http://pubsonline.informs.org/journal/mnsc/

MANAGEMENT SCIENCE
Articles in Advance, pp. 1–21
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

How Scheduling Can Bias Quality Assessment: Evidence from Food-Safety Inspections

Maria R. Ibanez,[a] Michael W. Toffel[b]

[a] Kellogg School of Management at Northwestern University, Evanston, Illinois 60208; [b] Harvard Business School, Boston, Massachusetts 02163

Contact: maria.ibanez@kellogg.northwestern.edu, http://orcid.org/0000-0002-0261-5397 (MRI); mtoffel@hbs.edu, http://orcid.org/0000-0002-1224-9159 (MWT)

Received: April 24, 2018
Revised: October 20, 2018
Accepted: January 9, 2019
Published Online in Articles in Advance: July 8, 2019

https://doi.org/10.1287/mnsc.2019.3318

Abstract. Accuracy and consistency are critical for inspections to be an effective, fair, and useful tool for assessing risks, quality, and suppliers—and for making decisions based on those assessments. We examine how inspector schedules could introduce bias that erodes inspection quality by altering inspector stringency. Our analysis of thousands of food-safety inspections reveals that inspectors are affected by the inspection outcomes at their prior-inspected establishment (*outcome effects*), citing more violations after they inspect establishments that exhibited worse compliance levels or trends. Moreover, consistent with negativity bias, the effect is stronger after observing compliance deterioration than improvement. Inspection results are also affected by when the inspection occurs within an inspector's day (*daily schedule effects*): Inspectors cite fewer violations after spending more time conducting inspections throughout the day and when inspections risk prolonging their typical workday. Overall, our findings suggest that currently unreported violations would be cited if the outcome effects—which increase scrutiny—were triggered more often and if the daily schedule effects—which erode scrutiny—were reduced. For example, our estimates indicate that if outcome effects were doubled and daily schedule effects were fully mitigated, 11% more violations would be detected, enabling remedial actions that could substantially reduce foodborne illnesses and hospitalizations. Understanding and addressing these inspection biases can help managers and policymakers improve not only food safety but also process quality, environmental practices, occupational safety, and working conditions.

History: Accepted by Serguei Netessine, operations management.
Funding: Financial support from Harvard Business School's Division of Research and Faculty Development is gratefully acknowledged.
Supplemental Material: The online appendix is available at https://doi.org/10.1287/mnsc.2019.3318.

Keywords: quality • scheduling • bias • discretion • inspection • behavioral operations • empirical • regulation • public health

1. Introduction

Many companies inspect their own and their suppliers' operations to ensure that they are meeting quality, labor, safety, and environmental standards. Various government agencies also inspect for regulatory compliance. The accuracy of quality audits and inspections is critical to their being a useful input to key managerial decisions, including how to allocate quality-improvement resources, which suppliers to source from, and how to penalize noncompliance. Inaccurate assessments can prevent managers, workers, customers, and neighbors from making well-informed decisions based on the risks imposed by an establishment's operations. Moreover, inspections that miss what they could have caught can undermine the inspection regime's ability to deter intentional noncompliance. In this study, we theorize and find evidence of several sources of bias that lead to inaccurate inspections. We also propose solutions—including alternative inspection-scheduling regimes—that can improve inspection accuracy without increasing inspection costs.

Several studies have revealed various sources of inspection inaccuracy, yet little is known about inspector bias. We consider an unexplored type of bias that results from an operational decision: scheduling. Building on work from the behavioral sciences, we hypothesize how the sequence of inspections might affect the number of violations cited. Specifically, inspector stringency on a particular inspection may be influenced by (a) the outcomes of the inspector's prior inspection (prior inspection outcome effects or, simply, outcome effects) and (b) its position within the inspector's day (daily schedule effects).

We study the influence of scheduling on inspection accuracy in the context of local health-department food-safety inspections of restaurants and other food-handling establishments. Although these inspections need to accurately assess compliance in order to protect consumer health, the number of violations cited in these reports is a function of both the facility's actual hygiene and the inspector's stringency in detecting and recording violations. Because citing violations requires supporting documentation, inspector bias takes the form of underreporting the violations that are actually present rather than reporting nonexistent ones. Using data on thousands of inspections, we find strong evidence that inspectors' schedules affect the number of violations cited.

We hypothesize three ways in which an inspector's experience at one inspection affects the number of violations cited at his or her next inspection (outcome effects). Throughout this paper, we refer to *an inspector's* preceding inspection as his or her "prior" inspection and *an establishment's* preceding inspection as its "previous" inspection. (Figure 1 illustrates this distinction and depicts the relationships we hypothesize.)

First, we hypothesize that an inspector's stringency will be influenced by the number of violations at his or her prior inspection. Those violations will affect the inspector's emotions and perceptions about the general compliance of the community of inspected establishments (via the salience of those recent inspection results), in turn altering his or her expectations and attitudes when inspecting the next establishment. This leads us to predict that having just conducted an inspection that cites more violations will lead the inspector to also cite more violations in the next establishment he or she inspects. As predicted, we find that *each additional violation* cited in the inspector's prior inspection (of a different establishment) increases by 1.7% the number of violations he or she cites at the next establishment.

Second, we hypothesize that trends matter, too: Discovering more compliance deterioration (or less improvement) at one inspection affects inspectors' emotions and perceptions in ways that lead them to cite more violations at the next establishment. Supporting this hypothesis, we find that inspectors cite 2.1% more (fewer) violations after having inspected another establishment whose violation trend worsened (improved) by one standard deviation.

Third, we hypothesize, based on negativity bias, that this trend effect will be stronger following an inspection that found deterioration than following one that found improvement. Indeed, we find empirical evidence that the trend effect is asymmetric, occurring when compliance at the inspector's prior establishment deteriorates, but not when it improves.

We then hypothesize two daily schedule effects. We first theorize that, over the course of a day, inspecting causes fatigue that erodes inspectors' stringency and leads them to cite fewer violations. We find empirical evidence to support this, observing that each subsequent hour an inspector conducts inspections during the day yields 3.7% fewer citations per inspection.[1] Second, we hypothesize that inspections that risk prolonging an inspector's workday will be conducted less stringently, which will lead to fewer violations being cited. We find empirical support for this, too, in that potentially shift-prolonging inspections yield 5.0% fewer citations.

Overall, our findings reveal that currently unreported violations would be cited if the outcome effects (which increase scrutiny) were triggered more often and the daily schedule effects (which erode scrutiny) were reduced. Our estimates suggest that, if the outcome effects were doubled (that is, amplified by 100%) and the daily schedule effects were fully mitigated (that is, reduced by 100%), the increase in inspectors' detection rates would result in their citing an average of 11% more violations.

Our work contributes to both theory and practice. By identifying factors that bias inspections, we contribute to the literature on monitoring and quality improvement (e.g., Mani and Muthulingam 2018). Our focus on how scheduling affects inspector stringency introduces the operational lens of scheduling to the literature on inspector bias, which has otherwise largely focused on experience or other sociological and economic factors (e.g., Short et al. 2016, Ball et al. 2017). Our examination of how operational decisions affect inspector behavior also contributes to the literature on behavioral operations, which emphasizes the importance of human behavior in operations-management decisions (Bendoly et al. 2006). By examining data from actual decisions with important consequences for public health, we contribute to the recent attempts to explore high-stakes decision making in field settings (e.g., Chen et al. 2016). We also go beyond previous work by estimating not only the magnitude of bias but also its real-world consequences. With managers across many different industries seeking to monitor and improve quality, our research suggests a cost-effective lever: exploiting the behavioral effects of the organization of work.

2. Related Literature

Our research builds on two streams of literature: (a) quality management and monitoring and (b) scheduling and task performance.

2.1. Quality Management and Monitoring

Scholars have for decades explored various approaches to ensuring that operations adhere to quality specifications. Research has, for example, examined total quality management (e.g., Lapré et al. 2000), programs that encourage self-disclosure of process errors and regulatory violations (e.g., Gawande and Bohara 2005, Kim 2015), and electronic monitoring systems (Staats et al. 2017). A primary approach remains physical inspections, such as internal quality-control departments assessing manufacturing processes (Shah et al.

2016), internal auditors assessing inventory records (Kök and Shang 2007), and third-party monitors assessing the conformance of supplier operations to buyers' codes of conduct (e.g., Handley and Gray 2013, Short and Toffel 2016) and to management standards such as International Organization for Standardization 9001 (Corbett 2006, Levine and Toffel 2010, Gray et al. 2015a).

An extensive literature has highlighted the role of inspections in fostering organizational learning (Short et al. 2019, Mani and Muthulingam 2018) and in promoting operational routines and adherence to Good Manufacturing Processes (e.g., Anand et al. 2012, Gray et al. 2015b), occupational health and safety regulations (e.g., Ko et al. 2010, Levine et al. 2012), and environmental regulations (for a review, see Shimshack 2014). Research has found compliance to be a function of an establishment's inspection history (including how many inspections it had undergone and the time lag between them) and inspector characteristics (including training, experience, and familiarity with an establishment) (Ko et al. 2010, Toffel et al. 2015). In contrast, we examine the extent to which an establishment's inspection report is influenced by the *inspector's* schedule, including (a) the inspector's experience at his or her prior inspection of a different establishment and (b) when during the inspector's day the inspection is conducted.

The usefulness of inspections is contingent on their accuracy. Researchers have long been interested in how to conduct quality-control inspections (e.g., Ballou and Pazer 1982), recognizing inspectors' fallibility and variability (Feinstein 1989). The limited number of studies of the heterogeneity across inspectors' propensity to report violations has identified the importance of their tenure, training, gender, and former exposure to the establishment (Macher et al. 2011, Short et al. 2016, Ball et al. 2017). Inspector accuracy among third-party inspection firms has been shown to be influenced by whether the establishment or its buyer hires the inspection firm and pays for the inspection (Ronen 2010, Duflo et al. 2013, Short and Toffel 2016), the level of competition among inspection firms (Bennett et al. 2013), and whether the inspecting firm has cross-selling opportunities (Koh et al. 2013, Pierce and Toffel 2013). In contrast to these demographic aspects of individual inspectors and structural dimensions of the relationship between the inspection firm and the inspected establishment, we explore a very different potential source of inspection bias: where the inspection falls within an inspector's schedule.

2.2. Scheduling and Task Performance

Our study also relates to research that has examined how work schedules affect task performance. This literature has, for example, proposed optimal scheduling of workforces (e.g., Green et al. 2013) and of periodic tasks such as machine inspections (e.g., Lee and Rosenblatt 1987). Studies of the sequencing of individual workers' tasks have shown that scheduling similar tasks consecutively to increase task repetition can improve performance by reducing delays incurred from switching tasks (e.g., Staats and Gino 2012, Ibanez et al. 2017) and that healthcare workers work more quickly later in a service episode (Deo and Jain 2019). We extend this work by focusing on the effects of work schedules on task quality in a setting that purports to provide inspections that are of consistent quality as the basis for a fair and objective monitoring regime.

A few studies have examined the relationship between work schedule and task quality. Dai et al. (2015) found that healthcare workers become less compliant with handwashing rules over the course of their shift. That study focused on adherence to a secondary task that was largely unobservable to others, where noncompliance was common and where fatigue might lead workers to shift their attention from this secondary task toward their primary tasks. In contrast, our study focuses on a primary task, the outcome of which (violations cited) is explicitly observable to others and the visibility of which could deter variation. Moreover, whereas Dai et al. (2015) measured adherence dichotomously, we use a more nuanced scalar measure. Another study examined the decisions of eight judges and found that they were more likely to deny parole as they issued more judgments since their last break (whether overnight or midday), suggesting that repeated decisions might have caused mental depletion (Danziger et al. 2011). Whereas judges became harsher as they made more decisions throughout the day, inspectors might behave differently, given that, for an inspector, greater harshness (manifested as stringency) requires more work.

Finally, two studies examined how workers adjust their decisions based on their prior decisions. A study of MBA application assessments found that the higher the cumulative average of the scores an interviewer had given to applicants at a given moment on a given day, the lower he or she scored subsequent applicants that day, suggesting that decision makers adjust their scores to maintain a consistent daily acceptance rate (Simonsohn and Gino 2013). Another study found that judges, loan reviewers, and baseball umpires were more likely to make "accept" decisions immediately after a "reject" decision (and vice versa), a form of decision bias (Chen et al. 2016). Whereas these two studies find that subsequent decisions typically oppose prior ones, inspectors do not have explicit or self-imposed quotas or targets, and, as we explain below, their emotions and perceptions may be affected by their prior tasks in ways that encourage subsequent decisions to be similar to prior ones. Additionally, we go beyond what prior work has considered by proposing that the magnitude of the effects from prior task outcomes will be asymmetric and will depend on whether the prior outcome was positive or negative.

3. Theory and Hypotheses

Quality-assurance audits and inspections have detailed procedures to be followed in pursuit of accuracy. Yet, in practice, behavioral biases may influence an inspector's stringency. Whereas inspections are typically assumed to yield the same results no matter when they occur on the inspector's schedule, we hypothesize that inspection results will indeed be influenced by the type of experience inspectors have at their immediately prior inspection (of a different establishment)—which we refer to as *prior inspection outcome effects*—and by when an inspection occurs during an inspector's daily schedule—which we refer to as *daily schedule effects*.

3.1. Prior Inspection Outcome Effects on Quality Assessment

3.1.1. Violation Level at the Inspector's Prior Inspected Establishment. We theorize that inspectors will be influenced by the results of prior inspections. One such outcome effect is driven by whether the establishment an inspector just visited had many or few violations. There are two reasons why inspecting an establishment with many violations can imbue inspectors with a negative attitude that leads them to inspect more diligently at their next inspection, whereas inspecting a more compliant establishment can lead them to be less stringent in their subsequent inspection. First, an inspector's prior inspection can affect him or her emotionally. When more violations are cited at that prior establishment, its personnel are more likely to be dissatisfied and resentful, which can lead to hostile interactions with inspectors that can erode their goodwill and thus heighten their stringency during the next inspection. Merely observing such dissatisfaction and resentfulness can similarly affect inspectors via emotional contagion (Barsade 2002). Conversely, finding fewer violations at the prior inspection is more likely to bolster an inspector's goodwill at the next inspection. Second, the experience at the inspector's prior inspection can shape his or her perceptions of the overall behavior of establishments, which can influence his or her stringency at the subsequent inspection. Recently experiencing an event (such as compliance) increases its salience and results in more rapid recall. An inspector may therefore use the results of that inspection to update his or her estimate of typical compliance levels, relying on the availability heuristic (Tversky and Kahneman 1974) and seeking evidence at his or her next inspected establishment that supports these expectations, consistent with confirmation bias (Nickerson 1998). This becomes a self-fulfilling prophecy, where experiencing poor (good) compliance at the prior establishment leads inspectors to heighten (reduce) scrutiny at their next inspection and therefore detect more (fewer) violations.[2]

We therefore hypothesize the following:

Hypothesis 1. *The more (fewer) violations an inspector cites at one establishment, the more (fewer) violations he or she will cite at the next establishment.*

3.1.2. Violation Trend at the Inspector's Prior Inspected Establishment. An inspector's behavior is shaped not only by the prior establishment's *level* of compliance, but also by its *change* in compliance relative to its previous inspection. This second type of outcome effect also results from how the prior inspection affects the inspector's emotions and perceptions.

The inspector's emotional response (through emotional contagion and interactions) at his or her prior establishment will depend on the trend there because the expectations of the establishment's personnel will be based on its previous inspection; they will be pleased or displeased according to whether their violation count has decreased or increased. After visiting an establishment with greater improvement, we predict that the inspector will exhibit a more positive temperament and approach his or her next inspection with greater empathy and less stringency.

An inspector's perceptions, too, may be biased by the change in violations at the prior establishment. Many inspectors view inspections as a cooperative endeavor with the regulated entity to help improve business operations and safeguard stakeholders (e.g., May and Wood 2003, Pautz 2009, Pautz 2010). Improved compliance may therefore be attributed to management taking the rules and regulations seriously—that is, cooperating—whereas worsened compliance may be attributed to management ignoring or deliberately flouting the rules—definitely not cooperating. Improved compliance therefore confirms a cooperative relationship that includes inspectors bestowing leniency (Hawkins 1983), which we assert can increase an inspector's faith that the overall community of inspected establishments is cooperating and thus lead him or her to be less stringent in the next inspection. Similarly, worsened compliance can lead an inspector to believe that the establishments are "defecting" from their commitment to compliance (Ayres and Braithwaite 1992), which we assert can lead the inspector to update beliefs about the overall community of inspected establishments as not being cooperative, triggering him or her to be more stringent in the next inspection. We therefore hypothesize the following:

Hypothesis 2. *The more an establishment's compliance has deteriorated (improved), the more (fewer) violations an inspector will record at the next establishment.*

3.1.3. Violation Trend at the Inspector's Prior Inspected Establishment: Asymmetric Effects of Deterioration vs. Improvement. According to the principle of *negativity bias*, negative events are generally more salient and dominant than positive events (Rozin and Royzman 2001). Negative events instigate greater information processing to search for meaning and justification, which in turn strengthens the memory and tends to spur stronger and more enduring effects in many psychological dimensions (Baumeister et al. 2001).

Negativity bias can affect the impact of the prior inspection's violation trend on the inspector's emotions and perceptions. First, negativity bias implies that for the inspected establishment's staff, the negative emotional effect of a drop in compliance will be stronger than the positive emotional effect of an improvement. This would result in stronger conveyance to inspectors of negative emotions associated with a drop in compliance and weaker conveyance of positive emotions associated with improvement. An inspector will then absorb more negative emotions after the negative finding than positive emotions after the positive finding. Moreover, as argued by Barsade (2002), mood contagion might be more likely for unpleasant emotions because of higher attention and automatic mimicry. These asymmetries in the extent to which declining versus improving conditions affect inspectors' emotions will lead, in turn, to asymmetric effects on the strength of the resulting positive or negative outcome effects.

Second, the salience of negative outcomes may have a stronger effect on inspectors' perceptions of how the establishments they monitor generally think about compliance, which can shape their stringency in a subsequent inspection. This is because of the *status-quo bias*: With the status quo acting as the reference point, negative changes are perceived as larger than positive changes of the same magnitude (Samuelson and Zeckhauser 1988, Kahneman 2003). We therefore hypothesize the following:

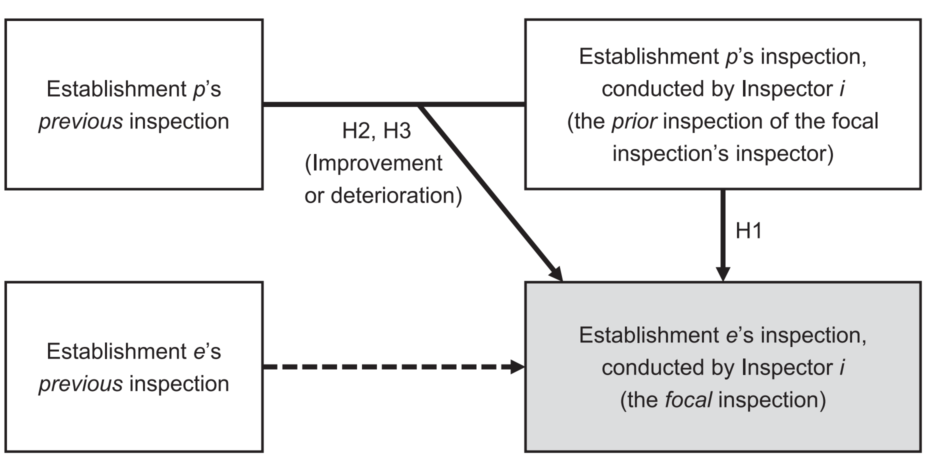
Hypothesis 3. *Observing deteriorated conditions at an establishment will increase the inspector's stringency at the next establishment to a greater extent than observing improved conditions will reduce his or her stringency.*

To summarize, Figure 1 depicts the relationships we theorize in our first three hypotheses.

3.2. Daily Schedule Effects on Quality Assessment

3.2.1. Inspector Fatigue. Inspectors are influenced not only by the results of prior inspections, but also by the sequencing of inspections within the day. Their work typically consists of a sequence of evaluative tasks that include physical tasks (such as manually examining the dimensions of a part or the temperature of a freezer) and mental tasks (such as interviewing an employee or determining whether a set of observations is within

Figure 1. Prior Inspection Outcome Effects



Notes. This diagram represents the history of Inspector *i* (downward arrow) and of two establishments, *p* and *e* (left to right). The shaded box represents the focal inspection, in which inspector *i* inspects establishment *e*. We refer to *an inspector's* preceding inspection as his or her "prior" inspection and *an establishment's* preceding inspection as its "previous" inspection. In this diagram, inspector *i* inspects establishment *p* and then establishment *e*, the focal inspection. H1 refers to the influence of the former on the latter. H2 and H3 refer to how the focal inspection is influenced by establishment *p*'s change in compliance compared with its previous inspection (Hypothesis 2) and propose that the effect is stronger when establishment *p* deterioriates than when it improves (Hypothesis 3). Prior research has focused, in contrast, on the relationship between an establishment's previous and focal inspections (depicted by the dashed arrow), such as an establishment's improvements as it undergoes successive inspections, the lag between inspections, inspectors' familiarity with an establishment from having inspected it before, and other factors related to the focal establishment's inspection history (e.g., Ko et al. 2010, Macher et al. 2011, Toffel et al. 2015, Ball et al. 2017, Mani and Muthulingam 2018).

acceptable standards). As these tasks are executed, physical and mental fatigue will increase (Brachet et al. 2012). Furthermore, experimental evidence indicates that mental fatigue itself increases physical fatigue (Wright et al. 2007, Marcora et al. 2009).

Over the course of a day, inspectors' physical and mental fatigue will reduce their physical and cognitive effort. This undermines stringency, which requires physical and cognitive efforts such as moving throughout the facility, interviewing personnel, waiting to observe work, executing procedures such as taking measurements, and conducting unpleasant tasks (such as observing storage practices in a walk-in freezer). Once an attribute is observed, inspectors need to recall and interpret the relevant standards to decide whether there is a violation and, if so, to document it. Each step must be executed according to rules that increase the complexity even of tasks that might appear simple to the untrained eye. Moreover, mental effort is required to make decisions against the status quo; as inspectors grow more tired during the day, they may become more willing to accept the status quo (Muraven and Baumeister 2000, Danziger et al. 2011), which, in the context of inspections, can take the form of passing inspection items. Finally, mental effort is required to withstand the social confrontations that can erupt when a finding of noncompliance is disputed by those working at the establishment, who may genuinely disagree and for whom, in any case, much may be at stake in terms of reputation and sales. Citing violations can also provoke threats of appeals and lawsuits. Anticipating such responses, inspectors who are growing fatigued may exert less effort and seek to avoid confrontation, both of which increase leniency. For all these reasons, we hypothesize the following:

Hypothesis 4. *Inspectors will cite fewer violations as they spend more time conducting inspections throughout the day.*

3.2.2. Potential Shift Prolonging. In many settings, workers have discretion over their pace, which can lead them to prolong tasks to fill the time available (Hasija et al. 2010) and to conduct work more quickly when facing higher workloads (KC and Terwiesch 2012, Berry Jaeker and Tucker 2017). Beyond these workload-related factors, we propose that inspectors will inspect less stringently when they expect to work later than usual (that is, beyond when they typically end work for the day). We hypothesize that inspectors' reluctance to suspend an inspection once under way—which would require them to bear the travel cost again the next day to finish the inspection—combined with a desire to finish at their typical time will create pressure to speed up and inspect less thoroughly. As workers approach their typical end-of-shift time, accomplishing whatever remaining work cannot be postponed can become increasingly pressing as their perceived opportunity cost of time increases. The desire to speed up in these circumstances can result in increased reliance on using workarounds and cutting corners (Oliva and Sterman 2001), which in turn can reduce the quality of the work performed. Because properly conducting inspections requires carefully evaluating a series of individual elements to identify whether each is in or out of compliance, omitting or expediting tasks to avoid prolonging the shift will result in a less comprehensive inspection with fewer violations detected and cited. We therefore hypothesize the following:

Hypothesis 5. *Inspectors will cite fewer violations at inspections when they are at risk of working beyond the typical end of their shift.*

4. Empirical Analysis

4.1. Empirical Context: Food-Safety Inspections

Our hypotheses are ideally tested in an empirical context in which inspectors work individually, which avoids the challenge of discerning individuals' behaviors from those of coinspectors. Food-safety inspections conducted by local health departments fulfill this criterion because environmental health officers are individually responsible for the inspection of restaurants, grocery stores, and other food-handling establishments to protect consumers by monitoring compliance and educating kitchen managers in their assigned geographical area. Moreover, food-safety inspections, commonly known as restaurant health inspections despite their broader scope, are designed to minimize foodborne illness; noncompliance can jeopardize consumer health. The quality of these assessments—and their ability to safeguard public health—depends on the accuracy of inspectors.

Foodborne disease in the United States is estimated to cause 48 million illnesses resulting in 128,000 hospitalizations and 3,000 deaths each year, imposing billions of dollars of medical costs and costs associated with reduced productivity and with pain and suffering (Scallan et al. 2011, Scharff 2012, Minor et al. 2015). Violations can affect firms' reputations and revenues and can trigger organizational responses that range from additional training to legal representation to refute citations.

Several prior studies have examined food-safety inspections. For example, Lehman et al. (2014) found that consumers are less concerned about food safety at restaurants that they perceive to be more "authentic." Others have investigated the extent to which restaurants improved hygiene practices once they were required to disclose their inspection results to consumers via restaurant grade cards (Jin and Leslie 2003, Simon et al. 2005, Jin and Leslie 2009). More recent studies have found that online customer reviews of restaurants contain text related to hygiene conditions that can predict health-inspection results (Kang et al. 2013) and can increase inspector effectiveness if health-inspection agencies take them into account when prioritizing establishments for inspection (Glaeser et al. 2016).

Because inspectors need evidence to justify citing violations (and thus can cite violations only if they are truly present), studies of inspection bias (e.g., Bennett et al. 2013, Duflo et al. 2013, Short et al. 2016) are based on the assumption that deviations from the true number of violations are due only to underdetection and that bias does not lead inspectors to cite nonexistent violations. This assumption was validated in our interviews with inspectors and underlies our empirical approach. Moreover, because violations are based on regulations that are based on science-based guidance for protecting consumers, each violation item is relevant.

We purchased data from Hazel Analytics, a company that gathers food-safety-inspection data from several local governments across the United States, processes the information to create electronic datasets, and sells these datasets to researchers and to companies—such as restaurant chains—interested in monitoring their licensees. These datasets include information about the inspected establishment (name, identification number, address, city, state, and ZIP code), the inspector, the inspection type, the date, the times when the inspection began and ended, the violations recorded, and, where available, the inspector's comments on those violations.

We purchased all of Hazel Analytics' inspection datasets that included inspection start and end times as well as unique identifiers for each inspector, all of which are necessary to observe inspector schedules. This included all food-safety inspections conducted in Lake County, Illinois, from September 4, 2013, to October 5, 2015; in Camden County, New Jersey, from September 4, 2012, to September 24, 2015; and in Alaska from December 8, 2007, to October 4, 2015. (The date range for each domain reflects the periods available at the time from Hazel Analytics.) From the 38,065 inspections in these raw data, excluding 46 exact duplicates believed to be data errors, we dropped 11,875 inspections (including the 64 Alaska inspections conducted in 2007) that were missing the lagged data necessary for us to construct the *prior inspected establishment's violation trend* variable. We then excluded 1,658 inspector-days (4,549 inspections) for which we cannot adequately calculate relevant variables based on what appear to be data-entry errors that we were unable to correct (for example, when there was ambiguity about inspection sequence). This yielded a clean sample of 21,641 inspections. The conditional fixed-effects Poisson regression models (described below) automatically dropped an additional 9,624 inspections corresponding to inspector–establishment combinations that had (a) just a single inspection or (b) multiple observations all citing zero violations. This results in an estimation sample containing 12,017 inspections of 3,399 establishments conducted by 86 inspectors on 6,880 inspector-days in Camden County (1,402 inspections), Lake County (8,962 inspections), and Alaska (1,653 inspections).[3]

To learn more about the setting, we observed and interviewed inspectors, environmental health department managers, store operators, and directors of food safety and quality assurance at retail companies. Our interviews with managers and inspectors at health-inspection departments represented in our data set indicate that inspectors have limited discretion over scheduling. Each inspector is responsible for inspecting all establishments within an assigned geographic territory. Inspectors rotate to different territories every 2–3 years. They are instructed to schedule their inspections by prioritizing establishments based on their due dates, which are computed for routine (and routine-education) inspections based on previous inspection dates and the required inspection frequency for an establishment type (based on the riskiness of its operations). Particular events (such as a consumer complaint or the need to verify that a severe violation has been rectified) may trigger more immediate due dates; we control for these in our models via the inspection-type dummies described below. To minimize travel time, inspectors are instructed to group inspections with similar due dates by geographic proximity.

Although inspectors also carry out many administrative duties (such as reviewing records, answering emails, and attending department meetings at the office), most of their work is inspections and the associated travel. As they prepare to conduct inspections, inspectors review the establishments' most recent inspections. Traveling between their office and establishments to inspect often accounts for a substantial portion of the day because of the geographical dispersion in the areas covered by our data. Inspectors are discouraged from working overtime.

When inspectors arrive at an establishment, they ask to speak to the person in charge and encourage this person to accompany them during their visit. During the inspection, they inspect the establishment (e.g., taking temperatures), observe workers' behaviors (e.g., whether and how they use gloves and wash their hands), and ask many questions to understand the processes (e.g., the procedures to receive food from suppliers). As they walk through the establishment, the inspectors point out the violations they find, explain the public-health rationale, and ask the personnel to correct them straightaway when possible. Although any immediately corrected violations are still marked as violations on the inspection form, this approach ensures that (a) the violations are corrected as soon as possible to improve food safety, and (b) the personnel learn how to be compliant. Because of (a) the immediate corrections, (b) the instruction about regulations and how to improve the processes in the future, and (c) the incentive for compliance resulting from effective monitoring and enforcement, the citing of each violation—or the failure to cite it—has a real impact on public health. Thus, reducing the underreporting of violations resulting from the effects we identify would improve actual compliance and health outcomes.

4.2. Measures

4.2.1. Dependent and Independent Variables. We measure *violations* as the number of violations cited in each inspection, a typical approach used by others (e.g., Helland 1998, Stafford 2003, Langpap and Shimshack 2010, Short et al. 2016).

Prior inspected establishment's violations is the number of violations the inspector cited at the establishment inspected prior to the focal inspection, whether minutes or days earlier. *Prior inspected establishment's violation trend* is calculated as the percentage change in the number of violations between the inspector's prior inspection and that establishment's previous inspection (adding 1 to the denominator to avoid dividing by 0).[4]

We create two indicator variables to distinguish whether the inspector's prior establishment had an improved, deteriorated, or not substantially changed violation count compared with its previous inspection. We classify an establishment's violation trend as *improved saliently* (or *deteriorated saliently*) if its current inspection yielded at least two fewer (more) violations than its previous inspection. (The intermediate case, in which the number of violations differed by only 1 or remained constant, is the baseline condition.) We create the dummy variables *prior inspected establishment saliently improved*, coded 1 when the inspector's prior inspected establishment *improved saliently* and 0 otherwise; and *prior inspected establishment saliently deteriorated*, coded 1 when the inspector's prior inspected establishment *deteriorated saliently* and 0 otherwise. An inspection conducted immediately after the inspection of an establishment whose performance change was only 1 or no violations is considered the baseline condition.

We measure an inspector's schedule-induced fatigue at a given inspection as *time inspecting earlier today*—computed as the cumulative number of hours (with minute precision) inspectors spent onsite in their prior inspections that day before the focal inspection—to better account for the fact that some inspections take longer than others and that longer (and not just more numerous) inspections are likely to cause more fatigue. (Our results are robust to measuring schedule-induced fatigue in several alternative ways, as described in the Robustness Tests section.)

To measure whether an inspection might reasonably be anticipated to conclude after the inspector's typical end-of-shift time, we calculated (a) the inspection's anticipated end time as the inspection start time plus the duration of that establishment's previous inspection and (b) the inspector's typical end-of-shift time as the inspector's running average daily clock-out time based on all of his or her preceding days in our sample. For both metrics, we use time stamps at the minute level, but convert them to fractions of hours. We created the indicator variable *potentially shift-prolonging*, coded 1 when the inspection's anticipated end time fell after the inspector's typical end-of-shift time and 0 otherwise.

4.2.2. Control Variables. We measure *inspector experience* as the number of inspections the inspector had conducted (at any establishment) since the beginning of our sample period by the time he or she began the focal inspection.

We create an indicator variable, *returning inspector*, coded 1 when the inspector of the focal inspection had inspected it before and 0 otherwise.

We create an indicator variable, *lunch period*, coded 1 when the inspection began between 11:00 a.m. and 3:59 p.m., the period that tends to be especially busy for kitchen operations.

We create a series of indicator variables to control for whether the inspection is the *establishment's* n*th inspection* (2*nd*–10*th or more*), each of which indicates whether an inspection is the establishment's second or third (and so on) inspection in our sample period; first is the omitted category.

We create a series of inspection-type dummies to indicate whether the inspection was (a) routine, (b) routine-education, (c) related to permitting, (d) due to a complaint, (e) an illness investigation, or (f) a follow-up. *Routine inspections* are conducted to periodically monitor establishments; *routine-education inspections* are similar to routine inspections but also include an educational presentation to train establishment staff. These two types make up 79% of the inspections in our estimation sample. *Permit inspections* are conducted when establishments change ownership or undergo construction, upgrades, or remodeling. *Complaint inspections* are triggered by the local health department receiving a complaint. Because Camden County does not classify particular inspections as triggered by complaints but does record complaint dates and the inspectors assigned to investigate them, *complaint risk inspections* refers to all inspections those inspectors conducted the day—and the day after—they were assigned to investigate a complaint. *Illness-investigation inspections* are those conducted to investigate a possible foodborne illness (food poisoning). A *follow-up inspection* (or reinspection) is conducted to verify that violations found in a preceding inspection have been corrected and is therefore of limited scope. *Other inspections* includes visits to confirm an establishment's deactivation/closure and inspections of mobile establishments, vending machines, and temporary events such as outdoor festivals; this is the omitted category in our empirical specifications.

Table 1 reports summary statistics. Additional descriptive statistics are provided in Online Appendix A. Establishments are inspected an average of 2.2 times per year. Inspections yield an average of 2.42 violations; an average of 0.54 of these violations were also cited in the establishment's previous inspection, and 1.88 (that is, 77%) are new violations.

4.3. Empirical Specification

We test our hypotheses by estimating (via quasi-maximum likelihood) a conditional fixed-effects Poisson model that predicts the number of *violations* cited in an inspection. Our specification exploits the double-panel structure of the data, where an inspection could be viewed as both the nth inspection of establishment e and the jth inspection of inspector i in our sample.

Our independent variables include (1) the inspector's *prior inspected establishment's violations*—that is, the number of violations that inspector i cited at his or her most recent inspection of any other establishment; (2) the *prior inspected establishment's violation trend* or, in some specifications, the two variables that indicate particular ranges of that variable: *prior inspected establishment saliently improved* and *prior inspected establishment saliently deteriorated*; (3) inspector i's *time inspecting earlier today*; and (4) an indicator for whether the inspection was *potentially shift-prolonging*.

The model also includes several controls. First, we control for *inspector experience* (Macher et al. 2011, Short et al. 2016). We control for *returning inspector* because inspectors who return to an establishment they had inspected before tend to behave differently

Table 1. Summary Statistics

Variable	Description	Mean	Standard deviation	Minimum	Maximum
Violations	Number of violations cited in the inspection	2.42	2.73	0	25
Prior inspected establishment's violations (Hypothesis 1)	Number of violations cited at the establishment inspected by the inspector immediately prior to the focal inspection	2.11	2.62	0	25
Prior inspected establishment's violation trend (Hypothesis 2)	Percentage change in the number of violations between the inspector's prior inspection and that establishment's previous inspection (adding 1 to the denominator to avoid dividing by 0)	0.42	1.58	–0.95	23
Prior inspected establishment saliently improved (Hypothesis 3)	Indicates if the inspector's prior inspected establishment *improved saliently* (i.e., its current inspection yielded at least two *fewer* violations than its previous inspection)	0.24	0.43	0	1
Prior inspected establishment saliently deteriorated (Hypothesis 3)	Indicates if the inspector's prior inspected establishment *deteriorated saliently* (i.e., its current inspection yielded at least two *more* violations than its previous inspection)	0.21	0.41	0	1
Time inspecting earlier today (Hypothesis 4)	Time (in hours) already spent conducting inspections on the same day before the focal inspection	0.87	1.06	0	8.42
Potentially shift-prolonging (Hypothesis 5)	Indicates if the anticipated end time of an inspection (calculated as the inspection start time plus the duration of that establishment's previous inspection conducted by any inspector) falls after the inspector's running average daily clock-out time based on all of that inspector's preceding days in our sample	0.26	0.44	0	1
Inspector experience	Number of inspections the inspector had conducted (at any establishment) since the beginning of our sample period by the time he or she began the focal inspection	520.09	303.30	1	1,429
Returning inspector	Indicates if the inspector of the focal inspection had inspected the establishment beforehand	0.84	0.37	0	1
Establishment's nth inspection (1st–10th or more)	Denotes whether an inspection is the establishment's *n*th (first, second, third, and so on) inspection in our sample period (modeled in our empirical specification as a series dummies for 2nd–10th or more, using the first as the baseline category)	4.02	2.06	1	10
Lunch period (11:00 a.m.–3:59 p.m.)	Indicates if the inspection began 11:00 a.m.–3:59 p.m.	0.66	0.47	0	1

Note. N = 12,017 inspections.

than inspectors who are there for the first time (Short et al. 2016, Ball et al. 2017).

We include *lunch period* to control for the possibility that an establishment's cleanliness might vary over the course of a day and because prior research indicates that many individual behaviors are affected by time of day (Linder et al. 2014, Dai et al. 2015). We also include two sets of fixed effects to denote the month and the year of the inspection.

We include a series of fixed effects to control for the *establishment's nth inspection* (2*nd*–10*th or more*) because research has shown that, in other settings, establishments improve compliance over subsequent inspections (Ko et al. 2010, Toffel et al. 2015).

Because different types of inspections might mechanically result in different numbers of violations (e.g., due to different scopes), the model includes *inspection-type* dummies.

Finally, we include fixed effects for every inspector–establishment combination. These inspector–establishment dyads control for all time-invariant inspector characteristics (such as gender, formal education, and other factors that might affect his or her average stringency) and all time-invariant establishment characteristics (such as cuisine type and neighborhood). Thus, our specification identifies changes in the number of violations that a particular inspector cited when inspecting a given establishment on different occasions. Including inspector–establishment fixed effects also avoids concerns that our results are driven by spatial correlation; specifically, the concern that proximate establishments that inspectors tend to visit sequentially might exhibit similar violation counts because they share neighborhood characteristics that might affect the supply of and demand for compliance. Including fixed effects for inspector–establishment dyads is more conservative than including separate sets of fixed effects for inspectors and for establishments; a robustness test that includes these separate sets of fixed effects yields similar results.

4.4. Identification

We took several steps to ensure that our empirical approach tests our hypothesized relationships, controlling for or ruling out alternative plausible explanations. For example, the positive correlation between the number of violations that inspectors cite at a focal establishment and at their prior establishment could result not only from the mechanism represented in Hypothesis 1, but also if inspectors clustered on their schedules the establishments they expected to yield many (or few) violations. Our inspector interviewees revealed that they in fact tended to cluster inspections of establishments near each other in order to minimize travel time. Although violations might be spatially correlated due to demographic clustering, our inclusion of fixed effects for inspector–establishment dyads controls for such time-invariant establishment characteristics.

We test our hypothesis that inspector fatigue reduces inspector stringency (Hypothesis 4) by looking for evidence that fewer violations are cited at inspections conducted later in an inspector's daily schedule. But that could have two other explanations. First, daily trends in customer visits, staffing levels, and staff-cleaning efforts could result in better hygiene conditions later in the day. Our inspector interviews indicated, however, that many violations reflect longer-term problems whose propensity does not change throughout the day (e.g., sinks functioning improperly) and that hygiene conditions often get worse (not better) as establishments serve more customers, which would bias against our hypothesized effect. Our specifications nonetheless include fixed effects for time of day to control for potential variation in establishments' cleanliness at different time periods of the day. Second, inspectors might intentionally schedule "dirtier" establishments—those with historically more violations and thereby expected to have more violations—earlier in their daily schedule, leaving "cleaner" establishments for later in their schedule. However, two supplemental analyses yielded no evidence for this. A simple correlation analysis reveals that an establishment's previous inspection violation count is not significantly related to when in an inspector's daily schedule its focal inspection is conducted (Pearson's $\chi^2 = 285$, $p = 0.93$). Moreover, Poisson regression results enable us to rule out that inspectors intentionally sequenced, to any meaningful degree, their day's inspections based on establishments' previous violations. Specifically, a Poisson regression that predicts how long an inspector has already been conducting inspections that day before he or she begins to inspect the focal establishment (*time inspecting earlier today*) based on the focal establishment's previous violation count (and including inspector-day fixed effects as controls) indicates that more violations in a previous inspection predicts that an establishment's subsequent inspection will be scheduled slightly *later* in the inspector's shift [$\beta = 0.016$, standard error (S.E.) = 0.006, with standard errors clustered by inspector-day], which would be a bias against our hypothesized effect.

Finally, we test our hypothesis that an inspection being potentially shift-prolonging reduces inspector stringency (Hypothesis 5) by assessing whether shift-prolonging inspections yield fewer violations. However, shift-prolonging inspections might also yield fewer violations if, as an inspector's normal shift end-time approaches, he or she intentionally chooses to inspect establishments anticipated to yield fewer violations in order to minimize how late he or she will need to work, presuming "cleaner" establishments can be inspected more quickly. Two supplemental analyses, however, rule that out. First, establishments with a potentially shift-prolonging inspection averaged 3.1 violations in their previous inspection, significantly *more* than the average of 2.3 average violations in the previous inspection of establishments whose inspections were not potentially shift-prolonging (Pearson's $\chi^2 = 243$, $p < 0.01$). Second, a logistic regression indicates that the probability of an establishment's inspection being potentially shift-prolonging slightly increases if its previous inspection yielded more violations. Specifically, regressing a dummy indicating whether an establishment's inspection is *potentially shift-prolonging* on the violation count from its previous inspection and inspector-day fixed effects yields a significant positive coefficient on the violation count ($\beta = 0.104$, S.E. = 0.013, clustered by inspector-day). Both results would bias against our hypothesized effect.

4.5. Results

4.5.1. Model Results. We estimate the count model using fixed-effects Poisson regression and report standard errors clustered by establishment (Table 2). Poisson panel estimators are consistent, even if the data are not Poisson-distributed, provided the conditional mean is correctly specified (Azoulay et al. 2010, Cameron and Trivedi 2010). Because of the weaker distributional assumption of the Poisson panel estimators, they may be more robust than negative binomial regression (Cameron and Trivedi 2010).

Our results are robust to several alternatives: clustering standard errors by inspector, estimating the model with negative binomial regression with conditional fixed effects, and estimating the model using ordinary least-squares (OLS) regression predicting log violations. Multicollinearity is not a serious concern, given that variance inflation factors are less than 1.68 for all hypothesized variables and less than 6.01 for all variables, except three of the inspection-type indicators. Because our specifications control for a variety of factors that affect the number of violations cited, we interpret coefficients on the hypothesized variables as evidence of bias, as done in prior studies (e.g., Chen et al. 2016, Short et al. 2016). Because deviations from the true number of violations are assumed to result only from underdetection (as described above), we interpret negative coefficients to indicate the extent of underdetection, whereas positive coefficients indicate

Table 2. How Inspectors' Schedules Influence Inspection Outcomes

			Dependent variable: *Violations*	
Type of schedule effect	Hypothesis	Independent variable	(1)	(2)
Prior inspection outcome effects	Hypothesis 1	*Prior inspected establishment's violations*	0.017*** (0.004)	0.015*** (0.004)
	Hypothesis 2	*Prior inspected establishment's violation trend*	0.013** (0.006)	
	Hypothesis 3	*Prior inspected establishment saliently improved*		0.011 (0.023)
	Hypothesis 3	*Prior inspected establishment saliently deteriorated*		0.079*** (0.027)
Daily schedule effects	Hypothesis 4	*Time inspecting earlier today*	−0.038*** (0.011)	−0.039*** (0.011)
	Hypothesis 5	*Potentially shift-prolonging*	−0.051** (0.025)	−0.050** (0.025)
		Inspector experience	0.001*** (0.000)	0.001*** (0.000)
		Returning inspector	−0.114*** (0.035)	−0.117*** (0.035)
		Lunch period (11:00 a.m.–3:59 p.m.)	−0.049** (0.024)	−0.049** (0.024)
		Month fixed effects	Included	Included
		Year fixed effects	Included	Included
		Establishment's nth inspection (2nd–10th or more) fixed effects	Included	Included
		Inspection-type fixed effects	Included	Included
		Inspector–establishment dyad fixed effects	Included	Included
		Number of observations (inspections)	12,017	12,017

Note. Poisson regression coefficients with robust standard errors clustered by establishment.
$^{**}p < 0.05$; $^{***}p < 0.01$.

the extent to which underdetection is avoided. We interpret effect sizes based on incidence rate ratios (IRRs).

We test Hypotheses 1, 2, 4, and 5 using model 1. We begin by interpreting the coefficients on our control variables. The estimated coefficient on *inspector experience* is positive and statistically significant, suggesting that, all else constant, the number of violations cited per inspection increases as the inspector conducts inspections over time, albeit by a small amount on an inspection-by-inspection level. The negative and statistically significant coefficient on *returning inspector* ($\beta = -0.114$, $p < 0.01$, IRR = 0.892) indicates that inspectors who return to an establishment cite 11% fewer violations[5] than inspectors who had not inspected that establishment before, which is consistent with prior studies. Considering time-of-day effects, we note that, on average, inspections conducted during the *lunch period* cite 5% fewer violations than inspections conducted at other times of the day ($\beta = -0.049$, $p < 0.05$, IRR = 0.952). The estimated coefficients on the *establishment's nth inspection* (not reported) indicate that fewer violations were cited at successive inspections of a given establishment, a result consistent with prior research on other types of inspections. For example, the estimated coefficient on the dummy variable denoting establishments' third inspection ($\beta = -0.210$, $p < 0.01$, IRR = 0.811) indicates that those inspections cite 18.9% fewer violations on average than establishments' initial inspection.

To explore the influence of the outcome at the inspector's prior inspected establishment, we first consider the number of violations cited in that inspection. The coefficient on *prior inspected establishment's violations* is positive and statistically significant ($\beta = 0.017$, $p < 0.01$, IRR = 1.017), which supports Hypothesis 1. Each additional citation at the establishment inspected immediately before the focal inspection increases the number of violations cited in the focal inspection by 1.7%.

The statistically significant positive coefficient on *prior inspected establishment's violation trend* ($\beta = 0.013$, $p < 0.05$, IRR = 1.013, IRR-1 = 1.3%) supports Hypothesis 2. A one-standard-deviation increase in this trend (that is, a 1.58 increase) increases the number of citations in the focal inspection by 2.07%. Note that this is in addition to the effect of the number of violations (Hypothesis 1).

To test Hypothesis 3, model 2 replaces prior inspected establishment's violation trend with the indicator variables *prior inspected establishment saliently improved* and *prior inspected establishment saliently deteriorated*. The baseline condition occurs when the prior inspected establishment had no more than one violation more or less than it had in its previous inspection. Compared with this baseline condition, we find that inspectors cite more violations after their prior inspected establishment exhibited salient deterioration ($\beta = 0.079$, $p < 0.01$, IRR = 1.082). The IRR indicates that, on average, an inspector who has just inspected an establishment with salient deterioration will report 8% more violations in the focal inspection. However, we find no evidence that observing salient improvement in the prior inspected establishment has any effect on the number of violations cited in the focal inspection. A Wald test indicates that these effects significantly differ (Wald $\chi^2 = 4.74$, $p < 0.05$), which supports Hypothesis 3: The spillover effect on the focal inspection of having observed salient deterioration in the prior inspected establishment is statistically significantly stronger than the spillover effect of having observed salient improvement.

Model 1 also supports both of our hypothesized daily schedule effects. The negative, statistically significant coefficient on *time inspecting earlier today* ($\beta = -0.038$, $p < 0.01$, IRR = 0.963) indicates that inspectors cite 3.7% fewer violations for each additional hour already spent conducting inspections that day, which supports Hypothesis 4. Applying this 3.7% effect to the 2.42 average violations per inspection yields an average marginal effect of 0.09 fewer violations being cited per inspection for each additional hour the inspector has been conducting inspections throughout the day. This amounts to 90 violations not being cited for every 1,000 inspections conducted by inspectors who had already inspected for 1 hour that day, 180 violations not being cited for every 1,000 inspections conducted by inspectors who had already inspected for 2 hours that day, and so on.

The negative, statistically significant coefficient on *potentially shift-prolonging* ($\beta = -0.051$, $p < 0.05$, IRR = 0.950) indicates that inspections that risked extending an inspector's workday result in 5.0% fewer citations, as predicted by Hypothesis 5. Applying this 5% decline to the 2.42 average violations yields an average marginal effect of 0.12 fewer violations cited in each potentially shift-prolonging inspection. This amounts to 120 violations not cited for every 1,000 potentially shift-prolonging inspections, a substantial number given that 26% of inspections in our sample are *potentially shift-prolonging*.

4.5.2. Results Interpretation. Our main results indicate that inspectors, despite their effort and training, are vulnerable to decision biases that lead them to underreport violations in predictable ways. Thus, these behavioral effects have real implications because they affect citation rates of actual violations. If inspectors' detection rates were improved so that they cited the violations that are currently going unreported due to the scheduling biases we identify, establishments could improve their food-safety practices in two ways. First, they can improve compliance immediately because many violations can be instantly rectified. Second, they can improve future compliance because citations not only motivate establishments to improve the processes that generated them but also more broadly motivate compliance to prevent other violations, which is the deterrent intent of monitoring and enforcement. Thus, citations prompt behavioral responses that improve compliance, which in turn prevents foodborne health incidents.

Prior research that reveals decision biases tends to focus on quantifying their magnitudes. Improving the accuracy of inspectors' citations of violations is in itself a very important outcome, one that organizations and governments care deeply about. We go beyond that typical approach by also estimating the real-world consequences. Our efforts to translate our primary findings (how scheduling affects the citations of violations) into their broader societal impacts (health consequences) would be equivalent to, for example, Chen et al. (2016) not only quantifying a source of bias among baseball umpires in calling balls and strikes (which that paper does), but also estimating how a team's win/loss record would be affected if umpires called pitches more accurately—that is, without the identified bias (which that paper does not consider). Similarly, it would be akin to Chen et al. (2016) not only revealing an important source of asylum judges' decision bias, but also estimating the resulting social injustice.

Thus, to better understand the potential benefits of addressing these biases, we develop nationwide estimates of how many fewer violations would be underreported—and the consequent healthcare outcomes and costs that would be avoided—if inspection managers implemented measures such as better awareness, new training, and different scheduling regimes that would somewhat take into account these biases. We consider the impact of interventions that would exploit outcome effects and ameliorate daily schedule effects, leading inspectors to cite violations that currently go underreported. We estimate the effects of such interventions on the average inspection based on our sample, scale up the results to estimate how many currently undetected violations would be cited nationwide, and then estimate how many fewer foodborne-illness cases and hospitalizations would result and the associated reduction in healthcare costs. Translating the effects into health outcomes and costs is difficult; what we offer are back-of-the-envelope calculations. Our methodology and results (including assumptions and caveats) are described in Online Appendix B, but we briefly report some key results here.

In the ideal scenario, the outcome effects (which increase scrutiny) would be fully triggered all the time, and the daily schedule effects (which erode scrutiny) would be entirely eliminated. In practice, different interventions would have different effectiveness. Figures B1–B3 in the online appendix illustrate a range of scenarios. If the drivers of outcome and daily schedule effects were, respectively, doubled (amplified by 100%) and fully mitigated (reduced by 100%), inspectors would cite 11% more violations, yielding 267,315 additional violations cited annually nationwide, which would result in 56,471 fewer foodborne-illness-related hospitalizations and 21.09 million fewer foodborne-illness cases and would reduce foodborne illness costs by \$15.75 billion to \$34.29 billion. Amplifying the outcome effects and mitigating the daily schedule effects by 50% would generate half of these gains: 127,352 additional violations cited, 26,903 fewer hospitalizations, 10.05 million fewer foodborne-illness cases, and savings of \$7.51 billion to \$16.34 billion in foodborne illness costs. Even a very conservative 10% scenario elicits substantial benefits: 24,536 additional violations cited annually nationwide, 5,183 fewer foodborne-illness-related hospitalizations, 1.94 million fewer foodborne-illness cases, and savings of \$1.45 billion to \$3.15 billion.

4.6. Robustness Tests

We conduct several analyses to confirm the robustness of our findings. Our primary results are based on a conservative approach that includes fixed effects for inspector–establishment dyads. We find similar results whether we instead include establishment fixed effects or separate sets of fixed effects for inspectors and for establishments [estimating the latter with Poisson regression led to convergence problems that led us to instead use OLS regression to predict $\log(violations + 1)$] or if we include the leave-out-means instead of individual fixed effects (Chen et al. 2016). Our results are robust to omitting the establishment's nth inspection (2nd–10th or more) fixed effects or replacing them with a continuous measure of the establishment's inspection sequence. All of our results also hold when we remove the inspection-type indicators and instead control for whether the inspection was routine. Our results are also robust to suppressing any of the other fixed effects.

To assess whether unusually busy days, which might make inspectors especially fatigued, might be driving our schedule-induced fatigue (Hypothesis 4) results, we reestimated our models on the subsample of inspector-days with no more than six inspections (the 99th percentile). Our hypothesized results are robust to this subsample test.

Our results regarding the effects of schedule-induced fatigue (Hypothesis 4) hold even when we measure it using any of the following four alternative approaches rather than the time spent conducting prior inspections on the day of the focal inspection. In our first alternative, we calculate the *number of prior inspections today*, coded 0 for an inspector's first inspection of the day, 1 for the second, and so on. Although this does not account for the fact that some inspections take longer than others and that longer (and not just more numerous) inspections are likely to cause more fatigue, it captures the setup costs of inspecting each establishment. In our second alternative approach, we use *time since starting the first inspection today* (*hours*), computed as the number of hours (measured with minute precision) elapsed since the inspector began his or her first inspection of the day and the starting time of the focal inspection. This incorporates the *time inspecting earlier today* (that is, the time spent actually conducting inspections) and the time that elapsed between those inspections, which could also add to fatigue. Our last two alternative approaches accommodate the concern that fatigue might have increased the duration of the inspector's prior inspections that day. First, instead of *actual* time onsite (*time inspecting earlier today*), we use the *anticipated* cumulative hours inspectors would spend onsite in their prior inspections that day, computed as the average of the durations of those establishments' previous two inspections (or the duration of their single previous inspection if only one is available). Second, we use the *predicted* cumulative hours inspectors would spend onsite in their prior inspections that day, calculated as the predicted durations derived from an ordinary least-squares regression model, with a log-transformed outcome variable and including the covariates from the corresponding main specification.

The evidence supporting our hypotheses is also robust to including, as additional controls in our primary models, indicator variables denoting the day of the week the inspection occurred. Our results are robust to substituting our control for *lunch period* with three time-of-day periods to designate when the inspection began: *breakfast period* (midnight–10:59 a.m.), *lunch period* (11:00 a.m.–3:59 p.m.), and *dinner period* (4:00 p.m.–11:59 p.m.), omitting one as the baseline category. Our results are mostly robust to substituting our control for *lunch* time with indicator variables for each hour of the day at which the inspection occurred. Only the *potentially shift-prolonging* coefficient is no longer statistically significant, likely due to the higher multicollinearity introduced by this approach. Moreover, our results are robust to not controlling for time of day at all by eliminating the control for *lunch* time.

Our results are robust to controlling for various aspects of the establishment's past inspections, using several measures from the literature (Macher et al. 2011, Anand et al. 2012, Mani and Muthulingam 2018), including the length of the focal establishment's previous inspection; the *years since the establishment's previous inspection* (with decimals; this averaged 0.42 in our sample); and four alternative approaches to measure the focal establishment's compliance history: (1) the number of violations cited in its previous inspection; (2) the running tallies of the number of the focal *establishment's previous inspections without violations* and the number of the focal *establishment's previous inspections with violations*; (3) separating the focal establishment's inspection history between the focal inspector and other inspectors, measured as the running tallies of the number of the focal *establishment's previous inspections* (a) *without violations conducted by the focal inspector*, (b) *with violations conducted by the focal inspector*, (c) *without violations conducted by other inspectors*, and (d) *with violations conducted by other inspectors*; and (4) whether the focal establishment's two previous inspections yielded no violations (a *good record*), at least one violation (a *bad record*), or when the focal establishment's previous inspection resulted in more violations than its inspection before that one (a *deteriorating record*).

Our main specification controls for inspectors' experience at the focal establishment via *inspector experience* and *returning inspector*, which are complemented by the establishment's inspection history (i.e., indicators for whether the inspection is the *establishment's* n*th inspection*—2nd–10th or more) and the indicators for the inspector–establishment dyad. Our results are robust to replacing *inspector experience* and *returning inspector* with (a) *site-specific experience*, measured as the number of times during our sample period that the inspector had inspected the focal establishment by the time he or she began conducting the focal inspection; and (b) *other experience*, measured as the number of inspections during our sample period that the inspector had conducted at all other establishments by the time he or she began conducting the focal inspection. (Our results are robust to both maintaining and removing the *establishment's* n*th-inspection* indicators.) The negative and statistically significant coefficient on *site-specific experience* suggests that, on average and holding everything else constant, fewer violations are cited when inspectors have more experience inspecting the establishment. The estimated coefficient on *other experience* suggests that more violations are cited, although only a small amount more, when inspectors' experience inspecting other establishments is higher. These results complement prior findings that site-specific experience is associated with greater hazard of future recalls and weakens the relationship between inspection outcomes and future recall hazards, consistent with inspector's complacency instead of learning (Ball et al. 2017). Our results provide evidence of this link, with site-specific experience leading to inspectors citing fewer violations.

Finally, our results are robust to controlling for *weekly workload* or *monthly workload,* measured as the number of inspections the inspector conducted the week or the month of the focal inspection. As an aside, the estimated coefficient on workload is not significant, suggesting that, despite the prevalence of workload effects in other settings, inspectors in our sample are resilient to them. This shows that inspection outcomes are difficult to influence and makes our identified effects even more impressive (Prentice and Miller 1992).

4.7. Extensions

Our primary results test the effects on inspector scrutiny of conducting inspections that risk prolonging the inspector's shift (Hypothesis 5). We also investigate whether the extent to which the shift might be prolonged matters (Online Appendix C). To measure the extent to which an inspection might reasonably be anticipated to extend beyond the inspector's typical end-of-shift time, we calculate *potential extent of shift-prolonging* (in fractions of hours) by subtracting the "inspector's typical end-of-shift time" from the "inspection's anticipated end time" (both defined in Section 4.2.1), using the difference when it is positive and otherwise coding this variable as 0. For both metrics, we use time stamps at the minute level but convert them to hours (and fractions of hours). Including this continuous variable instead of the binary *potentially shift-prolonging* lets us examine how the magnitude of shift-prolonging affects inspector scrutiny. As reported in the online appendix (columns 3–4 of Table C1), the negative, statistically significant coefficient on *potential extent of shift-prolonging* ($\beta = -0.068$, $p < 0.01$, IRR = 0.934) indicates that inspectors cite 6.6% fewer violations for each hour that the inspection risked going past his or her typical end-of-shift time. We also pursued a more flexible approach by creating a series of dummy variables denoting the following ranges of *potential extent of shift-prolonging*: (a) up to 0.5 hours (the omitted baseline category), (b) above 0.5 hours and up to 1 hour, (c) above 1 hour and up to 1.5 hours, and (d) above 1.5 hours (columns 5–6 of Table C1 in the online appendix). These specifications yield the same inferences as our main models.

We conduct additional analysis to examine the persistence of some of our outcome effects (Online Appendix D). To explore whether these outcome effects persist beyond the next inspection, we added two variables to our models: the inspector's *penultimate inspected establishment's violations* (that is, two establishments ago) and then also the inspector's *antepenultimate inspected establishment's violations* (three establishments ago). The significant positive coefficients on both indicate that the number of violations cited is significantly affected not only by the violations at the inspector's immediately preceding inspection but also by each of the two inspections before that; the declining magnitudes of these coefficients indicates that the effect dissipates (columns 1–2 of Table D1 in the online appendix). Second, we assessed whether the outcome effects attenuated if an inspector's successive inspections occur across different days, rather than on the same day. We replaced *prior inspected establishment's violations* with two variables: *prior inspected establishment's violations for the first inspection of the day* (coded as *prior inspected establishment's violations* for the first inspection of the day and 0 otherwise) and *prior inspected establishment's violations for the second+ inspection of the day* (coded 0 for the first inspection of the day and as *prior inspected establishment's violations* otherwise). Finding on both variables nearly identical significant negative coefficients that are statistically indistinguishable (Wald $\chi^2 = 0.09$, $p = 0.76$) indicates that an overnight break did not attenuate the outcome effects (column 3 of Table D1 in the online appendix). These results reduce the likelihood that the inspector's mood or other temporary factors are driving the outcome effects.

We explore other ways that breaks could affect inspector scrutiny (Online Appendix E). Our primary results that tested Hypothesis 4 indicate that inspectors exhibit more scrutiny following an overnight break. This finding is similar in some ways to the finding of Danziger et al. (2011) that parole judges' decisions were affected by overnight breaks. But although those judges' decisions were similarly affected after they took each of two food breaks, we find no evidence that the length of an inspector's break between inspections during the day affects scrutiny. We reach this conclusion by estimating a model that includes two variables measuring the length of two types of breaks: (a) *overnight break length*, which measures the amount of time elapsed between the end of the inspector's prior inspection on a preceding day and the start of the focal inspection that is the inspector's first inspection of a day; and (b) *within-day break length*, which measures the amount of time elapsed between the end of the inspector's prior inspection on the same day as the focal inspection and the start of the focal inspection. Both of these variables are measured in tens of hours and are top-coded at their 99th percentiles to avoid outliers influencing results. Adding these two variables to our primary specification (column 1 of Table E1 in the online appendix) yields a positive significant coefficient on *overnight break length*, which indicates that break length matters for the first inspection of the day (specifically, longer overnight breaks increase inspector's violation-detection rate), but a nonsignificant coefficient on *within-day break length*, which provides no evidence that break length matters for breaks within the day. We also find no evidence of within-day break length affecting inspectors when we add to our primary specification a set of dummy variables for breaks of different lengths (a break of less than 1 hour, a break of 1–2 hours, a break of 2–3 hours, and a break longer than 3 hours, with overnight breaks as the omitted category).

We highlight a few other ways in which our study and that of Danziger et al. (2011) resemble and differ from each other. First, the directions of the effects: Overnight breaks in our context lead inspectors to *restore* their "harshness" by overlooking fewer violations, whereas breaks in the Danziger et al. (2011) study lead judges to *reduce* their "harshness," with greater tendency to grant parole. The apparent disparity resolves, however, if breaks are viewed as leading both types of decision makers to reduce status-quo bias associated with fatigue. Specifically, fatigue leads inspectors to increasingly overlook violations that require effort to discover, document, and defend and thus to accept the status quo of not citing a violation and allowing the establishment to continue operating as it currently does; similarly, fatigue leads judges to be "more likely to accept the default, status quo outcome: deny a prisoner's request" (Danziger et al. 2011: 6889). In both studies, breaks serve to counter this status-quo bias resulting from fatigue. Second, although judges' and inspectors' decisions are both affected by overnight breaks, only judges appear to be affected by breaks within the day.[6]

We assess whether the outcome effects on inspector scrutiny are affected by how much time had lapsed since the inspector's prior inspection by adding to our primary model *overnight break length* and *within-day break length* and their interactions with the two outcome-effect variables, *prior inspected establishment's violations* and *prior inspected establishment's violation trend*. None of the coefficients on these four interaction terms is statistically significant, yielding no evidence that break time (whether overnight or within-day) attenuates either of these outcome effects. (Results are reported in column 2 of Table E1 of the online appendix.)

We also assess whether our hypothesized daily schedule effect regarding inspector fatigue (Hypothesis 4) is moderated by break length. That main result indicated that inspectors cited fewer violations after having inspected for more time on that day. Although that provides evidence that overnight breaks enhance inspector scrutiny, we explore whether the within-day decline in inspector scrutiny was affected by within-day break length. We add to our primary specification *break length* [the amount of time that had lapsed since the inspector's prior inspection, measured as tens of hours since the inspector's prior inspection (top-coded at its 99th percentile to avoid outliers influencing results)] and its interaction with *time inspecting earlier today* (our Hypothesis 4 measure). The coefficient on the interaction term is not statistically significant, yielding no evidence that break length during the day affects the rate at which inspector scrutiny declines throughout their day. (Results are reported in column 3 of Table E1 of the online appendix.)

We also investigate the extent to which our hypothesized effects influence the citing of two types of violations: (a) *critical violations*, which are related to food-preparation practices and employee behaviors that more directly contribute to foodborne illness or injury; and (b) *noncritical violations*, which are overall sanitation and preventative measures to protect foods, such as proper use of gloves, that are less risky but also important for public health. We find that the four schedule effects (*prior inspected establishment's violations, prior inspected establishment's violation trend, time inspecting earlier today*, and *potentially shift-prolonging*) significantly affect noncritical violation citations and that three of them (all but *potentially shift-prolonging*) significantly affect critical violation citations. (Results are reported in Online Appendix F.)

We also examine whether our hypothesized effects influence other aspects of inspections that might be linked to scrutiny (see Online Appendix G). We find that inspectors conduct inspections more quickly as they progress through their shifts: Inspection duration decreases by 3.1% for each additional hour of inspection earlier in the day. Moreover, the inspector's citation pace—*violation citations per hour*, a measure of productivity in this setting, representing the net of the effects on violations and inspection duration—decreases by 1.9% for each additional hour of prior inspections that day. *Potentially shift-prolonging* inspections are conducted 4.3% more quickly, but citation pace remains largely unaffected; thus, our main finding that *potentially shift-prolonging* inspections result in fewer violations is likely due to inspectors' desire to avoid working late, rather than to fatigue eroding their citation pace. We find little to no evidence of outcome effects on inspection duration and conclude that our main outcome-effect findings—that more violations and worsening trends at an inspector's prior establishment increase the inspector's citations at his or her next inspection—mostly result from inspectors increasing their citation pace rather than from spending more time onsite.

Finally, we examine whether our hypothesized effects are associated with documentation effort. We find no evidence that average violation comment length (in characters or words) is influenced by the *time inspecting earlier today*, an inspection being *potentially shift-prolonging*, or the *prior inspected establishment's violations* (results not reported). However, we do find that an increase in the *prior inspected establishment's violation trend* is associated with a decrease in the focal inspection's comment length (in characters and words).

That is, on average, an inspector documents the focal inspection with shorter comments when the prior establishment exhibited worsening violation trends. Thus, a potential mechanism by which such trends might increase citation pace at the focal inspection (that is, improve inspectors' productivity in citing violations) is by shifting some effort from documentation to detection.

5. Discussion

We find strong evidence that inspectors' evaluations are affected by their experience at the prior establishment they inspected. We also find that inspectors' scrutiny is influenced by their daily schedules: As inspectors conduct inspections throughout their workday, their scrutiny is eroded by increasing fatigue and by the perceived time pressure to finish before the typical end of their shift. The effect magnitudes that we identify, ranging from 1.3% to 8.2% individually and 11% overall, are large compared with decision bias among professionals in other field settings—such as the 0.5% effect size regarding decision bias exhibited by judges, 0.9% by baseball umpires, and 2.1%–6.9% by social auditors—and compared with experimental results yielding biases of 0–8 percentage points by loan-review officers (Chen et al. 2016, Short et al. 2016).

5.1. Contributions

This study contributes to three literature streams. First, it is among the first to bring an operational lens to the literature on monitoring and assessment of standards adherence. In particular, we identify important scheduling effects on the scrutiny and thus the accuracy of those who monitor establishments' adherence to standards. We contribute to this literature's focus on improving monitoring schemes' effectiveness by analyzing how inspection outcomes are affected by outcomes of prior inspections at other establishments and by inspectors' daily schedules.

Second, by identifying spillover effects between inspections, our findings contribute to a related literature on the spillover effects of regulatory sanctions (e.g., Cohen 2000, Shimshack and Ward 2005). Although that literature focuses on how an inspection agency's monitoring and enforcement affect its reputation for stringency, which has a spillover influence on other establishments' compliance, our study focuses on how inspectors' experiences at one establishment have spillover effects on their scrutiny at others. Ours is thus the first study of which we are aware that identifies spillover effects on inspector stringency associated not only with the outcomes of the immediately preceding inspection, but also with how many prior inspections an inspector had already conducted that day and with the inspector's apparent desire to avoid working late. We contribute to the nascent literature on the accuracy of inspections—specifically, of regulatory regimes and third-party monitoring of labor conditions in supply chains—that has largely focused on inspector bias due to economic conflicts of interest, team composition, and site-specific experience (e.g., Duflo et al. 2013, Short and Toffel 2016, Short et al. 2016, Ball et al. 2017). To our knowledge, our study is the first to bring the operational lens of scheduling to this literature by showing how work schedules can drive inaccuracies.

Third, we contribute to the literature on the performance implications of scheduling and task sequencing. By examining actual decisions with important consequences for consumers, we contribute to the recent attempts to explore high-stakes decision making in field settings (e.g., Chen et al. 2016). The idiosyncrasies of quality-evaluation decisions result in biases that are different from those for other types of decisions. In contrast to a prior study that finds that judges, loan reviewers, and baseball umpires are more likely to make an "accept" decision following a "reject" decision (and vice versa) (Chen et al. 2016), we find the opposite relationship among inspectors' assessments of subsequent establishments. This disparity could be due to inspectors being less susceptible to the causes of the negative autocorrelation found in those other settings. First, the law of small numbers and the gambler's fallacy, whereby the decision maker underestimates the likelihood of sequential streaks occurring by chance, are ameliorated because inspectors know the establishments they sequentially inspect often share external factors such as competition that can explain their exhibiting similar compliance (that is, sequential streaks). Second, sequential contrast effects, whereby the decision maker's perception of the quality of the current establishment is negatively biased by the quality of the previous one, might be mitigated by inspectors' targeted training to evaluate quality consistently based on what they observe and evidence they can document, and also by the longer time between inspections (including traveling from their prior inspected establishment). Third, quotas or targets (in terms of violations cited or overall assessments of an establishment) that could lead an individual's decisions to be influenced by his or her recent decisions are rare in the context of inspections. Further research is needed to identify circumstances under which decisions are similar or opposite to prior decisions. Furthermore, we find that this effect was asymmetric: Prior negative outcomes are much more influential than prior positive ones. To the best of our knowledge, our findings provide the first evidence of how sequential decision making is influenced by negativity bias, whereby negative events are generally more salient and dominant than positive events (Rozin and Royzman 2001).

In contrast to prior research that finds judges becoming *more* stringent as they make more decisions since their last break (whether overnight or midday) (Danziger et al. 2011), we find that inspectors become *less* stringent. One way to resolve this apparent contradiction is to consider that in both studies, decisions made over the course of the day tend toward yielding the same result that would occur in the absence of a decision being made, a form of status-quo bias. In particular, judges tend toward denying parole, which results in the same situation that would have occurred in the absence of a hearing; inspectors tend to avoid citing violations, the same situation that would have occurred in the absence of an inspection. Differing relationships between effort and stringency might also contribute to the opposing effects observed between judges and inspectors. Although judges can exert stringency by denying parole without justification and thus with little effort, for inspectors to exhibit stringency, they must find proof of violations, which requires physical and mental effort to interact with establishment staff. Fatigue associated with additional inspections can thus impede violation detection. Finally, although both parole decisions and inspections result in mental fatigue, inspections also trigger physical fatigue; it is plausible that mental and physical fatigue affect stringency differently.

Our daily schedule-effects findings also complement the literature that has found that increased worker fatigue after long hours led to accidents among nuclear and industrial plant operators, airline pilots, truck drivers, and hospital workers (Dinges 1995, Landrigan et al. 2004). In response to such findings, industry standards and regulations have capped the number of consecutive work hours in some of these professions; our results indicate that such policies might also improve inspection accuracy. We contribute to this debate by providing evidence of the negative effects of fatigue on work quality during normal shifts (rather than the very long work periods others have examined) in a different setting (health inspections), focusing on primary tasks (rather than secondary ones). Moreover, we investigate a different performance dimension (accuracy of quality assessments) and identify potential remedies. To the best of our knowledge, we are the first to provide evidence of the negative impact on quality assessment of within-day fatigue and potentially shift-prolonging tasks. The results of our extension analysis suggest that inspectors themselves might attempt to ameliorate these effects by focusing on critical violations at the expense of detecting fewer noncritical violations and producing less documentation.

In addition, our finding that inspectors inspect less stringently as they approach the time they typically end their workday contributes to a broader understanding of how workers alter their procedures as they approach the end of their shift [e.g., Chan's (2018) finding that hospital physicians concluding their shifts accept fewer patients and make different decisions about patient care]. More broadly, our work shows that even workers who lack formal shifts and have some flexibility to schedule their own work hours behave differently as they approach the typical end of their workday, suggesting that work is done differently toward the end of a workday in more settings than previously conceived. Our work also responds to the call for behavioral research in the operations-management field (Bendoly et al. 2006) by identifying ways in which task sequencing affects worker behavior. Our finding that inspectors' experiences at prior inspections bias their subsequent inspections shows that the outcome of tasks can affect how humans—unlike machines—perform their next task.

5.2. Managerial Implications

Extrapolating our study's results to the approximately 1 million food-handling establishments monitored annually across the United States suggests that hundreds of thousands of violations of food-safety regulations are likely being systematically overlooked each year. The effect is far larger if one considers food-safety inspections conducted around the world, as well as inspections worldwide in other domains such as environmental, quality, and financial management. Moreover, overlooking even a single food-safety violation matters beyond the correction of that violation because citing that violation could have increased the salience of food safety to the establishment's personnel and thus spurred other improvements. Personnel can get very upset if even one violation is cited (Rossen 2017). In our observations of food-safety inspections, we witnessed several occasions in which establishment staff were frustrated when an inspection yielded even one violation. Compared with the average of 2.42 violations cited per inspection, citing one fewer violation constitutes a 41% decrease, a large change that creates unfairness across facilities and impedes accurate decisions being taken in response to inspection reports. Thus, more accurate inspections that result in fewer violations being overlooked could prompt more effort to fully comply with food-safety standards. For example, franchisors could be better equipped to interpret inspection reports so as to know which franchisees require more (or less) oversight.

Moreover, regulators and private-sector inspectors across industries can take steps to mitigate these biases in order to create more accurate inspection reports, which would yield fairer and more comparable results across inspected establishments, generate more reliable information for consumers, and better motivate compliance. For example, our identified outcome effects imply that increasing the salience of noncompliance and thus the need to enforce regulation could increase the number of violations detected. This suggests that reminders of noncompliance to inspectors—or other ways to increase such salience to them—could be a lever for inspection managers to increase inspectors' stringency, even if the information is already available to those inspectors and despite their innate desire to protect consumers.

With respect to daily schedule effects, one way to reduce the extent to which these biases erode inspection accuracy is to limit fatigue effects by smoothing the number of inspections per day or, if inspection capacity needs allow it, by capping the number of inspections a given inspector can conduct each day. Another approach, which can be used at the same time, is to minimize the number of shift-prolonging inspections by reallocating an inspector's weekly schedule to reduce variation in the predicted completion time of their final inspection each day or by shifting administrative tasks (such as office meetings) from the beginning to the end of the day. Reorganizing inspectors' schedules could eliminate these negative outcomes and might—according to our interviews with health inspectors in the areas covered in our data as well as in other areas across the United States—be possible without adding cost.

Managers can also use our findings to develop policies to reduce the *consequences* of inspector biases eroding inspection accuracy. For example, understanding that scrutiny typically declines as inspectors (a) conduct successive inspections during the day and (b) conduct inspections that risk prolonging their shift, the inspectors themselves could be required to schedule establishments that pose greater risks earlier in the shift. Such changes could reduce risk to public health.

5.3. Limitations and Future Research

Our study has several limitations that could be explored in future research. Although our data contain details of inspections and citations, we do not observe inspectors' beliefs or their on-site interactions. We find that they cite fewer violations after inspecting establishments that had fewer violations. Perhaps they make less effort to find hidden violations and are more willing to take a coaching approach for borderline violations—training operators to operate with better hygiene rather than writing citations. Possible extensions of our study could use observations of these actions to quantify how they are affected by scheduling. In addition, although our research context—food-safety inspections—is common worldwide, it is just one of many types of inspections conducted by companies and governments. Future research should examine whether the relationships we identified hold in other contexts.

Acknowledgments

The authors thank numerous health officers and coordinators, store operators, directors of food safety and quality assurance, Xiang Ao, Ryan Buell, Matthew Grochowski, Matthew Johnson, Andrew Marder, Melissa Ouellet, Ananth Raman, Brad Staats, Ioannis Stamatopoulos, and participants in the Harvard Business School TOM Workshop for helpful comments and support.

Endnotes

[1] Our supplemental analysis demonstrates that this effect is not attributable to inspectors' scheduling "presumably cleaner establishments" later in their workday.

[2] Other types of decisions might exhibit the opposite bias, whereby successive decisions are negatively autocorrelated, akin to the law of small numbers, the gambler's fallacy, sequential contrast effects, or quotas (Chen et al. 2016). These effects are likely weak in the case of inspections. First, the law of small numbers and the gambler's fallacy, in which the decision maker underestimates the likelihood of sequential streaks occurring by chance, are less likely to apply to inspectors. Instead, inspectors can be expected to predict a high likelihood that the establishments they sequentially inspect will exhibit similar compliance (that is, sequential streaks) because they share external factors that affect their compliance, including their competition, regulatory knowledge, and requirements about whether they must disclose their inspection results (such as restaurants in Los Angeles, New York, and Boston being required to post restaurant grade cards). Second, sequential contrast effects, in which the decision maker's perception of the quality of the current establishment is negatively biased by the quality of the previous one, are ameliorated because inspectors are extensively trained to evaluate quality based on what they observe and thus have well-defined evaluation criteria that reduce the influence of prior inspections as temporary reference points. Moreover, each inspection takes significant time and often involves additional time traveling across inspected entities, so decisions are farther apart than sequential instantaneous decisions that may lead to unconscious contrasts of establishments. Third, quotas for the number of positive (or negative) decisions (in terms of violations cited or overall assessments of an establishment) would imply that fewer positive (or negative) decisions could be made after a prior positive (or negative) decision. Although the immediate prior decision would not directly matter, the cumulative prior decisions could. However, inspectors typically lack quotas or targets.

[3] These sample restrictions do not affect our inferences, as all of our hypothesized results hold when using alternative specifications estimated on all inspections in the raw data set (results not reported).

[4] Let V_2 denote the number of violations in the inspector's prior inspection, and let V_1 denote the number of violations in the previous inspection of that same establishment. The violation trend is computed as the percentage change in violations, except that we add 1 to the denominator: *Prior inspected establishment's violation trend* = $[(V_2 - V_1)/(V_1 + 1)]$.

[5] This percentage is calculated as IRR-1 = 0.892 − 1 = −0.108 = −10.8%.

[6] There are several reasons why midday breaks might have influenced parole judges in the context of Danziger et al. (2011), but not the inspectors we study. First, judges hear cases in their courtroom continuously, except when they take breaks, whereas inspectors travel to different establishments to conduct successive inspections, resulting in breaks between every inspection. Second, judges' breaks between cases might bestow more mental relief than inspector's breaks between inspections, because inspectors need to be mentally engaged while driving to their next inspection. Third, breaks within the day might not meaningfully restore physical fatigue, and inspectors (but not judges) exert a lot of physical energy throughout the day (e.g., examining kitchens, inspecting freezers, and traveling). Fourth, the judges' midday breaks always included food, whereas (presumably) not all of the inspectors' breaks did. Fifth, the decisions of parole judges and food-safety inspectors require different

amounts of documentation to justify and face different appeals processes. Sixth, cultural differences might play a role; for example, the differences between judges and inspectors and between Israel (the setting of Danziger et al.) and the United States (our setting). Future research should explore how different types of breaks affect various types of decisions.

References

Anand G, Gray J, Siemsen E (2012) Decay, shock, and renewal: Operational routines and process entropy in the pharmaceutical industry. *Organ. Sci.* 23(6):1700–1716.

Ayres I, Braithwaite J (1992) *Responsive Regulation: Transcending the Deregulation Debate* (Oxford University Press, Oxford).

Azoulay P, Graff Zivin JS, Wang J (2010) Superstar extinction. *Quart. J. Econom.* 125(2):549–589.

Ball G, Siemsen E, Shah R (2017) Do plant inspections predict future quality? The role of investigator experience. *Manufacturing Service Oper. Management* 19(4):534–550.

Ballou DP, Pazer HL (1982) The impact of inspector fallibility on the inspection policy in serial production systems. *Management Sci.* 28(4):387–399.

Barsade SG (2002) The ripple effect: Emotional contagion and its influence on group behavior. *Admin. Sci. Quart.* 47(4):644–675.

Baumeister RF, Bratslavsky E, Finkenauer C, Vohs KD (2001) Bad is stronger than good. *Rev. General Psych.* 5(4):323–370.

Bendoly E, Donohue K, Schultz KL (2006) Behavior in operations management: Assessing recent findings and revisiting old assumptions. *J. Oper. Management* 24(6):737–752.

Bennett VM, Pierce L, Snyder JA, Toffel MW (2013) Customer-driven misconduct: How competition corrupts business practices. *Management Sci.* 59(8):1725–1742.

Berry Jaeker JA, Tucker AL (2017) Past the point of speeding up: The negative effects of workload saturation on efficiency and patient severity. *Management Sci.* 63(4):1042–1062.

Brachet T, David G, Drechsler AM (2012) The effect of shift structure on performance. *Amer. Econom. J. Appl. Econom.* 4(2):219–246.

Cameron AC, Trivedi PK (2010) *Microeconometrics Using Stata*, revised edition (Stata Press, College Station, TX).

Chan DC (2018) The efficiency of slacking off: Evidence from the emergency department. *Econometrica* 86(3):997–1030.

Chen DL, Moskowitz TJ, Shue K (2016) Decision making under the gambler's fallacy: Evidence from asylum judges, loan officers, and baseball umpires. *Quart. J. Econom.* 131(3):1181–1242.

Cohen MA (2000) Empirical research on the deterrent effect of environmental monitoring and enforcement. *Environ. Law Reporter News Anal.* 30(4):10245–10252.

Corbett CJ (2006) Global diffusion of ISO 9000 certification through supply chains. *Manufacturing Service Oper. Management* 8(4): 330–350.

Dai H, Milkman KL, Hofmann DA, Staats BR (2015) The impact of time at work and time off from work on rule compliance: The case of hand hygiene in healthcare. *J. Appl. Psych.* 100(3):846–862.

Danziger S, Levav J, Avnaim-Pesso L (2011) Extraneous factors in judicial decisions. *Proc. Natl. Acad. Sci. USA* 108(17):6889–6892.

Deo S, Jain A (2019) Slow first, fast later: Temporal speed-up in service episodes of finite duration. *Production Oper. Management*. Forthcoming.

Dinges DF (1995) An overview of sleepiness and accidents. *J. Sleep Res.* 4(S2):4–14.

Duflo E, Greenstone M, Pande R, Ryan N (2013) Truth-telling by third-party auditors and the response of polluting firms: Experimental evidence from India. *Quart. J. Econom.* 128(4):1499–1545.

Feinstein JS (1989) The safety regulation of U.S. nuclear power plants: Violations, inspections, and abnormal occurrences. *J. Political Econom.* 97(1):115–154.

Gawande K, Bohara AK (2005) Agency problems in law enforcement: Theory and application to the U.S. Coast Guard. *Management Sci.* 51(11):1593–1609.

Glaeser EL, Hillis A, Kominers SD, Luca M (2016) Crowdsourcing city government: Using tournaments to improve inspection accuracy. *Amer. Econom. Rev.* 106(5):114–118.

Gray JV, Anand G, Roth AV (2015a) The influence of ISO 9000 certification on process compliance. *Production Oper. Management* 24(3):369–382.

Gray JV, Siemsen E, Vasudeva G (2015b) Colocation still matters: Conformance quality and the interdependence of R&D and manufacturing in the pharmaceutical industry. *Management Sci.* 61(11):2760–2781.

Green LV, Savin S, Savva N (2013) "Nursevendor problem": Personnel staffing in the presence of endogenous absenteeism. *Management Sci.* 59(10):2237–2256.

Handley SM, Gray JV (2013) Inter-organizational quality management: The use of contractual incentives and monitoring mechanisms with outsourced manufacturing. *Production Oper. Management* 22(6):1540–1556.

Hasija S, Pinker E, Shumsky RA (2010) Work expands to fill the time available: Capacity estimation and staffing under Parkinson's Law. *Manufacturing Service Oper. Management* 12(1):1–18.

Hawkins K (1983) Bargain and bluff: Compliance strategy and deterrence in the enforcement of regulation. *Law Policy Quart.* 5(1): 35–73.

Helland E (1998) The enforcement of pollution control laws: Inspections, violations, and self-reporting. *Rev. Econom. Statist.* 80(1):141–153.

Ibanez MR, Clark JR, Huckman RS, Staats BR (2017) Discretionary task ordering: Queue management in radiological services. *Management Sci.* 64(9):4389–4407.

Jin GZ, Leslie P (2003) The effect of information on product quality: Evidence from restaurant hygiene grade cards. *Quart. J. Econom.* 118(2):409–451.

Jin GZ, Leslie P (2009) Reputational incentives for restaurant hygiene. *Amer. Econom. J. Microeconom.* 1(1):237–267.

Kahneman D (2003) Maps of bounded rationality: Psychology for behavioral economics. *Amer. Econom. Rev.* 93(5):1449–1475.

Kang JS, Kuznetsova P, Choi Y, Luca M (2013) Where not to eat? Improving public policy by predicting hygiene inspections using online reviews. *Proc. 2013 Conf. Empirical Methods Natural Language Processing* (Association for Computational Linguistics, Seattle), 1443–1448.

KC DS, Terwiesch C (2012) An econometric analysis of patient flows in the cardiac intensive care unit. *Manufacturing Service Oper. Management* 14(1):50–65.

Kim S-H (2015) Time to come clean? Disclosure and inspection policies for green production. *Oper. Res.* 63(1):1–20.

Ko K, Mendeloff J, Gray W (2010) The role of inspection sequence in compliance with the US Occupational Safety and Health Administration's (OSHA) standards: Interpretations and implications. *Regulation Governance* 4(1):48–70.

Koh K, Rajgopal S, Srinivasan S (2013) Non-audit services and financial reporting quality: Evidence from 1978 to 1980. *Rev. Accounting Stud.* 18(1):1–33.

Kök AG, Shang KH (2007) Inspection and replenishment policies for systems with inventory record inaccuracy. *Manufacturing Service Oper. Management* 9(2):185–205.

Landrigan CP, Rothschild JM, Cronin JW, Kaushal R, Burdick E, Katz JT, Lilly CM, et al (2004) Effect of reducing interns' work hours on serious medical errors in intensive care units. *New England J. Medicine* 351(18):1838–1848.

Langpap C, Shimshack JP (2010) Private citizen suits and public enforcement: Substitutes or complements? *J. Environ. Econom. Management* 59(3):235–249.

Lapré MA, Mukherjee AS, Van Wassenhove LN (2000) Behind the learning curve: Linking learning activities to waste reduction. *Management Sci.* 46(5):597–611.

Lee HL, Rosenblatt MJ (1987) Simultaneous determination of production cycle and inspection schedules in a production system. *Management Sci.* 33(9):1125–1136.

Lehman DW, Kovács B, Carroll GR (2014) Conflicting social codes and organizations: Hygiene and authenticity in consumer evaluations of restaurants. *Management Sci.* 60(10):2602–2617.

Levine DI, Toffel MW (2010) Quality management and job quality: How the ISO 9001 standard for quality management systems affects employees and employers. *Management Sci.* 56(6): 978–996.

Levine DI, Toffel MW, Johnson MS (2012) Randomized government safety inspections reduce worker injuries with no detectable job loss. *Science* 336(6083):907–911.

Linder JA, Doctor JN, Friedberg MW, Reyes Nieva H, Birks C, Meeker D, Fox CR (2014) Time of day and the decision to prescribe antibiotics. *JAMA Internal Medicine* 174(12):2029–2031.

Macher JT, Mayo JW, Nickerson JA (2011) Regulator heterogeneity and endogenous efforts to close the information asymmetry gap. *J. Law Econom.* 54(1):25–54.

Mani V, Muthulingam S (2018) Does learning from inspections affect environmental performance? Evidence from unconventional well development in Pennsylvania. *Manufacturing Service Oper. Management* 21(1):177–197.

Marcora SM, Staiano W, Manning V (2009) Mental fatigue impairs physical performance in humans. *J. Appl. Physiol.* 106(3):857–864.

May PJ, Wood RS (2003) At the regulatory front lines: Inspectors' enforcement styles and regulatory compliance. *J. Public Admin. Res. Theory* 13(2):117–139.

Minor T, Lasher A, Klontz K, Brown B, Nardinelli C, Zorn D (2015) The per case and total annual costs of foodborne illness in the United States. *Risk Anal.* 35(6):1125–1139.

Muraven M, Baumeister RF (2000) Self-regulation and depletion of limited resources: Does self-control resemble a muscle? *Psych. Bull.* 126(2):247–259.

Nickerson RS (1998) Confirmation bias: A ubiquitous phenomenon in many guises. *Rev. General Psych.* 2(2):175–220.

Oliva R, Sterman JD (2001) Cutting corners and working overtime: Quality erosion in the service industry. *Management Sci.* 47(7):894–914.

Pautz MC (2009) Trust between regulators and the regulated: A case study of environmental inspectors and facility personnel in Virginia. *Political Policy* 37(5):1047–1072.

Pautz MC (2010) Front-line regulators and their approach to environmental regulation in Southwest Ohio. *Rev. Policy Res.* 27(6):761–780.

Pierce L, Toffel MW (2013) The role of organizational scope and governance in strengthening private monitoring. *Organ. Sci.* 24(5):1558–1584.

Prentice DA, Miller DT (1992) When small effects are impressive. *Psych. Bull.* 112(1):160–164.

Ronen J (2010) Corporate audits and how to fix them. *J. Econom. Perspect.* 24(2):189–210.

Rossen J (2017) *12* Secrets of restaurant health inspectors. Accessed March 27, 2018, http://mentalfloss.com/article/500853/12-secrets-restaurant-health-inspectors.

Rozin P, Royzman EB (2001) Negativity bias, negativity dominance, and contagion. *Pers. Soc. Psych. Rev.* 5(4):296–320.

Samuelson W, Zeckhauser R (1988) Status quo bias in decision making. *J. Risk Uncertainty.* 1(1):7–59.

Scallan E, Griffin PM, Angulo FJ, Tauxe RV, Hoekstra RM (2011) Foodborne illness acquired in the United States—Unspecified agents. *Emerging Infectious Diseases* 17(1):16–22.

Scharff RL (2012) Economic burden from health losses due to foodborne illness in the United States. *J. Food Protection* 75(1):123–131.

Shah R, Ball GP, Netessine S (2016) Plant operations and product recalls in the automotive industry: An empirical investigation. *Management Sci.* 63(8):2439–2459.

Shimshack JP (2014) The economics of environmental monitoring and enforcement. *Annual Rev. Resource Econom.* 6(1):339–360.

Shimshack JP, Ward MB (2005) Regulator reputation, enforcement, and environmental compliance. *J. Environ. Econom. Management* 50(3):519–540.

Short JL, Toffel MW, Hugill AR (2019) Code contingencies: Designing monitoring regimes to promote improvement in supply chain working conditions. Working Paper No. 17-001, Harvard Business School, Boston.

Short JL, Toffel MW (2016) The integrity of private third-party compliance monitoring. *Admin. Regulatory Law News* 42(1):22–25.

Short JL, Toffel MW, Hugill AR (2016) Monitoring global supply chains. *Strategic Management J.* 37(9):1878–1897.

Simon PA, Leslie P, Run G, Jin GZ, Reporter R, Aguirre A, Fielding JE (2005) Impact of restaurant hygiene grade cards on foodborne-disease hospitalizations in Los Angeles County. *J. Environ. Health* 67(7):32–36.

Simonsohn U, Gino F (2013) Daily horizons: Evidence of narrow bracketing in judgment from 10 years of M.B.A. admissions interviews. *Psych. Sci.* 24(2):219–224.

Staats BR, Gino F (2012) Specialization and variety in repetitive tasks: Evidence from a Japanese bank. *Management Sci.* 58(6): 1141–1159.

Staats BR, Dai H, Hofmann D, Milkman KL (2017) Motivating process compliance through individual electronic monitoring: An empirical examination of hand hygiene in healthcare. *Management Sci.* 63(5):1563–1585.

Stafford SL (2003) Assessing the effectiveness of state regulation and enforcement of hazardous waste. *J. Regulatory Econom.* 23(1): 27–41.

Toffel MW, Short JL, Ouellet M (2015) Codes in context: How states, markets, and civil society shape adherence to global labor standards. *Regulation Governance* 9(3):205–223.

Tversky A, Kahneman D (1974) Judgment under uncertainty: Heuristics and biases. *Science* 185(4157):1124–1131.

Wright RA, Junious TR, Neal C, Avello A, Graham C, Herrmann L, Junious S, Walton N (2007) Mental fatigue influence on effort-related cardiovascular response: Difficulty effects and extension across cognitive performance domains. *Motivation Emotion* 31(3): 219–231.

How Scheduling Can Bias Quality Assessment

Online Supplement to

How Scheduling Can Bias Quality Assessment:

Evidence from Food Safety Inspections

Maria R. Ibanez
Kellogg School of Management at Northwestern University, Evanston, IL 60208,
maria.ibanez@kellogg.northwestern.edu

Michael W. Toffel
Harvard Business School, Boston, MA 02163, mtoffel@hbs.edu

Appendix A. Supplemental Descriptive Statistics and Correlations

Table A1. Inspection Sequence within the Day

1st inspection of the day	5,328
2nd inspection of the day	3,618
3rd inspection of the day	1,971
4th inspection of the day	763
5th inspection of the day	248
6th inspection of the day	61
7th+ inspection of the day	28
Total number of inspections	12,017

Table A2. Number of Inspector-days by Inspection Workload

Inspector-days with 1 inspection	1,790
Inspector-days with 2 inspections	2,226
Inspector-days with 3 inspections	1,637
Inspector-days with 4 inspections	801
Inspector-days with 5 inspections	295
Inspector-days with 6 inspections	83
Inspector-days with 7+ inspections	48
Total number of inspector-days	6,880

An *inspector-day* refers to a particular day during which a given inspector conducts at least one inspection.

Table A3. Inspections by Hour Begun and Corresponding Meal Period

7 am or earlier	39	3,856 during *breakfast period* (inspection began midnight to 10:59 am)
8 am	222	
9 am	972	
10 am	2,623	
11 am	1,986	7,888 during *lunch period* (inspection began 11:00 am–3:59 pm)
12 pm	1,331	
1 pm	2,331	
2 pm	1,653	
3 pm	587	
4 pm	171	273 during *dinner period* (inspection began 4:00 pm–11:59 pm)
5 pm	59	
6 pm or later	43	
Total number of inspections:	12,017	

Table A4. Extensions: Summary Statistics

Variable	Obs	Mean	SD	Min	Max
Potential extent of shift-prolonging (in hours)	12,017	0.25	0.52	0	1.79
Potential extent of shift-prolonging is >0.5 & <=1 hours (binary)	12,017	0.07	0.25	0	1
Potential extent of shift-prolonging is >1 & <=1.5 hours (binary)	12,017	0.05	0.21	0	1
Potential extent of shift-prolonging is >1.5 & <= max of 1.79 hours (binary)	12,017	0.07	0.25	0	1
Penultimate inspected establishment’s violations	12,014	2.10	2.60	0	25
Antepenultimate inspected establishment’s violations	12,008	2.07	2.57	0	25
Prior inspected establishment’s violations for the first inspection of the day	12,017	0.96	2.05	0	25
Prior inspected establishment’s violations for the second+ inspection of the day	12,017	1.15	2.21	0	23
Overnight break length	12,017	2.88	7.36	0	76.40
Within-day break length	12,017	0.04	0.08	0	1.10
Break length	12,017	2.92	7.35	0	76.40
Critical violations	12,017	0.93	1.27	0	11
Noncritical violations	12,017	1.49	1.94	0	16
Inspection duration (minutes)	12,017	62.87	32.01	5	270
Log inspection duration	12,017	4.01	0.55	1.61	5.60
Violation citations per hour	12,017	2.35	2.66	0	42
Log (violation citations per hour + 1)	12,017	0.96	0.71	0	3.76

Table A5. Correlations

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1)	Violations	1.00									
(2)	Prior inspected establishment’s violations (H1)	0.18*	1.00								
(3)	Prior inspected establishment’s violation trend (H2)	0.11*	0.56*	1.00							
(4)	Prior inspected establishment saliently improved (H3)	0.01	-0.17*	-0.37*	1.00						
(5)	Prior inspected establishment saliently deteriorated (H3)	0.12*	0.60*	0.69*	-0.29*	1.00					
(6)	Time inspecting earlier today (H4)	0.01	0.10*	0.08*	-0.02	0.08*	1.00				
(7)	Potentially shift-prolonging (H5)	-0.01	0.03*	0.01	0.00	0.01	0.38*	1.00			
(8)	Inspector experience	-0.05*	-0.01	0.02	-0.04*	0.01	-0.03*	-0.02*	1.00		
(9)	Returning inspector	-0.14*	-0.03*	-0.01	-0.01	-0.01	-0.05*	0.10*	0.33*	1.00	
(10)	Establishment’s *n*th inspection (2nd through 10th)	0.05*	0.01	0.02*	-0.01	0.02*	-0.02*	0.06*	0.48*	0.35*	1.00
(11)	Lunch period (11:00 am–3:59 pm)	-0.01	0.02*	0.02*	-0.01	0.02	0.39*	0.34*	-0.01	0.01	-0.01
(12)	Potential extent of shift-prolonging (in hours)	-0.03*	0.02	0.00	0.00	0.00	0.35*	0.8*	-0.05*	0.07*	0.05*
(13)	Potential extent of shift-prolonging is >0.5 & <=1 hours (binary)	-0.01	0.01	0.01	-0.01	0.00	0.15*	0.44*	0.01	0.05*	0.02*
(14)	Potential extent of shift-prolonging is >1 & <=1.5 hours (binary)	0.00	0.00	0.01	0.02	0.00	0.16*	0.37*	0.00	0.05*	0.04*
(15)	Potential extent of shift-prolonging is >1.5 & <= max of 1.79 hours (binary)	-0.03*	0.01	-0.01	0.00	0.00	0.23*	0.46*	-0.06*	0.02*	0.02*
(16)	Penultimate inspected establishment’s violations	0.15*	0.17*	0.07*	0.05*	0.1*	0.03*	-0.01	-0.02*	-0.04*	0.01
(17)	Antepenultimate inspected establishment’s violations	0.15*	0.16*	0.07*	0.05*	0.08*	0.03*	-0.01	-0.02*	-0.03*	0.01
(18)	Prior inspected establishment’s violations for the first inspection of the day	0.11*	0.58*	0.31*	-0.1*	0.33*	-0.39*	-0.15*	-0.05*	-0.02*	0.00
(19)	Prior inspected establishment’s violations for the second+ inspection of the day	0.12*	0.65*	0.38*	-0.11*	0.41*	0.47*	0.17*	0.03*	-0.01	0.02
(20)	Overnight break length	0.00	0.00	-0.01	0.00	-0.01	-0.32*	-0.08*	-0.12*	-0.04*	-0.05*
(21)	Within-day break length	0.01	-0.02*	0.00	-0.02*	-0.01	0.26*	0.27*	0.03*	0.01	0.01
(22)	Break length	0.00	0.00	-0.01	0.00	-0.01	-0.32*	-0.07*	-0.11*	-0.04*	-0.05*
(23)	Critical violations	0.76*	0.13*	0.08*	0.01	0.09*	0.00	0.01	-0.08*	-0.16*	-0.01
(24)	Noncritical violations	0.91*	0.17*	0.10*	0.01	0.11*	0.01	-0.02*	-0.01	-0.09*	0.07*
(25)	Inspection duration (minutes)	0.43*	0.02	0.01	-0.01	0.01	-0.03*	-0.03*	-0.19*	-0.14*	-0.01
(26)	Log inspection duration	0.40*	0.00	0.01	-0.02*	0.01	-0.05*	-0.04*	-0.16*	-0.11*	-0.03*
(27)	Violation citations per hour	0.75*	0.18*	0.11*	0.02	0.12*	0.03*	0.02	0.03*	-0.06*	0.04*
(28)	Log (violation citations per hour + 1)	0.78*	0.17*	0.10*	0.01	0.12*	0.02*	0.00	0.00	-0.09*	0.01

N = 12,017 inspections

(continued on next page)

Table A5. Correlations (continued)

	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)	(25)	(26)	(27)	(28)
(11)	1.00																	
(12)	0.21*	1.00																
(13)	0.19*	0.26*	1.00															
(14)	0.15*	0.43*	-0.06*	1.00														
(15)	0.05*	0.79*	-0.07*	-0.06*	1.00													
(16)	-0.02*	-0.01	0.01	-0.01	-0.01	1.00												
(17)	0.00	-0.01	0.00	-0.01	0.00	0.16*	1.00											
(18)	-0.25*	-0.13*	-0.07*	-0.07*	-0.06*	0.1*	0.09*	1.00										
(19)	0.26*	0.13*	0.08*	0.06*	0.07*	0.12*	0.11*	-0.25*	1.00									
(20)	-0.17*	-0.05*	-0.04*	-0.04*	-0.01	0.01	-0.01	0.22*	-0.20*	1.00								
(21)	0.26*	0.27*	0.12*	0.1*	0.19*	-0.02*	0.01	-0.24*	0.20*	-0.20*	1.00							
(22)	-0.16*	-0.04*	-0.03*	-0.04*	-0.01	0.01	-0.01	0.22*	-0.20*	1.00*	-0.19*	1.00						
(23)	-0.02*	0.00	0.00	-0.01	0.01	0.11*	0.11*	0.09*	0.07*	0.04*	0.01	0.04*	1.00					
(24)	0.00	-0.04*	-0.01	0.00	-0.04*	0.14*	0.14*	0.09*	0.12*	-0.03*	0.01	-0.03*	0.42*	1.00				
(25)	-0.09*	-0.04*	-0.03*	-0.02*	-0.02*	0.03*	0.02*	0.09*	-0.07*	0.10*	-0.04*	0.10*	0.41*	0.33*	1.00			
(26)	-0.10*	-0.05*	-0.02*	-0.03*	-0.03*	0.01	0.01	0.09*	-0.07*	0.09*	-0.04*	0.09*	0.37*	0.32*	0.92*	1.00		
(27)	0.05*	0.00	0.01	0.02*	-0.02	0.15*	0.15*	0.05*	0.16*	-0.04*	0.03*	-0.04*	0.51*	0.72*	-0.03*	-0.04*	1.00	
(28)	0.04*	-0.01	0.00	0.01	-0.02*	0.14*	0.14*	0.06*	0.15*	-0.03*	0.03*	-0.03*	0.57*	0.72*	0.08*	0.10*	0.89*	1.00

Appendix B. Interpretation of Results

To illustrate the magnitude of the estimated effects, we consider interventions that exploit outcome effects and ameliorate daily schedule effects, both of which would lead inspectors to cite violations that currently go underreported. In particular, we consider various scenarios that both (a) *amplify the outcome effects* in order to more routinely trigger the heightened inspector scrutiny that ensues after inspections reveal many violations and worsening compliance trends, and (b) *mitigate the daily schedule effects* in order to attenuate the reduced scrutiny that accompanies successive inspections and potentially shift-prolonging inspections. We estimate the effects of such interventions on the average inspection based on our sample, scale up the results to estimate the impact across the entire United States, and translate how such an increase in cited violations would translate to fewer foodborne illness cases and their associated healthcare costs.

In the best-case scenario, outcome effects (which increase scrutiny) would be fully triggered all the time and daily schedule effects (which erode scrutiny) would be entirely eliminated. The full consequence of these biases is reflected by the difference in inspection outcomes between this best-case scenario and the status quo, which quantifies the number of unreported violations and excess illnesses and costs that could be avoided if steps were taken to address these biases. Our discussions with inspectors suggest that some interventions are feasible—such as limiting or smoothing the number of inspections each inspector conducts per day—often without imposing any additional costs. We estimate a range of scenarios that consider the impacts associated with the daily schedule effects being attenuated by—and the outcome effects being actuated by—varying amounts.

We first consider the average impact on violations cited per inspection. Specifically, we compare the status quo (that is, the current practice with its associated scheduling effects) with alternative scenarios that consider various percentage changes (10% to 100% in 10% increments) of the effects we identified that would increase inspectors' detection rate (that is, decrease by 10% the daily schedule effects and increase by 10% the outcome effects). We make all these comparisons based on Model 1 in Table 2. Specifically, we calculate average predicted values for each scenario based on the model's estimates after recoding the estimated coefficients on *time inspecting earlier today*, *potentially shift-prolonging*, *prior inspected establishment's violations*, and *prior inspected establishment's violation trend* by the percentage specified; we report results in Column 1 of Table B1. We obtain equivalent results if we preserve the estimated coefficients and instead recode the values of the variables by that same percentage; thus, the results can be interpreted as altering the per-unit bias represented by the estimated coefficients (for example, raising inspectors' detection rates to the heightened levels associated with the identified outcome effects) or as altering the factors that generate the bias (for example, reducing the number of prior inspections conducted by the inspector per day to reduce the *time inspecting earlier* *today*). For the status quo, we use the model's estimates to calculate the average predicted number of violations per inspection, based on actual values of all variables, to be 2.42365. Column 2 reports the percent change (from the status quo) in the average predicted violations for each of these scenarios. This shows the average percent change in violations cited per inspection compared to the status quo.

For example, consider the very conservative "10% scenario" depicted in the second row of Table B1. In this scenario, we estimate the effects of (1) amplifying the outcome effects by increasing by 10% the actual values of *prior inspected establishment's violations* and *prior inspected establishment's violation trend* while also (2) mitigating the daily schedule effects by decreasing by 10% the actual values of *time inspecting earlier today* and *potentially shift-prolonging*. Applying these recoded values to the coefficient estimates from Model 1 of Table 2, we calculate the average predicted number of violations to be 2.448 (Column 1). This indicates that this "10% scenario" would result in 1.01% more violations being cited per inspection than the status quo of 2.42365 (Column 2).[1]

We then estimate the potential nationwide implications of our calculations, based on the assumptions that the estimated one million food establishments that are monitored by state, local, and tribal agencies in the United States (US Food and Drug Administration 2016) are each inspected annually and that our sample of inspections is representative of those conducted across the country. To calculate a nationwide figure, we take the difference between the average predicted values from each scenario and the status quo (that is, the Column 1 figure minus 2.42365) and multiply that by the one million inspections conducted annually across the country; results are reported in Column 3. Those figures can be interpreted in the context of an estimated 2.4 million violations cited in the status quo scenario.[2] Continuing the 10% scenario, we scale the difference in average predicted violations per inspection that arise in this scenario compared to the status quo (2.44819 - 2.42365) by the one million inspections conducted annually nationwide to estimate that this scenario would yield 24,536 additional violations (currently undetected) being cited nationwide per year (Column 3).

Citing violations leads establishments to improve their food safety practices, which in turn mitigates foodborne illness cases and resulting hospitalizations. To calculate the health impacts if these undetected violations were being cited, we translate the estimated nationwide changes in violation counts into health outcomes and their associated costs. We attempt to be as conservative as possible but acknowledge that there are uncertainties associated with these conversions.[3]

---

[1] The 1.01% figure is calculated as (2.44819 - 2.42365) / 2.42365.
[2] The 2.4 million figure is calculated by multiplying 2.424 violations cited per inspection in the status quo scenario by the one million inspections conducted annually nationwide.
[3] The estimates we construct should be considered as an illustration of the possible implications of the biases. We acknowledge the possibility that our estimates might *overestimate* the effects if the conversion factors we use overestimate the benefits of citing a particular violation and that they might *underestimate* the effects because we do not incorporate the spillovers and system-wide benefits of citing a particular violation, as each citation may encourage establishments to improve health practices more broadly. Failing to cite one violation thus not only carries the health risks associated with that violation but may also encourage noncompliance—an effect similar to the broken window phenomenon. That said, while developing a more

First, we consider how the increased violations cited per inspection beyond the status quo translates into fewer hospitalizations for foodborne illness (Columns 4 and 5). We multiply the percent change in the average predicted number of violations between the scenario and the status quo (Column 2) by the ratio of a 20% decrease in hospitalizations per 5% improvement in restaurant compliance scores based on prior research on Los Angeles restaurants (Jin and Leslie 2003).[4] To calculate the impact of the 10% scenario, we multiply the 1.01% increase in the number of violations cited per inspection (Column 2) by the ratio of 20% decrease in hospitalizations per 5% improvement in restaurant compliance scores, which indicates that hospitalizations would decrease by 4.05% (Column 4). This would correspond to 5,183 fewer hospitalizations for foodborne illness each year across the United States (Column 5), based on applying the 4.05% decline to the estimated 128,000 annual US hospitalizations for foodborne illness under the status quo (Scallan et al. 2011).

We also estimate the impact of citing more violations on annual nationwide foodborne illness cases (Column 6). Based on the ratio of Scallan et al.'s (2011) two nationwide annual estimates of 47.8 million foodborne illness cases and 128,000 annual foodborne illness hospitalizations, there are 373.4 foodborne illness cases per foodborne illness hospitalization. Therefore, the 10% scenario, estimated earlier to reduce foodborne illness hospitalizations by 5,183, would also reduce foodborne illness cases by 1.94 million cases (calculated as 5,183 * 373.4).

Finally, we estimate the impact of citing more violations on the costs associated with foodborne illness cases based on two alternative estimates of the average cost per foodborne illness case of $747 (Minor et al. 2015) and $1,626 (Scharff 2012), which we use to construct the lower and upper bounds of our cost estimates (Columns 7 and 8). In the 10% scenario, applying these figures to the estimated 1.94 million fewer foodborne illness cases compared to the status quo yields a $1,446-million-to-$3,147-million drop in the annual costs associated with foodborne illness cases nationwide.

As noted, there are many assumptions and caveats associated with these analyses and one can consider alternative scenarios. Our estimations above assume that mitigating bias would yield citations of violations that are as correlated with foodborne incidents as the violations currently cited. But what if newly cited violations are less "important," meaning they impose less health risk? For example, suppose

comprehensive methodology to estimate the health impacts of citing more food safety violations is a necessary and worthy endeavor, it is beyond the scope of this paper.

[4] We are aware of little research that has estimated the effect of each food safety violation on health outcomes and we rely on Jin and Leslie (2003), which we believe presents the best estimate. They show that introducing restaurant grade cards—signs posted outside restaurants that report the establishment's letter grade based on its most recent food safety inspection results—affects food safety inspection violation scores and health outcomes, so restaurant grade cards can be viewed as an instrument that reveals the relationship between violations cited and health outcomes. Because violations are supposed to be corrected when cited, we assume that the new citations resulting from reducing the bias translate into fewer actual violations. (To be conservative, we are not accounting for how citations motivate compliance more broadly.) The relationship Jin and Leslie (2003) identified between compliance and health outcomes applies to our setting because it is based on a similar type of inspection and a compliance measure based on total violations, which implicitly controls for the heterogeneous effects of different types of violation on health.

that remediating a newly cited violation would prevent half as many foodborne incidents as remediating a currently cited violation. Estimating the health impacts would then require adjusting Jin and Leslie's (2003) finding that a 5% improvement in restaurant compliance yields a 20% decline in foodborne illness hospitalizations to a 10% decline. In that scenario, if the drivers of outcome effects were doubled (that is, amplified by 100%) and the drivers of daily schedule effects were fully mitigated (that is, reduced by 100%), the 11.03% increase in citations (last row, Column 2) would translate into a 22.06% decline in hospitalizations [=11.03*(-10/5)] (compared to our original estimate of a 44.12% decline, calculated as 11.03*(-20/5) and reported in the last row of Column 4), which nationwide would result in 28,235 fewer foodborne-illness–related hospitalizations and 10.54 million fewer foodborne illness cases, saving $7.88 billion to $17.14 billion in foodborne illness costs.

Table B1. Estimates of Nationwide Effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Impact on citations of violations			Impact on health and associated costs				
Bias-reduction scenario	Average predicted number of violations cited per inspection	Percent change in average predicted number of violations cited per inspection compared to the status quo	Change in nationwide annual number of violations cited compared to the status quo	Percent change in foodborne illness hospitalizations compared to the status quo	Change in nationwide annual number of foodborne illness hospitalizations compared to the status quo	Change in nationwide annual number of foodborne illness cases compared to the status quo, **in millions**	Change in nationwide annual costs of foodborne illness cases compared to status quo, **in $millions**	
							Lower estimate	Upper estimate
0% (status quo)	2.424	0.00%	0	0.00%	0	0.00	$0	$0
10%	2.448	1.01%	24,536	-4.05%	-5,183	-1.94	-$1,446	-$3,147
20%	2.473	2.04%	49,528	-8.17%	-10,463	-3.91	-$2,919	-$6,353
30%	2.499	3.09%	74,987	-12.38%	-15,841	-5.92	-$4,419	-$9,619
40%	2.525	4.16%	100,924	-16.66%	-21,320	-7.96	-$5,947	-$12,946
50%	2.551	5.25%	127,352	-21.02%	-26,903	-10.05	-$7,505	-$16,336
60%	2.578	6.37%	154,283	-25.46%	-32,593	-12.17	-$9,092	-$19,790
70%	2.605	7.50%	181,732	-29.99%	-38,391	-14.34	-$10,710	-$23,311
80%	2.633	8.65%	209,711	-34.61%	-44,302	-16.54	-$12,358	-$26,900
90%	2.662	9.83%	238,234	-39.32%	-50,327	-18.79	-$14,039	-$30,559
100%	2.691	11.03%	267,315	-44.12%	-56,471	-21.09	-$15,753	-$34,290

Each row represents a bias-reduction scenario. For example, the 10% scenario depicted in the second row illustrates the results of reducing bias if the outcome effects (which increase scrutiny) were amplified by 10% and the daily schedule effects (which erode scrutiny) were mitigated by 10%.

Column 1 is the average predicted number of violations per inspection, based on Model 1 of Table 2, under each scenario.

Column 2 is calculated as the percent change in the average predicted number of violations per inspection, comparing each scenario (Column 1) to the status quo value of 2.42365.

Column 3 is calculated as the difference in the average predicted number of violations per inspection, comparing each scenario (Column 1) to the status quo value of 2.42365 and multiplying this by the one million food safety inspections conducted nationwide each year.

Column 4 is calculated by multiplying the percent change in average predicted number of violations compared to the status quo (Column 2) by the ratio of the change in hospitalizations to the change in compliance (derived from the 20% decline in hospitalizations per 5% improvement in restaurant compliance relationship reported by Jin and Leslie (2003); that is, -20%/5% = -4).

Column 5 is calculated as the difference in hospitalizations between (a) the estimated number that would have occurred under each scenario and (b) the 128,000 that actually occurred (Scallan et al. 2011). Specifically, we multiply the percent change in hospitalizations (Column 4) by the 128,000 nationwide annual hospitalizations.

Column 6 is calculated by multiplying the change in nationwide annual number of foodborne illness hospitalizations compared to the status quo (Column 5) by 373.4, the number of illness cases per hospitalization (calculated as the ratio between Scallan et al.'s (2011) two estimates of the 47.8 million annual foodborne illnesses and the resulting 128,000 hospitalizations).

Columns 7 and 8 are calculated by multiplying the estimated change in illness cases (Column 6) by $747 (the weighted average from Minor et al. (2015)) and by $1,626 (the enhanced model estimate from Scharff (2012)) in estimated costs per illness case, respectively.

Figure B1. Estimated nationwide increase in food safety violations being cited as biases are attenuated

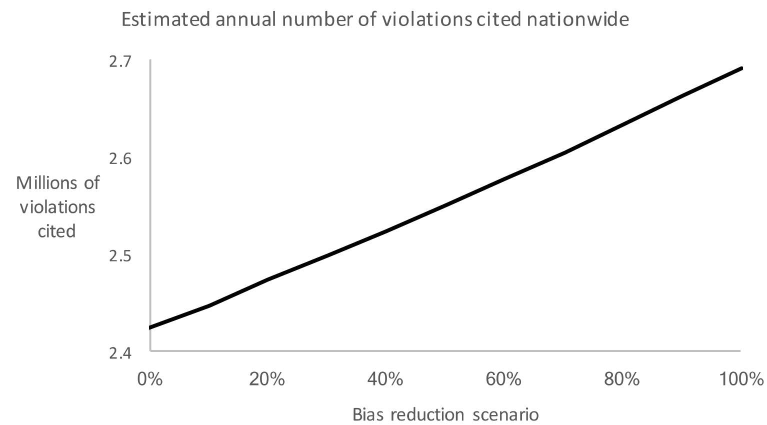


Figure B2. Estimated reduction in healthcare cases associated with more food safety violations being cited as biases are attenuated

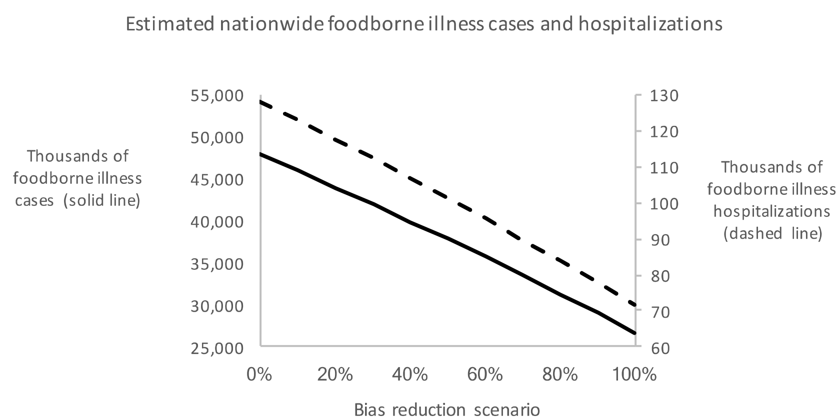
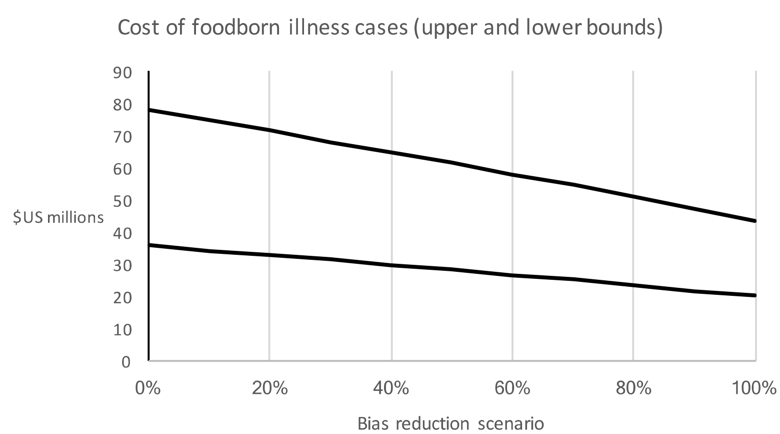


Figure B3. Estimated cost reductions associated with improved health impacts resulting from more food safety violations being cited as biases are attenuated



Notes: These figures graph data from Table B1, which is based on the methodology described in Appendix B. The horizontal axes represent different bias-reduction scenarios. For example, the 20% scenario illustrates the results of reducing bias by amplifying by 20% the outcome effects (which increase scrutiny) and mitigating by 20% the daily schedule effects (which erode scrutiny).

Appendix C. Supplemental Analysis: Alternative Measures of Potentially Shift-prolonging (H5) Construct

Our primary results test the effects on inspector scrutiny of conducting inspections that risk prolonging the inspector's shift (H5). As discussed in the paper (Section 4.7, "Extensions"), we also investigate whether the extent to which the shift might be prolonged matters. Results are reported in Table C1 below. First, we use a continuous variable instead of the binary *potentially shift-prolonging* to examine how the magnitude of shift-prolonging affects inspector scrutiny (Columns 3–4 of Table C1). We also pursued a more flexible approach by creating a series of dummy variables denoting ranges of *potential extent of shift-prolonging* (Columns 5–6 of Table C1). These specifications yield the same inferences as our main models, which are reported in the paper in Table 2 and are reproduced in Columns 1–2 of Table C1 as the basis of comparison for the following robustness tests.

.

Table C1. Alternative Measures of Potentially Shift-prolonging (H5) Construct

Measure of potentially shift-prolonging (H5):	***Potentially shift-prolonging* (binary) (our primary specification)**		***Potential extent of shift-prolonging* (in hours)**		***Potential extent of shift-prolonging* indicators**	
Dependent variable:	*violations*	*violations*	*violations*	*violations*	*violations*	*violations*
	(1)	(2)	(3)	(4)	(5)	(6)
Prior inspected establishment's violations (H1)	0.017***	0.015***	0.017***	0.015***	0.017***	0.015***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Prior inspected establishment's violation trend (H2)	0.013**		0.013**		0.013**	
	(0.006)		(0.006)		(0.006)	
Prior inspected establishment saliently improved (H3)		0.011		0.012		0.012
		(0.023)		(0.023)		(0.023)
Prior inspected establishment saliently deteriorated (H3)		0.079***		0.079***		0.079***
		(0.027)		(0.027)		(0.027)
Time inspecting earlier today (H4)	-0.038***	-0.039***	-0.034***	-0.035***	-0.036***	-0.036***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Potentially shift-prolonging (binary) (H5)	**-0.051****	**-0.050****				
	(0.025)	**(0.025)**				
Potential extent of shift-prolonging (in hours)			**-0.068*****	**-0.068*****		
			(0.023)	**(0.023)**		
Potential extent of shift-prolonging is >0.5 & <=1 hours (binary)					**-0.050**	**-0.051**
					(0.040)	**(0.040)**
Potential extent of shift-prolonging is >1 & <=1.5 hours (binary)					**-0.078***	**-0.078***
					(0.045)	**(0.045)**
Potential extent of shift-prolonging is >1.5 & <= max of 1.79 hours (binary)					**-0.109****	**-0.109****
					(0.044)	**(0.044)**
Inspector experience	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Returning inspector	-0.114***	-0.117***	-0.115***	-0.117***	-0.116***	-0.118***
	(0.035)	(0.035)	(0.035)	(0.035)	(0.035)	(0.035)
Lunch period (11:00 am–3:59 pm)	-0.049**	-0.049**	-0.054**	-0.053**	-0.055**	-0.054**
	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)
Month fixed effects	Included	Included	Included	Included	Included	Included
Year fixed effects	Included	Included	Included	Included	Included	Included
Establishment's-*n*th-inspection (2nd through 10th or more) fixed effects	Included	Included	Included	Included	Included	Included
Inspection-type fixed effects	Included	Included	Included	Included	Included	Included
Inspector-establishment dyad fixed effects	Included	Included	Included	Included	Included	Included
Number of observations (inspections)	12,017	12,017	12,017	12,017	12,017	12,017

Notes: Poisson regression coefficients with robust standard errors clustered by establishment. *** p < 0.01, ** p < 0.05, * p < 0.10.

Appendix D. Supplemental Analysis: Persistence of Outcome Effects

We conduct additional analysis to examine the persistence of some of our outcome effects (Table D1; see Section 4.7, "Extensions").

Table D1. Persistence of Outcome Effects

Dependent variable:	*violations*		
	(1)	(2)	(3)
Prior inspected establishment's violations	0.017***	0.017***	
	(0.004)	(0.004)	
Penultimate inspected establishment's violations	**0.010*** **	**0.010*** **	
	(0.003)	**(0.003)**	
Antepenultimate inspected establishment's violations		**0.007* **	
		(0.004)	
Prior inspected establishment's violations for the first inspection of the day			**0.016*** **
			(0.006)
Prior inspected establishment's violations for the second+ inspection of the day			**0.018*** **
			(0.005)
Prior inspected establishment's violation trend	0.013**	0.013**	0.014**
	(0.006)	(0.006)	(0.006)
Time inspecting earlier today	-0.038***	-0.039***	-0.040***
	(0.011)	(0.011)	(0.012)
Potentially shift-prolonging	-0.055**	-0.056**	-0.051**
	(0.024)	(0.024)	(0.025)
Inspector experience	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)
Returning inspector	-0.118***	-0.120***	-0.114***
	(0.035)	(0.034)	(0.035)
Lunch period (11:00 am–3:59 pm)	-0.046*	-0.046*	-0.051**
	(0.024)	(0.024)	(0.025)
Month fixed effects	Included	Included	Included
Year fixed effects	Included	Included	Included
Establishment's-nth-inspection (2nd through 10th or more) fixed effects	Included	Included	Included
Inspection-type fixed effects	Included	Included	Included
Inspector-establishment dyad fixed effects	Included	Included	Included
Number of observations (inspections)	12,011	12,000	12,017

Notes: Poisson regression coefficients with robust standard errors clustered by establishment. Columns 1 and 2 have fewer observations than our main estimation sample because inspectors' *Penultimate inspected establishment's violations* is missing for the inspectors' first and second inspections and inspectors' *Antepenultimate inspected establishment's violations* is missing for the inspectors' first, second and third inspections. Most of those observations were already dropped by other sample restrictions.
*** p < 0.01, ** p < 0.05, * p < 0.10.

Appendix E. Supplemental Analysis: Breaks

We conduct additional analysis to explore how breaks could affect inspector scrutiny (Table E1; see Section 4.7, "Extensions").

Table E1. Breaks

	Dependent variable:	*violations*		
		(1)	(2)	(3)
H1	Prior inspected establishment's violations (H1)	0.017***	0.014***	0.017***
		(0.004)	(0.005)	(0.004)
H2	Prior inspected establishment's violation trend (H2)	0.014**	0.018**	0.014**
		(0.006)	(0.007)	(0.006)
H4	Time inspecting earlier today (H4)	-0.034***	-0.034***	-0.036***
		(0.012)	(0.012)	(0.012)
H5	Potential extent of shift-prolonging (H5)	-0.055**	-0.055**	-0.056**
		(0.025)	(0.025)	(0.025)
	Inspector experience	0.001***	0.001***	0.001***
		(0.000)	(0.000)	(0.000)
	Returning inspector	-0.117***	-0.117***	-0.117***
		(0.035)	(0.035)	(0.035)
	Lunch period (11:00 am–3:59 pm)	-0.051**	-0.050**	-0.049**
		(0.025)	(0.025)	(0.024)
	Overnight break length	**0.003***	**0.002**	
		(0.002)	**(0.002)**	
	Within-day break length	**0.089**	**0.051**	
		(0.124)	**(0.148)**	
	Prior inspected establishment's violations (H1) × × Overnight break length		**0.000**	
			(0.001)	
	Prior inspected establishment's violations (H1) × × Within-day break length		**0.032**	
			(0.054)	
	Prior inspected establishment's violation trend (H2) × × Overnight break length		**-0.001**	
			(0.001)	
	Prior inspected establishment's violation trend (H2) × × Within-day break length		**-0.061**	
			(0.093)	
	Break length			**0.003***
				(0.002)
	Time inspecting earlier today (H4) × Break length			**0.062**
				(0.095)
	Month fixed effects	Included	Included	Included
	Year fixed effects	Included	Included	Included
	Establishment's-*n*th-inspection (2nd through 10th or more) fixed effects	Included	Included	Included
	Inspection-type fixed effects	Included	Included	Included
	Inspector-establishment dyad fixed effects	Included	Included	Included
	Number of observations (inspections)	12,017	12,017	12,017

Notes: Poisson regression coefficients with robust standard errors clustered by establishment.
*** p < 0.01, ** p < 0.05, * p < 0.10.

Appendix F. Supplemental Analysis: Critical versus Noncritical Violations

To assess whether our hypothesized relationships differentially influence inspectors' behavior across different types of violation, we estimated our models on two subsets of violations. First, we predict the number of *critical violations*, which are related to food preparation practices and employee behaviors that more directly contribute to foodborne illness or injury. These factors are prioritized in Alaska and in Camden County by being displayed on the first page of the inspection report and in Lake County by being tagged in the reports. Second, we estimated our models on the number of *noncritical violations* (that is, violations of procedures often referred to as "good retail practices"). While less risky than the other type, these are also important for public health and include overall sanitation and preventative measures to protect foods, such as proper use of gloves. Inspections averaged 0.93 *critical violations* and 1.49 *noncritical violations*.

Fewer noncritical violations are cited in inspections conducted during the lunch period than in other periods, but the results yield no evidence that time of day affects critical violations (see Table F1). The latter finding is consistent with critical violations being related to longer-term establishment practices that are insensitive to the number of customers being served or to the staff's busyness and thus ability to respond to the inspector's presence.

Outcome effects are ubiquitous, affecting critical and noncritical violations alike. Each additional violation cited at the inspector's prior inspected establishment is associated with 1.92% more critical violations (Column 1: $\beta = 0.019$, $p < 0.01$) and 1.61% more noncritical violations (Column 3: $\beta = 0.016$, $p < 0.01$) cited in the focal inspection.

As with total violations, there is no evidence of critical and noncritical violations being affected when the *prior inspected establishment saliently improved*. When the *prior inspected establishment saliently deteriorated*, inspections yield, on average, 7.57% more critical violations (Column 2: $\beta = 0.073$, $p < 0.10$) and 8.22% more noncritical violations (Column 4: $\beta = 0.079$, $p < 0.05$).

Turning to daily schedule effects, we find that fatigue affects inspectors' ability to discover and report both types of violations. Specifically, the estimated coefficients on *time inspecting earlier today* indicate that each additional hour conducting prior inspections during the day results, on average, in 2.86% fewer critical violations cited (Column 1: $\beta = -0.029$, $p < 0.10$) and 4.30% fewer noncritical violations cited (Column 3: $\beta = -0.044$, $p < 0.01$).

These results also indicate that the *potentially shift-prolonging* effects identified in our primary results are driven by noncritical violations rather than critical ones. In particular, *potentially shift-prolonging* inspections result in 6.29% fewer citations (Column 3: $\beta = -0.065$, $p < 0.01$). However, we find no evidence that citations of critical violations are affected by whether the inspection risks prolonging the shift: the coefficient on *potentially shift-prolonging* is not statistically significant when predicting critical violations (Columns 1 and 2). This suggests that avoiding prolonging the shift does not affect inspectors' ability to discover and report critical violations.

Overall, these results indicate that inspectors' schedules have somewhat different effects on citing critical versus noncritical violations. Citing noncritical violations appears to be influenced by all types of daily schedule effects and outcome effects, while citing critical violations appears to be influenced by all but the *potentially shift-prolonging* effects.

Table F1. Critical and Noncritical Violations

	Dependent variable:	***critical violations***		***noncritical violations***	
		(1)	(2)	(3)	(4)
H1	Prior inspected establishment's violations	0.019***	0.016***	0.016***	0.015***
		(0.005)	(0.006)	(0.005)	(0.005)
H2	Prior inspected establishment's violation trend	0.014*		0.013*	
		(0.008)		(0.007)	
H3	After salient improvement		-0.017		0.028
			(0.031)		(0.028)
H3	After salient deterioration		0.073*		0.079**
			(0.039)		(0.031)
H4	Time inspecting earlier today	-0.029*	-0.030*	-0.044***	-0.045***
		(0.015)	(0.015)	(0.013)	(0.013)
H5	Potentially shift-prolonging	-0.026	-0.025	-0.065**	-0.064**
		(0.034)	(0.034)	(0.029)	(0.029)
	Inspector experience	0.001***	0.001***	0.000***	0.000***
		(0.000)	(0.000)	(0.000)	(0.000)
	Returning inspector	-0.113**	-0.115**	-0.102**	-0.104**
		(0.049)	(0.049)	(0.041)	(0.041)
	Lunch period (11:00 am–3:59 pm)	-0.038	-0.037	-0.057**	-0.056**
		(0.032)	(0.032)	(0.028)	(0.028)
	Month fixed effects	Included	Included	Included	Included
	Year fixed effects	Included	Included	Included	Included
	Establishment's-*n*th-inspection (2nd through 10th or more) fixed effects	Included	Included	Included	Included
	Inspection-type fixed effects	Included	Included	Included	Included
	Inspector-establishment dyad fixed effects	Included	Included	Included	Included
	Number of observations (inspections)	10,298	10,298	10,624	10,624

Notes: Poisson regression coefficients with robust standard errors clustered by establishment.
*** p < 0.01, ** p < 0.05, * p < 0.10.

Appendix G. Supplemental Analysis: Inspector Speed and Citation Pace

Our primary results show how inspections of prior establishments and daily schedules are associated with the number of *violations* cited. To assess whether such results might be driven by inspectors spending more or less time and exhibiting more or less scrutiny in the subsequent (focal) inspection, we estimate our primary models on the log of *inspection duration*, the number of minutes between an inspection's start time and end time. Moreover, to assess the net effect of the changes in *violations* cited and *inspection duration*, we explore the inspector's citation pace—a measure of productivity in this setting—and estimate our primary models on the log (after adding 1) of *violation citations per hour*. The results are reported in Table G1.

Considering potential outcome effects, we find that inspectors spend only slightly more time conducting inspections succeeding inspections in which more violations were cited (Column 1: *prior inspected establishment's violations* $\beta = 0.004$, $p < 0.10$) and find no evidence that the violation trend of the inspector's prior inspection affects inspection duration (Columns 1 and 2: the estimated coefficients on the *prior inspected establishment's violation trend*, *prior inspected establishment saliently improved* and *prior inspected establishment saliently deteriorated* are not statistically significant). Recall that our primary results found that more violations or worsening trends at an inspector's prior establishment predicted more violations cited at the focal inspection. Results in Column 3 indicate that citation pace increases by 1.0% for each additional violation at the prior establishment ($\beta = 0.010$, $p < 0.01$) and by 1.9% for each one-standard-deviation increase in the *prior inspected establishment's violation trend* ($\beta = 0.012$, $p < 0.05$). Column 4 indicates that, as was the case with the number of violations, this effect is asymmetric and driven by negative trends: whereas we find no change in citation pace after inspecting an establishment with salient improvement, it does increase by 3.9% after inspecting an establishment with salient deterioration (Column 4: $\beta = 0.039$, $p < 0.10$). This indicates that our main outcome-effect findings—that more violations and worsening trends at an inspector's prior establishment increase the inspector's citations at his or her next inspection—result mostly from inspectors increasing their citation pace rather than spending more time onsite.

We next consider potential daily schedule effects. We find that inspectors conduct inspections more quickly as they progress through their shift: *inspection duration* decreases by 3.1% for each additional hour already spent conducting inspections that day (Column 1: *time inspecting earlier today* $\beta = -0.031$, $p < 0.01$). For context, recall that our primary results indicate that each additional hour inspecting during the day cites an average of 3.73% fewer violations. The model reported in Column 3 indicates that the net effect is that inspector citation pace decreases by 1.9% for each subsequent inspection of the day (*time inspecting earlier today* $\beta = -0.019$, $p < 0.05$).

Turning to *potentially shift-prolonging* inspections, recall that our primary results indicated that these had 5.0% fewer citations. Column 1 reveals that inspectors conduct such inspections 4.3% more quickly (*potentially shift-prolonging* $\beta = -0.044$, $p < 0.01$). Column 3 reveals that the effect of *potentially shift-prolonging* on citation pace is not statistically significant. These results jointly suggest that the diminishment in citations results from shorter inspection duration rather than slower inspector speed, with inspectors' citation pace remaining largely unaffected by the risk of working late. This, in turn, suggests that our earlier finding that *potentially shift-prolonging* inspections result in fewer violations is likely due to inspectors' desire to avoid working late, rather than to fatigue eroding their citation pace.

Table G1. Effects of Inspectors' Schedules on Speed and Citation Pace

	Inspector speed		**Inspector citation pace**	
Dependent variable:	**log *inspection duration***		**log (*violation citations per hour* + 1)**	
	(1)	(2)	(3)	(4)
Prior inspected establishment's violations	0.004*	0.003*	0.010***	0.010***
	(0.002)	(0.002)	(0.003)	(0.003)
Prior inspected establishment's violation trend	-0.000		0.012**	
	(0.003)		(0.005)	
Prior inspected establishment saliently improved		0.009		-0.017
		(0.009)		(0.016)
Prior inspected establishment saliently deteriorated		0.006		0.039*
		(0.012)		(0.020)
Time inspecting earlier today	-0.031***	-0.031***	-0.019**	-0.020***
	(0.005)	(0.005)	(0.008)	(0.008)
Potentially shift-prolonging	-0.044***	-0.044***	-0.016	-0.015
	(0.011)	(0.011)	(0.018)	(0.018)
Inspector experience	0.000	0.000	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Returning inspector	0.082***	0.082***	-0.100***	-0.100***
	(0.016)	(0.016)	(0.024)	(0.024)
Lunch period (11:00 am–3:59 pm)	-0.034***	-0.034***	0.010	0.010
	(0.010)	(0.010)	(0.016)	(0.016)
Month fixed effects	Included	Included	Included	Included
Year fixed effects	Included	Included	Included	Included
Establishment's-*n*th-inspection (2nd through 10th or more) fixed effects	Included	Included	Included	Included
Inspection-type fixed effects	Included	Included	Included	Included
Inspector-establishment dyad fixed effects	Included	Included	Included	Included
Number of observations (inspections)	12,017	12,017	12,017	12,017
R-squared	0.45	0.45	0.20	0.20

Notes: Ordinary least squares coefficients with robust standard errors clustered by establishment.
*** p < 0.01, ** p < 0.05, * p < 0.10.

References

Jin G.Z., Leslie P. 2003. The effect of information on product quality: Evidence from restaurant hygiene grade cards. *Quarterly Journal of Economics* **118**(2) 409-451.

Scallan E., Griffin P.M., Angulo F.J., Tauxe R.V., Hoekstra R.M. 2011. Foodborne illness acquired in the United States—unspecified agents. *Emerging Infectious Disease journal* **17**(1) 16.

Scharff R.L. 2012. Economic burden from health losses due to foodborne illness in the United States. *Journal of Food Protection* **75**(1) 123-131.