SIEPR | STANFORD INSTITUTE FOR ECONOMIC POLICY RESEARCH

DID RESTAURANT HYGIENE GRADING IN LOS ANGELES IMMEDIATELY REDUCE FOODBORNE ILLNESS BY 20% ACROSS ALL OF SOUTHERN CALIFORNIA? A RESPONSE TO JIN & LESLIE

Daniel E. Ho
Stanford University

Cassandra Handan-Nader
Stanford University

November, 2019

Working Paper No. 19-030

Did Restaurant Hygiene Grading in Los Angeles Immediately Reduce Foodborne Illness by 20% Across All of Southern California? A Response to Jin & Leslie

Cassandra Handan-Nader* Daniel E. Ho†

November 12, 2019

We appreciate the engagement by Jin and Leslie (2019) (J&L) with our study (Ho, Ashwood, and Handan-Nader, 2019) (HAH). Understanding the causal effect of restaurant grading on health outcomes remains a pivotal case in scholarship and policy debate about information disclosure (see, e.g., Ben-Shahar and Schneider, 2011, 2014; Fung et al., 2007; Loewenstein et al., 2014; Weil et al., 2006), and we are grateful for their attention to this issue.

Our paper made several advances in understanding the evidence around Los Angeles's sanitation grading policy. First, we collected substantially more data than originally analyzed in Jin and Leslie (2003). We expanded hospitalizations fivefold, increasing the original observation period of 1995-99 to 1983-2000. We also hand collected mandatory reports of illnesses from 1990-2015 and specifically salmonella from 1964-2015. Collecting such data is critical to assessing the parallel trends assumption that looms large in a difference-in-difference (or triple differences) design (Imbens and Wooldridge, 2009). Second, this new data and longer observation window revealed that Southern California experienced the largest salmonella outbreak in state history in the years immediately prior to J&L's observation period. The outbreak poses significant challenges to the original research design because Jin and Leslie (2003) compared Los Angeles (LA) to the rest of California (CA) before and after LA adopted grading in 1998, when Northern CA was unaffected

*Ph.D. Student, Stanford University, 559 Nathan Abbott Way, Stanford, CA 94305; Email: slnader@stanford.edu.
†William Benjamin Scott and Luna M. Scott Professor of Law; Professor of Political Science; Senior Fellow at Stanford Institute for Economic Policy Research; Stanford University, 559 Nathan Abbott Way, Stanford, CA 94305; Tel: 650-723-9560; Fax: 650-725-0253; Email: dho@law.stanford.edu.

by the outbreak (Passaro et al., 1996). Applying the same original model to Southern CA (but excluding LA) hence estimates the same "treatment effects" for restaurant grading, even though none of these Southern CA counties changed any disclosure policies during the relevant observation window. Third, we showed that mitigating the role of the salmonella outbreak – by using Southern CA as a control group or including the most important foodborne pathogen next to salmonella (campylobacter) – revealed no evidence of the health effects of LA's grading system. These results undercut the influential finding that LA's grading system caused a 20% reduction in foodborne hospitalizations.

Jin and Leslie's response underscores the importance of getting this finding right, and we welcome the opportunity for this exchange. At the outset, we would like to highlight several points of agreement, which illustrate the progress made in understanding this important case. First, as HAH discussed, the original design in Jin and Leslie (2003) was neither a difference-in-differences (DID) nor a triple differences design, making a restrictive assumption that foodborne and non-foodborne (digestive system) hospitalizations had the same time trend, but only in CA not LA. This is important because time trends in fact diverge sharply during the 1995-99 observation period. Monte Carlo simulations showed that the restrictive assumption increases Type I error by 22-51%, depending on the sample, at $\alpha = 0.05$ (see HAH Appendix G). In response, J&L agrees "that our original assumption of common time fixed effects might be violated" and "acknowledge [our] point about the interaction term and believe results [including the interaction term] are more trustworthy." Second, J&L agrees "LA and the rest of Southern CA are similar in food-related hospitalizations even after LA adopted grade cards," which means applying the original model in a placebo test to Southern CA (excluding LA) generates the same treatment effects. Third, we agree with J&L's ultimate conclusion that isolating the causal effect of disclosure is empirically challenging. As J&L notes, "media attention, political pressure, the LA-specific assessment system, and the corresponding educational efforts" may each confound the LA design. This inferential challenge is precisely why we designed a randomized controlled trial with the Public Health – Seattle & King

County (Handan-Nader et al., 2018). As our article sought to show, there is now agreement that researchers and policymakers must be careful about understanding and interpreting the originally reported 20% reduction in foodborne hospitalizations.

We write to clarify three points in response to the new analyses reported by J&L. First, J&L's response insufficiently engages with the role of the Southern CA salmonella outbreak. We show that because their analyses are only conducted within the 1995-99 window, they are unable to account for the peak of the salmonella outbreak in 1994. Extending their analysis with only two more years of pre-treatment data confirms that there is no evidence that grading affected hospitalizations and invalidates CA (or Northern CA) as a control group. Second, we discuss the potential for spillover effects, which would also threaten the original design in Jin and Leslie (2003). The dramatic and nearly identical drops in hospitalizations across all of Southern CA, beginning from the salmonella crisis in 1994 through 2001, are inconsistent with spillover effects from a local, retail-level disclosure policy. Third, we consider evidence based on non-salmonella pathogens, which show substantial violations of parallel treatment trends and reveal no evidence of any decrease in hospitalizations or illnesses around 1998.

1 Salmonella Outbreak

While J&L reports many additional regression analyses, the response insufficiently addresses the most important public health factor: the largest recorded salmonella outbreak in state history occurring in Southern CA prior to when LA adopted grading. Evidence of the outbreak becomes clear when expanding the observation window just 2 years before the short period of 1995-99 for hospitalizations and examining mandated illness reports. Each of J&L's additional tests rely exclusively on the short observation window of 1995-99, which is unable to capture the peak of the outbreak in 1994, and some models continue to draw an inference omitting the interaction term.[1] While we are sympathetic to the challenges of gathering a long pre-treatment time series,

[1] Section 6 and Appendix F of HAH explains the methodological reasons why the Foodborne × post-1998 interaction term is required, and how its omission inflates the size of treatment coefficients while deflating their standard errors.

we do not think that considering only 2 more years of data imposes an undue burden or unrealistic expectations on researchers, as J&L's reply suggests. That data was available at the time J&L originally conducted the study and provides a straightforward check of a critical assumption in a difference-in-differences approach.

Consider J&L's analysis presented in J&L's Table 2. We note first that adding the interaction term cuts treatment effect point estimates in half, which suggests how important it is to account for the divergent trends in foodborne and digestive disorder diseases. J&L then fits separate DID regressions to foodborne and digestive disorder hospitalizations, which, as we originally pointed out, is exactly the kind of estimation strategy that combined leads to a triple difference (i.e., a difference in DIDs). Although the effect of voluntary placards is no longer statistically significant in the foodborne DID, J&L highlights that the treatment effect for mandatory placards is statistically significant. But this conclusion drawn from the short observation window misses the point about the salmonella outbreak: the dramatic outbreak will confound the foodborne DID model.

We replicate these results with our data in the top panel of Table 1. Curiously, J&L does not include the missing interaction term in the LA-only model, which compares foodborne to non-foodborne hospitalizations within ZIPs partially or fully in LA county. The last column shows that evidence of grading effects disappears when adding the necessary interaction term. To illustrate the role of the salmonella outbreak, the bottom panel of Table 1 conducts placebo tests, fitting the same models but substituting Southern CA (excluding LA) as the treatment group. Again, we find the same "treatment effects."

Notwithstanding the dramatic evidence of the salmonella outbreak (see Figures 2, 3, 4, 5, and 9 of HAH), J&L conducts tests for lead effects on the 1995-99 data and concludes that "no control group has a significantly different pre-treatment trend from foodborne hospitalizations in LA." Yet this is largely an artifact of the short observation window. To show this, we replicate the lead DID models and extend the beginning of the observation period to 1993.[2] For ease of comparison,

[2]This corresponds to the earliest period with no changes in hospitalization discharge codes.

	LA Treated					
	J&L (2003) Table VI	With interaction	Foodborne DID	Digestive DID	LA only	LA only w/interaction
LA mandatory	0.04^{*}	-0.01		0.00	0.25^{**}	0.05
	(0.03)	(0.02)		(0.02)	(0.10)	(0.03)
LA voluntary	0.08^{**}	0.02		0.00	0.27^{**}	0.05
	(0.04)	(0.03)		(0.03)	(0.11)	(0.04)
Foodborne × post-1998		-0.10^{**}				-0.39^{***}
		(0.04)				(0.13)
Foodborne × LA mandatory	-0.31^{***}	-0.20^{**}	-0.22^{***}		-0.31^{***}	0.10
	(0.07)	(0.08)	(0.08)		(0.07)	(0.13)
Foodborne × LA voluntary	-0.27^{***}	-0.15^{*}	-0.11		-0.27^{***}	0.18
	(0.08)	(0.09)	(0.09)		(0.08)	(0.16)
R^2	0.99	0.99	0.73	0.99	0.99	0.99
N	2,280	2,280	1,140	1,140	720	720

	Southern CA "Treated" (Placebo Test)					
	J&L (2003) Table VI	With interaction	Foodborne DID	Digestive DID	LA only	LA only w/interaction
S. Cal. × 1998	0.03	-0.01		0.01	0.17^{*}	-0.03
	(0.04)	(0.04)		(0.04)	(0.09)	(0.06)
S. Cal. × 1999	0.09^{*}	0.04		0.02	0.26^{**}	0.06
	(0.05)	(0.05)		(0.05)	(0.13)	(0.07)
Foodborne × post-1998		-0.09^{**}				-0.37^{**}
		(0.04)				(0.15)
Foodborne × S. Cal. × 1998	-0.32^{***}	-0.22^{**}	-0.26^{***}		-0.32^{***}	0.08
	(0.09)	(0.10)	(0.09)		(0.09)	(0.18)
Foodborne × S. Cal. × 1999	-0.31^{***}	-0.21^{**}	-0.15		-0.31^{***}	0.09
	(0.10)	(0.10)	(0.09)		(0.10)	(0.18)
R^2	0.99	0.99	0.73	0.99	0.99	0.99
N	2,280	2,280	1,140	1,140	800	800

Table 1: Replication of Table 2 in Jin and Leslie (2019), using LA as the treated area (top) and Southern CA excluding LA as the treated area (bottom). The variable LA represents the proportion of a three-digit ZIP code's population in LA, while the variable S. Cal. represents the proportion of a three-digit ZIP code's population in Southern CA excluding LA. Coefficients 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 add a sixth model to illustrate the necessity of adding in the missing interaction term Foodborne × post-1998 to the model comparing foodborne and non-foodborne illnesses within ZIPs fully or partially in LA (or Southern CA) only, which J&L omits from its version of Table 2. $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

	CA Control		S. CA Control		N. CA Control	
	J&L	HAH	J&L	HAH	J&L	HAH
1994 × treated (θ_1)	0.21^{***}	0.19^{**}	0.19^{**}	0.08	0.23^{***}	0.25^{**}
	(0.07)	(0.09)	(0.08)	(0.12)	(0.08)	(0.10)
1995 × treated (θ_2)	0.11	0.15^{*}	0.06	0.01	0.15^{*}	0.25^{***}
	(0.08)	(0.09)	(0.08)	(0.11)	(0.08)	(0.09)
1996 × treated (θ_3)	0.13^{*}	0.12	0.09	−0.01	0.17^{**}	0.21^{**}
	(0.07)	(0.09)	(0.08)	(0.12)	(0.08)	(0.10)
1997 × treated (θ_4)	0.09	0.17	0.02	−0.02	0.13	0.30^{**}
	(0.08)	(0.11)	(0.09)	(0.13)	(0.09)	(0.12)
1998 × treated (θ_5)	−0.12	−0.05	−0.11	0.04	−0.14	−0.08
	(0.09)	(0.11)	(0.10)	(0.13)	(0.09)	(0.11)
1999 × treated (θ_6)	−0.16	−0.05	-0.17^{*}	−0.05	−0.15	−0.02
	(0.10)	(0.11)	(0.10)	(0.14)	(0.10)	(0.11)
1994 × foodborne (θ_7)		0.04		0.15		−0.02
		(0.06)		(0.09)		(0.07)
1995 × foodborne (θ_8)		−0.06		0.08		-0.15^{**}
		(0.05)		(0.08)		(0.06)
1996 × foodborne (θ_9)		0.00		0.13		−0.07
		(0.06)		(0.09)		(0.06)
1997 × foodborne (θ_{10})		-0.13^{*}		0.05		-0.24^{***}
		(0.07)		(0.10)		(0.08)
1998 × foodborne (θ_{11})		-0.12^{**}		-0.21^{**}		−0.08
		(0.06)		(0.10)		(0.07)
1999 × foodborne (θ_{12})		-0.18^{***}		−0.17		-0.19^{***}
		(0.06)		(0.11)		(0.07)
F-test of H_0: $\{\theta_{1:4}\} = 0$	2.72	2.11	2.22	0.42	3.21	4.40
p-value of F-test	0.03	0.08	0.07	0.79	0.01	0.00
F-test of H_0: $\{\theta_{5:6}\} = 0$	3.20	0.27	3.06	0.51	3.00	0.63
p-value of F-test	0.04	0.76	0.05	0.60	0.05	0.54
F-test of H_0: $\{\theta_{7:12}\} = 0$		5.62		7.47		4.30
p-value of F-test		0.00		0.00		0.00
R^2	0.99	0.99	0.99	0.99	0.99	0.99
N	3,192	3,192	1,848	1,848	2,352	2,352

Table 2: Panel A of Table 3 of Jin and Leslie (2019) adding two more pre-treatment years to the observation window (1993 and 1994). Coefficients 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. Though J&L's version of the table omitted the two-way interactions between year and foodborne, F-tests restricting these parameters to zero strongly reject and we included them here as the "HAH" specification. The parameter 1994 × treated (θ_1) represents the peak of salmonella outbreak in LA, and is highly statistically significant in comparison to areas that did not experience the outbreak. CA and CA excluding Southern CA strongly reject F-tests for parallel pre-treatment trends with LA. Southern CA retains the F-test for both parallel treatment trends and no treatment effect in 1998-1999. *p<0.1; **p<0.05; ***p<0.01.

Panel B: food-only, "treated" = % of 3-digit ZIP in LA

	CA Control	S. CA Control	N. CA Control
1994 × treated (θ_1)	0.19**	0.08	0.25**
	(0.09)	(0.12)	(0.10)
1995 × treated (θ_2)	0.15*	0.01	0.25***
	(0.09)	(0.11)	(0.09)
1996 × treated (θ_3)	0.12	−0.01	0.21**
	(0.09)	(0.12)	(0.10)
1997 × treated (θ_4)	0.17	−0.02	0.30**
	(0.11)	(0.13)	(0.12)
1998 × treated (θ_5)	−0.05	0.04	−0.08
	(0.11)	(0.13)	(0.11)
1999 × treated (θ_6)	−0.05	−0.05	−0.02
	(0.11)	(0.14)	(0.11)
F-test of H$_0$: $\{\theta_{1:4}\} = 0$	1.12	0.22	2.31
p-value of F-test	0.35	0.92	0.06
F-test of H$_0$: $\{\theta_{5:6}\} = 0$	0.14	0.27	0.33
p-value of F-test	0.87	0.76	0.72
R^2	0.72	0.74	0.75
N	1,596	924	1,176

Panel C: digestive disorders only, "treated" = % of 3-digit ZIP in LA

	CA Control	S. CA Control	N. CA Control
1994 × treated (θ_1)	−0.01	0.00	−0.02
	(0.02)	(0.02)	(0.03)
1995 × treated (θ_2)	−0.03	−0.03*	−0.02
	(0.02)	(0.02)	(0.03)
1996 × treated (θ_3)	−0.05*	−0.06**	−0.04
	(0.03)	(0.03)	(0.03)
1997 × treated (θ_4)	−0.05*	−0.07	−0.04
	(0.03)	(0.05)	(0.03)
1998 × treated (θ_5)	−0.04	−0.05	−0.02
	(0.03)	(0.06)	(0.04)
1999 × treated (θ_6)	−0.06	−0.08	−0.04
	(0.04)	(0.07)	(0.04)
F-test of H$_0$: $\{\theta_{1:4}\} = 0$	2.94	3.79	2.18
p-value of F-test	0.02	0.00	0.07
F-test of H$_0$: $\{\theta_{5:6}\} = 0$	4.64	5.55	3.32
p-value of F-test	0.01	0.00	0.04
R^2	0.99	0.99	0.99
N	1,596	924	1,176

Table 3: Panels B and C of Table 3 of Jin and Leslie (2019) adding two more pre-treatment years to the observation window (1993 and 1994). Coefficients shown with standard errors, clustered by three-digit ZIPs, in parentheses. Each model is estimated with fixed effects for three-digit ZIPs and year-quarters. Because foodborne and non-foodborne hospitalizations are modeled in separate regressions, there is no need for additional interaction terms. As in Table 2, the parameter 1994 × treated (θ_1) in Panel B represents the peak of salmonella outbreak in LA, and is highly statistically significant in comparison to areas that did not experience the outbreak. Non-foodborne hospitalizations reject lead tests for parallel trends with every control group, suggesting they are not appropriate controls for foodborne hospitalizations. *p<0.1; **p<0.05; ***p<0.01.

we retain the same organization as that of J&L in Tables 2 and 3 presented here. We observe substantial evidence of lead effects. In Table 2, the F-test on lead effects rejects the null for all specifications except for when Southern CA counties are used as the comparison group.[3] Panel B of Table 3 shows that the evidence of lead effects is strongest when Northern CA is used as the control group, which makes sense as Northern CA's food system is quite distinct from that of Southern CA (Starrs and Goin, 2010, pp. 17-22).[4] Panel C shows that digestive system diseases are a poor additional comparison group, as F-tests also reject the null of no lead effects.

Extending the pre-treatment time series by only two years shows that using J&L's own proposed lead tests, the parallel trends assumption is not met with CA or Northern CA as the control group. Contrary to J&L's conclusion that all control groups are comparable to LA, the only plausible control group appears to be Southern CA.

2 Interpretation of Spillovers

J&L's analysis confirms that Southern CA's foodborne hospitalizations are indistinguishable from LA's. J&L argues that spillover effects from LA's restaurant grading may account for why Southern CA exhibits the same sharp drop.

We agree, of course, that spillover effects would violate the independence assumptions of a DID and triple differences design. Yet spillovers would equally invalidate the original design of Jin and Leslie (2003), biasing coefficient estimates and standard errors. The original design was oriented around comparing LA with neighboring counties — hence a treatment indicator that incorporates the percentage of the population within three-digit ZIP codes that resides in LA — and variation

[3] We note that, though J&L's version of the table omitted the two-way interactions between year and foodborne diseases, F-tests restricting these parameters to zero strongly reject ($p < 0.001$), showing that disease $\times$ year interactions are required for each of the individual lead effects to complete the triple differences specification. For completeness, we show lead tests with both specifications to illustrate the poor model fit of the J&L specification. Section 6 and Appendix F of HAH detail why these terms are necessary.

[4] Though the joint F-test for all lead coefficients does not reject for foodborne illnesses with the rest of CA as a control group, the individual lead coefficient on treated $\times$ 1994 (the year of the salmonella outbreak) is statistically significant at $p < 0.05$, while the treatment coefficients (treated $\times$ 1998 and treated $\times$ 1999) reject neither individually nor jointly. The insignificance of the F-test is an artifact of J&L's testing batches of lead and lag coefficients separately against a saturated model.

of municipal adoptions within LA. That design cannot be sustained if J&L posits that restaurant grading would affect all of Southern CA.

The key question is whether it more likely (a) that the largest recorded salmonella outbreak confounds original estimates or (b) that the protective effect of restaurant grading extended across 10 counties? There are strong reasons to believe the former. First, Southern CA spans over 55,000 square miles. This geography makes it implausible that nonindependence is driven by dining across Southern CA counties. As we noted, San Diego is roughly the same distance from LA as New York City is from Hartford. Second, while restaurant grading was a retail level initiative isolated to LA, the salmonella outbreak is well documented to have affected the *regional* food supply (Passaro et al., 1996). Sharp increases in salmonella (enteriditis) of 700% to 1,782% across five Southern CA counties were reported. The principal culprit appeared to be a single large egg ranch, producing roughly 4.5 million eggs in six months. The response was substantial reform at the food supply level: eggs were pasteurized at that ranch, CA expanded its Egg Quality Assurance Plan, and required salmonella testing, vaccination, and more stringent safety protocols. Third, the decrease in LA observed from 1995-1998 is merely a continuation of the same decrease from the peak of the salmonella crisis in 1994. This can be seen clearly in the left panel of Figure 1, the top left panel of Figure 2, and the second left panel of Figure 3. Fourth, if spillovers from LA's grading system affected the rest of Southern CA, we would expect the effect to be highest and most immediate in LA, and more gradually emerge across other Southern CA counties. The reason is that LA had over 20,000 restaurants establishments, each of which would be inspected 1-2 times a year (Satzman, 1997). Mechanistically, it takes time to conduct the inspection to placard establishments. What we observe instead is that grading effects are detected in J&L's specification within the first quarter of 1998 and at the same magnitude for all of Southern CA.

These reasons strongly suggest that whatever spillover effects might exist do not explain the dramatic declines in foodborne illness hospitalizations from 1994 to 2001. We next show that the non-salmonella diseases corroborate this explanation, with no evidence for a protective halo across

Southern CA.

3 Disease Discharge Codes

In our study, we highlighted the exclusion of campylobacter from J&L's disease codes. We showed that foodborne illness hospitalizations are overwhelmingly driven by salmonella and campylobacter (see HAH Table 1), and the latter, if anything, diminished more rapidly for CA than LA around 1998. In response, J&L cites a recent study that asserts that "campylobacter rarely causes outbreaks in restaurant settings" (Firestone and Hedberg, 2018). The public health literature refutes this claim. Friedman et al. (2004), for instance, uses FoodNet data to estimate that two largest population attributable fractions of campylobacter by far were poultry and non-poultry consumed in restaurants, explaining nearly half of laboratory confirmed instances of campylobacter. Raw CDC data show that restaurant settings were the predominant identifiable source of hospitalizations due to campylobacter outbreaks between 1998 and 2017.[5] Indeed, Firestone and Hedberg (2018) relies on a citation that discusses cross-contamination as the primary route of campylobacter transmission in domestic and catering kitchens, but makes no mention of the likelihood of campylobacter outbreaks originating from restaurants (Silva et al., 2011).

More directly related to the econometric specification, J&L argues that "campylobacter alone follows a quite different trend from the other non-food conditions." This claim is misguided in several respects.

First, hospitalizations for non-food digestive system disorders fundamentally follow different time trends from foodborne hospitalizations: they increase from 1995-1999, while foodborne hospitalizations decrease (see Figure 1). This divergence introduces bias when the Foodborne $\times$ post-1998 interaction term is not included. Control illnesses include a wide range of gastrointestical disorders, such as appendicitis, hemorrhoids, peptic ulcers, esophageal carcinoma, and Crohn's disease. Substantively, there is hence not much reason to believe these illnesses controls for anything

[5] 19% of hospitalizations were linked directly back to restaurants, followed by 18.6% categorized as "Other," and 18.4% with unidentifiable sources. 13.5% were identified as originating from private homes.

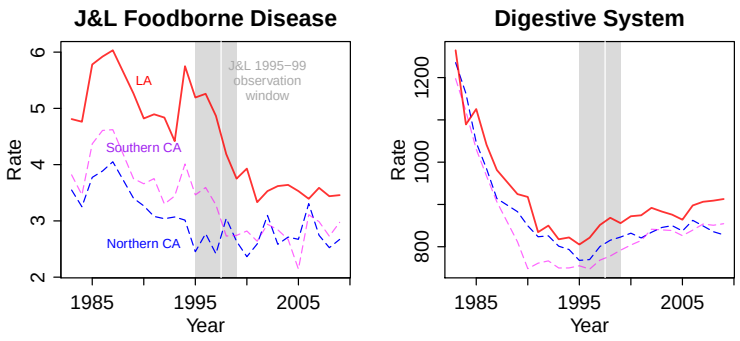

Figure 1: Trends in LA (solid red), Southern CA (dashed purple), and Northern CA (dashed blue) of hospitalizations for foodborne and (non-foodborne) digestive system disorders, in left and right panels respectively. Northern CA is shorthand for the rest of CA excluding Southern CA and LA. The time period under study by J&L (1995-99) is highlighted in grey, with the white vertical line representing the adoption of grading.

meaningful, such as food risk.

Second, the pathogen that appears to be an outlier is, if anything, salmonella. While J&L use lead tests with the short 1995-99 observation window (omitting the Foodborne $\times$ year interaction terms) to support parallel pre-treatment trends, an assessment of the violation of parallel trends is easily gleaned visually with a longer time series. Figure 2 displays time trends for primary foodborne hospitalizations, with LA in solid red lines, Southern CA in dashed purple lines, and Northern CA in dashed blue lines. The y-axes are fixed across panels to facilitate understanding the relative prevalence in the aggregated outcome data analyzed by J&L. In terms of violation of parallel trends, the most acute outlier around the observation window is salmonella due to the outbreak. Salmonella spiked in LA and Southern CA in 1994, but not Northern CA.

Figure 2 also shows that parallel trends may be violated for other pathogens. A single outbreak in 1985 caused unspecified food poisoning hospitalization rates to increase by 50% in a single year in LA. The only evidence of any drop in LA relative to CA around 1998 is in E. coli, but the pre-period evidence shows that LA and Northern CA have sharply different pretreatment trends.[6] More importantly, we observe no evidence of the 20% reduction across any of the diseases with comparable pre-treatment series.[7]

[6]Communicable disease reports in LA confirm that, if anything, LA was experience sharp *increases* in E. coli during this observation period. See, e.g., Haughton, 1998, p. 47 ("Annual incidence of E. coli O157:H7 has been steadily increasing since 1995.").

[7]One might argue that cysticercosis exhibits a drop, but cysticercosis also rises much more sharply for LA than

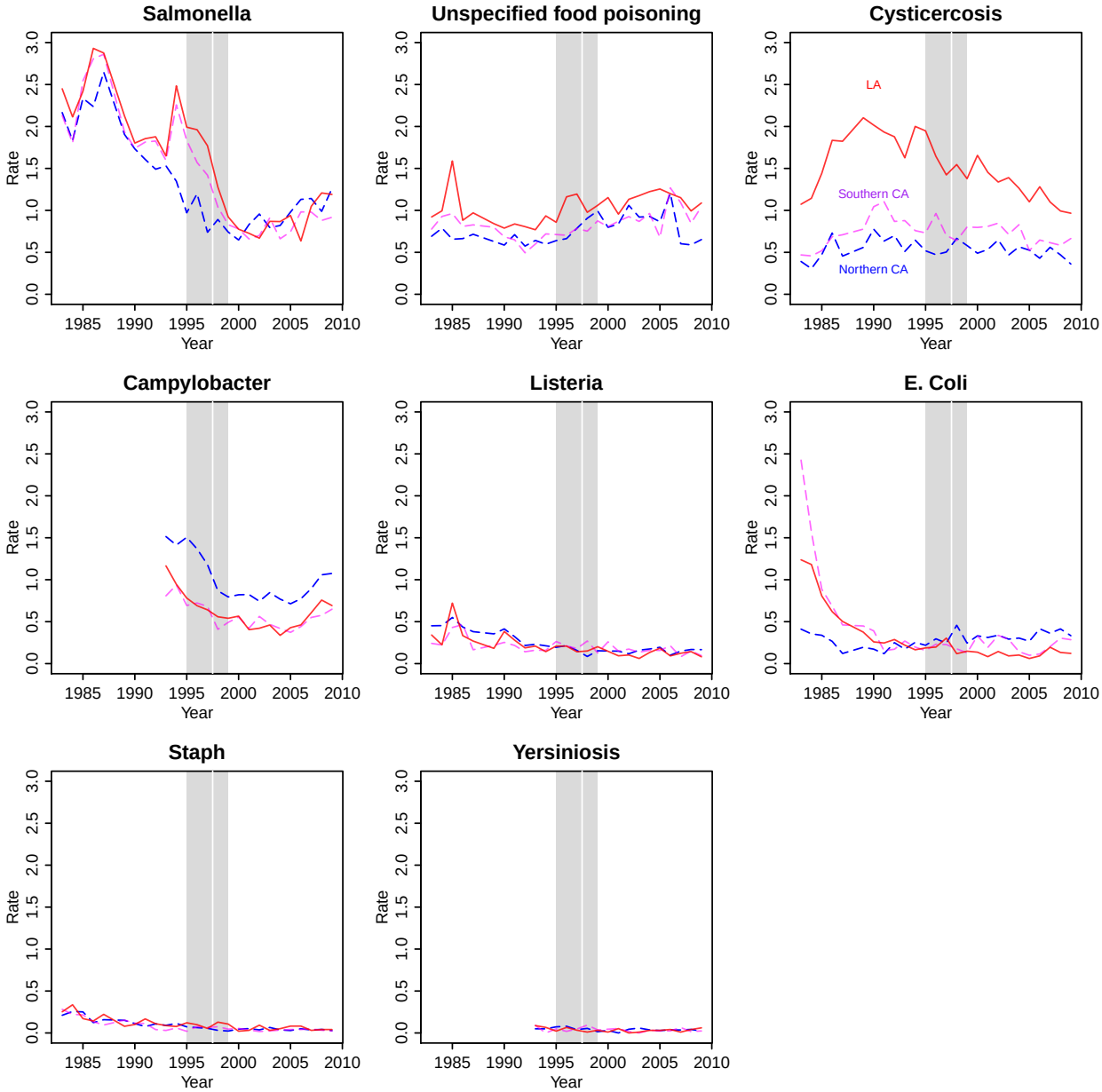

Figure 2: Hospitalization rates for the most prevalent foodborne pathogens from 1985-2009. Salmonella data show a clear violation of the parallel trends assumption when comparing LA to the rest of CA excluding Southern CA. There is no evidence of the 20% reduction across any disease with comparable pre-treatment trends.

Figure 3 plots mandatory reports of illnesses, a richer source of foodborne illness data, with y-axes floating in the left column and fixed in the right column. The fixed axes plots confirm the extant understanding in public health that campylobacter is the most important foodborne pathogen next to salmonella in prevalence (Mead et al., 1999; Scallan et al., 2011; Silva et al., 2011). Furthermore, we again observe no evidence of a disproportionate drop in foodborne illnesses around 1998 for any pathogen except for salmonella. This evidence across diseases corroborates the explanation that the salmonella outbreak – not spillover effects from LA restaurant grading to all of Southern CA – explains the effects identified by Jin and Leslie (2003).

Last, Table 4 uses the expanded observation period to fit the models presented in J&L's Table 5. For simplicity, and because only the major pathogens make a large difference in the aggregate foodborne hospitalization counts, we simply study the addition of campylobacter to J&L's original disease selection. Even if we exclude Southern CA as a possible control group, the results show that (a) there are significant lead effects when using CA or Northern CA as the control group, (b) lead effects are mitigated for the rest of CA by the inclusion of campylobacter, and (c) for models with the least evidence of lead effects (i.e., those with campylobacter included), treatment effects are statistically insignificant (see Panel C).

4 Conclusion

We reiterate our points of agreement with J&L. Southern CA and LA do not have distinguishable trends in foodborne illness before and after LA adopted restaurant grading. The open question is one of interpretation, namely whether this descriptive fact is because (a) LA's restaurant grading system protected all 10 Southern CA counties within the first quarter of implementation, or (b) the largest salmonella outbreak in state history led to sharp drops in salmonella in Southern CA,

elsewhere in the 1980s. As our research documented extensively, J&L's original disease coding was questionable. Cysticercosis is a parasitic pork tapeworm that is primarily travel-related, with sharply different pre-treatment trends between LA and CA, and so there is no reason to expect it to drive the restaurant grading effect. Of the top six diseases by incidence, J&L *excluded* campylobacter, botulism, and listeriosis, but *included* cysticercosis. In a later public health analysis of the LA effect, J&L included campylobacter and excluded cysticercosis. The agreement in disease selection across the two articles was 31%.

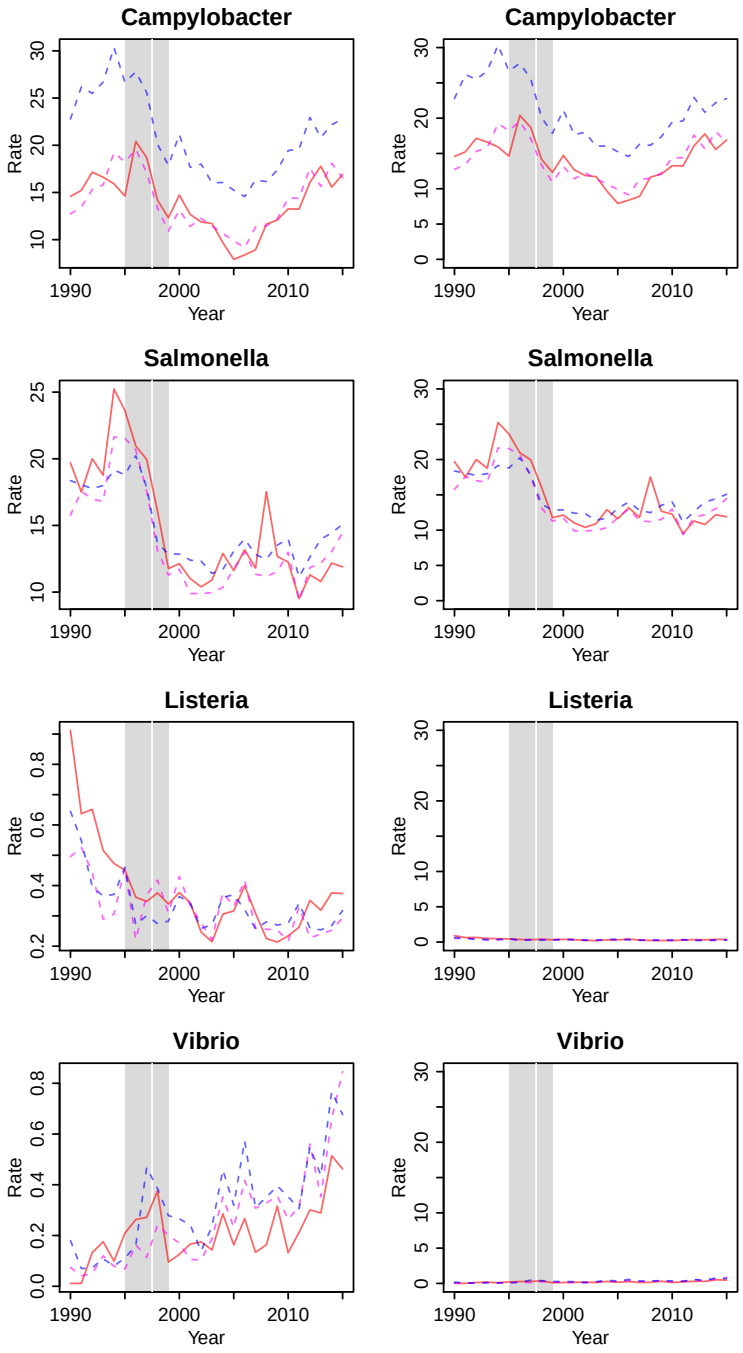

Figure 3: Illness rates for the most prevalent reportable foodborne pathogens from 1990-2009, with floating axes to facilitate visual inspection of parallel trends (left) and fixed axes to facilitate comparison of prevalence across pathogens (right). As with hospitalization rates, we observe no eviddence of a disproportionate decline in LA for any other pathogen except salmonella, which was affected by the outbreak across Southern CA counties in 1994.

Panel A: full sample, "treated" = food × % of 3-digit ZIP in LA

	J&L Disease Selection		J&L plus Campylobacter	
	CA Control	N. CA Control	CA Control	N. CA Control
1994 × treated (θ_1)	0.19^{**}	0.25^{**}	0.14	0.21^{**}
	(0.09)	(0.10)	(0.09)	(0.10)
1995 × treated (θ_2)	0.15^{*}	0.25^{***}	0.09	0.15^{*}
	(0.09)	(0.09)	(0.08)	(0.09)
1996 × treated (θ_3)	0.12	0.21^{**}	0.08	0.15
	(0.09)	(0.10)	(0.08)	(0.09)
1997 × treated (θ_4)	0.17	0.30^{**}	0.10	0.20^{*}
	(0.11)	(0.12)	(0.10)	(0.11)
1998 × treated (θ_5)	−0.05	−0.08	−0.01	−0.04
	(0.11)	(0.11)	(0.10)	(0.11)
1999 × treated (θ_6)	−0.05	−0.02	−0.02	0.02
	(0.11)	(0.11)	(0.10)	(0.10)
1994 × foodborne (θ_7)	0.04	−0.02	0.05	−0.01
	(0.06)	(0.07)	(0.05)	(0.06)
1995 × foodborne (θ_8)	−0.06	-0.15^{**}	−0.04	−0.10
	(0.05)	(0.06)	(0.05)	(0.06)
1996 × foodborne (θ_9)	0.00	−0.07	0.00	−0.05
	(0.06)	(0.06)	(0.06)	(0.07)
1997 × foodborne (θ_{10})	-0.13^{*}	-0.24^{***}	-0.14^{**}	-0.22^{***}
	(0.07)	(0.08)	(0.06)	(0.08)
1998 × foodborne (θ_{11})	-0.12^{**}	−0.08	-0.22^{***}	-0.19^{**}
	(0.06)	(0.07)	(0.06)	(0.08)
1999 × foodborne (θ_{12})	-0.18^{***}	-0.19^{***}	-0.27^{***}	-0.29^{***}
	(0.06)	(0.07)	(0.06)	(0.07)
F-test of H0: $\{\theta_{1:4}\} = 0$	2.11	4.40	1.12	2.34
p-value of F-test	0.08	0.00	0.35	0.05
F-test of H0: $\{\theta_{5:6}\} = 0$	0.27	0.63	0.06	0.26
p-value of F-test	0.76	0.54	0.94	0.77
F-test of H0: $\{\theta_{7:12}\} = 0$	5.62	4.30	12.71	7.18
p-value of F-test	0.00	0.00	0.00	0.00
R^2	0.99	0.99	0.99	0.99
N	3,192	2,352	3,192	2,352

Panel C: Triple difference (J&L specification plus interaction term)

	J&L Disease Selection		J&L plus Campylobacter	
	CA Control	N. CA Control	CA Control	N. CA Control
LA mandatory	−0.02	−0.01	−0.02	0.00
	(0.03)	(0.02)	(0.03)	(0.02)
LA voluntary	0.01	0.01	0.00	0.00
	(0.03)	(0.03)	(0.03)	(0.03)
Foodborne × post-1998	-0.13^{***}	−0.06	-0.23^{***}	-0.18^{***}
	(0.04)	(0.04)	(0.04)	(0.04)
Foodborne × LA mandatory	-0.17^{*}	-0.25^{***}	−0.08	−0.14
	(0.09)	(0.09)	(0.08)	(0.09)
Foodborne × LA voluntary	−0.11	-0.20^{*}	−0.09	−0.14
	(0.11)	(0.11)	(0.12)	(0.12)
R^2	0.99	0.99	0.99	0.99
N	3,192	2,352	3,192	2,352

Table 4: Panels A and C of Table 5 of Jin and Leslie (2019) adding two more pre-treatment years to the observation window (1993 and 1994). Coefficients 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. In Panel A, we include Foodborne × year interactions, which strongly reject F-tests for null effects because of the divergent trends between foodborne and non-foodborne hospitalizations over the observation window. Neither CA nor Northern CA exhibit parallel trends with LA in the lead tests, invalidating them as control groups. Adding campylobacter to the disease selection mitigates the parallel trends violation for CA, but also causes treatment effects to diminish to statistical insignificance. *p<0.1; **p<0.05; ***p<0.01

given substantial protective measures at the food supply level implemented in response.

The case of LA illustrates the severe challenges of drawing causal inferences in observational settings. As J&L notes, media coverage and political pressure may confound the intervention. Restaurant grading was implemented along with a new certification program, making it difficult to isolate the effect of disclosure. Given these inferential challenges, we believe it is of utmost importance to lay bare identifying assumptions, and are glad that this thorough exchange has led to a basic agreement that disclosure may have played a more limited role.

We conclude by noting that we discovered the issues with Jin and Leslie (2003) in the context of designing an RCT calibrated around the originally reported 20% treatment effects. Fortunately, we now have evidence from that RCT in Seattle and King County, where we powered a stepped-wedge randomized trial at 85% to detect for a 20% effect on illnesses (not hospitalizations). That RCT found no evidence for any health benefits of restaurant grading (Handan-Nader et al., 2018).

References

Ben-Shahar, Omri, and Carl E. Schneider. 2011. "The failure of mandated disclosure." *University of Pennsylvania Law Review* pp. 647–749.

Ben-Shahar, Omri, and Carl E. Schneider. 2014. "The Futility of Cost-Benefit Analysis in Financial Disclosure Regulation." *The Journal of Legal Studies* 43(S2):S253–S271.

Firestone, Melanie J., and Craig W. Hedberg. 2018. "Restaurant Inspection Letter Grades and Salmonella Infections, New York, New York, USA." *Emerging Infectious Diseases* 24(12):2164–2168.

Friedman, Cindy R., Robert M. Hoekstra, Michael Samuel, Ruthanne Marcus, Jeffrey Bender, Beletshachew Shiferaw, Sudha Reddy, Shama Desai Ahuja, Debra L. Helfrick, Felicia Hardnett, Michael Carter, Bridget Anderson, and Robert V. Tauxe. 2004. "Risk Factors for Sporadic Campylobacter Infection in the United States: A Case-Control Study in FoodNet Sites." *Clinical Infectious Diseases* 38(Supplement_3):S285–S296.

Fung, Archon, Mary Graham, and David Weil. 2007. *Full disclosure: The perils and promise of transparency*. Cambridge University Press.

Handan-Nader, Cassandra, Daniel E. Ho, and Becky Elias. 2018. "Feasible Policy Evaluation by Design:A Randomized Synthetic Stepped-Wedge Trial in King County." *Unpublished manuscript* .

Haughton, James G. 1998. "Annual Morbidity Report."

Ho, Daniel E., Zoe C. Ashwood, and Cassandra Handan-Nader. 2019. "New Evidence on Information Disclosure through Restaurant Hygiene Grading." *American Economic Journal: Economic Policy* 11(4):404–28.

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

Jin, Ginger, and Phillip Leslie. 2003. "The Effect of Information on Product Quality: Evidence from Restaurant Hygiene Grade Cards." *The Quarterly Journal of Economics* 118(2):409–451.

Jin, Ginger Zhe, and Phillip Leslie. 2019. "New Evidence on Information Disclosure through Restaurant Hygiene Grading: Reply." *American Economic Journal: Economic Policy* 11(4):429–43.

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

Mead, PS, L Slutsker, V Dietz, LF McCaig, JS Bresee, C Shapiro, PM Griffin, and RV Tauxe. 1999. "Food-related illness and death in the United States." *Emerging Infectious Diseases* 5(5):607–625.

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):126–130.

Satzman, Darrell. 1997. "Restaurant Inspections; Each week, more than 150 health inspectors check up on the region's 20,000 plus eateries looking for violations that could literally make you sick." *Los Angeles Times* .

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.

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.

Starrs, Paul F., and Peter Goin. 2010. *Field Guide to California Agriculture*. Berkeley, CA: University of California.

Weil, David, Archon Fung, Mary Graham, and Elena Fagotto. 2006. "The Effectiveness of Regulatory Disclosure Policies." *Journal of Policy Analysis and Management* 25(1):155–181.