

Government Under Review

BR bostonreview.net/editors-picks-us-books-ideas/daniel-e-ho-becky-elias-government-under-review

6/14/2016



Even More Bureaucrats (1993) by Synnøve Anker Aurdal / photo by Bosc d'Anjou

Like many good arguments, this one started over a stiff drink. An Earl Grey MarTEAni, to be precise.

In January 2010, Nathalie Louissaint, a New York City health inspector, visited Pegu Club, an upscale cocktail bar. She watched as the bartender mixed the signature tea-infused drink. Borrowing a technique from the nineteenth century, the bartender added raw egg whites, which give the drink a silky body and an alluring layer of foam. Louissaint decided that the raw egg warning on the menu was insufficient and cited the bar for a health code violation.

The citation outraged many. Paul Clarke, a Seattle-based food writer, was perplexed by the department's rigid position on raw eggs, writing on the website *Serious Eats*, "Does this mean the health department will begin targeting restaurants that serve raw eggs in a Caesar salad?" Others decried the health department's seeming mandate to use pasteurized eggs, but those, said Pegu Club owner Audrey Sanders, "impart this really funky wet-diaper nose." One bartender, who insisted on anonymity for fear of reprisal, told the *New York Times*, "If they make it illegal to serve egg-white drinks, that would be Hurricane Katrina for us." In response to the uproar, the health department overruled the inspector.

In our experiment, health inspectors observing identical conditions disagreed 60 percent of the time.

This confusion is no outlier. Nationwide, implementation of health codes varies dramatically across inspectors and health departments. In Seattle, two inspectors observed Caesar salad dressing prepared with raw (unpasteurized) eggs in the same restaurant, but disagreed about whether to cite a violation. Contrary to New York City health department guidelines, New York State's website doesn't mention menu warnings, instead admonishing, "Consider using commercially pasteurized eggs in recipes that use eggs or consider removing the item from your menu." The Centers for Disease Control and Prevention (CDC) document that 80 percent of restaurants nonetheless use unpasteurized eggs.

When it comes down to it, the marriage fight is not so much about eggs as it is an endemic challenge across government. From airport security checkpoints and routine traffic stops to home construction permits, citizens and government interact frequently through individual officials. At times, the decisions of these frontline government officials can seem disturbingly arbitrary.

A bar can always take a drink off its menu, but sometimes the arbitrariness can have more serious impact. In 2013 the Administrative Conference of the United States reported that administrative law judges grant Social Security Disability claims at rates ranging from 4 percent to 98 percent. In asylum adjudication, New York immigration judges vary in their grant rates from 6 percent to 91 percent when cases are assigned irrespective of merits, leading to a denunciation of the process as “refugee roulette.” A study of Illinois child-welfare case managers found substantial differences in decisions to place children in foster care based on allegations of abuse or neglect. It is no wonder that scholars assail the child-welfare system as a form of institutional “chaos, oppression, and tragic ineffectiveness.” In nuclear safety, violation-detection rates can vary from less than 10 percent to more than 60 percent depending on the inspector. The regulatory requirements are so complex that one nuclear official conceded, “Nondetection is endemic.”

Inconsistency breeds mistrust. A city analyst in New York described the city’s restaurant inspection system as “arbitrary” and concluded, “If we can’t trust the Health Department to provide real scientific data . . . then we can’t trust any agency.” Businesses feel this too. An owner of a fast-food chain with many locations across Washington state observed sharp differences in how individual restaurants were scored and how the same restaurant was scored over time, despite his chain’s uniform food-safety protocol: “We always thought we were doing a great job in putting safety first; but it turns out in many, if not most, cases, inspectors were either not being as thorough as they could have been or were only verbally coaching” rather than writing violations. One Virginia bartender, responding to the Pegu Club episode, said, “I’m not 100 percent sure what the law is.” At a staff meeting in King County, Washington, health inspectors expressed similar misgivings. They hoped for better consistency in order to build “credibility and trust.”

The pervasiveness of these challenges leads some to point fingers, some to throw up their hands, and others to bemoan government altogether. If you are on the right, blame public sector unions, civil service protections, or listless bureaucrats. If you are on the left, blame underfunding, deregulation, or the lack of federal oversight. Governments generally cannot resolve these underlying ideological battles, but they must find ways to address the consistency of frontline decision making nonetheless. The question is, how?

The Peer Review Experiment

One possible solution is peer review. If frontline government officials could review and deliberate over each other’s work, the quality and consistency of decision making might improve. While isolated examples of such peer review exist, we have regrettably little systematic evidence of peer review’s effectiveness in the public sector. The reason is understandable. Due to perceived costs, logistics, and ethical and political concerns, rigorous experiments can be difficult to design and implement when it comes to regulation.

Beginning in 2014, we designed a randomized, controlled trial to test the effectiveness of peer review with the food safety staff of King County, where Seattle is located. Half of the inspection staff was randomly assigned to engage in peer review. For sixteen weeks, these inspectors spent one day per week with a randomly selected fellow inspector, taking turns conducting inspections and independently scoring health code violations. We then used information from these peer inspections to identify and train for violations that cause the most confusion.

The results were remarkable. We discovered that, when observing identical conditions in restaurants, health inspectors disagreed nearly 60 percent of the time. Inspectors differed in their assessments of risk magnitude and in interpretations and applications of the health code to particular circumstances, resulting in varying citations for the

same condition. Food science is evolving, and the FDA model food code spans nearly 800 pages, so it may not be surprising that implementation varies so much. As one inspector put it, “In the beginning, we [thought] we kn[e]w the code,” but comparing assessments with others provided a “wake-up call.”

The peer review process caused an average increase in violation detections of 17–19 percent. Because the increase was driven by inspectors who had previously been loath to cite violations, the net effect was to reduce the variability across inspectors in the peer review group. By comparison, the control group conducted inspections in the same fashion as before. The bottom line: peer review improved consistency across inspectors.

The most unexpected change was cultural. At the outset, a major concern within the inspection department was that peer reviewing would lead to tension within the staff. Perhaps inspectors would argue and criticize each other. As one inspector described an earlier attempt at “standardization,” review can be “torture” and “buil[d] resentment.”

But we found that the opposite happened. Rather than challenge one another, inspectors began to observe and learn from their partnerships. “Seeing the other person do their inspection helped highlight where my weaknesses are,” one inspector told us. “This is a very good thing because we usually go out by ourselves and can get stuck in our own way of thinking instead of expanding it by seeing other people and how they do things.” During weekly meetings, the peer review group clarified the code and risk assessments and discussed how to exercise their discretion effectively. While earlier trainings presented materials largely in lecture format, the peer-review trainings became more interactive. Inspectors evaluated pictures from the field and broke down code items and food science within teams.

Staff reported learning a wide range of skills from peers: how to defuse tension with an angry restaurateur, how to inquire about the food-preparation process to overcome a conventional criticism that inspections provide only a snapshot in time, and how to handle specialized code items for vulnerable populations, such as the elderly, children, and pregnant women. For instance, three inspectors noted what they learned from their fellow inspector Ray Solis, who had been on the job for eighteen years. When visiting delis, Solis made a habit of asking managers whether they could take apart their meat slicer. This is critical to food safety: just a year earlier, the failure to clean the inside of a superficially spotless meat slicer contributed to a multi-venue salmonella outbreak. By encouraging managers to disassemble equipment, Solis also gained an opportunity to build trust. If the manager couldn’t disassemble the unit, Solis could teach them.

High-performing athletes have coaches, so why not surgeons and government officials?

Peer review promoted a sense of professionalism, replacing a perception that inspections are mindless applications of checklists. One staff member captured this ethos, reporting to us, “A good inspector also should know many subjects and disciplines so that he or she can help to trouble shoot and provide a solution for operators like cooking, HVAC, plumbing, people skills, psychology, project management, construction materials, mechanics, proper cleaning techniques.” There were other benefits too. One inspector, who had been on staff more than seven years, didn’t realize until peer-review day that another person he’d seen around was also a food inspector. And for others the gains were downright existential. Wrote one inspector, “I do not feel so alone, which is nice.”

In light of all these benefits, Seattle and King County have now adopted peer review as an ongoing practice.

The Washington experiment proves that peer review can address what has long been a seemingly intractable problem of arbitrary bureaucracy. Even if results are limited to food safety, the experiment has significant implications. The Government Accountability Office describes the cobwebbed food safety system as “high-risk” due to “inconsistent oversight, ineffective coordination, and inefficient use of resources.” Each year, the CDC estimates some 128,000 individuals are hospitalized due to foodborne illness. While our experiment was underway, King County itself was investigating numerous large-scale outbreaks, one resulting in the recall of more than 500,000 pounds of salmonella-infected pork. Getting inspections right matters, and peer review helps health departments do that.

Beyond Food Safety

Peer review can be adapted to the myriad of federal, state, and local government units with decentralized decision making. Consider a few examples.

In the 1960s, federal district court judges in Detroit, Chicago, and Brooklyn experimented with peer review, meeting once a week to confer about sentencing decisions. “My limited experience in criminal work appalled me when I became a judge,” Judge Theodore Levin of Michigan’s Eastern District wrote. “Soon after taking the bench I realized that it was helpful to exchange views with my colleagues.” Judge Levin reported, based on data from internal deliberations, that the meetings of his “Sentencing Council” of peers affected sentencing decisions in a third of cases. These early experiments provide a historical counterfactual of promoting consistency of sentences by judicial peers, instead of the much-maligned later imposition of sentencing guidelines to promote consistency by rule.

In the late 1960s and early ’70s, the Oakland Police Department experimented with peer review. Officers who received complaints or were involved in a significant number of violent incidents could choose to discuss their cases with senior officers in place of formal disciplinary review. Discussions were raw and revelatory. One officer, after being confronted with his peers’ advice and criticism, admitted, “I had a problem and didn’t even know it . . . anytime there were insurmountable odds against me I was tearing-ass into it.” Among participants, incidents of aggression appeared to fall, but budget cuts killed the program before it could make a lasting impact.

In the case of child welfare, courts have stepped in to mandate a form of peer review. William Simon and Charles Sabel, leading researchers concerned with peer review in governance, studied the systems in Alabama and Utah, where peer-review teams take samples of case files and provide feedback to frontline officials. Simon and Sabel argue that such review promotes mutual understanding of how to apply the law to different cases.

In education, two constitutional law scholars, Rebecca Brown and Lee Epstein, decided one semester to sit in on each other’s classes and review each other’s pedagogy. “We taught each other in both substance and style,” Brown reported. Epstein’s “perspective on Justice [Robert] Jackson’s iconic opinion in the *Youngstown* case changed the way I understand and teach it,” she added, referring to an important Supreme Court case concerning executive power. Despite the fact that they taught the same cases, they discovered eye-opening differences in their interpretation and instruction of the material and hailed the semester as transformative.

To be sure, conducting and evaluating peer review takes time and resources. But bureaucracies make wholesale and costly changes all the time. By doing so for a (randomized) pilot group, it becomes possible to assess whether the intervention is effective, thereby potentially saving public dollars from fruitless initiatives. And while the above examples show peer review’s perceived value and feasibility outside the food safety context, the Washington randomized controlled trial offers the first powerful, systematic evidence of peer review’s wide-ranging effects.

Going It Alone is Not the Answer

In a well-known *New Yorker* story, the surgeon Atul Gawande related his sense of having plateaued in his profession. So he asked a fellow surgeon to observe and coach him. Gawande’s coach noticed a range of little problems, each of which could affect his patients. Gawande’s elbow was too high, impeding precise movement; for a half hour, the operating light wasn’t focused on the wound. Gawande concluded, “Coaching done well may be the most effective intervention designed for human performance.”

High-performing athletes have coaches, so why not surgeons and government officials? Going it alone is not the answer. Just as his coach enhanced Gawande’s skill and competence, peer review can improve the frontline administration of law. And just as coaching of physicians has broader implications in promoting patient health, peer

review in governance has far-reaching potential down the road: restoring citizen trust in government.

The authors wish to acknowledge Stanford Law Review, publishers of their complete research, "Does Peer Review Work? An Experiment of Experimentalism," in Issue 69, forthcoming.