American Economic Journal: Economic Policy 2019, 11(4): 404–428
https://doi.org/10.1257/pol.20180230

New Evidence on Information Disclosure through Restaurant Hygiene Grading†

By Daniel E. Ho, Zoe C. Ashwood, and Cassandra Handan-Nader*

The case of restaurant hygiene grading occupies a central role in information disclosure scholarship. Comparing Los Angeles, which enacted grading in 1998, with California from 1995–1999, Jin and Leslie (2003) found that grading reduced foodborne illness hospitalizations by 20 percent. Expanding hospitalization data and collecting new data on mandatorily reported illnesses, we show that this finding does not hold up under improvements to the original data and methodology. The largest salmonella outbreak in state history hit Southern California before Los Angeles implemented grading. Placebo tests detect the same treatment effects for Southern California counties, none of which changed restaurant grading. (*JEL* D83, H75, I12, I18, L83, L88)

A core question of information economics surrounds the conditions under which mandatory information disclosure improves welfare (Dranove and Jin 2010). To date, the empirical literature has yielded mixed results. Werner et al. (2012), for instance, finds nursing home report cards led patients to choose higher rated homes, but that the magnitude of the effect was "minimal." Figlio and Lucas (2004) reports that grading of schools affected short-term housing prices, but that the stochastic nature of grades diminished effects over time. Dranove et al. (2003) concludes that cardiac surgery report cards decreased social welfare due to hospital selection effects. One of the few studies finding uniformly large and positive effects on public health is Jin and Leslie (2003)—henceforth, JL. JL found that the enactment of restaurant hygiene grades in Los Angeles County in 1998 caused an increase in restaurant health inspection scores, a change in consumer demand for restaurant

*Ho: William Benjamin Scott and Luna M. Scott Professor of Law, Professor of Political Science, and Senior Fellow at Stanford Institute for Economic Policy Research; Stanford University, 559 Nathan Abbott Way, Stanford, CA 94305 (email: dho@law.stanford.edu). Ashwood: PhD Student, Computer Science, Princeton University, 35 Olden Street, Princeton, NJ 08540 (email: zashwood@cs.princeton.edu). Handan-Nader: PhD Student, Political Science, Stanford University, P.O. Box 18655, Stanford, CA 94309 (email: slnader@stanford.edu). Kate Ho was coeditor for this article. We thank Taylor Cranor for astute research assistance, Philip Leslie for sharing disease coding notes, Jenny Lloyd (King County) and Karen Wong (Centers for Disease Control and Prevention) for insights into foodborne disease discharge codes, Alain Kelder and Ron Nakao for data recovery of hospitalization data from 8mm tapes, Sonia Moss for help in locating historical communicable disease reports, John Blodgett for help on historical ZIP code census information, and Glenn Carroll, Taylor Cranor, Dick Craswell, John Donohue, Jacob Goldin, Mark Kelman, Rob MacCoun, Michelle Mello, Alison Morantz, Sam Sherman, David Studdert, and workshop participants at Irvine and Stanford for helpful comments.

†Go to https://doi.org/10.1257/pol.20180230 to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

hygiene, and a 20 percent decrease in the number of hospitalizations for foodborne illnesses in LA compared to the rest of California (CA).[1]

JL occupies a pivotal role in the scholarship on information disclosure. Loewenstein, Sunstein, and Golman (2014, 392, 403), for instance, reviews the literature on information disclosure, finding a "paucity of data supporting the efficacy of such policies," but points out JL as "a paradigm for successful disclosure." JL's study of restaurant grading is the cornerstone of the leading and influential synthesis of information disclosure policies (Fung, Graham, and Weil 2007; Weil et al. 2006). Based exclusively on the LA evidence, Fung, Graham, and Weil (2007) posits restaurant grading as the archetypal disclosure and model of "targeted transparency," as it is simple and relevant at the time of decision making. As JL puts it, "It seems remarkable that simply providing a standard format for disclosure … would be sufficient to change the equilibrium from zero disclosure and low hygiene quality, to high hygiene quality with potentially full disclosure" (Jin and Leslie 2003, 450). The findings intensified scholarly focus on the simplification of information disclosure (e.g., Ayres, Raseman, and Shih 2013; Ben-Shahar and Schneider 2014b; Feng Lu 2012; Loewenstein, Sunstein, and Golman 2014; Marotta-Wurgler 2012; and Robertson 2015). Even the toughest skeptics of disclosure, who argue that information disclosure has been a "spectacular failure," described JL at one point as an effective "emblem of simplification" (Ben-Shahar and Schneider 2011, 647, 743; Ben-Shahar and Schneider 2014a, S260).

The foodborne illness finding in particular has surfaced in a wide range of policy debates over the enactment of restaurant grading regimes.[2] Since publication, nearly 30 jurisdictions across the globe, from New York City to Seattle and from the United Kingdom to South Korea, have adopted restaurant grading. Such enactment efforts can be quite contentious—as public health practitioners have long harbored skepticism of the reliability of grading systems (Seiver and Hatfield 2000, Wiant 1999)—and consume considerable resources of local health agencies with limited budgets. New York City, for instance, allocated $3.2 million for the implementation of grading, an increase of roughly 19 percent in the food safety program's budget.[3] In 61 percent of enactment debates we identified, officials and commentators specifically referenced the foodborne illness finding in support of grading. Yet despite the ubiquity of this finding in public discourse, its internal empirical validity has never been reexamined.[4]

[1] A different version of the foodborne illness analysis was also published as Simon et al. (2005). Online Appendix A shows that the same issues identified here affect that analysis.

[2] JL's findings have also frequently appeared in policy contexts outside of food safety. We provide more details of the study's appearance in policy debates in online Appendix B.

[3] Collins, G. (2010). "For Restaurants It's Like Going Back to School." *New York Times*, July 28, A15; New York City, 2009 Executive Budget.

[4] To be sure, some have raised questions. Winston (2008) argues that isolating the effects of grading in LA is challenging given considerable changes in the food safety system in the 1990s. Ho (2012) concludes that high inter-inspector variability undermines the reliability of New York's grading system and that grading distorted enforcement resources, but does not examine the evidence in LA. Characterizing JL as a "landmark study," Bubb (2015, 1040–41) questions why LA's grading system appeared more successful at reducing foodborne illness than grading systems in other jurisdictions. Ben-Shahar and Schneider (2014a, S260) expresses skepticism of the "sensational 20 percent decline in foodborne illnesses."

We sought to build on JL's foodborne illness results by choosing foodborne illness hospitalizations as a primary outcome for a resource-intensive randomized controlled trial of a restaurant grading system we designed for Seattle and King County. Yet we found that even with a posited 20 percent effect size, the stochastic nature of foodborne hospitalizations severely impeded statistical power. To understand why power might be so distinct from LA, we expanded hospitalization data from the same source as JL by fivefold to cover 1983–2009, and collected new data on mandatorily reported foodborne illnesses from 1990–2015 and all cases of salmonella from 1964–2015. We did not reexamine the other two JL findings, and do not offer evidence to question their validity. Although we were able to replicate the foodborne illness results reported in JL, we find that the expanded hospitalization and illness data do not support the inference that LA's grading system reduced foodborne illness.

First, Southern CA experienced one of the state's largest and most acute outbreaks of salmonella in the years immediately preceding LA's enactment of restaurant grading. Because JL used all of California (excluding LA) as a control group and observed hospitalizations only from 1995–1999, the study was unable to capture the sharp drops in salmonellosis that Southern CA counties experienced starting before 1995 and most sharply around 1998. Raw data reveal this pattern across Southern CA counties that had no changes in grading policies. Consequently, JL's research design detects nearly identical and statistically significant grading effects for Southern CA counties that did not introduce restaurant grading.

Second, we show how bias resulting from JL's model specification worsens when adjusting the control group to account for the salmonella outbreak. JL's specification imposes a restrictive assumption that digestive and foodborne hospitalizations follow the same time trend in the geographic control group, while allowing these time trends to vary in LA. This assumption is not borne out in the original comparison to CA, producing moderate bias. When using Southern CA as a control group to account for the salmonella outbreak, bias becomes more substantial. Adding the missing interaction term for a full triple difference specification reduces grading effects to null. Similarly, we keep CA as a control group but reduce the weight of salmonella by including campylobacter, one of the two leading foodborne diseases. Pretreatment trends become more parallel between LA and CA as a result, but the interaction effect once again reduces grading effects to null.

Our paper proceeds as follows. Section I briefly outlines JL's research design. Section II details our much more expansive data sources on foodborne illnesses and hospitalizations. Section III examines how the Southern CA salmonella outbreak explains the effects JL attributed to grading. Section IV performs placebo tests with JL's design, showing statistically identical grading effects on hospitalizations for Southern CA counties that did not adopt restaurant grading. Section V discusses the treatment effect bias introduced by JL's model specification. Section VI concludes with implications for our understanding of information disclosure and replication efforts in economics. For brevity and to focus only on the core issues, we relegate many of the complexities and auxiliary issues to the online Appendix.

I. Extant LA Evidence

A. *Motivating Research Design*

Using discharge data between 1995 and 1999 from CA's Office of Statewide Health Planning and Development (OSHPD), JL motivates the analysis with a difference-in-differences (DID) framework, comparing foodborne illness hospitalizations between LA and CA (excluding LA)[5] before and after the grading policy took effect on January 16, 1998. The DID estimator subtracts the difference in average (logged) foodborne illness hospitalizations in CA before and after grading in 1998 ($\bar{F}^{CA}_{Before}$ and $\bar{F}^{CA}_{After}$, respectively) from the corresponding difference in LA before and after grading in 1998 ($\bar{F}^{LA}_{Before}$ and $\bar{F}^{LA}_{After}$, respectively):

$$\delta_{Foodborne} = \left(\bar{F}^{LA}_{After} - \bar{F}^{LA}_{Before}\right) - \left(\bar{F}^{CA}_{After} - \bar{F}^{CA}_{Before}\right). \tag{1}$$

JL adds a second control group of all other (non-foodborne) digestive system disorders. For simplicity, we refer to these categories as "foodborne" and "digestive" disorders. The DID estimate for digestive outcomes would subtract the difference in average (logged) hospitalizations CA before and after grading in 1998 ($\bar{D}^{CA}_{Before}$ and $\bar{D}^{CA}_{After}$, respectively) from the corresponding difference in LA before and after grading in 1998 ($\bar{D}^{LA}_{Before}$ and $\bar{D}^{LA}_{After}$, respectively):

$$\delta_{Digestive} = \left(\bar{D}^{LA}_{After} - \bar{D}^{LA}_{Before}\right) - \left(\bar{D}^{CA}_{After} - \bar{D}^{CA}_{Before}\right). \tag{2}$$

Subtracting equation (2) from equation (1) comprises a difference-in-difference-in-differences (or triple differences) identification strategy, where $\delta_{Grading}$ is the average treatment effect of grading:

$$\delta_{Grading} = \delta_{Foodborne} - \delta_{Digestive}. \tag{3}$$

JL describes its slightly different identification strategy in the text: "identification is based on time-series variation and cross-sectional variation provided by the presence of two control groups: California outside of LA and admissions for nonfood-related digestive disorders" (Jin and Leslie 2003, 439). As we show in Section V, the specific implementation deviates from triple differences in a consequential way.

Figure 1 plots the four trends described above, independently replicated from the same OSHPD data source.[6] Panel A presents the foodborne hospitalization rate (per 100,000) for LA (solid) and CA excluding LA (dashed).[7] The vertical line indicates

[5] Hereafter, the shorthand CA denotes CA excluding LA County.

[6] The correlation coefficients between our tabulated counts and JL's for the LA foodborne, CA foodborne, LA digestive system disorder, and CA digestive system disorder series are $\approx$1.00, 0.98, 0.99, and 0.99, respectively. The small remaining differences are likely due to different masking rules employed by OSHPD over time. See online Appendix J for details, showing that quarter masking leads to a loss of only 2–2.7 percent of hospitalizations.

[7] JL restricts hospitalizations to those for digestive system disorders in which the patient was admitted from home as part of an unscheduled visit. Food-related digestive disorders include diagnoses such as salmonella gastroenteritis classified as food-related in over 90 percent of cases. JL relied on Mead et al. (1999) and a medical researcher to make this classification. Section V and online Appendix D discuss the validity of this disease selection.

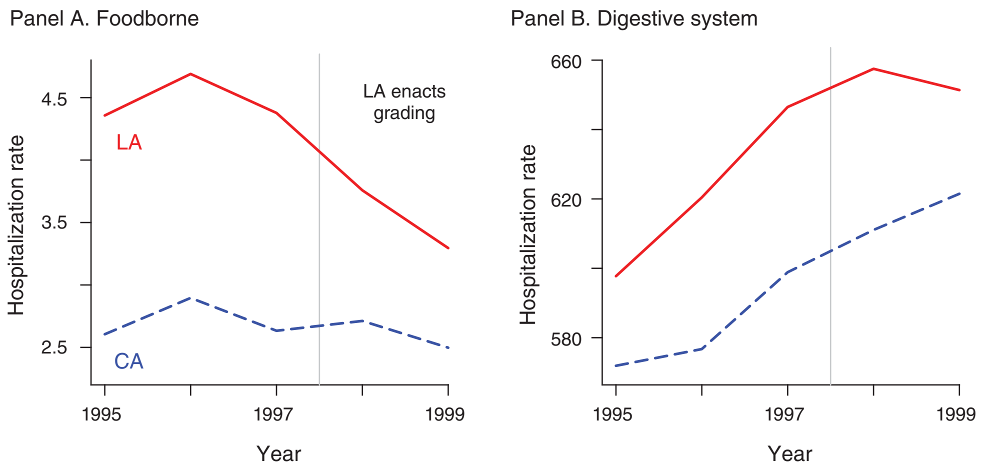

Figure 1. Trends in LA (Solid) and the Rest of CA (Dashed) of Hospitalizations for Foodborne and (Non-foodborne) Digestive System Disorders from 1995–1999

Notes: The adoption of grading is denoted by the vertical line. This figure plots the same data as Table V in Jin and Leslie (2003, 437).

when LA County grading went into effect in 1998. Foodborne hospitalization rates drop more substantially in LA compared to the rest of CA. Panel B plots analogous time trends of digestive system hospitalizations, which show minimal differences in trends between LA and CA. Because these control illnesses include anything pertaining to the digestive system (e.g., hemorrhoids, ulcers, Crohn's disease), the hospitalization rate is substantially higher and the time trend is distinct from that of foodborne illnesses.

Based on these descriptive data, JL characterizes the sharp decrease in LA's foodborne hospitalizations in 1998, in contrast to CA and digestive hospitalizations, as "basic and compelling evidence in favor of hygiene grade cards causing an improvement in actual health outcomes" (Jin and Leslie 2003, 437).

B. *JL Implementation*

After motivating the analysis of LA's grading system in a way that resembles a triple difference, JL implements the analysis as follows.

JL separates the treatment effect $\delta_{Grading}$ into two distinct mechanisms at the municipal level: *mandatory* and *voluntary* disclosure. Grading went into effect in unincorporated LA County on January 16, 1998,[8] but the 88 incorporated LA municipalities had to separately adopt grading to make posting mandatory in each city. For these incorporated cities, JL hence denotes January 1998 as imposing voluntary disclosure in incorporated cities, whereby restaurants could voluntarily disclose placards, and each subsequent municipal enactment as imposing mandatory

[8] The grading ordinance was adopted by the LA Board of Supervisors on December 16, 1997, with an effective date of January 16, 1998.

disclosure.[9] (In unincorporated LA, JL denote January 1998 as imposing mandatory disclosure.)

While these mechanisms operate at the municipal level, JL implements the analysis at the three-digit ZIP code level.[10] (Three-digit ZIP codes aggregate conventional five-digit ZIP codes to the first three digits.[11]) The key treatment variables m and v represent the population proportion of a ZIP code subject to mandatory or voluntary disclosure, respectively. We make two preliminary remarks here. First, three-digit ZIP codes align with neither municipal nor county boundaries and hence pose nontrivial implementation challenges articulated in online Appendices E, L, and M. Second, because most large LA municipalities quickly adopted grading in 1998, m and v largely stand in for the post-1998 period in LA. As we show in Section IV, v is principally identified by LA in 1998 and m by LA in 1999.

Formally, let the outcome a_{ijt} represent the number of hospital admissions for illness type $j \in \{0,1\}$, where j equals 1 if foodborne and 0 if not, in the three-digit ZIP code $i \in \{1, \ldots, 57\}$ in month t. Let the indicator variable $Food_j$ equal 1 for foodborne hospitalizations and 0 for digestive hospitalizations. The proportion of a ZIP code subject to mandatory and voluntary disclosure are calculated as m_{it} and v_{it}, respectively.[12] JL fit the following linear regression to explain logged hospitalizations:

$$(4) \qquad \ln(a_{ijt}) = \alpha_{ij} + \tau_t + \beta_1 m_{it} + \beta_2 v_{it} + \gamma_1 m_{it} \times Food_j + \gamma_2 v_{it} \times Food_j + \epsilon_{ijt},$$

where α_{ij} are fixed effects for each illness-ZIP combination, τ_t are time fixed effects, and β_1 and β_2 control for illness-invariant, time-specific effects in mandatory and voluntary disclosure areas in LA, respectively. The causal effects of mandatory and voluntary grading, γ_1 and γ_2, are each estimated to be a roughly 20 percent reduction in foodborne hospitalizations.[13]

II. Expanded Data

It is widely accepted that the credibility of DID (or triple differences) hinges on parallel pretreatment time trends between comparison groups (Wing, Simon, and Bello-Gomez 2018; Greene 2012; Imbens and Wooldridge 2009; and Angrist and Pischke 2008). Because the original observation period was relatively short (1995–1999), we begin our investigation by substantially expanding the data in three ways.

[9] The only LA County cities not ever subject to voluntary grading are Long Beach, Pasadena, and Vernon, as these cities operated inspections independently of the county and adopted neither voluntary nor mandatory grading.

[10] While OSPHD data include a county field, JL likely used three-digit ZIP codes because OSHPD does not include municipality identifiers.

[11] For instance, ZIP codes 94306 and 94305 are contained in the 943 ZIP code.

[12] While m and v sum to unity for interior ZIP codes, m and v can sum to less than one for ZIP codes that are partially outside of LA. For details, see online Appendix L.

[13] Specifically, $\gamma_1 = -0.22$ (standard error $= 0.04$) and $\gamma_2 = -0.21$ (standard error $= 0.04$). JL interprets the "net effect [as] the sum of the coefficients" β and γ, but as we explain in online Appendix F, this does not cohere with the triple difference design. In later descriptions, JL correctly interprets the causal effect as a "20 percent decrease in foodborne illness hospitalizations" (Jin and Leslie 2009, 238).

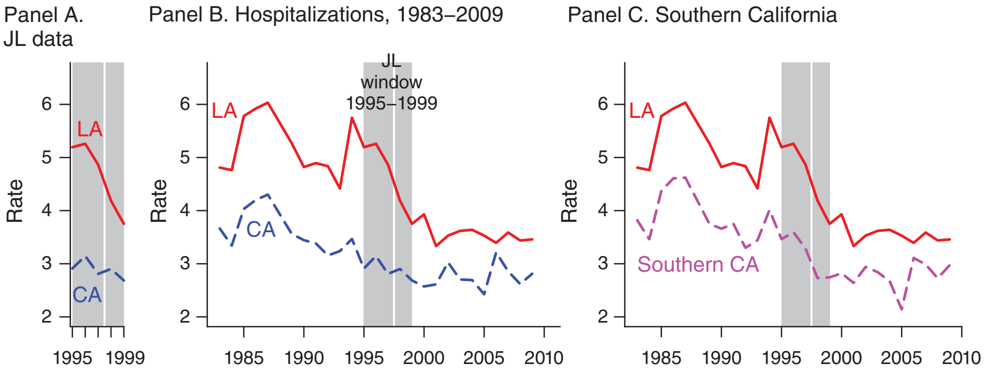

Figure 2. Expanding the Observation Period for Hospitalization Rates to 1983–2009

Notes: The gray box and left panel indicate the original JL observation window, with the vertical line indicating the adoption of grading. The salmonella outbreak causes a sharp spike in LA around 1994 compared to the rest of CA. There is little evidence of parallel trends in the pretreatment period. The right panel plots LA against Southern CA (excluding LA). Spikes in the late 1980s reflect a salmonella newport outbreak.

Hospitalizations, 1983–2009.—We obtain discharge data from OSHPD to expand JL's 5 years of hospitalization data to 26 years (1983–2009).[14] This increases the total number of hospitalizations by more than fivefold. Panel A of Figure 2 plots, for reference, the comparable foodborne hospitalization data from Figure 1.[15] Using the original 1995–1999 observation period, we successfully replicate the JL model, finding comparable (statistically significant) effects for mandatory and voluntary disclosure (see Model 1 of Table 3).[16] Our coefficients are slightly larger in magnitude due to small differences in aggregation and masking, but our coefficients are statistically indistinguishable from those reported by JL (p-values are 0.33 and 0.63).

Panel B expands JL's observation window (denoted by the shaded gray box) and plots hospitalization rates for 1983–2009. Several trends become apparent. First, the decrease in LA hospitalizations started from a proximate peak in 1994 well *before* grading was adopted. Second, the data refute the assumption of parallel trends. Hospitalizations, for instance, spike much more sharply (and statistically significantly) in LA in 1994 than in CA overall.[17] Third, the panel C plots a separate time series for Southern CA (excluding LA). For simplicity and to maintain a comparable number of treated ZIP codes between LA and Southern CA,[18] we

[14] We could not recover one year of data in 1988.

[15] The only difference is that we do not apply the hospitalization filter for unscheduled admissions from home, as OSHPD does not contain these fields before 1995. As online Appendix J shows, this choice is inconsequential.

[16] While our data source is identical, we use a version of OSHPD data that contains five-digit ZIP code aggregated at the quarterly level. Because JL does not explicitly discuss the treatment of zero counts for the outcome ln (a_{ijt}), we take a standard approach by modeling $\ln(a_{ijt} + 1)$.

[17] A test for lead effects on hospitalizations in the pretreatment period confirms that trends are not parallel. The p-value < 0.05 on the treatment coefficient in a DID with 1994 as the treatment year, an observation window of 1990–1997, fixed effects for three-digit ZIPs and year-quarters, and standard errors clustered by three-digit ZIP. For a fuller set of placebo tests, see online Appendix G.

[18] Because of LA's large population, out of 57 total three-digit ZIP codes, there are 18 fully or partially in LA county and 20 fully or partially in Southern CA (excluding LA).

use a widely accepted definition of Southern CA as the ten southernmost counties excluding LA.[19] It is worth noting that Southern California is geographically quite dispersed: San Diego is 120 miles from Los Angeles, roughly the same distance as between New York City and Hartford, CT. The panel shows that LA and Southern CA trends are substantially similar throughout the longer time series. As we explain in Section III, the similarity around 1995–1999—when no other Southern CA county other than LA adopted restaurant grading—is in fact driven by a major salmonella outbreak in Southern CA.

Illnesses, 1990–2015.—While the expanded hospitalization data are useful, the epidemiological consensus is that hospitalizations capture only a small part of foodborne illness incidence (Mead et al. 1999, Scallan et al. 2011). The Institute of Medicine reports: "hospital discharge summary records … significantly under-report specific [foodborne] infections, as laboratory diagnoses may often not be reflected in the discharge … coding" (Institute of Medicine 2006, 77–78), a point widely echoed in the public health field.[20] Only 27 percent of laboratory-confirmed salmonella illnesses, for instance, result in hospitalization (Scallan et al. 2011). Table 1 presents estimates of total annual hospitalizations and laboratory-confirmed illnesses for foodborne pathogens from an authoritative review of foodborne pathogens (Scallan et al. 2011). The left columns present national estimates and the right columns present average annual counts for LA. The actual number of hospitalizations in LA is quite low. From 1995–1999 in LA, we observe an annual average of only 213 hospitalizations for salmonella, compared to 1,701 laboratory-confirmed salmonella illnesses. CDC data similarly show that while the average number of hospitalizations for an outbreak is 0.95, the average number of reported illnesses is 19.5.[21] Hospitalization data may be particularly sensitive to outbreaks. In 2014, for instance, a salmonella outbreak at Foster Farms sickened 490 Californians, with 38 percent hospitalized. This single outbreak might account for nearly 20 percent of the CA's total foodborne hospitalizations in a year.[22]

Fortunately, Table 1 also points to a much more complete second data source: laboratory-confirmed illnesses. CA law requires that health practitioners, administrators, laboratories, and schools report a wide range of communicable diseases, including the major foodborne illnesses studied by JL, to local health departments (Kim-Farley 2000). This rich source of data (typically referred to as "notifiable"

[19] See https://en.wikipedia.org/wiki/Southern_California ("Southern California … is a geographic and cultural region that generally comprises California's 10 southernmost counties.") Excluding LA, this area is comprised of Imperial, Kern, Orange, San Bernardino, San Diego, San Luis Obispo, Santa Barbara, Riverside, and Ventura counties. Counties that did not engage in restaurant grading during the observation period are Imperial, Orange, San Luis Obispo, Santa Barbara, and Ventura counties. As mentioned above, San Diego and Riverside have been grading restaurants since 1947 and 1963, respectively. Two counties enacted restaurant grading outside of JL's observation period but within our longer observation window: San Bernardino in 2005 and Kern in 2006. Online Appendix I shows that the results are insensitive to accounting for grading adoptions.

[20] See, e.g., Roberts, Jensen, and Unnevehr (1995, 53) ("ICD [hospital discharge] codes are currently incomplete and provide insufficient detail for analyzing sources of foodborne illness."); Boehmer et al. (2011, 106) ("Using discharge diagnoses alone is practical only for diseases that have simple case definitions and nonspecific symptoms that require laboratory confirmation before making a diagnosis.").

[21] This is based on data on all reported outbreaks to the CDC from 1998–2015.

[22] This is comparing hospitalizations (0.38×490) against the annual average observed by JL of just above 1,000 annual hospitalizations for CA from the 1995–1999 data. See Table 6 in online Appendix D.

Table 1—Estimated Annual US Hospitalizations and Lab-Confirmed Illnesses for Top Bacterial Foodborne Pathogens

	National				Los Angeles	
	Hospitalizations		Illnesses		Hospitalizations	Illnesses
Pathogen	Count	Prop.	Count	Prop.	Count	Count
Salmonella	19,366	0.57	41,930	0.44	213	1,701
Campylobacter	8,463	0.25	43,696	0.46	59	1,479
E. Coli	2,138	0.06	3,704	0.04	<18	18
Listeria	1,455	0.04	808	0.01	17	35
Staphylococcus	1,064	0.03	323	<0.01	9	–

Notes: These estimates are based on 18 pathogens reported in Scallen et. al (2011, 11–12, Tables 2 and 3) that are estimated to be 50 percent or more foodborne. US estimates adjust for underreporting and underdiagnosis and are based on data from 2000–2008. Rows are sorted by hospitalization cases. Salmonella is non-typhoidal; E. Coli is Shiga-toxin producing E. Coli O157; and staphylococcus is staphylococcus aureus. Prop. represents the proportion out of a total of 34,239 hospitalizations or 95,726 lab-confirmed illnesses for 18 bacterial pathogens. This table shows the importance of a small number of pathogens, particularly salmonella and campylobacter, in explaining variability in foodborne hospitalizations and illnesses. By comparison, we also present average annual hospitalization and illness counts for LA between 1995–1999. We use the ICD 9 codes described in Table 1 of Mead et al. (1999) to match pathogens to hospitalizations. Because foodborne pathogens do not perfectly align with ICD 9 codes, an upper bound is presented for E. Coli O157 using all applicable ICD codes.

or "reportable conditions") is the most common tool for foodborne illness surveillance.[23] Because there is nothing in the theory of restaurant grading to suggest it should exclusively affect hospitalizations,[24] these data provide an alternative way to test the impact of restaurant grading. We hence digitize mandatorily reported foodborne illnesses from 1990–2015 using state communicable disease reports. We begin the observation period in 1990, as it corresponds to when campylobacter and vibrio became subject to mandatory reporting.[25] Figure 3 plots illness rates over time based on these data. The solid lines plot LA compared against the rest of CA in panel A and against Southern CA (excluding LA) in panel B. While illnesses drop dramatically in LA during the 1995–1999 observation period (in gray), this drop is evident in the rest of the state as well. Most striking is that the Southern CA series, which does not include LA, is nearly indistinguishable from the LA series.

Salmonella, 1964–2015.—The patterns in Figure 3 led us to further investigate the drop in foodborne illnesses by examining specific pathogens. While JL aggregates over a dozen pathogens, Table 1 shows that salmonella is by far the most prevalent. Fifty-seven percent of nationwide hospitalizations for foodborne bacterial infections are for salmonella (Scallan et al. 2011, 12). In JL's data, 61 percent of

[23] See Dodd et al. (2017, 35) ("While disease surveillance can include the ongoing collection and monitoring of morbidity and mortality data from a variety of sources (i.e., physician visits, hospital records, death certificate data), most foodborne disease surveillance data rely on the monitoring of laboratory-diagnosed cases of infection.") and Schweitzer, Reza Zali and Jackson (2006, 44) ("Mandatory reporting of selected diagnoses and laboratory test results is a pillar of our system.").

[24] Indeed, JL itself contemplates an impact on illnesses, not hospitalizations: "grade cards should cause … decreases in *illnesses*" (Jin and Leslie 2003, 414, emphasis added).

[25] Campylobacter became reportable on March 30, 1989 (Hastings et al. 1991, 3), but because we collect data at an annual level, we begin in 1990. Vibrio became mandatorily reportable in 1988 (Kizer 1994), and we first observe cases compiled in 1989. The foodborne diseases that were mandatorily reportable from 1990–2015 were salmonella, campylobacter, listeria, and vibrio. E. Coli O157 became reportable in 1996 (Belshé et al. 2003), and online Appendix K shows that results are comparable when including E. Coli.

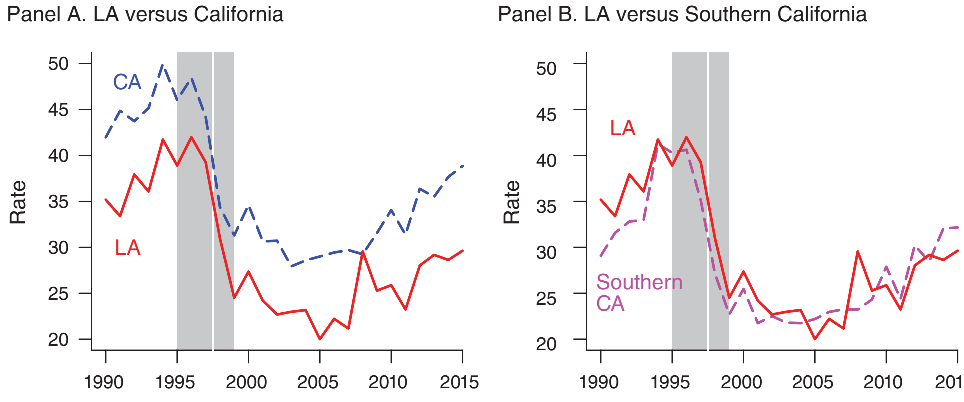

Figure 3. Mandatorily Reported Foodborne Illnesses, 1990–2015

Notes: This figure plots foodborne illness rates (per 100,000 individuals) over time. Panel A compares CA excluding LA (dashed) against LA (solid) and panel B compares Southern CA excluding LA (dashed) against LA (solid). The drop in reported illnesses around 1998 is comparable for the CA series, and Southern CA counties exhibit the same pattern as LA. The four conditions that were mandatorily reported from 1990–2015 were salmonella, campylobacter, listeria, and vibrio.

foodborne hospitalizations are for salmonella. We hence hand-collect salmonella illness reports from weekly and yearly state reports from 1964, which corresponds to the beginning of the salmonella national reporting system, to 2015 (Swaminathan, Barrett, and Fields 2006).

III. Southern CA Salmonella Outbreak

We document here that one of the largest recorded salmonella outbreaks in state history affected Southern CA counties prior to LA's implementation of restaurant grading. Following the outbreak, salmonella cases sharply dropped from 1994–1999 across Southern CA counties, with none other than LA implementing restaurant grading. Because JL's analysis compares LA to the rest of CA in an observation window too short to capture the full course of the outbreak, it misattributes the entire salmonellosis drop to restaurant grading, without consideration of the time- and region-specific shock of the outbreak and its response.

Figure 4 displays illness and hospitalization rates for salmonellosis over time in Southern CA counties.[26] Illness rates are significantly higher than hospitalization rates, confirming the value of this richer dataset. Most striking is that each county experiences a sharp spike in illness and hospitalization rates in the years immediately prior to JL's observation period (shaded in gray). The salmonella (enteritidis) outbreak was widely recognized. The *LA Times* reported, "Health officials from five Southern California counties [Los Angeles, Orange, Riverside, San Bernardino, and San Diego] are reporting dramatic increases—ranging from 700% to 1,782%—in

[26] The unusually high reported cases of salmonellosis in Riverside from 1965 is accurate. In that year, there was an outbreak of salmonella typhimurium affecting an estimated 16,000 of 133,000 residents (Riverside County Health Department 1971).

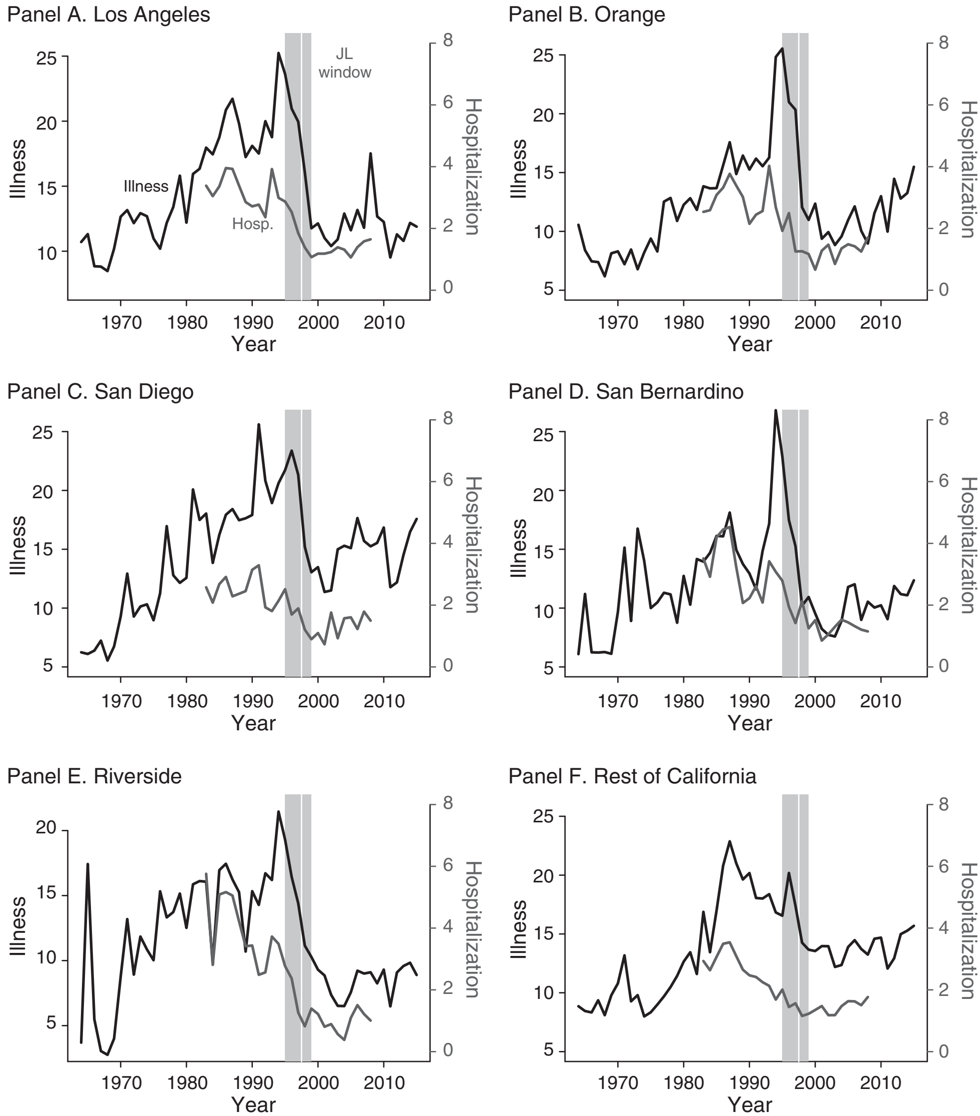

Figure 4. Salmonella Time Series for Reported Illness Rates (per 100,000) (Black) and Hospitalization Rates (per 100,000) (Gray) for Five Southern CA Counties and the Rest of CA Excluding Those Counties

Notes: The time period under study by JL (1995–1999) is highlighted in gray, with the white vertical line representing the adoption of grading. This figure shows that the dramatic salmonella outbreak in the mid-1990s preceded restaurant grading in 1998. Reported illnesses and a longer observation period show that the drop in salmonellosis occurred before restaurant grading in 1998 and is comparable across each of the Southern CA counties known to have been affected by the outbreak, but which exhibited no changes in grading policy.

illnesses linked to contaminated eggs over the past six years."[27] One state epidemiologist opined, "Salmonella enteritidis has become No. 1 on the hit parade" in

[27] These are rates for the specific serotypes linked to eggs (Salmonella enteritidis), and are hence even more dramatic than the increases in Figure 4, which pool all salmonella serotypes. Salmonella enteriditis, typically found in eggs and poultry, is the salmonella serotype most commonly responsible for outbreaks (Jackson et al. 2013).

Table 2—Comparison of Illness Rates from 1994–1999 between LA and Orange County

	Salmonella rate		Percent change	
	LA	Orange	LA	Orange
1994	25.2	24.8	34.5	52.6
1995	23.6	25.5	−6.4	2.9
1996	21.0	21.0	−11.3	−17.8
1997	20.0	20.3	−4.8	−3.1
1998	16.1	12.0	−19.5	−40.8
1999	11.8	11.0	−26.8	−8.9

Notes: The illness rate is the reported salmonella rate per 100,000. Percent change indicates the percent change relative to the preceding year. This table shows that, if anything, Orange County exhibited a sharper drop in salmonellosis than LA in 1998.

Southern CA.[28] Public health reports noted that the salmonella enteritidis "epidemic … involved much of Southern California [but that] northern California counties experienced no overall increase during this period" (Passaro et al. 1996).

The result of the outbreak was that other Southern CA counties, which made no changes in restaurant grading policies during this time,[29] experienced salmonellosis declines as large, if not larger than, LA. Most importantly, those decreases began *before* the 1995–1999 observation period and continued throughout that observation period. Consider the case of Orange County, reported by the *LA Times* as most similarly affected by the outbreak. Table 2 provides salmonellosis rates from 1994–1999, as well as annual percent changes. Both LA and Orange County exhibit peak rates in 1994–1995, and both rates begin to drop before 1998, continuing through the end of JL's observation period. If anything, Orange County, with no introduction of restaurant grading, exhibits a larger decrease (40 percent) in 1998 than LA County (20 percent).

In response to the outbreak, LA, Orange, Riverside, San Bernardino, and San Diego counties took "unprecedented joint action" to warn consumers about raw eggs.[30] The major cause of the outbreak appeared to be contamination at a single, large Southern CA egg ranch (Kinde et al. 1996, Passaro et al. 1996). Egg producers were particularly threatened by the outbreak and responded by instituting strict protocols for industrial safe practices (e.g., salmonella testing and vaccination, "Hazard Analysis and Critical Control Points" (HACCP) protocols). This egg safety program likely played a substantial role in decreasing salmonella in southern CA.[31] Regardless of which response might have helped, the richer data reveal that the tail end of the outbreak coincided nearly perfectly with the advent of restaurant grading in LA.

It is worth emphasizing that the illness rates plotted in Figure 4 stem directly from communicable disease reports issued by the state of CA. These raw data depend on no modeling assumptions and plainly show confounding in the large LA grading effect

[28] Nicolosi, M. 1994. "Salmonella Called Epidemic in O.C." *Orange County Register*, December 6, B1.

[29] San Diego and Riverside have been grading restaurants since 1947 and 1963, respectively. Orange County has not graded restaurants in modern history. San Bernardino began grading restaurants in 2005, and Figure 4 does not suggest any health benefits close to the 20 percent magnitude reported for LA.

[30] Puzo, D.P. 1995. "Illnesses tied to tainted eggs increase sharply." *Los Angeles Times,* March 7.

[31] See Mumma et al. (2004, 1782) (egg quality assurance programs adopted in response to the "epidemic" likely "played a major role in reducing of S. Enteritidis"). Other CA regions had not experienced major problems with contaminated eggs before the program was in place (Breitmeyer 1997).

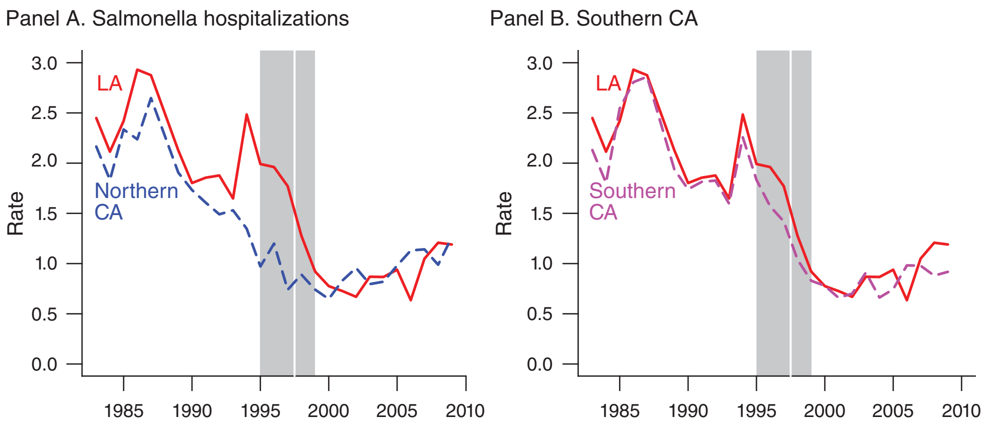

Figure 5. Salmonella Time Series for Hospitalization Rates in LA, Southern CA, and Northern CA (all other CA counties excluding LA and southern CA counties)

Table 3—Replication of JL's Analysis with Placebo Tests Substituting Southern CA Counties (Excluding LA) as Placebo Treated Units

	Continuous treatment				Binary treatment			
	LA treated		S. Cal. treated		LA treated		S. Cal. treated	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Foodborne × mandatory disclosure post-1998	−0.31	−0.33	−0.28	−0.39	−0.31	−0.31	−0.31	−0.37
	(0.07)	(0.05)	(0.11)	(0.09)	(0.06)	(0.04)	(0.10)	(0.08)
Foodborne × voluntary disclosure post-1998	−0.27	−0.30	−0.32	−0.36	−0.26	−0.27	−0.32	−0.32
	(0.08)	(0.08)	(0.12)	(0.12)	(0.05)	(0.06)	(0.09)	(0.09)
Mandatory disclosure post-1998 (digestive)	0.04	0.05	0.07	0.22	0.05	0.04	0.09	0.21
	(0.03)	(0.03)	(0.06)	(0.08)	(0.03)	(0.03)	(0.05)	(0.07)
Voluntary disclosure post-1998 (digestive)	0.08	0.09	0.07	0.09	0.04	0.04	0.03	0.07
	(0.04)	(0.03)	(0.04)	(0.06)	(0.03)	(0.03)	(0.04)	(0.05)
R^2	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
Observations	2,280	7,752	2,280	7,752	2,280	7,752	2,280	7,752
Observation period	1995–1999	1993–2009	1995–1999	1993–2009	1995–1999	1993–2009	1995–1999	1993–2009

Notes: Coefficients are shown with standard errors, clustered by three-digit ZIP and illness-type combinations, in parentheses. Each model is estimated with fixed effects for three-digit ZIP and illness-type combinations and year-quarters. We present results for the original observation period (1995–1999) as well as an expanded observation period (1993–2009) for comparability with later models. Continuous treatment indicates m and v as the fraction of a ZIP code subject to mandatory or voluntary grading. Binary treatment for v equals the proportion of the three-digit ZIP in LA if the year is 1998 and 0 otherwise, and for m equals the proportion of the three-digit ZIP in LA if the year is 1999 or later and 0 otherwise. The continuous treatment for Southern CA is calculated from 1,000 random draws from LA's observed m and v values, adjusting for boundary populations. Models 3 and 4 present representative models out of the 1,000 draws based on the lowest sum of squared distances from the median t-statistics for each parameter. For details, see online Appendix L. The rejection rate for a significant m effect in the Southern California continuous models is 81.5 percent for 1995–1999 and 100 percent for 1993–2009. The rejection rate for a significant v effect in the Southern California continuous models is 86.2 percent for 1995–1999 and 89.6 percent for 1993–2009.

JL infer from the drop in LA hospitalizations (relative to CA) from 1995–1999.[32] Figure 5 confirms with the expanded hospitalization data that LA and Southern CA

[32] As JL writes, in 1998, "there was a 13.3 percent decrease in hospital admissions for food-related digestive disorders in Los Angeles, relative to the year before … In addition, if one looks at food-related digestive disorders

are nearly indistinguishable during this period, while Northern CA fails to capture the sharp drop in salmonella hospitalizations between 1994 and 1999.

IV. Simple Placebo Tests

We show here that because of the salmonella outbreak, JL's model generates identical treatment effects for Southern CA counties, excluding LA.

We fit JL's model in equation (4), first replicating the effects for LA and then substituting Southern CA as treated units. Table 3 presents results. Model 1 shows the replicated results from JL on the quarterly OSHPD data with three-digit ZIP codes from 1995–1999.[33] The first two rows of Table 3 present treatment effect estimates: coefficients on the interaction term between foodborne hospitalization and either mandatory m (first row) or voluntary v (second row) disclosure. Because m and v are positive only starting in 1998 for the treated units, we label them "post-1998" for clarity.[34] Both are statistically significant and negative, indicating that restaurant grading reduced foodborne hospitalizations by more than 20 percent. Coefficients on m and v (β_1 and β_2 from equation (4)), reported in the third and fourth rows, allow for digestive hospitalizations to vary over time within LA. (For clarity, we label it as "Digestive" to indicate that it represents the time trend for digestive disorders in LA.) Model 2 fits the same model using a longer observation period from 1993–2009, with comparable treatment effect estimates.[35]

Models 3 and 4 fit the same model using Southern CA counties as treated units, with LA in the control group. To assign placebo m and v values, we randomly sample m and v values from LA, adjusting only for ZIP codes at the boundary.[36] We find strikingly similar results. Southern CA is estimated to have comparable treatment effects.

Because most large LA municipalities adopted restaurant grading within a year, mandatory disclosure (m) and voluntary disclosure (v) measures may proxy simply for time in LA. Figure 6 displays the average m and v values over time in LA ZIP codes. We find that v spikes in 1998 and m rises to nearly 0.7 in 1999. (Due to boundary ZIP codes, m and v do not necessarily sum to unity in a ZIP code.) These coefficients may hence reveal less about mandatory enactment by municipalities per se

in the rest of California in 1998, there was a 3.2 percent increase in hospitalizations from the prior year … This is basic and compelling evidence in favor of hygiene grade cards causing an improvement in actual health outcomes."

[33] Although the data source is identical, JL uses a version aggregated at the three-digit ZIP code monthly level, while our version is aggregated at the five-digit ZIP code quarterly level. For replicability, we retain the three-digit ZIP code unit of analysis in this section, but use quarterly time units. We spell out some of the limitations of using three-digit ZIP codes in online Appendix E. A three-digit ZIP code, for instance, can span from LA to Yosemite, requiring fractional assignment of the treatment. Our data allow us to assign hospitalizations to a specific county (rather than a three-digit ZIP) in online Appendix K. Most importantly, because hospitalizations are sparse, aggregating from months to quarters is unlikely to provide any additional leverage with these data. For instance, JL report 910 total foodborne hospitalizations in CA in 1999. Given 57 three-digit ZIP codes, the expected number of hospitalizations at the monthly level is only 1.3 cases and at the quarterly level is 4. Table 3 shows that the aggregation does not matter, yielding coefficients that are statistically indistinguishable from the ones reported by JL. Online Appendix E also shows that because municipal adoptions largely track quarters, there is little to be gained from monthly data.

[34] As we explain in Section V, these coefficients are in actuality triple interaction terms between time, unit location, and illness type.

[35] For comparability with models presented later, we use 1993 as the beginning date, as that is the first full year after a separate diagnosis code for campylobacter was developed. Online Appendix I shows that these results are insensitive to observation period.

[36] For details, see online Appendix L.

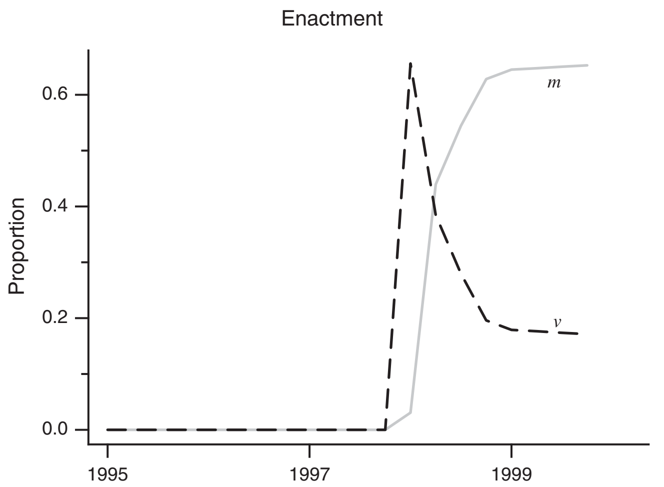

FIGURE 6. AVERAGE VALUES OF m AND v OVER ZIP AGAINST TIME

Notes: v indicates the proportion of LA subject to "voluntary" grading, defined as the passage of grading in the county, without municipal enactment. m indicates the proportion of LA subject to "mandatory" grading, meaning after the enactment of grading amongst 88 municipalities. This figure shows that because most of the LA population lived in municipalities that quickly adopted grading, m and v largely proxy for the first and second year after 1998.

than the continuing decline in LA foodborne hospitalizations in 1998 and 1999.[37] To test this, models 5–8 in Table 3 replace the continuous variables m and v with binary indicators for whether the year is 1998 and whether the year is 1999 or later, respectively.[38] The results are comparable: the binary treatment yields effect estimates very close to models 1–4, including the statistically significant Southern CA placebo effect.

In sum, using the same model as JL, Southern CA exhibits statistically identical grading effects, even though no Southern CA county adopted grading in this time period.[39] This confirms the substantive account of Section III: any detectable effects of restaurant grading are confounded with the sharp drops in salmonella across Southern CA after a multiyear regional epidemic.

V. Model Bias in Treatment Effect

In this section, we show how specification bias in JL's model is exacerbated with an improved geographic control group or an improved illness selection that reduces the weight of salmonella.

[37] The version of the analysis by the same researchers published as Simon et al. (2005) uses a dummy variable representation of the posttreatment years and finds no statistically significant change in the decline from 1998 to 1999.

[38] To account for boundary ZIP codes, just as in JL, we multiply the binary treatment indicator with the proportion of each ZIP code's population that is in LA County or Southern CA for the placebo analysis.

[39] The following table reports p-values for pairwise comparisons of treatment effect coefficients:

		S. CA			
		Mandatory		Voluntary	
LA	1995–1999	0.80	0.99	0.73	0.61
	1993–2009	0.58	0.46	0.68	0.63

Table 4—Marginal Effects of Restaurant Grading in LA with Southern CA as the Control Group

	1995–1999			1993–2009		
	(1)	(2)	(3)	(4)	(5)	(6)
Foodborne × LA mandatory disclosure post-1998	−0.04	−0.31	0.00	−0.08	−0.33	0.03
	(0.10)	(0.07)	(0.10)	(0.06)	(0.05)	(0.09)
Foodborne × LA voluntary disclosure post-1998	0.11	−0.27	0.06	0.02	−0.30	0.07
	(0.10)	(0.08)	(0.11)	(0.12)	(0.09)	(0.12)
Foodborne × S.Cal. post-1998			−0.29			−0.33
			(0.07)			(0.07)
LA mandatory disclosure post-1998 (digestive)		0.13	−0.02		0.07	−0.10
		(0.05)	(0.04)		(0.05)	(0.06)
LA voluntary disclosure post-1998 (digestive)		0.19	0.02		0.15	−0.03
		(0.06)	(0.05)		(0.05)	(0.06)
Model	DID	JL	DDD	DID	JL	DDD
R^2	0.75	0.99	0.99	0.73	0.99	0.99
Observations	660	1,320	1,320	2,244	4,488	4,488
F-ratio			30.74			116.15

Notes: Coefficients shown with standard errors, clustered by three-digit ZIP and illness-type combinations, are in parentheses. Each model is estimated with fixed effects for three-digit ZIP and illness-type combinations and year-quarters. We present results for the original observation period (1995–1999) as well as an expanded observation period (1993–2009). Disease selection is consistent with JL's original design.

A. *Southern CA*

The results of Section IV suggest that we can leverage Southern CA as a control group to identify the marginal effect of restaurant grading in LA net of the salmonella outbreak. At the outset, we acknowledge that spillover effects in the region could threaten this design. However, the assumption of no spillovers follows directly from JL's original design, which assumes away spillovers in order to identify mandatory and voluntary effects between closely neighboring areas. As an additional check, online Appendix H offers statistical tests for regional spillover effects and finds no significant results. As we mention in Section II, the dispersed geography of the region makes it implausible that non-independence is driven primarily by dining across Southern CA counties.

Table 4 shows the results of using Southern CA as the control group with a difference-in-differences specification on foodborne illnesses alone (DID);[40] the JL specification in equation (4) that includes a disease control group (JL); and a fully specified triple differences model, which simply adds a two-way interaction between digestive disorders and time to equation (4) (DDD). Regardless of observation period, the DID and DDD models find no marginal effect of restaurant grading net of other regional factors that also influenced Southern CA counties. Visualizing

[40] Motivated by equation (1), this model is equivalent to the JL specification without a disease control group,

$$\ln(a_{it}) = \alpha_i + \tau_t + \gamma_1 m_{it} + \gamma_2 v_{it} + \epsilon_{it}, \tag{5}$$

where α_i are fixed effects for each three-digit ZIP, τ_t are time fixed effects, and γ_1 and γ_2 represent the treatment effects for mandatory and voluntary disclosure areas in LA, respectively.

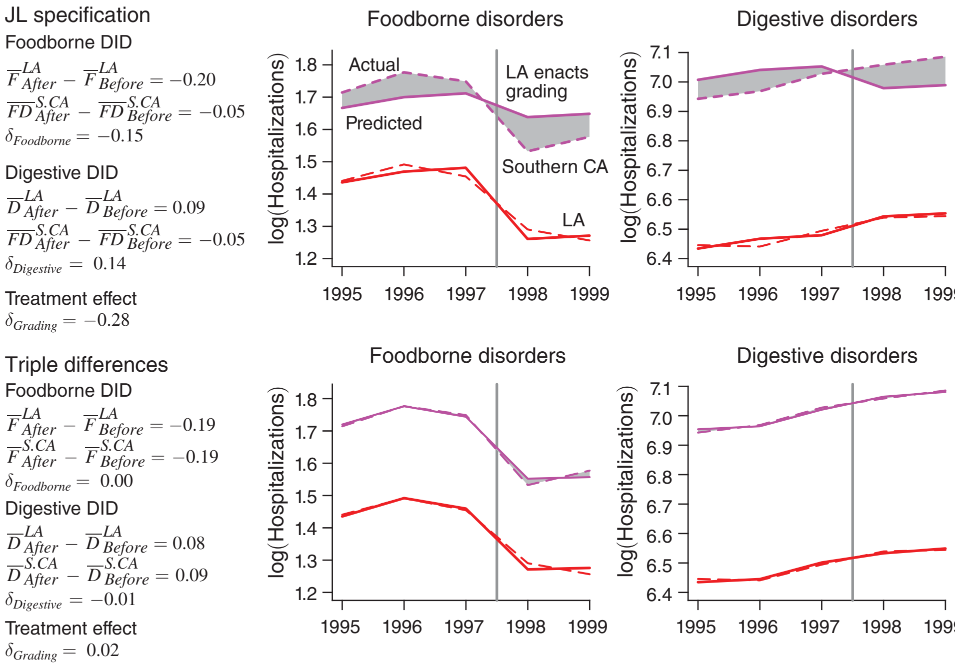

FIGURE 7. BIAS FROM OMITTING TWO-WAY INTERACTION TERMS WITH SOUTHERN CA AS A CONTROL GROUP IN THE ORIGINAL OBSERVATION PERIOD, 1995–1999

Notes: The top row presents model predictions against observed data from JL's specification with a single treatment indicator averaged over mandatory and voluntary disclosure. The bottom row presents model predictions against observed data from a fully specified triple difference, modeling the *Food* × *After* interaction with illness-by-time fixed effects. Dashed lines indicate observed data and solid lines indicate predicted values. For consistency with JL's specification, outcomes are presented as log counts of hospitalizations. Because the JL specification assumes time trends for foodborne and digestive disorders are the same for Southern CA (but not LA), when foodborne disorders in fact decrease and digestive disorders in fact increase after 1998, the top row shows that Southern CA's predicted slope is attenuated toward zero. Shading indicates prediction error. The foodborne DID estimate is biased downward because the model does not fit the drop in CA after 1998. The digestive DID is biased upward because the model does not fit Southern CA's increase after 1998. Both of these biases magnify the treatment effect. Calculations are shown on the left: the triple difference estimate is −0.28 for the JL specification and 0.02 for the fully specified triple difference.

these findings, Figure 7 shows that foodborne illnesses drop, and digestive disorders rise, by nearly identical amounts in LA and Southern CA between 1995 and 1999.

In stark contrast to these models, the JL specification finds significant effects because it does not fit the data for Southern CA. Figure 7 illustrates how this occurs. Digestive disorders trend upward and foodborne illnesses trend downward for both LA and Southern CA, but the JL specification predicts the average over these opposing trends for Southern CA alone. As shown in gray, this specification results in substantial prediction error. Highly statistically significant F-ratios in Table 4 confirm that the effects reported in models 2 and 5 are artifacts of poor model fit.

To clarify the specification issue, we rewrite JL's equation (4) with a single average treatment effect γ_1 (corresponding to $\delta_{Grading}$ in Section I), which can be interpreted as the effect averaged over mandatory and voluntary disclosure. This is based on the fact that m_{it} and v_{it} sum to the proportion of LA county ZIP codes subject to any form of grading after January 16, 1998. The combined average effect of mandatory and voluntary grading is a triple interaction between LA ZIP codes, time after 1998, and illness type:

$$(6) \qquad \ln(a_{ijt}) = \alpha_{ij} + \tau_t + \beta_1(LA_i \times After_t) + \gamma_1(LA_i \times After_t \times Food_j).$$

Three parameters capture temporal variation: β_1 for LA digestive disorder hospitalizations after grading, γ_1 for LA foodborne hospitalizations after grading, and $\tau_{t\in 1998,1999}$ for the geographic control group's hospitalizations after grading. Unlike canonical triple differences (e.g., Angrist and Pischke 2008, 242–43; Gruber 1994, 630–31; Imbens and Wooldridge 2007, 2–3; and Khandker, Koolwal, and Samad 2010, 80–82), there is no explicit parameterization of the geographic control group's *foodborne* hospitalizations after grading (e.g., $\beta_3 Food_j \times After_t$).[41] The result is that the average trend of foodborne and digestive hospitalizations in the geographic control group ($\overline{FD}^{CA}_{After} - \overline{FD}^{CA}_{Before}$) stands in for both the control group's trend in foodborne hospitalizations ($\bar{F}^{CA}_{After} - \bar{F}^{CA}_{Before}$) and the control group's trend in digestive disorder hospitalizations ($\bar{D}^{CA}_{After} - \bar{D}^{CA}_{Before}$). In contrast, the fully specified triple differences in equation (7) allows foodborne trends to vary distinctly from digestive trends in both LA and its control group by adding the bolded missing interaction term to equation (4):

$$(7) \qquad \ln(a_{ijt}) = \alpha_{ij} + \tau_t + \beta_1 m_{it} + \beta_2 v_{it} + \boldsymbol{\beta_3 Food_j \times After_t} + \gamma_1 m_{it} \times Food_j + \gamma_2 v_{it} \times Food_j + \epsilon_{ijt}.$$

The consequences of omitting the interaction term are twofold. First, the treatment effect point estimates are biased to the extent that the trend averaged over digestive and foodborne illnesses diverges from the trends within each disease category in the geographic control group. Figure 8 shows that time trends for foodborne and digestive hospitalizations move in opposite directions in LA, CA, and Southern CA.[42] But since LA and Southern CA also experience a steep secular drop in foodborne illnesses, averaging over the foodborne and digestive trends is most problematic for these geographies. The better the match on pretreatment foodborne illness trends between LA and its control group (holding digestive disorder trends constant), the worse the model fits the data.

Second, omitting the interaction term downwardly biases the (cluster-robust) standard errors of treatment effects. Absent the interaction term, the JL model pools time fixed effects across foodborne and digestive hospitalizations for the geographic control group. Foodborne hospitalizations have much more within-cluster temporal

[41] For applications of triple differences, each of which include the full set of second-level interactions, see Gruber and Poterba (1994), Acs and Nelson (2004), Figlio (2006), and de Carvalho Filho (2008).

[42] The divergence may also call into question the validity of the digestive disorders as a disease control group. This control group, after all, comprises nearly 1,000 diagnostic codes for any disorder of the digestive system, ranging from "foreign body in colon" to "unspecified open wound of abdominal wall" to "malignant carcinoid tumor of the appendix" (Centers for Medicare and Medicaid Services 2016).

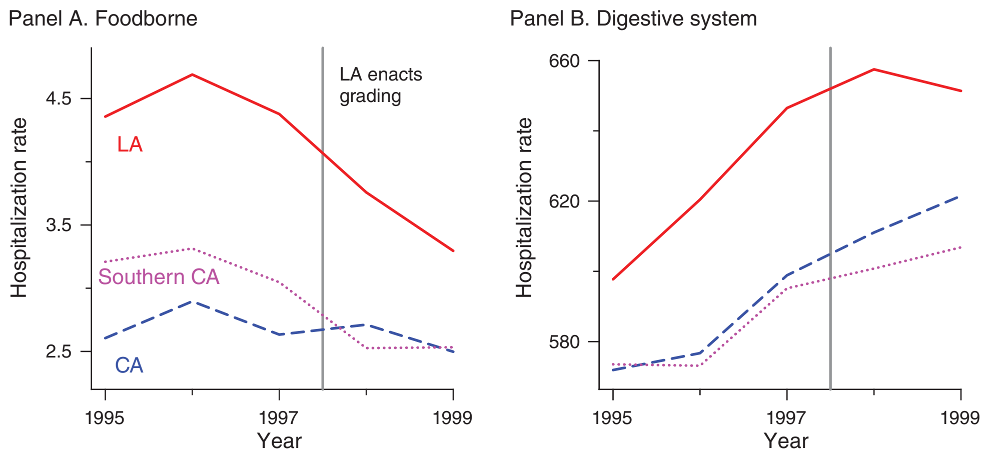

Figure 8. Trends in LA (Solid), the Rest of CA (Dashed), and Southern CA (Dotted) in Hospitalizations for Foodborne and (Non-foodborne) Digestive System Disorders from 1995–1999

Note: Digestive disorders trend in the opposite direction of foodborne illnesses for all three geographic groups.

variability than digestive disorder hospitalizations, both because foodborne illnesses are outbreak-driven and the digestive category includes a much higher volume of cases, adding stability to the trends. Pooling time fixed effects across the two categories hence underestimates the variance of the effect on foodborne hospitalizations.[43] Table 4 shows that the standard errors of the treatment effects uniformly inflate compared to the JL specification when adding in the interaction term to decouple the temporal variances of foodborne and digestive disorders. In model 6, the standard error of the mandatory grading effect nearly doubles. Monte Carlo simulations in online Appendix G reveal that the downward bias in the standard errors, coupled with the upward bias in the absolute value of the point estimates, results in Type I error of 22 percent in the shortest observation window and up to 51 percent with a longer observation window.

B. *CA*

Similar patterns emerge when we improve CA as a control group by reducing the influence of salmonella in the aggregate foodborne illness trends. We do so by including a highly prevalent and widely studied foodborne illness, campylobacter, to the set of foodborne illnesses. Campylobacter is universally recognized in the medical literature as a dominant source of bacterial foodborne illness in the United States (Mead et al. 1999, Scallan et al. 2011). The authoritative synthesis on foodborne illness in the United States describes campylobacter and salmonella as the

[43] More formally, the within-cluster covariance of foodborne and digestive ZIP clusters in the geographic control group are given equal weight by the ZIP-and-time-demeaned treatment indicator in the meat of the sandwich matrix $\sum_{g=1}^{G} \mathbf{X}_g \hat{\mathbf{u}}_g \hat{\mathbf{u}}_g^T \mathbf{X}_g^T$, where G is the number of ZIP-hospitalization type clusters, N_g is the number of time points per cluster, $\mathbf{X}_g$ is an $N_g \times 2$ matrix of demeaned regressors for cluster g representing parameters β_1 and γ_1 in equation (6), and $\hat{\mathbf{u}}_g$ is an $N_g \times 1$ vector of OLS residuals for cluster g from equation (6).

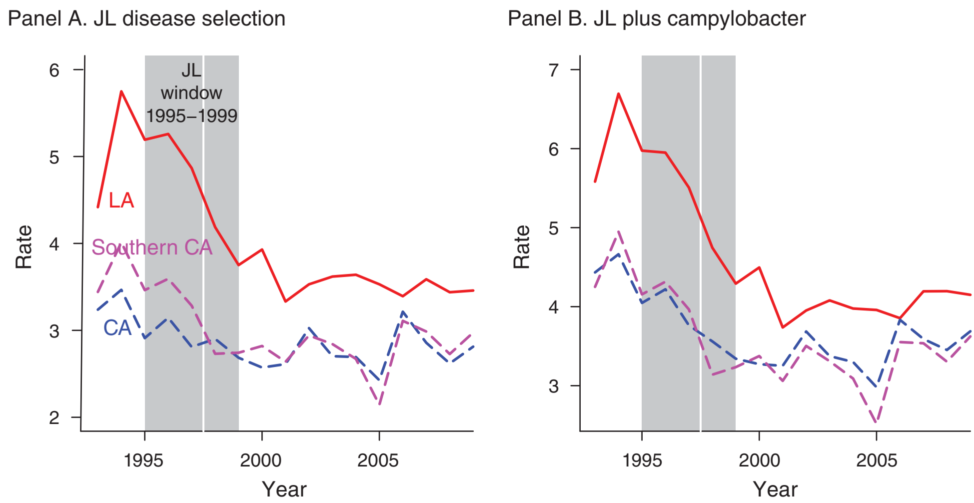

Figure 9. Foodborne Hospitalizations per 100,000 People Comparing JL's Disease Selection with and without Campylobacter in LA, Southern CA, and CA excluding LA (CA)

Notes: When campylobacter is added into the disease selection, pretreatment trends become more parallel between LA and CA. Trends remain parallel between LA and Southern CA regardless of disease selection.

"leading causes of foodborne illnesses" in the United States (Scallan et al. 2011, 13). As Table 1 shows, campylobacter and salmonella are responsible for 80 percent of foodborne hospitalizations and 90 percent of foodborne illnesses. In many years, cases of campylobacter outstrip cases of salmonella, leading one review to conclude that campylobacter "is well recognized as the leading cause of bacterial foodborne diarrheal disease worldwide" (Silva et al. 2011).[44] The transmission of campylobacter is linked to poultry (similar to salmonella), unpasteurized milk, and cheese (Bryan and Doyle 1995, USDA 2012), and restaurant food preparation practices affect the risk of campylobacter infection (Friedman et al. 2004, Jones et al. 2016).[45]

Figure 9 shows that when adding campylobacter to the aggregate foodborne disease selection, the pretreatment trends between LA and the rest of CA become *more* parallel.[46] Lead (difference-in-differences) tests between LA and CA before

[44] Campylobacter was dubbed the "the King of Foodborne Disease in the US" by Marler Clark, the nation's leading food safety law firm, which represented plaintiffs in the Jack in the Box E. Coli case. See http://www.foodpoisonjournal.com/food-poisoning-information/campylobacter-the-king-of-foodborne-disease-in-the-us.

[45] JL appears to have excluded campylobacter because of an inclusion threshold that diseases must be 90 percent foodborne. However, this omission is inconsistent with JL's selection of other diseases. Roughly 80 percent of campylobacter cases are foodborne (Mead et al. 1999), though sources disagree on the precise percentage, but JL appears to select multiple diseases beneath the 90 percent threshold. For instance, all forms of E. Coli appear to be included, but the same public health source on which JL relies indicates that none of the three E. Coli serotypes would meet the 90 percent threshold. Online Appendix D goes into more detail about foodborne disease selection, informed by extensive research and conferrals with a King County epidemiologist and a representative from the Centers for Disease Control and Prevention.

[46] Since the modeled outcome is the aggregated foodborne illness counts, not campylobacter alone, the concern with parallel trends must be in the aggregate counts. More generally, online Appendix D documents that the wide-ranging selection of 24 diseases would make it challenging to model disease specific trends and effects. Figure 9 in online Appendix D demonstrates that there is substantial evidence that trends are not parallel between LA and CA for other specific discharge codes (e.g., food poisoning (unspecified) and cysticercosis), with evidence

Table 5—Bias from Omitting Two-Way Interaction Terms with CA as the Control Group, with and without Campylobacter in the Disease Selection

	JL disease selection				Campylobacter added			
	1995–1999		1993–2009		1995–1999		1993–2009	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Foodborne × LA mandatory disclosure post-1998	−0.31	−0.20	−0.33	−0.13	−0.30	−0.10	−0.36	−0.08
	(0.07)	(0.08)	(0.05)	(0.06)	(0.07)	(0.08)	(0.05)	(0.06)
Foodborne × LA voluntary disclosure post-1998	−0.27	−0.15	−0.30	−0.09	−0.33	−0.11	−0.36	−0.06
	(0.08)	(0.09)	(0.08)	(0.10)	(0.10)	(0.10)	(0.09)	(0.10)
Foodborne × CA post-1998		−0.10		−0.18		−0.19		−0.26
		(0.04)		(0.04)		(0.04)		(0.04)
LA mandatory disclosure post-1998 (digestive)	0.04	−0.01	0.05	−0.05	0.10	0.00	0.09	−0.05
	(0.03)	(0.02)	(0.03)	(0.04)	(0.03)	(0.02)	(0.03)	(0.04)
LA voluntary disclosure post-1998 (digestive)	0.08	0.02	0.09	−0.01	0.12	0.01	0.14	−0.01
	(0.04)	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)
Model	JL	DDD	JL	DDD	JL	DDD	JL	DDD
R^2	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
Observations	2,280	2,280	7,752	7,752	2,280	2,280	7,752	7,752
F-ratio		8.71		82.13		33.93		177.14

Notes: Coefficients shown with standard errors, clustered by three-digit ZIP and illness-type combinations, are in parentheses. Each model is estimated with fixed effects for three-digit ZIP and illness-type combinations and year-quarters. We present results for the original observation period (1995–1999) as well as an expanded observation period (1993–2009).

and after 1994 in the pretreatment period 1993–1997 confirm this visual finding ($p < 0.05$ without campylobacter, $p = 0.87$ with campylobacter). As a result of conforming CA's pretreatment foodborne illness trends to LA's trends, holding the digestive disorder trends constant, bias from the missing interaction term increases. Table 5 compares the original JL disease selection (left) with the enhanced disease selection including campylobacter (right), using CA as a control group. Even in the shortest time span with JL's disease selection as shown in model 1, omitting the interaction term leads to poor model fit ($F = 8.71, p < 0.01$). However, after adding campylobacter aligns CA's pretreatment foodborne illness trends closer to LA's in model 5, bias from omitting the interaction term substantially increases ($F = 33.93, p < 0.001$).[47] With the longest time span shown in model 7, the bias is even larger ($F = 177.14, p < 0.001$). The addition of the two-way interaction term brings all grading effects to null in models with campylobacter included.[48]

for year- and disease-specific shocks. One plausible rationale for the specification is that aggregating these diseases can average out small shocks to potentially make the pretreatment time series in the aggregate credible. But when a large outbreak occurs for the dominant disease category like salmonella, this approach may still fall short, as this paper demonstrates.

[47] This is likely the reason why JL reports finding similar results with a sensitivity check for diagnoses that are over 50 percent foodborne, which would have included campylobacter.

[48] A later version of the work by the same authors, along with several authors trained in public health, makes several changes to the analysis and includes campylobacter (Simon et al. 2005). Online Appendix A shows that, in addition to failing to account for the salmonella outbreak, this version reports statistically significant grading effects because of the omission of the two-way interaction term.

VI. Conclusion

We conclude with several implications. First, our findings show that there is little credible evidence for the impact of restaurant grading—what many regard as the archetypal disclosure—on foodborne illness. Although we do not address JL's findings on consumer behavior and restaurant hygiene scores, our analysis suggests that much work remains to be done to empirically ground our understanding of information disclosure. The fact that it has taken nearly 15 years for anyone to reexamine the foodborne illness finding stems in part from data limitations in public health. Hospitalization data, even if anonymized, remain difficult to access. Reporting of foodborne illness and food safety enforcement is highly decentralized in the US system, with limited efforts at unifying surveillance data. This system is not only problematic for enforcement, but also poses serious impediments to developing a rigorous evidence base for health interventions and the economic understanding of disclosure regimes.

Second, we note a sharp divergence between the popularization of JL and its actual findings. In policy circles, the LA finding, popularized by the synthesis of Fung, Graham, and Weil (2007), has inspired calls for simplified, mandatory disclosure like letter grading across fields. Yet JL originally found no statistically significant differences between voluntary and mandatory restaurant grading. Although the study carefully couched its findings as being uninformative about informational unraveling, nearly every jurisdiction contemplating grading has applied these findings to support mandatory grading, even though the evidence at the time may have also supported informational unraveling.[49]

Third, our study illustrates the need to go beyond direct replication of observational studies in economics. While many have discussed the challenges of direct replication (McCullough and Vinod 2003, Camerer et al. 2016), our results show that the validity of an inference may hinge on more than merely replicating a regression (Rosenbaum 2001). Our deeper reassessment required expanding the data to uncover the salmonella outbreak, implementing a full triple differences specification, studying the power of the design, and engaging with medical science and epidemiology to assess robustness to disease selection and outbreaks. We fear that the incentives for researchers to engage in such resource-intensive efforts may be lacking. (We did so only after having spent three years working on food safety, assuming the validity of the LA hospitalization effect.)

Last, while the quantitative examination of public health initiatives is a sign of the positive and growing influence of econometric policy evaluation, our analysis shows the need for cross-fertilization. JL's study could have benefited from substantive insights from public health: that reported illnesses are a richer source of foodborne illness information than hospitalization data; that two pathogens are responsible for the bulk of foodborne illnesses; and that acute outbreaks make

[49] This confusion persists in the scholarly literature. Loewenstein, Sunstein, and Golman (2014), for instance, describes JL as showing that "mandatory disclosure was more effective than voluntary disclosure." Benoît and Dubra (2006, 182) infer that "unraveling occurred in the disclosure of grades obtained by restaurants for their hygiene." But see Jin and Leslie (2003, 425–26) ("caution should be used when interpreting the similarity between the effects of mandatory and voluntary disclosure.").

evaluation of food safety interventions challenging. By the same token, rapid developments in econometrics have much to add to public health policy and the study of information disclosure. Focused attention to research design (see, e.g., Angrist and Pischke 2010), improved approaches for intra-cluster dependence (e.g., Cameron, Gelbach, and Miller 2008), and synthetic control methods (e.g., Abadie, Diamond, and Hainmueller 2010) are considerably more developed in the economic literature. These methodological advancements can help researchers determine whether it is possible to evaluate a policy of limited scope, such as restaurant grading, using an outcome as stochastic as foodborne hospitalizations. Yet much work remains to be done to complete that revolution in credibility.

REFERENCES

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2010. "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program." *Journal of the American Statistical Association* 105 (490): 493–505.

Acs, Gregory, and Sandi Nelson. 2004. "Changes in Living Arrangements during the Late 1990s: Do Welfare Policies Matter?" *Journal of Policy Analysis and Management* 23 (2): 273–90.

Angrist, Joshua D., and Jörn-Steffen Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press.

Angrist, Joshua D., and Jörn-Steffen Pischke. 2010. "The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con Out of Econometrics." *Journal of Economic Perspectives* 24 (2): 3–30.

Ayres, Ian, Sophie Raseman, and Alice Shih. 2013. "Evidence from Two Large Field Experiments That Peer Comparison Feedback Can Reduce Residential Energy Usage." *Journal of Law, Economics, and Organization* 29 (5): 992–1022.

Belshé, S. Kimberley, Diana M. Bontá, Mark Starr, Duc Vugia, and Stanley R. Bissell. 2003. *Communicable Diseases in California, 1999–2000*. Sacramento: State of California Department of Health Services.

Benoît, Jean-Pierre, and Juan Dubra. 2006. "Information Revelation in Auctions." *Games and Economic Behavior* 57 (2): 181–205.

Ben-Shahar, Omri, and Carl E. Schneider. 2011. "The Failure of Mandated Disclosure." *University of Pennsylvania Law Review* 159: 647–749.

Ben-Shahar, Omri, and Carl E. Schneider. 2014a. "The Futility of Cost-Benefit Analysis in Financial Disclosure Regulation." *Journal of Legal Studies* 43 (S2): S253–71.

Ben-Shahar, Omri, and Carl E. Schneider. 2014b. *More Than You Wanted to Know: The Failure of Mandated Disclosure*. Princeton: Princeton University Press.

Boehmer, Tegan K., Jennifer L. Patnaik, Steven J. Burnite, Tista S. Ghosh, Ken Gershman, and Richard L. Vogt. 2011. "Use of Hospital Discharge Data to Evaluate Notifiable Disease Reporting to Colorado's Electronic Disease Reporting System." *Public Health Reports* 126 (1): 100–106.

Breitmeyer, R.E. 1997. "Proceedings from the Third Annual Federal/State Conference on Food Safety November 20–21, 1997: California Partnerships—California Egg Quality Assurance Program." https://www.fsis.usda.gov/OPPDE/animalprod/Proceedings/California%20Partnerships.html.

Bryan, Frank L., and Michael P. Doyle. 1995. "Health Risks and Consequences of Salmonella and Campylobacter jejuni in Raw Poultry." *Journal of Food Protection* 58 (3): 326–44.

Bubb, Ryan. 2015. "TMI? Why the Optimal Architecture of Disclosure Remains TBD." *Michigan Law Review* 113 (6): 1021–42.

Camerer, Colin F., Anna Dreber, Eskil Forsell, Teck-Hua Ho, Jürgen Huber, Magnus Johannesson, Michael Kirchler, et al. 2016. "Evaluating Replicability of Laboratory Experiments in Economics." *Science* 351 (6280): 1433–36.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller. 2008. "Bootstrap-Based Improvements for Inference with Clustered Errors." *Review of Economics and Statistics* 90 (3): 414–27.

Centers for Medicare and Medicaid Services. 2016. "Draft ICD-10-CM/PCS MS- DRGv28 Definitions Manual, MDC 6 Assignment of Diagnosis Codes." US Department of Health and Human Services. https://www.cms.gov/icd10manual/fullcode_cms/P0148.html (accessed September 17, 2019).

de Carvalho Filho, Irineu Evangelista. 2008. "Old-Age Benefits and Retirement Decisions of Rural Elderly in Brazil." *Journal of Development Economics* 86 (1): 129–46.

Dodd, Christine E.R., Tim Aldsworth, Richard A. Stein, Dean O. Cliver, and Hans P. Riemann, eds. 2017. *Foodborne Diseases*. 3rd ed. Amsterdam: Elsevier.

Dranove, David, and Ginger Zhe Jin. 2010. "Quality Disclosure and Certification: Theory and Practice." *Journal of Economic Literature* 48 (4): 935–63.

Dranove, David, Daniel Kessler, Mark McClellan, and Mark Satterthwaite. 2003. "Is More Information Better? The Effects of 'Report Cards' on Health Care Providers." *Journal of Political Economy* 111 (3): 555–88.

Feng Lu, Susan. 2012. "Multitasking, Information Disclosure, and Product Quality: Evidence from Nursing Homes." *Journal of Economics and Management Strategy* 21 (3): 673–705.

Figlio, David N. 2006. "Testing, Crime and Punishment." *Journal of Public Economics* 90 (4–5): 837–51.

Figlio, David N., and Maurice E. Lucas. 2004. "What's in a Grade? School Report Cards and the Housing Market." *American Economic Review* 94 (3): 591–604.

Friedman, Cindy R., Robert M. Hoekstra, Michael Samuel, Ruthanne Marcus, Jeffrey Bender, Beletshachew Shiferaw, Sudha Reddy, et al. 2004. "Risk Factors for Sporadic Campylobacter Infection in the United States: A Case-Control Study in FoodNet Sites." *Clinical Infectious Diseases* 38 (S3): S285–96.

Fung, Archon, Mary Graham, and David Weil. 2007. *Full Disclosure: The Perils and Promise of Transparency*. Cambridge, UK: Cambridge University Press.

Greene, William H. 2012. *Econometric Analysis*. 7th ed. Upper Saddle River, NJ: Prentice Hall.

Gruber, Jonathan. 1994. "The Incidence of Mandated Maternity Benefits." *American Economic Review* 84 (3): 622–41.

Gruber, Jonathan, and James Poterba. 1994. "Tax Incentives and the Decision to Purchase Health Insurance: Evidence from the Self-employed." *Quarterly Journal of Economics* 109 (3): 701–33.

Hastings, John, Shu Sebesta, Terry Thompson, and John Williamson. 1991. *California Morbidity 1990*. California: US Department of Health Services.

Ho, Daniel E. 2012. "Fudging the Nudge: Information Disclosure and Restaurant Grading." *Yale Law Journal* 122 (3): 574–688.

Ho, Daniel E., Zoe C. Ashwood, and Cassandra Handan-Nader. 2019. "New Evidence on Information Disclosure through Restaurant Hygiene Grading: Dataset." *American Economic Journal: Economic Policy*. https://doi.org/10.1257/pol.20180230.

Imbens, Guido W., and Jeffrey Wooldridge. 2007. "What's New in Econometrics? Differences-in-Differences Estimation." NBER Summer Institute Lecture Notes. https://www.nber.org/WNE/lect_10_diffindiffs.pdf.

Imbens, Guido W., and Jeffrey Wooldridge. 2009. "Recent Developments in the Econometrics of Program Evaluation." *Journal of Economic Literature* 47 (1): 5–86.

Institute of Medicine. 2006. *Addressing Foodborne Threats to Health: Policies, Practices, and Global Coordination, Workshop Summary*. Washington, DC: National Academies Press.

Jackson, Brendan R., Patricia M. Griffin, Dana Cole, Kelly A. Walsh, and Shua J. Chai. 2013. "Outbreak-Associated Salmonella enterica serotypes and Food Commodities, United States, 1998–2008." *Emerging Infectious Diseases* 19 (8): 1239–44.

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

Jin, Ginger Zhe, and Phillip Leslie. 2009. "Reputational Incentives for Restaurant Hygiene." *American Economic Journal: Microeconomics* 1 (1): 237–67.

Jones, Anna K., Dan Rigby, Michael Burton, Caroline Millman, Nicola J. Williams, Trevor R. Jones, Paul Wigley, et al. 2016. "Restaurant Cooking Trends and Increased Risk for Campylobacter Infection." *Emerging Infectious Diseases* 22 (7): 1208–15.

Khandker, Shahidur R., Gayatri B. Koolwal, and Hussain A. Samad. 2010. *Handbook on Impact Evaluation: Quantitative Methods and Practices*. Washington, DC: World Bank Publications.

Kim-Farley, Robert. 2000. "Regulations for Communicable Disease Reporting." County of Los Angeles Department of Public Health B-82.

Kinde, H., D.H. Read, R.P. Chin, A.A. Bickford, R.L. Walker, A. Ardans, R.E. Breitmeyer, et al. 1996. "Salmonella enteritidis, Phage Type 4 Infection in a Commercial Layer Flock in Southern California: Bacteriologic and Epidemiologic Findings." *Avian Diseases* 40 (3): 665–71.

Kizer, Kenneth W. 1994. "Vibrio Vulnificus Hazard in Patients with Liver Disease." *Western Journal of Medicine* 161 (1): 64–65.

Loewenstein, George, Cass R. Sunstein, and Russell Golman. 2014. "Disclosure: Psychology Changes Everything." *Annual Review of Economics* 6 (1): 391–419.

Marotta-Wurgler, Florencia. 2012. "Does Contract Disclosure Matter?" *Journal of Institutional and Theoretical Economics* 168 (1): 94–119.

McCullough, B.D., and H.D. Vinod. 2003. "Verifying the Solution from a Nonlinear Solver: A Case Study." *American Economic Review* 93 (3): 873–92.

Mead, P.S., L. Slutsker, V. Dietz, L.F. McCaig, J.S. Bresee, C. Shapiro, P.M. Griffin, and R.V. Tauxe. 1999. "Food-Related Illness and Death in the United States." *Emerging Infectious Diseases* 5 (5): 607–25.

Mumma, Gerald A., Patricia M. Griffin, Martin I. Meltzer, Chris R. Braden, and Robert V. Tauxe. 2004. "Egg Quality Assurance Programs and Egg-Associated Salmonella Enteritidis Infections, United States." *Emerging Infectious Diseases* 10 (10): 1782–89.

Passaro, Douglas J., Roshan Reporter, Laurene Mascola, Liga Kilman, George B. Malcolm, Henry Rolka, S. Benson Werner, and Duc J. Vugia. 1996. "Epidemic Salmonella enteritidis Infection in Los Angeles County, California. The Predominance of Phage Type 4." *Western Journal of Medicine* 165 (3): 126–30.

Riverside County Health Department, California State Department of Public Health Berkeley, Center for Disease Control Atlanta, and National Center for Urban and Industrial Health Cincinnati. 1971. "A Waterborne Epidemic of Salmonellosis in Riverside, California, 1965. Epidemiologic Aspects. A Collaborative Report." *American Journal of Epidemiology* 93 (1): 33–48.

Roberts, Tanya, Helen Jensen, and Laurian Unnevehr. 1995. *Tracking Foodborne Pathogens from Farm to Table*. Washington, DC: US Department of Agriculture.

Robertson, Christopher T. 2015. "New DTCA Guidance—Enough to Empower Consumers?" *New England Journal of Medicine* 373 (12): 1085–87.

Rosenbaum, Paul R. 2001. "Replicating Effects and Biases." *American Statistician* 55 (3): 223–27.

Scallan, Elaine, Robert M. Hoekstra, Frederick J. Angulo, Robert V. Tauxe, Marc-Alain Widdowson, Sharon L. Roy, Jeffery L. Jones, and Patricia M. Griffin. 2011. "Foodborne Illness Acquired in the United States—Major Pathogens." *Emerging Infectious Diseases* 17 (1): 7–15.

Schweitzer, Glenn, Mohammad Reza Zali, and George Jackson, eds. 2006. *Food Safety and Foodborne Disease Surveillance Systems: Proceedings of an Iranian-American Workshop*. Washington, DC: National Academies Press.

Seiver, Owen H., and Thomas H. Hatfield. 2000. "Grading Systems for Retail Food Facilities: A Risk-Based Analysis." *Journal of Environmental Health* 63 (3): 22–27.

Silva, Joana, Daniela Leite, Mariana Fernandes, Cristina Mena, Paul Anthony Gibbs, and Paula Teixeira. 2011. "Campylobacter spp. as a Foodborne Pathogen: A Review." *Frontiers in Microbiology* 2: 200.

Simon, Paul A., Phillip Leslie, Grace Run, Ginger Zhe Jin, Roshan Reporter, Arturo Aguirre, and Jonathan E. Fielding. 2005. "Impact of Restaurant Hygiene Grade Cards on Foodborne-Disease Hospitalizations in Los Angeles County." *Journal of Environmental Health* 67 (7): 32–36.

Swaminathan, Bala, Timothy J. Barrett, and Patricia Fields. 2006. "Surveillance for Human Salmonella Infections in the United States." *Journal of AOAC International* 89 (2): 553–59.

United States Department of Agriculture (USDA). 2012. "Campylobacter Questions and Answers." USDA. https://www.fsis.usda.gov/wps/portal/fsis/topics/food-safety-education/get-answers/food-safety-fact-sheets/foodborne-illness-and-disease/campylobacter-questions-and-answers/ct_index. (accessed September 17, 2019).

Weil, David, Archon Fung, Mary Graham, and Elena Fagotto. 2006. "The Effectiveness of Regulatory Disclosure Policies." *Journal of Policy Analysis and Management* 25 (1): 155–81.

Werner, Rachel M., Edward C. Norton, R. Tamara Konetzka, and Daniel Polsky. 2012. "Do Consumers Respond to Publicly Reported Quality Information? Evidence from Nursing Homes." *Journal of Health Economics* 31 (1): 50–61.

Wiant, Chris J. 1999. "Scores, Grades, and Communicating about Food Safety." *Journal of Environmental Health* 61 (9): 37–39.

Wing, Coady, Kosali Simon, and Ricardo A. Bello-Gomez. 2018. "Designing Difference in Difference Studies: Best Practices for Public Health Policy Research." *Annual Review of Public Health* 39 (1): 453–69.

Winston, Clifford. 2008. "The Efficacy of Information Policy: A Review of Archon Fung, Mary Graham, and David Weil's Full Disclosure: The Perils and Promise of Transparency." *Journal of Economic Literature* 46 (3): 704–17.