operators. Criterion (b) ruled out a range of sophisticated statistical adjustments (e.g., item response theory). After several meetings with stakeholders, which presented grading systems in other jurisdictions, several other recommendations followed. First, to counteract the conventional critique of inspections as presenting merely snapshots-in-time, stakeholders articulated the desire to have a grade based on more than just a single inspection.<sup>156</sup> Second, the political compromise in many jurisdictions was to introduce an additional reinspection for grade resolution with the grading system.<sup>157</sup> In New York, adjudicative appeals for grade resolution also skyrocketed.<sup>158</sup> Rates of appeal, particularly with attorney representation, are likely to exacerbate demographic differences across restaurateurs.<sup>159</sup> Because the evidence shows that grade reinspections tend to draw resources away from the highest health risks, the county ruled out regrading based on reinspections.<sup>160</sup> Third, the county rightly focused on the first-order challenge of improving the accuracy and consistency of the inspection process.<sup>161</sup> Our randomized controlled trial revealed that peer review increased violations cited and scored by roughly 17–19% and increased consistency between inspectors.<sup>162</sup> As a result, the county generalized the peer review program to the entire staff beginning in September 2015.<sup>163</sup> While peer review is very promising, it is also a continuing process, as inconsistencies and disagreements about the health code can only gradually be resolved.

With these criteria in mind, we aimed to place the grading system on a better evidence basis using historical inspection data.

---

156. See, e.g., *Restaurant Reporting Subcommittee Meeting Notes*, PUB. HEALTH (Sept. 15, 2014), [http://www.kingcounty.gov/healthservices/health/ehs/foodsafety/~/\\_/media/health/publichealth/documents/foodsafety/RestaurantReportingSubCommittee140915Notes.ashx](http://www.kingcounty.gov/healthservices/health/ehs/foodsafety/~/_/media/health/publichealth/documents/foodsafety/RestaurantReportingSubCommittee140915Notes.ashx) [<https://perma.cc/A63A-Y32S>].

157. See Ho, *supra* note 23, at 650.

158. *Id.* at 647.

159. See Chao, *supra* note 16 (“[M]any owners and operators in Chinatown, when confronted by fines or settlements in the mail, paid them without question. They also tended to show up at the tribunals by themselves, with no lawyer, and without any documents in hand.”).

160. See, e.g., *Restaurant Reporting Subcommittee Meeting Notes*, *supra* note 156.

161. See, e.g., *id.*

162. See Ho, *supra* note 17, at 8.

163. See *id.* at 71.

First, the peer review inspections allowed us to identify violations that were the most inconsistently applied across frontline inspectors. Figure 13 plots the baseline citation rate and the deviation rate for individual citations from 378 peer inspections conducted from January–April 2015. The deviation rate represents how frequently two inspectors observing identical conditions on the ground disagreed on whether to cite a specific violation. Each dot represents one type of violation, with red and blue dots corresponding to red and blue violations respectively. The bands represent simply linear fits, with 95% confidence intervals. What becomes apparent from peer review inspections is that blue violations are much more inconsistently applied. This finding makes substantive sense, as state training, marking instructions, and county trainings focus disproportionately on red (critical) violations. Coupled with the fact that red violations are also the ones that pose higher health risks, we concluded that in contrast to other jurisdictions, which overwhelmingly include minor violations as the basis for grading, King County’s grading system should be based on red violations.<sup>164</sup>

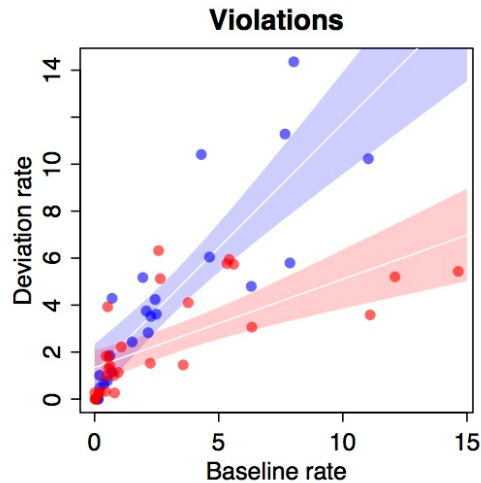


Figure 13: Peer review evidence of deviations. Each dot represents one of fifty-two violations, with colors corresponding to red (critical) and blue (noncritical) violations. The *x*-axis represents the baseline rate at which violations are cited, and the deviation rate indicates the rate at which two inspectors observing the same conditions disagreed on whether or not to cite that violation. This evidence shows that blue violations, which receive relatively little training, are much more inconsistently cited.

Second, to develop a sense of how many inspections cycles to use, we studied the predictive power of past inspection scores and found that the predictive validity drops off considerably beyond five inspections. We hence illustrate an adjusted grading system based on the average of up to four routine inspections.

---

164. One alternative would be to conduct inverse weighting based on the deviation rate, but this would both change over time and be more difficult to explain.

Third, we studied area rotations and differences in inspection styles, concluding that rotations had only minor effects on scoring behavior of inspectors (see Figure 12). To illustrate, Figure 14 plots performance of the same establishments operating in Redmond before and after the area rotation in January 2014. The left panel plots red points assigned by a lenient inspector on the  $x$ -axis compared to red points assigned by more stringent inspectors on the  $y$ -axis after the area rotation. In 2013, nearly 80% of establishments were assigned zero red points, but many of those establishments fared much worse after the area rotation, with only 34% scoring zero red points.

The middle panel plots the distribution of scores for establishments, averaged across four routine inspections, before the area rotation. The dashed vertical lines represent unadjusted grading system cutoffs, uniformly imposed across the county. What unadjusted cutoffs miss is that earning twelve red points from a lenient inspector places the establishment in the bottom 10% of restaurants in that ZIP code. Unadjusted grading would hence result in a false negative. The right panel presents the distribution of average scores after the area rotation. For the tougher inspectors, twelve red points actually place the establishment in the top half of establishments. Unadjusted grading would hence result in a false positive.

To reduce such error, we thus studied an adjustment based on relative ZIP code performance. We kept the overall proportion of A's, B's, and C's the same as in the unadjusted system (50%, 40%, and 10%, respectively when conducting grading at the end of 2015), but identify the identical percentile cutoffs *within* each ZIP code. This has two major advantages. Because inspectors are assigned principally based on ZIP codes, this system adjusts for differences in inspection styles.<sup>165</sup> In addition, because most food choices are local, the relative performance is more meaningful for (a) consumers choosing where to eat, and (b) providing establishments with incentives to improve. The system is also easy to explain. Grades represent the relative performance in an area for up to four routine inspections. The colors on the bars in the middle and right panels of Figure 14 represent the adjusted grades, with green, orange, and red corresponding to A's, B's, and C's, respectively.<sup>166</sup> These demonstrate how, for Redmond, it is possible to use the information generated under widely heterogeneous inspections styles to meaningfully identify differences among establishments.

---

165. Two practical complications arise here. First, some ZIP codes are sufficiently large that they share multiple inspectors. Second, inspectors can be assigned to multiple ZIP codes, not all of which are geographically contiguous. We hence studied a fuller adjustment, identifying percentile cutoffs in each unique inspector area. The ZIP code adjustment appeared to perform close to this ideal adjustment, but the ideal adjustment has the downside of being harder to explain, as unique inspector areas do not correspond with otherwise recognizable areas. While area rotations also present a transition issue, the facts that the grade is based on up to the last four routines and that new inspectors take over the whole area tend to mitigate this transition cost.

166. We also proposed a "D" category of high health hazards, based on multiple return inspections or a recent closure. This category has the virtue of minimizing false positives, and providing an inspector with additional leverage to induce corrective action, when one return visit was insufficient. Because this category is constant across the adjusted and unadjusted schemes, we do not focus on it here.

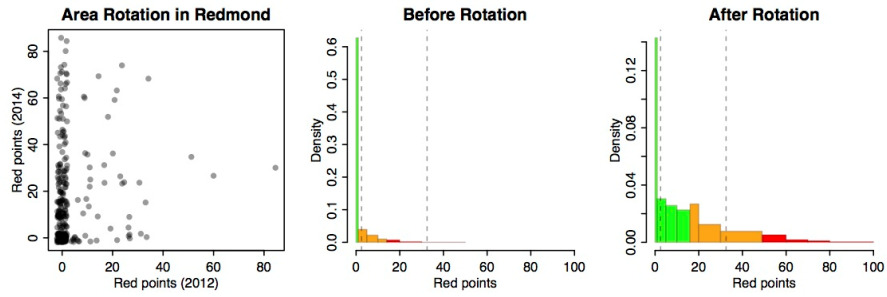


Figure 14: Effect of area rotation in Redmond. The left panel plots red points assigned during a single inspection round on the  $x$ -axis for 2013, when the area was primarily inspected by a lenient inspector, and on the  $y$ -axis for 2015, when the area was assigned by several tough inspectors. The middle panel plots the average red points for four routines pre-area rotation. The dashed vertical lines represent grade cutoffs for unadjusted grading, leading to 80% of establishments earning A's. The colors represent grade assignments for the adjusted system (green for A, orange for B, and red for C), leading to 56% of establishments earning A's. The right panel plots the same establishments after the area rotation. Unadjusted grading would lead to 34% of establishments earning A's, but this is because tough inspectors are now assigned to the area. Adjusted grading leads to 49% of establishments earning A's.

Figure 15 displays the geography of A grades before and after the area rotation in rows and for the unadjusted and adjusted grading systems in columns. In the left column, Redmond goes from appearing to be a top performer to a bottom performer across area rotations. But this is entirely an artifact of inspector rotation. The right panels show how the adjusted grading system smoothens out these artificial geographic differences.

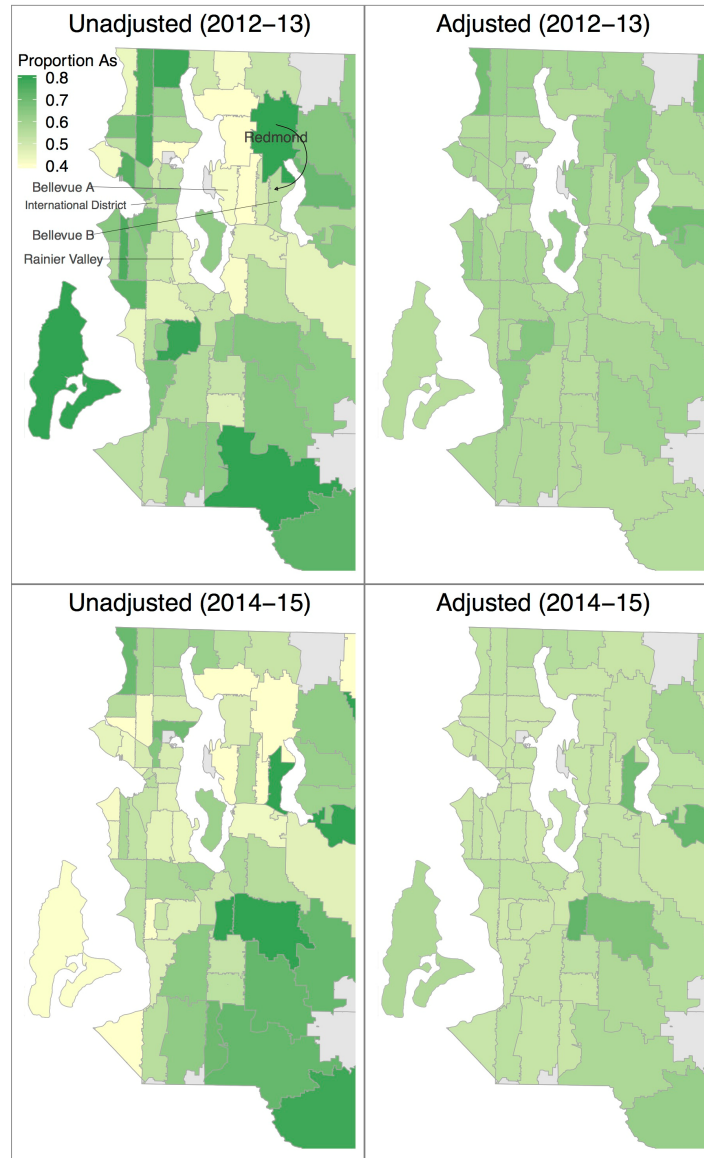


Figure 15: Adjusted and unadjusted grades across area rotations. The top row plots the proportion of A grades in a ZIP code based on inspection results from before the area rotation in January 2014, as if King County restaurants had been graded on December 31, 2013. The bottom row plots the proportion of A grades in a ZIP code based primarily on inspection results from after the rotation, as if King County restaurants had been graded on December 31, 2015. The left column plots unadjusted grades, based on a single routine inspection. The right panel plots grades based on an average of up to four routine inspections (some restaurants do not have results for as many as four inspections), based on percentile cutoffs within each ZIP code. Each colored ZIP code has at least ten restaurants.

	Points	Unadjusted			Adjusted		
		A's	B's	C's	A's	B's	C's
Overall (2013)	8.47	0.56	0.38	0.07	0.56	0.38	0.06
Lenient Inspector Rotates Overall (2015)	11.25	0.50	0.40	0.10	0.50	0.40	0.10
Redmond (2013)	4.08	0.78	0.19	0.03	0.56	0.38	0.06
Redmond (2015)	19.61	0.34	0.45	0.21	0.49	0.41	0.11
Bellevue A (2013)	10.94	0.50	0.40	0.10	0.54	0.40	0.06
Bellevue A (2015)	4.04	0.85	0.11	0.04	0.53	0.36	0.11
Bellevue B	16.66	0.36	0.46	0.18	0.47	0.43	0.10
International District	11.75	0.53	0.36	0.11	0.50	0.40	0.10
Rainier Valley	15.43	0.47	0.35	0.19	0.50	0.40	0.11
Asian	19.06	0.32	0.46	0.21	0.25	0.51	0.24

Table 1: Effect of adjustment for different subsets of establishments. Redmond (2013) and Redmond (2015) represent the same area before and after an area rotation (with simulated grading occurring on December 31, 2013 and December 31, 2015). The lenient inspector assigned to Redmond in 2013 moved to Bellevue A in 2014. Bellevue A represents ZIP 98008 and Bellevue B represents ZIP 98004, showing that there are considerable differences in the same city (results for Bellevue B, the International District, Rainier Valley and for Asian establishments are based grades calculated at the end of 2015). The bottom row presents estimates for Asian establishments (based on the subset that could be merged with Yelp cuisine categories): the adjusted grading system underscores that Asian establishments systematically underperform relative to other establishments in the same area.

Table 1 displays the breakdown of grades for a sample of areas. The first row shows that overall, the proportions of grades are held constant across the county as a whole. The third and sixth rows show how the movement of the lenient inspector from Redmond to Bellevue A (one ZIP code in Bellevue) leads (a) Redmond to appear to drop markedly in performance, and (b) Bellevue A to appear to improve markedly in performance. Horizontal inequity emerges in 2015 as well, as another ZIP code in Bellevue (Bellevue B) appears to perform much more poorly in the unadjusted system. In other words, inequities arise both inter-temporally for the same area and contemporaneously across neighboring areas. Adjusted grades moderate all of these differences, which are attributable primarily to differences in inspection styles. Indeed, using the same counterfactual grade simulation as above, the grade distribution is by

construction robust to whether a tough or lenient inspector was to inspect all King County restaurants. While moving from the tenth to ninetieth percentile inspector would change the grade for over *half* of establishments in the unadjusted system, inspector toughness has no effect on simulated grades in the adjusted system.

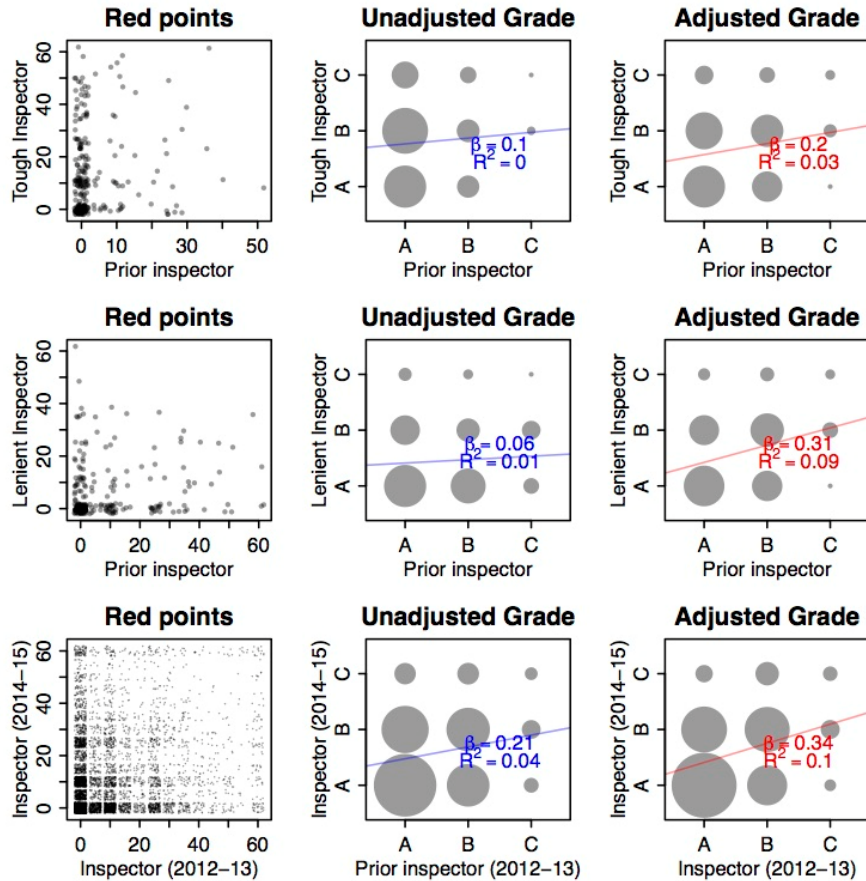


Figure 16: Illustration of how adjusted grade improves information content of grades. The left column represents raw scores across area rotations. The top and middle rows represent establishments assigned to tough and lenient inspectors after January 1, 2014, respectively. The bottom row represents all establishments. The right two columns depict how unadjusted and adjusted grading classifies establishments across area rotations, with dots proportional to the number of establishments. For instance, the largest dot in the top middle panel indicates that the modal category is establishments receiving an A under the prior inspector and B's after the assignment to a tough inspector. Lines represent least squares fits, with regression coefficients ( $\beta$ ) and coefficients of determination ( $R^2$ ) presented as summary measures of predictive validity. In all instances, adjusted grades improve the predictive power of grades.

Last, by making the measurement of food risk more accurate, the adjusted system increases predictive power over future performance. The left column of panels in Figure 16 presents the correlation of point performance over area rotation. The middle column



presents the classifications into grades in the unadjusted system and the right column presents classifications based on the adjusted system. The top row illustrates how even when a tough inspector takes over an area, the adjustment can improve predictive validity. The top middle panel shows that establishments previously assigned A's in the unadjusted system have a higher probability of earning a B or C with the tough inspector. The adjustment in the top right panel reverses this: an A with the prior inspector has the highest probability of earning an A under the new inspector. The lines represent simple linear measures of association, and in all instances, the slope of the line (characterizing predictive validity) increases with the adjusted grade, indicating that the adjusted grading power has better predictive validity than unadjusted grading.<sup>167</sup>

While adjusted grades are more consistent across time and geography, equity implications along demographic variables and cuisine types are more complex. Our initial hypothesis had been that due to geographic clustering (e.g., with far more Asian restaurants located in the International District), the adjusted grading system might improve the grade performance of Asian establishments. If anything, the adjusted grading system decreases the proportion of Asian restaurants earning A's by roughly 7%, as seen in Table 1. To understand the intuition behind this, in nearly every ZIP code, Asian establishments perform relatively worse compared to non-Asian counterparts. By introducing more noise due to differences in inspection styles, the unadjusted grading system attenuates the Asian-grade relationship. With a more accurate, adjusted grading system, it becomes clearer that Asian restaurants systematically are performing worse in inspections *across all inspectors*, a result consistent with evidence from other jurisdictions.<sup>168</sup> On the other hand, Table 1 also shows that for Rainier Valley, one of King County's most diverse and poorest neighborhoods, the adjusted grading system decreases the proportion of C's by roughly 8%. Because geographic mobility is likely lower in poorer neighborhoods, this shows one of the possible equity benefits of designing a locally meaningful grading system. Yet because the grading system is intended to protect the public, higher grades in a neighborhood are not necessarily better per se for that community.

While this research demonstrated ways to place grading on a better evidence basis, there remains substantial unexplained variability in inspection performance. Grading can be improved, but public health benefits remain nonobvious. Our policy analysis also challenges open-ended invocations of "equity." King County, for instance, appears to recognize inequity along "race, ethnicity, income, immigration status and ZIP code."<sup>169</sup> While the adjusted grading system is certainly more predictive of future inspection

---

167. We use linear fits for simplicity, but obtain comparable results using measures of association for categorical variables.

168. See Joe Satran, *New York Restaurant Health Grades Vary by Cuisine*, HUFFINGTON POST (last updated June 12, 2013), [http://www.huffingtonpost.com/2013/06/11/new-york-restaurant-health-grades\\_n\\_3392571.html](http://www.huffingtonpost.com/2013/06/11/new-york-restaurant-health-grades_n_3392571.html) [<https://perma.cc/ZR52-SYQ6>] (noting that the ten worst performing cuisines in New York were Japanese, Soul Food, African, Latin American, Chinese, Korean, Indian, Creole, Pakistani, and Bangladeshi); Chao, *supra* note 16 ("Chinatown Business Improvement District (BID) announced that the community has 20% fewer restaurant A ratings than citywide.").

169. KING CTY. EXEC., *supra* note 15, at 1 (emphasis added).

performance and more equitable geographically (across ZIP codes), it also highlights the Asian cuisine food-safety gap. And just as with Chilangos, it is unclear how the department should weight equity concerns with respect to establishments or patrons.

To borrow Judge Friendly's phrase, equity is a "verbal coat of many colors."<sup>170</sup>

#### IV. IMPLICATIONS

These three case studies illustrate the pervasive challenges of rigorously evaluating the equity dimensions in regulatory enforcement. Addressing the disparity between the food inspection performances of Asian and non-Asian establishments in King County remains a major challenge, with little research pinpointing root causes.<sup>171</sup> The difference may stem from underlying differences across cuisine types. Asian cuisines, for instance, may engage in more raw food preparation and raw butchering, activities that are either inherently riskier or have more difficulty conforming to the food code. Alternatively, the relative poverty level of certain immigrant communities may simply make it infeasible to upgrade the physical infrastructure (e.g., storage, food preparation area, or refrigeration) to comply fully with the health code. While our evidence cannot directly address these causes of the inspection disparity, our evidence does suggest several other affirmative policy interventions if other mechanisms are at stake.

##### *A. Peer Review and Data-Driven Training*

Lack of inspector knowledge about the food preparation process and historical tension between the food program and immigrant communities may contribute to the disparity, and agencies should confront such issues directly with training via peer review.

---

170. See The Honorable Henry Friendly, *Indiscretion About Discretion*, 31 EMORY L.J. 747, 763 (1982).

171. See Kimberly J. Harris et al., *Food Safety Inspections Results: A Comparison of Ethnic-Operated Restaurants to Non-Ethnic-Operated Restaurants*, 46 INT'L J. OF HOSPITALITY MGMT. 190, 197–98 (2015).

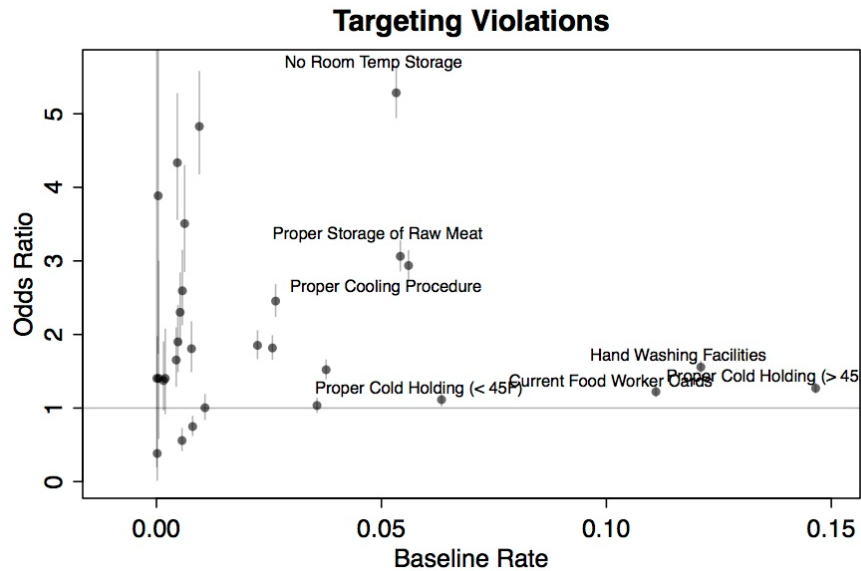


Figure 17: Baseline citation rate for a red violation on the *x*-axis against the odds ratio for an Asian establishment on the *y*-axis, with 95% confidence interval. The horizontal line represents identical citation rates across Asian and non-Asian establishments.

While frontline inspectors may have divergent priors on whether a particular practice (e.g., cooling large vats of broth at room temperature) is acceptable, there is often a scientific answer based on food science, which can help to resolve disagreements. Evidence from our peer review experiment suggests that the process may be a supportive environment to develop common guidelines for the department.

Figure 17 shows how simple analysis of historical inspection information might aid in this process, much in the same way that peer review information allowed us to design targeted training materials. The *x*-axis plots the baseline rate at which violations are cited, and the *y*-axis plots the odds ratio for Asian establishments, with a ratio above one indicating that Asian establishments are disproportionately cited for that violation. This simple analysis reveals that room temperature storage, raw meat storage, and cooling procedures are violations that are commonly cited but with very high odds ratios for Asian establishments. For instance, numerous inspectors mentioned substantial disagreement about cooling requirements for pho broth. One inspector noted that because the broth had been cooked for hours and would be reheated in the morning, cooling a big bucket should not pose a health risk. Other inspectors stated that such cooling procedures violated the health code and the broth was required to be discarded. Other commonly disputed code items involve rice noodles kept at room temperature, hanging roasted duck at room temperature, and the use of hot boxes (or cambros) to store food. Training materials addressing these violations and food items would help to (a) establish what the actual risks are to such practices, and (b) identify more feasible food preparation techniques.

Of course, not all issues have a clear answer in food science. It is undisputed that proper handwashing reduces food risk.<sup>172</sup> Yet given potential cultural sensitivities, how should inspectors most effectively educate and persuade operators to engage in proper handwashing? When the physical infrastructure needs to be upgraded, inspectors might shut down an establishment, with a penalty fee, to force it to upgrade. But how should an inspector deal with an establishment when the penalty fee may put it out of business? How can inspectors determine whether the failure to remedy a violation is due to income constraints or defiance? And how can inspectors effectively educate operators who may have limited familiarity with the food code? Peer review may be one of the only mechanisms to allow inspectors to acquire these soft skills that are invaluable to resolve equity issues in food enforcement.

### *B. Information Reporting and Information Disclosure*

Our research also has two concrete policy implications for the role of information.

First, targeting based on social media as currently practiced has limited use, but efforts to broadly improve the reporting of foodborne illnesses can be beneficial.<sup>173</sup> Big data cannot ignore principles of statistical inference and the normative basis for regulation.<sup>174</sup> On the other hand, Chicago's development of an online reporting system may present a worthwhile means of curing the fundamental issue of low reporting rates. Instead of investing in highly technical, gimmicky interventions, data scientists should invest resources to enable—rather than distract—agencies to conduct their core mission.

Second, our study shows how to engage in more meaningful information disclosure. More accurate, consistent, and equitable grading systems can mitigate conventional criticisms of inspections as merely a snapshot-in-time, by leveraging inspection information over multiple cycles, by using peer review information to focus on high-risk and consistently implemented violations, and by adjusting for inter-inspector differences. Any jurisdiction applying letter grading should place its system on a firmer evidence basis, using our research results.

### *C. Language Access*

Language and cultural barriers appear substantial in the inspection system. From the perspective of restaurateurs, “the biggest challenge . . . is simply dealing with the bureaucracy.”<sup>175</sup>

---

172. *Id.*

173. This does not rule out the possibility of more sophisticated targeting based on social media that adjusts, for instance, for sampling bias.

174. See, e.g., David Lazer et al., *The Parable of Google Flu: Traps in Big Data Analysis*, 343 SCI. 1203 (2014) (showing how failure to account for changing predictors invalidated Google Flu Trends).

175. Chao, *supra* note 16.

Much may be done to improve language access to the regulatory system.<sup>176</sup> First, agencies should continue to hire staff who are bilingual and, ideally, versed in different cultures around food.<sup>177</sup> Such hiring, as exhibited by King County's efforts, may not solve the problem of the ethnic gap in food-safety performance, but would permit a wider range of engagement between operators and inspectors. For instance, during a 2015 salmonella outbreak, the public health nurse noticed that most of the infected individuals appeared to be of East African origin. An inspector fluent in Amharic conducted the subsequent interviews and convened a series of community meetings to contain the scope of the outbreak.

Second, agencies may need to invest more resources into making the bureaucracy easier to navigate, including multilingual permitting processes<sup>178</sup> and food-safety course materials,<sup>179</sup> and in ways that are sensitive to a range of cuisines.<sup>180</sup> For instance, lower perceived reporting rates of foodborne illness from immigrant communities<sup>181</sup> may be

176. Currently, King County has three language tiers, with mandatory translation only for Spanish. See *Appendix C: Language Tiers*, KING CTY., WASH. GOV., [http://www.kingcounty.gov/~media/operations/policies/documents/inf142aeco\\_appxc\\_languagetiers\\_intro.ashx?la=en](http://www.kingcounty.gov/~media/operations/policies/documents/inf142aeco_appxc_languagetiers_intro.ashx?la=en) [https://perma.cc/T65C-9A9T] (last visited Feb. 29, 2016).

177. See Office for Civil Rights, *Guidance to Federal Financial Assistance Recipients Regarding Title VI and the Prohibition Against National Origin Discrimination Affecting Limited English Proficient Persons- Summary*, U.S. DEP'T OF HEALTH & HUMAN SERVS., <http://www.hhs.gov/civil-rights/providers/laws-regulations-guidance/guidance-federal-financial-assistance-title-VI/index.html> [https://perma.cc/4YZ5-6BZC] (last visited Feb. 28, 2016).

178. Online information about the permitting process in King County, for instance, is exclusively in English, although the site does note “[a]lternative formats [are] available upon request.” PUB. HEALTH SEATTLE & KING COUNTY, PLAN REVIEW AND PERMITTING GUIDELINES FOR THE NEW CONSTRUCTION OR REMODELING OF A FOOD SERVICE ESTABLISHMENT 1 (Sep. 2015), <http://www.kingcounty.gov/healthservices/health/ehs/foodsafety/FoodBusiness/~media/health/publichealth/documents/foodsafety/Plan-Guide-Food-Service-Plan-Review.ashx> [https://perma.cc/687J-PAHK].

179. For instance, New York's “online food protection course is available only in English, Spanish, and Chinese.” Chao, *supra* note 16. There is some evidence that food-training programs are effective. See M. Martinez-Tome, A.M. Vera & M.A. Murcia, *Improving the Control of Food Production in Catering Establishments with Particular Reference to the Safety of Salads*, 11 FOOD CONTROL 437 (2000); Denise Worsfold & Christopher J. Griffith, *A Survey of Food Hygiene and Safety Training in the Retail and Catering Industry*, 33 NUTRITION FOOD SCI. 68 (2003).

180. See Po et al., *supra* note 16.

181. King County officials suspect a sharp distinction among immigrant and nonimmigrant populations in reporting behavior. In their assessment, nonimmigrant populations appeared more likely to report milder symptoms (e.g., for norovirus). With outbreaks in 2014, however, large numbers of immigrants were reported as infected via hospital confirmations, with much more serious symptoms. Interviews in those communities suggested that vomiting, nausea, and diarrhea were deemed insufficient for reporting an illness and visiting a hospital, and that only dehydration and bloody stool led to a hospital visit. The reporting differential would be consistent with beliefs generally that immigrant populations have a lower reporting rate to law enforcement or are less likely to seek health care. See, e.g., ROBERT C. DAVIS & EDNA EREZ, U.S. DEP'T OF JUSTICE, IMMIGRANT POPULATIONS AS VICTIMS: TOWARD A MULTICULTURAL CRIMINAL JUSTICE SYSTEM (1998) (finding that 67% of officials believed that reporting rates were lower among immigrants); Erum Nadeem et al., *Does Stigma Keep Poor Young Immigrant and U.S.-Born Black and Latina Women from Seeking Mental Health Care?*, 58 PSYCH. SERVS. 1547 (2007) (finding underutilization of mental health care by ethnic minority groups); Ilhong Yun & David Mueller, *A Study of the Determinants of Reporting Crime to the Police Among Chinese Immigrants*, 35 INT'L J. COMP. & APPLIED CRIM. JUST. 53, 55 (2011) (noting consensus

explained in part by the fact that the King County nurse by default conducts the phone intake process in English. An online reporting system could be translated into other languages, and non-English reports routed to inspectors with corresponding language skills, thereby taking fuller advantage of language capacities among staff. (Language access on social media platforms may similarly become important if agencies were to actually rely on such platforms to target enforcement resources.)

Third, counties should consider translating inspection and violation comments into the languages used by the operators and their staff. This would be quite easy, for instance, for stock comments in the inspection tablet system in King County. Recall that the peer review process also revealed that one inspector attempted to simplify her comments to make them more accessible in light of the language barrier.<sup>182</sup> Such simplifications, even if in English, may also help to promote broader access and understanding of food-safety principles.

#### CONCLUSION

King County's equity initiative was a bold step. It led the food-safety program to offer workshops on equity issues for staff, to develop community engagement events, to distribute food worker manuals in a wider range of languages, and to increase cultural and linguistic diversity among inspectors. While these efforts are to be applauded, they nonetheless fall short of identifying a "system[] to engage and empower all county employees to advance equity through their daily work."<sup>183</sup>

This Essay has demonstrated how conducting collaborative research with academics, conducting a randomized controlled trial, and augmenting county data with census and social media information can clarify equity dimensions and identify potential solutions. Prior to this research, actual policy analyses conducted under the equity initiative remained quite limited.<sup>184</sup> The food-safety program had many intuitions about equity dimensions, but little clarity as to scope and fact. Our results underscore that while inspectors grapple with equity issues in varying ways, there remains a considerable gap in performance across Asian and non-Asian restaurants. Peer review may be a critical tool for inspection staff to bridge differences in enforcement and to develop a broader skillset to cure the ethnic cuisine gap. We have also shown that targeting via social media would distort enforcement in highly undesirable ways, and that restaurant letter grading can be improved to reduce unjustifiable differences. Such research can give content and meaning to otherwise diffuse notions of "equity."

More generally, equity interventions should themselves be placed on a firmer evidence base. Do staff workshops on equity issues (e.g., implicit bias training) actually affect awareness and work conduct? Outside the food program, King County also

---

of immigrant underreporting).

182. Ho, *supra* note 17.

183. KING CTY. EXEC., *supra* note 15.

184. See *Equity & Soc. Just.: Tools and Resources*, KING CTY., WASH. GOV., <http://www.kingcounty.gov/elected/executive/equity-social-justice/tools-resources.aspx> [<https://perma.cc/4L8F-5XV6>] (last visited Feb. 28, 2016).

proposed numerous interventions under the equity initiative, such as universal access to developmental screenings for young children, mentoring and job-training programs for youth, and artistic programs as alternatives to juvenile detention.<sup>185</sup> Yet to invest resources wisely, these programs should be rigorously evaluated, as the peer review program was. There is a long and well-established tradition of designing such interventions with evaluations in mind.

Equity in theory should not be blind to equity in fact.

---

185. *See* KING CTY. EXEC., KING CTY. EQUITY & SOC. JUST. REP. 11, 16 (2015).

## Appendix: Tables

		National	Yelp	Twitter
<i>Age</i>	18-34	0.23	0.43***	0.49***
	35-54	0.28	0.35***	0.31
	55+	0.25	0.21**	0.07***
<i>Education</i>	No college	0.43	0.27***	0.12***
	College	0.45	0.59***	0.53*
	Graduate	0.11	0.14**	0.33***
<i>Income</i>	\$0-59k	0.58	0.36***	0.31***
	\$60-99k	0.21	0.27***	0.34***
	\$100k+	0.21	0.37***	0.13***
<i>Race</i>	White	0.64	0.68**	0.55**
	Black	0.13	0.09***	0.22***
	Hispanic	0.16	0.12***	0.15

Table 2: Demographic differences between the U.S. population and Yelp and Twitter users. Due to data availability constraints, income brackets used for the Twitter data are distinct from the brackets in the columns, and are as follows: \$0-50k; \$50-100k; \$100k+. All reported income data refer to household income. Tests of significance were performed on the difference between Yelp/Twitter sample proportions and the National proportions given in the first column. \*/\*\*/\*\* denote statistical significance at  $\alpha$ -levels of 0.10, 0.05, and 0.01, respectively.<sup>186</sup>

186. See *comScore Media Metrix Description of Methodology: Unified Digital Measurement™* (Nov. 2013), COMSCORE, <http://www.journalism.org/files/2014/03/comScore-Media-Metrix-Description-of-Methodology.pdf> [<https://perma.cc/Y7U8-FW7A>] (estimating demographics from a user base of roughly two million internet users with computer tracking software); *Yelp.com Demographics*, QUANTCAST, <https://www.quantcast.com/yelp.com#/demographicsCard> [<https://perma.cc/3H3U-ZQ2R>] (last visited Feb. 9, 2016) (using Yelp viewer data from the past thirty days to provide race statistics for Yelp); *Annual Estimates of the Resident Population by Single Year of Age and Sex for the United States: April 1, 2010 to July 1, 2014*, U.S. CENSUS BUREAU (Jul. 1, 2015), <https://www.census.gov/popest/data/national/asrh/2015/files/NC-EST2015-AGESEX-RES.pdf> [<https://perma.cc/8NGF-H56P>] (providing national age statistics); *Educational Attainment: Five Key Data Releases from the U.S. Census Bureau*, U.S. CENSUS BUREAU (Feb. 24, 2012), [https://www.census.gov/newsroom/cspan/educ/educ\\_attain\\_slides.pdf](https://www.census.gov/newsroom/cspan/educ/educ_attain_slides.pdf) [<https://perma.cc/FN4N-4PEK>] (providing national education statistics); *Income Distribution to \$250,000 or More for Households*, U.S. DEP'T OF COM. (2012), <http://www.census.gov/data/tables/time-series/demo/income-poverty/cps-hinc/hinc-06.2011.html> [<https://perma.cc/6CKE-5ED8>] (providing national income distribution numbers to serve as the basis for national income statistics); *QuickFacts United States: Race and Hispanic Origin*, U.S. CENSUS BUREAU, <http://www.census.gov/quickfacts/table/RHI125215/00> [<https://perma.cc/A9R5-QUN5>] (last visited Feb. 9, 2016) (providing national race statistics); Tom Webster, *The Social Habit II: The Edison Research/Arbitron Internet and Multimedia Study 2011*, EDISON RESEARCH (May 29, 2011), <http://www.slideshare.net/webby2001/the-social-habit-2011-by-edison-research> [<https://perma.cc/Q5J3-W6F8>] (providing age and racial statistics for Twitter from a 2011 national random sample of 2020 interviews conducted with U.S. adults aged twelve and over); Tom Webster, *Twitter Usage in America: 2010 The Edison Research/Arbitron Internet and Multimedia Study*, EDISON RESEARCH, (April 29, 2010) [http://www.onecommunity.org/wp-content/uploads/2010/04/Twitter\\_Usage\\_In\\_America\\_2010.pdf](http://www.onecommunity.org/wp-content/uploads/2010/04/Twitter_Usage_In_America_2010.pdf) [<https://perma.cc/C3NV-ZLWD>] (providing education and income data came from a 2010 report on Twitter usage statistics, using a nationally representative survey of 1753 Americans aged twelve and over); *A Simple Guide to*



	Presence	Reviews
White (prop.)	0.15** (0.07)	35.01*** (9.70)
Asian (prop.)	-0.26** (0.11)	-53.61*** (15.18)
Hispanic (prop.)	-0.07 (0.16)	-85.38*** (19.69)
Black (prop.)	-0.22 (0.16)	-15.84 (22.26)
Foreign-born (prop.)	-0.19* (0.11)	-63.64*** (13.57)
Limited English proficiency (prop.)	-0.36 (0.23)	-98.83*** (31.18)
High school (prop.)	-0.20** (0.08)	-53.67*** (10.18)
Unemployment rate	-0.49*** (0.16)	-89.86*** (20.23)
Median income (\$10k)	-0.01*** (0.00)	0.47 (0.56)
Population density (per square mile)	0.01*** (0.00)	0.77*** (0.25)
Median age	-0.01*** (0.00)	0.40 (0.31)

Table 3: Correlates with presence on Yelp and Reviews per restaurant. Each cell represents a coefficient from a separate linear regression with the dependent variables of (1) the number of restaurants identified on Yelp divided by the number of permitted food establishments in the census tract (Presence), and (2) the number of Yelp reviews per restaurant identified on Yelp in the census tract (Reviews). “Prop.” indicates that the variable is a proportion. \*/\*\*/\*\* denote statistical significance at  $\alpha$ -levels of 0.10, 0.05, and 0.01, respectively.