New Evidence on Information Disclosure through Restaurant Hygiene Grading: Online Appendix

Daniel E. Ho* Zoe C. Ashwood† Cassandra Handan-Nader‡

December 5, 2018

*Corresponding Author; William Benjamin Scott and Luna M. Scott Professor of Law, Professor (by courtesy) 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

†Ph.D. Student, Computer Science, Princeton University.

‡Ph.D. Student, Political Science, Stanford University.

Appendices

Appendix A Extension to Later Analysis by Simon (2005)

In this Appendix, we show that the issues with J&L apply equally to those reported in Simon et al. (2005) ("Simon"). While Simon was separately published in the *Journal of Environmental Health*, the study comes from the same research team, as J&L are coauthors along with staff from the LA County Department of Health Services. Details for the precise specification are lacking, but Simon at core uses the same research design.

Simon estimates a simple DID model and specification similar to Equation 4, comparing LA and CA before and after the adoption of grading in 1998. There are three principal differences compared to J&L. First, the observation window is lengthened to 1993-2000. Because the fields for the type of admission differ, the analysis does not utilize a filter to include only unscheduled admissions from home (see Appendix J). Second, Simon no longer uses the mandatory disclosure (m) and voluntary disclosure (v) fractional treatment parameters, and instead uses a simple binary treatment indicator for the proportion of the three-digit ZIP in LA post-1998. Third, the most substantial difference is that Simon uses a radically different set of diseases (see Appendix D). For instance, in contrast to J&L, Simon includes campylobacter, botulism, listeria, yersinia, and c. perfringens, but excludes cysticercosis, amoebic dysentery, and gastrointestinal anthrax. Such differences underscore our point above that the disease selection used in J&L does not cohere with a public health understanding of how restaurant sanitation practices affect foodborne illness. In addition, instead of using all (non-foodborne) digestive system disorder discharge codes as a control group, Simon uses hospitalizations for appendicitis.

As before, it does not appear that Simon included all two-way interactions in the model specification, again imposing the assumption that foodborne and appendicitis hospitalizations follow the same average trend in CA, but not LA. Simon's DID model yields a 13% treatment effect and the model with appendicitis control illnesses yields a 12% treatment effect.

We hence estimate the same models (reported as models 1 and 1 in Simon) to show that the same

	Diff-in-Diff		J&L Spec.		Triple Diff.	
	LA	S. Cal.	LA	S. Cal.	LA	S. Cal.
Grade Cards	−0.04	−0.05	0.13***	0.16**	−0.01	0.02
	(0.06)	(0.07)	(0.03)	(0.06)	(0.03)	(0.06)
Grade Cards × Foodborne			−0.31***	−0.34***	−0.03	−0.07
			(0.06)	(0.09)	(0.07)	(0.09)
Foodborne × post-1998					−0.26***	−0.25***
					(0.04)	(0.03)
R^2	0.70	0.70	0.96	0.96	0.96	0.96
N	1,824	1,824	3,648	3,648	3,648	3,648

Table 1: Replication of Simon analysis with falsification tests substituting Southern CA (excluding LA) as placebo treated units. All models use Simon foodborne disease selection (see Appendix D) and appendicitis as the control illness group. As in Simon, the variable Grade Cards equals the proportion of the three-digit ZIP code population within LA for observations post-1998, and 0 otherwise. Coefficients shown with standard errors, clustered by three-digit ZIP and illness type combinations (or simply ZIPs for the DID model) in parentheses. Each model is estimated with fixed effects for year-quarters and either (a) fixed effects for three-digit ZIPs in the DID model or (b) fixed effects for three-digit ZIP by illness type combinations for the J&L specification and triple differences models. We present results for the original observation period (1993-2000). p-values for pairwise comparisons of LA and Southern CA coefficients in each set of models from left to right are 0.95, 0.74, and 0.73. *p<0.10; **p<0.05; ***p<0.01

limitations apply, using the 1993-2000 data with the new disease selection and binary treatment. Table 1 presents both the placebo models for Southern CA from Section IV, as well as the fully specified triple difference models from Section V. The first two columns provide estimates for DID models for LA and Southern CA. Curiously, these models do not yield a statistically significant treatment effect, which we explore below. The middle columns present the J&L specification, using appendicitis as a control illness, but imposing the assumption that trends for appendicitis and foodborne illnesses are the same for CA, but not LA. Point estimates are statistically comparable to those presented in Simon. Again, we find statistically significant grading effects for Southern CA, showing that the results are in part an artifact of the salmonella outbreak. The last two columns complete the triple difference specification, adding only the foodborne × post-1998 interaction term. Substantively, this allows for appendicitis time trends to be distinct from foodborne illness time trends in CA. As before, the treatment effects become statistically insignificant in the fully specified triple difference regression.

We also investigated why our DID estimate was statistically insignificant in contrast to Simon's. One potential explanation is that Simon may have omitted zero count observations. Because of the sparseness of foodborne hospitalizations, a non-trivial number of observations have counts of zero. For instance, as a lower bound for the prevalence in monthly data, roughly 7% of the quarterly-ZIP data is comprised of zero counts. J&L appears to have included zero count observations (e.g., by calculating log(a + 1)). J&L's reported sample size is 6,840, corresponding to 12 months × 5 years × 57 three-digit ZIP codes × 2 types of illnesses. Simon's reported sample size, however, falls short of the balanced panel. We would expect 10,944 observations, given the three additional years (= 12 months × 8 years × 57 three-digit ZIP codes × 2 types of illnesses), but Simon reports only 7,972 observations. As this number is not divisible by 12 or 8, one explanation may be that zero count observations are dropped, but some other unarticulated sample omission is possible.

To test this hypothesis, we approximate what the DID model would look like if we dropped zero count observations. We do so by dropping all observations with fewer than three hospitalizations in a quarter, which necessarily implies that at least one observation at the monthly level would be dropped. Figure 2 presents results, showing that we can replicate a 14% treatment effect with a DID model by dropping such observations. Doing so introduces obvious bias.

	All Obs.	Dropped
Grade Cards	−0.04	−0.14***
	(0.06)	(0.05)
R^2	0.70	0.64
N	1,824	1,132

Table 2: Comparison of DID model with LA as the treated county using all observations (All Obs.) and dropping quarters with fewer than 3 hospitalization counts (Dropped). All models use Simon foodborne disease selection. As in Simon, the variable Grade Cards equals the proportion of the three-digit ZIP code population within LA County for observations post-1998, and 0 otherwise. Coefficients shown with standard errors, clustered by three-digit ZIPs in parentheses. Each model is estimated with fixed effects for year-quarters and three-digit ZIPs. We present results for the original observation period (1993-2000). *p<0.10; **p<0.05; ***p<0.01

In sum, the salmonella outbreak and omission of the two-way interaction term affect the Simon results in the same way as the J&L results, pointing towards severe confounding in the evidence of grading effects.

Appendix B Appearance of J&L in Policy Debates

The LA foodborne illness findings have appeared frequently in public debates on disclosure. The public health community has traditionally harbored serious reservations about restaurant grading (Seiver and Hatfield, 2000; Wiant, 1999). One of the principal concerns was that the underlying inspection scores are quite stochastic, because (a) conditions can fluctuate considerably in restaurants over days, and (b) inspectors can apply dramatically different standards for scoring violations (Boehnke and Graham, 2000; Wiant, 1999). Indeed, after including restaurant grading in its Model Food Code for several decades, the Food and Drug Administration abandoned it in 1976 due to skepticism of the utility of grading (Ho, 2012, pp. 588-91). The evidence for the association between restaurant inspection scores and foodborne illness outbreaks is quite mixed (see, e.g., Cruz et al., 2001; Irwin et al., 1989; Jones et al., 2004; Serapiglia et al., 2007; NYCDHMH, 2012).

To assess the policy relevance of the LA findings, we performed newspaper searches of jurisdictions that considered restaurant hygiene disclosure since J&L's publication in 2003. We searched all top 20 US metropolitan areas by population and also recorded any other jurisdictions that our search revealed to have considered restaurant hygiene disclosure. Although our search is by no means exhaustive, we quickly identified 44 jurisdictions spanning 7 different countries, including 10 out of the United States' 20 largest cities, that underwent substantial deliberation about restaurant hygiene disclosure. (We do not include LA, which considered J&L's evidence in expanding the scheme to food trucks, in this group of 10.) As shown in Table 3, these jurisdictions top a combined population of 205 million people. Out of the top 19 metropolitan areas (excluding LA), 63% of residents lived in jurisdictions considering restaurant disclosure since 2003. 50% of residents experienced the adoption of restaurant grading since 2003.

	Jurisdiction contemplating restaurant disclosure	**Largest city**	**Population**	**J&L cited**	**Enactment / Amendment year**
United States	State of Florida	Jacksonville	18,801,310	Yes	
	Los Angeles County, CA	Los Angeles	9,818,605	Yes	2010
	New York City, NY	New York City	8,174,962	Yes	2010

	Harris County, TX	Houston	4,092,459		
	Maricopa County, AZ	Phoenix	3,817,117		2011
	Orange County, CA	Anaheim	3,010,232	Yes	
	San Bernardino County, CA	San Bernardino	2,035,210	Yes	2004
	Clark County, NV	Las Vegas	1,951,269		2010
	King County, WA	Seattle	1,931,249	Yes	2017
	Santa Clara County, CA	San Jose	1,781,642	Yes	2014
	Alameda County, CA	Oakland	1,510,271		2012
	Sacramento County, CA	Sacramento	1,418,788	Yes	2007
	San Antonio, TX	San Antonio	1,327,551		2016
	Cuyahoga County, OH	Cleveland	1,280,122	Yes	
	Allegheny County, PA	Pittsburgh	1,223,348	Yes	2015
	Contra Costa County, CA	Concord	1,049,025	Yes	2016
	Pima County, AZ	Tucson	980,263	Yes	2002
	Fulton County, GA	Atlanta	920,581		2013
	Mecklenburg County, NC	Charlotte	919,628		2012
	Kern County, CA	Bakersfield	839,631	Yes	2006
	Ventura County, CA	Oxnard	823,318	Yes	
	San Francisco, CA	San Francisco	805,195	Yes	
	Columbus, OH	Columbus	788,792		2006
	San Mateo County, CA	Daly City	718,451	Yes	2016
	Baltimore, MD	Baltimore	621,143		
	Boston, MA	Boston	617,680	Yes	2016
	Stanislaus County, CA	Modesto	514,453	Yes	
	Sonoma County, CA	Santa Rosa	483,878		2016
	Minneapolis, MN	Minneapolis	382,599	Yes	
	Albany County, NY	Albany	304,204		2011
	Weld County, CO	Greeley	252,825		2014
	Marin County, CA	San Rafael	252,409	Yes	2015
	Butte County, CA	Chico	220,000		2014
	Muskegon County, MI	Muskegon	172,188	Yes	
	Napa County, CA	Napa	136,484		2005
	Hartford, CT	Hartford	124,775	Yes	2012
	San Angelo, TX	San Angelo	93,227		
	Newton, MA	Newton	85,174		2015
Foreign	United Kingdom	London	62,300,000	Yes	2008
	South Korea	Seoul	49,410,366		2017
	New South Wales, Australia	Sydney	7,230,000	Yes	2010
	Hong Kong, China	Hong Kong	7,071,576	Yes	
	New Zealand	Auckland	4,362,000	Yes	
	Hamilton, Ontario, Canada	Hamilton	721,053	Yes	2014

Table 3: Selective sample of jurisdictions contemplating restaurant grading since publication of J&L in 2003. 10 of the 20 largest cities in the United States have enacted restaurant grading, excluding LA's amendment to cover food trucks. Commentators and government reports specifically cited J&L's LA evidence in 61% of enactment debates. In some cases, jurisdictions are coded as amending hygiene disclosure when the amendment debate referenced J&L. For instance, the evidence played into the consideration by LA County to extend the grading system to food trucks in 2010, the consideration by Pima County to change grading from voluntary to mandatory in 2008, and Allegheny County's conversion to letter grades in 2015. In three cases (Baltimore, Harris County, and San Angelo), jurisdictions considered restaurant grading, but did not cite the LA evidence and did not enact grading. In San Francisco, the LA evidence was cited in support of a partnership with Yelp to disclose inspection results online. In Stanislaus County, the LA evidence supported an argument for increased funding to post inspection results online. Population for domestic jurisdiction is from the 2010 census. Population for international jurisdictions is from the 2010 or 2011 local census, except for South Korea, which is from the World Bank for 2010.

We also searched for whether the LA foodborne illness evidence was cited in the enactment debates. The fourth column of Table 3 indicates whether the effect identified by J&L was explicitly referenced. In 27 amendment or enactment debates, the LA evidence was specifically referenced. For instance, the J&L evidence appeared in the debate about implementing restaurant grading in Hamilton, Ontario, the ninth largest city in Canada, in the following news report:

> Influential research from the United States suggests the threat of damage to business is exactly what makes full disclosure systems work, and work well. Stanford University's Leslie co-authored a groundbreaking study published in 2003 in the prestigious *Quarterly Journal of Economics*. The study examined the impact of the sign system introduced in Los Angeles County in 1998. It gives restaurants a grade of A, B or C depending on the results of inspections. Perhaps most revealingly, the year after the program was introduced, hospitalizations for food-borne illnesses dropped by 20 per cent, a trend not reflected in other areas of the state without the grade cards.

Hamilton's health officer expressed skepticism based on costs to implement the system and the ability to isolate the effect of grading: "I think when anybody's got a program, they are 100 per cent sold on the fact their program's the best, so I am not at all surprised that L.A. [is] saying, 'what we did was right.' They're not going to tell you it's a dud." The skepticism expressed by the Hamilton medical officer, however, is somewhat of an outlier. In nearly all reports we discovered, the LA foodborne illness findings were stated with certainty. Such citations exist not just in newspapers, but also in government reports (e.g., the United Kingdom's Food Standard Agency), official hearing records (e.g., Allegheny County Board of Public Health minutes), and grand jury reports (e.g., Orange County).

Beyond restaurant hygiene disclosure, J&L's findings as a whole also appeared to bolster arguments in favor of targeted disclosure in a variety of policy debates nationwide. Table 24 provides some examples. In a 2007 speech, for instance, then-governor of the Federal Reserve System Randall S. Kroszner invoked J&L as "systematic evidence that in practice, changes in disclosure affect both consumer and supplier behavior in a number of consumer product markets." Policy areas in which the findings were cited run the gamut from consumer protection (e.g., privacy) to environmental safety (e.g., toxic harms) and consumer finance (e.g., credit card disclosure) and communications technology (e.g., emergency coverage of wireless phones). Our analysis, however, does not call into question any finding in J&L except the one concerning foodborne illness.

Appendix C Nonrandom Enactment

C.1 Municipal Enactment

In November 1997, the local television station KCBS ran a multipart exposé, wherein journalists went undercover to document sanitation practices in LA restaurants. This exposé was the impetus for the county restaurant grading ordinance and subsequent municipal adoptions. J&L posit that adoption differences across LA's 88 incorporated municipalities were largely due to "bureaucratic delays rather than the influence of restaurants" (p. 419), and hence exogenous. Anecdotal news reports suggest that the influence of the restaurant lobby plays a major role in adoption of restaurant grading. The city of Long Beach, for instance, faced similar political pressure after the KCBS exposé, but "decided to eschew letter grades or numeric scores after a year of discussion and input from the local restaurant industry."[1] On the other hand, San Bernardino adopted restaurant grading in 2004, "[d]espite ardent opposition from restaurant owners."[2] To rule out nonrandom adoption, J&L estimates a duration model, with restaurant and city attributes as predictors. Because "coefficients on the restaurant characteristics are insignificantly different from zero," J&L concludes that the assumption of "exogenous city adoption dates" is well-founded (p. 420).

While adoption rates may be uncorrelated with restaurant attributes, we show here that adoption is strongly (and statistically significantly) related to a wide range of city demographics. Despite testing for such associations, J&L does not report results of city demographics in the duration model. We collect city demographics from a 1990 Census report for 81 of the 85 incorporated municipalities under the jurisdiction of the LA Department of Public Health.[3] In addition, we augment the data with information on the number of hospitals and hospital beds. The left three columns of Table 4 present conditional means for ten covariates for municipalities that adopted grading in 1998, municipalities that adopted later than 1998, and municipalities that never adopted grading.

[1]Dickerson, M. (1999, April 21). Long Beach's grading system gets low marks from restaurateurs. Los Angeles Times.

[2]Hugo, M. (2004, June 9). San Bernardino county to rate restaurants. Los Angeles Times.

[3]Four municipalities were not included in the 1990 Census either because their populations were under 1,000 or they were not yet incorporated in 1990.

Covariate	1998 Adopters Mean	Post-1998 Adopters Mean	Never Adopters Mean	F-test p-value	Survival Model exp(coef)	p-value
Population (1000s)	123.29	52.57	6.41	0.16	1.15	0.00
Hospitals	1.54	1.04	0.14	0.02	1.14	0.00
Hospital beds	2.54	3.51	0.59	0.00	1.00	0.91
Households (HHs, 1000s)	41.66	18.67	2.40	0.18	1.48	0.00
Median HH income ($1000s)	43.23	37.95	83.09	0.00	0.22	0.01
Children / HH	0.96	0.75	0.66	0.00	2.62	0.00
Female (%)	50.33	50.79	50.95	0.00	0.88	0.10
African-American (%)	4.29	8.54	2.24	0.00	1.00	0.86
Asian-American (%)	12.05	10.89	9.69	0.00	1.01	0.41
Hispanic (%)	39.20	31.89	14.04	0.00	1.01	0.00
No. of cities	48	26	7			
Perc. of LA pop.	81%	19%	1%			

Table 4: City attributes are correlated with time to municipal enactment of restaurant grading. We analyze 81 out of 85 incorporated cities under the jurisdiction of the LA Department of Public Health that had available 1990 census data. This excludes three cities that did not contract with LA county for food safety enforcement (Long Beach, Pasadena, Vernon), unincorporated areas, and cities that had populations under 1,000 people at the time of the 1990 census. Percent of LA population for each adopter group is calculated out of this subset. All covariates are from the 1990 Census except for number of hospitals and available hospital beds, which are both from OSPHD's Hospital Quarterly Financial and Utilization Report in Q4 of 1997 `http://www.oshpd.ca.gov/HID/Hospital-Quarterly.html`. For population, household, and median income, averages are expressed in units of 1000s and exponentiated coefficients are expressed in units of 100000s for readability. Available hospital beds are per 1,000 people. Missing percent female values were imputed from the 2000 census for Avalon, Westlake, Signal Hill, and Rolling Hills Estates. We report average covariate values for three groups: adopters in 1998, adopters post-1998, and never adopters. ANOVA p-value is from an F-test of difference-in-means between the three adoption groups (1998, post-1998, and never). Exponentiated coefficients and p-values are from a Cox proportional hazards regression on days to enactment using each covariate as the sole predictor.

We present simple p-values from F-tests of equivalence of means, with eight of ten tests rejecting the null. 48 of 81 municipalities adopted in 1998, covering 80% of LA's incorporated population under the jurisdiction of LA Department of Public Health. The differences are the most stark between adopters and never-adopters, with never adopting cities being smaller, more affluent, and much less ethnically diverse. The last two columns present results from Cox survival models, using each covariate as a separate predictor for duration until adoption. Six of these tests reject the null, again strongly refuting the notion that adoption was truly exogenous.[4] Such nonrandom adoption may, of course, bias effects. If competitive pressure is higher in large cities, for instance, the pressure to respond directly to the KCBS exposé independent of grading may induce greater sanitation

[4]It is possible that J&L interpreted tests from a saturated duration model. Due to collinearity, test statistics on individual coefficients may not be statistically significant. But a Wald test of a Cox model with all 10 predictors strongly rejects the null hypothesis that all coefficients are jointly equal to zero, with a pseudo-R^2 of 0.33.

improvements.

C.2 County Enactment

J&L posits that the “key feature of our data is the introduction of hygiene grade cards.” The paper describes the adoption as rapid and unanticipated (p. 417), justifying exogeneity. J&L addresses two other changes, namely the removal of a subjective inspection component in July 1997 and the addition of violations in March 1998, by examining effects over time.

What this characterization misses is that restaurant grading was one component of a litany of enforcement reforms. The sensational KCBS coverage placed intense political pressure on the Department of Health, leading to substantial and simultaneous changes in food safety enforcement beyond restaurant grading. Indeed, the head of the Department himself described these interventions as the “Marshall Plan for restaurant inspections.”[5]

At least three interventions distinct from restaurant grading — none of which were discussed by J&L — may confound restaurant grading. First, the same county ordinance that established restaurant grading also mandated that all managers become certified as food handlers.[6] The ordinance devoted equal space to the creation of a four-hour food handler training course, display of food handler’s certificate, and revocation procedures as to restaurant grading. Prior empirical work, albeit mixed in rigor, suggests that such training and certification requirements can improve sanitation practices (Anding et al., 2007; Cotterchio et al., 1998; Egan et al., 2007). Because this is a contemporaneous change, the research design offers no way to disentangle the effects of certification vs. restaurant grading. Second, the Department instituted a “zero tolerance” sanitation policy, engaging in an “unprecedented crackdown, closing eateries for code violations at three times the usual rate.”[7] This crackdown occurred at the same time that restaurant grading was implemented, so that any observed gains in sanitation practices may have been due to the direct effect of this zero

[5] Haefele, M.B. (1997, December 7). The state; dirty kitchens in the county’s health department. Los Angeles Times.

[6] Los Angeles County Ordinance 97-0071.

[7] Tobar, H., and Leeds J. (1998, January 29). Restaurants get a taste of tough county health policy. Los Angeles Times.

tolerance policy. Indeed, LA officials themselves stated that the crackdown "produced dramatic results" independent of grading.[8] Third, the Department implemented a new policy that a restaurant would be shut down if it exhibited critical violations on successive routine inspections.[9] To support the more aggressive enforcement policy, the Department hired dozens of additional inspectors, potentially increasing deterrence not just through sanctions but also inspection probability.

The fact that grading was part of a package of reforms does not invalidate efforts to assess the effects of the compound intervention. But the research design cannot isolate the effects of grading per se.

[8] Meyer, J. (1998, February 4). Loophole hampers restaurant crackdown. Los Angeles Times.

[9] Meyer, J. (1997, November 26). County crackdown on dirty restaurants OK'd. Los Angeles Times.

Appendix D Foodborne Disease

D.1 Foodborne Outbreaks and Panel Design

We now show how time trends in the major diseases selected by J&L are heavily outbreak driven, thereby undercutting the credibility of a DID (or triple difference) research design. This challenge is likely endemic to foodborne illnesses in the U.S. context. Figure 1 displays hospitalizations from 1983-2009 for the three most prevalent discharge codes studied by J&L.

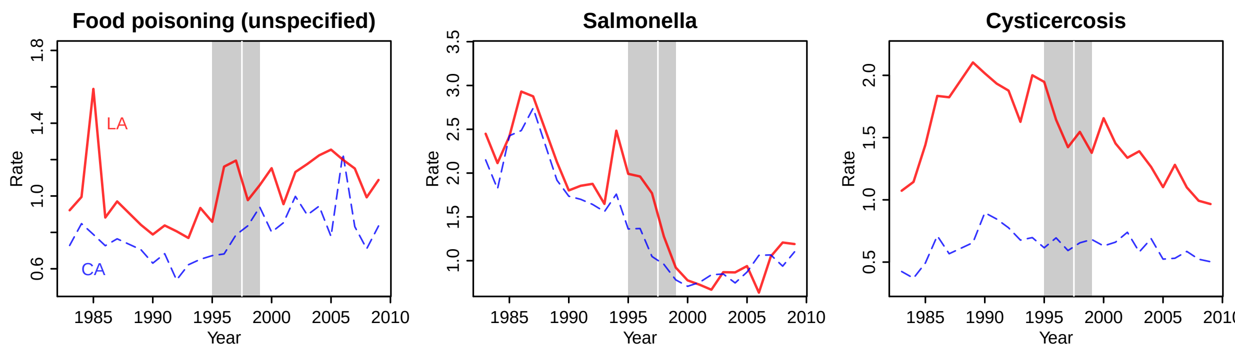

Figure 1: Foodborne hospitalizations from 1983-1999 for the three most prevalent discharge codes studied by J&L. LA is plotted in solid lines and the rest of CA in dashed lines. J&L observation window is highlighted in grey. Each of these time series suggests limitations to the original research design.

The left panel plots the LA and CA time series for "food poisoning (unspecified)." (This is a large residual category when physicians do not enter a more specific diagnosis code.) Most notable is that hospitalizations in LA spike in 1985, corresponding to a large outbreak from pesticide-contaminated watermelons in Central Valley (Goldman et al., 1990). Nearly 50 hospitalizations in LA were attributable to the outbreak (and 1,376 illnesses were reported), which is substantial considering that the average annual LA hospitalizations in J&L's data was 380.

The middle panel depicts the longer salmonella time series. From 1985-87, a salmonella newport outbreak from cattle in Tulare and San Bernardino counties caused a 4.9-fold increase in newport cases statewide, but LA was reported to have been particularly affected (Spika et al., 1987; Puzo, 1986). The longer time series also shows how the Southern CA salmonella enteritidis outbreak in the early 1990s is inconsistent with parallel trends. Peak salmonella rates occurred in 1994 and declined rapidly years *before* LA enacted restaurant grading.

The third panel plots cysticercosis, a discharge code responsible for nearly 20% of hospitalizations. This tapeworm infection is largely border-related. The panel shows that, as a result, cysticercosis exhibits in LA (a) a much steeper increase in the 1980s, and (b) a more pronounced secular decline from the mid-1990s to 2010 than CA.

These hospitalization trends illustrate why conventional panel methods do not appear to work well with foodborne hospitalization data.

D.2 Replication

We now spell out how we replicated J&L's disease selection. J&L focuses on hospitalizations for disorders of the digestive system falling in the "Major Diagnostic Category" (MDC) 6. J&L distinguishes foodborne and non-foodborne MDC 6 illnesses "based on the principal diagnosis" (i.e., the chief cause of admission) from International Classification of Diseases, Ninth Revision, Clinical Modification Code (ICD-9-CM codes or "discharge codes"). Although the paper purports to rely on Mead et al. (1999) and a medical researcher to classify discharge codes as over 90% foodborne, Table 5 of J&L appears to measure foodborne illnesses exclusively "based on the definition by the medical researcher." Simon, on the other hand, classifies foodborne illnesses principally based on whether Mead et al. (1999) indicates that more than 70% of cases are foodborne. Simon excludes two diseases (Salmonella Typhi and Vibrio cholerae) for being travel-related, as well as hospitalizations of children less than five years of age.

To replicate the disease selection, we obtained a spreadsheet from J&L including the discharge code and notes from their medical researcher. Out of the 41 illnesses in that spreadsheet, nine are marked by the medical researcher as being over 90% foodborne. This judgment is not strictly correct, as E. Coli, for instance, is marked as over 90% foodborne. (As noted in Subsection D.3, 85% of shiga-toxin producing E. Coli (e.g., O157:H7), 70% of enterotoxigenic E. Coli, and only 30% of other diarrheagenic E. Coli are estimated to be foodborne (Mead et al., 1999).) Even this information, however, is insufficient to determine the exact disease set. The spreadsheet, for instance, included only the first three ICD digits, but the first three digits (a) can correspond to multiple diagnostic categories (i.e., beyond MDC 6), and (b) can correspond to more specific diseases, not all of which are primarily foodborne (e.g., diarrheagenic E. Coli). We hence collected all ICD codes starting with these nine three-digits and subset to MDC 6 categories.

Table 5 displays diseases ostensibly studied by J&L and Simon. The first column lists the disease and the second column lists the count of total hospitalizations in CA from 1995-99. While there is some uncertainty about the inclusion of rare diseases, the major diseases are readily identified, because only a small number of diseases drive outcomes. J&L, for instance, studies a total of 5,068 CA foodborne hospitalizations. The only set of diseases yielding that volume include salmonella, food poisoning unspecified, and cysticercosis, which together comprise 85% of foodborne hospitalizations. Table 5 also merges in estimates for the percentage of cases that are foodborne from Mead et al. (1999) and Scallan et al. (2011), as well as the estimated percentage that is travel-related, based on matching ICD codes to pathogens. The J&L and Simon columns indicate whether each study included the particular discharge code.

We successfully replicate J&L's hospitalization counts. Table 6 compares counts between those reported in J&L and replicated here. While there are small differences, these are readily explained by differences in masking rules across versions of OSHPD data. Because masking rules have become stricter over time, J&L's earlier version of the dataset may include a few more counts than ours. In addition, because we use the 5-digit ZIP code, quarterly dataset and J&L uses the 3-digit ZIP code monthly dataset, some cases may be masked in J&L's dataset, but not ours (and vice versa). It's also unclear how J&L addresses instances of admissions occurring in one calendar year, but discharged in a subsequent calendar year. Our 1999 data, for instance, includes discharges occurring in 2000. Table 7 tabulates the ratio of replicated counts against those reported by J&L, with all ratios very close to 1.

Illness	Hosp.	Pathogen	Foodborne % (Mead)	Foodborne % (Scallan)	Travel %	Inclusion		ICD-9 Code
						J&L	**Simon**	
Salmonella gastroenteritis	1,809	Nontyphoidal Salmonella spp.	95	94	11	✓	✓	0030
Unspecified	1,273					✓	✓	0059
Cysticercosis	1,233	Taenia spp.				✓		1231
Campylobacteriosis	1,202	Campylobacter spp.	80	80	20		✓	00843
Botulism	242	Clostridium botulinum	100	100	1		✓	0051
Listeriosis	228	Listeria monocytogenes	99	99	3		✓	0270
Unspecified E. coli infection	173	Diarrheagenic E. Coli	30	30	1	✓	✓	00800
Staphylococcal	97	Staphylococcus aureus	100	100	1	✓	✓	0050
Other bacterial	77					✓	✓	00589
Other E. coli infection	69	Diarrheagenic E. Coli	30	30	1	✓	✓	00809
Yersiniosis	67	Yersinia enterocolitica	90	90	7		✓	00844
EHEC E. coli infection	52	STEC O157 and non-O157	85	68	4	✓	✓	00804
Amoebiasis	26	Entamoeba histolytica				✓		0060
ETEC E. coli infection	23	ETEC	70	100	55	✓	✓	00802
EPEC E. coli infection	16	Diarrheagenic E. Coli	30	30	1	✓	✓	00801
V. parahaemolyticus	7	V. parahaemolyticus	65	86	10		✓	0054
Amebic nondysenteric colitis	6	Entamoeba histolytica				✓		0062
Hymenolepiasis	5	Hymenolepiasis				✓		1236
V. vulnificus	3	V. vulnificus	50	47	2		✓	00581
Taenia solium infection	2	Tania spp.				✓		1230
Unspecified cestode infection	2					✓		1239
EIEC E. coli infection	1	Diarrheagenic E. Coli	30	30	1	✓	✓	00803
Unspecified Taeniasis	1	Taenia spp.				✓		1233
C. perfringens	0	C. perfringens	100	100	1		✓	0052
Clostridia	0	Clostridium spp.				✓	✓	0053
Enteric tularemia	0	Francisella tularensis				✓		0211
Gastrointestinal anthrax	0	Bacillus cereus				✓		0222
Taenia saginata infection	0	Taenia spp.				✓		1232
Diphyllobothriasis	0	Diphyllobothrium spp.				✓		1234
Sparganosis	0	Spirometra spp.				✓		1235
Other cestode infection	0					✓		1238

Table 5: Comparison of J&L and Simon disease selection. Each row represents a unique ICD-9 code either included by Simon or J&L. Hosp. indicates the total number of CA hospitalizations from 1995-99 in OHSPD data for that discharge code. Travel percentage indicates the estimated percentage of cases that are acquired through travel (Scallan et al., 2011). A checkmark indicates whether J&L or Simon likely included the discharge code. Rows shaded in light grey are discharge codes included only by J&L and rows shaded in dark grey are included only by Simon. Across discharge codes, the two studies agree for only 35% of 31 discharge codes, illustrating sharp differences in disease selection across these studies.

	Los Angeles				California			
	Foodborne		Digestive		Foodborne		Digestive	
Year	J&L	Replication	J&L	Replication	J&L	Replication	J&L	Replication
1995	401	396	54,412	54,328	607	589	128,849	129,295
1996	431	428	56,692	56,627	675	663	131,623	132,021
1997	405	403	59,585	59,520	634	613	139,645	139,415
1998	351	350	61,305	61,237	654	642	145,261	144,662
1999	309	311	60,915	61,472	601	601	148,338	149,542

Table 6: Hospital admissions for foodborne and non-foodborne digestive disorders (MDC 6). The J&L columns indicate counts reported in J&L's Table 5. The Replication columns indicate counts based on replicating their protocol with our version of the OSHPD data.

	LA		CA	
Year	Foodborne	Digestive	Foodborne	Digestive
1995	0.990	0.998	0.970	1.003
1996	0.993	0.999	0.980	1.003
1997	0.995	0.999	0.970	0.998
1998	0.997	0.999	0.980	0.996
1999	1.006	1.009	1.000	1.008

Table 7: Ratio between replicated counts and J&L counts. Ratios close to 1 indicate perfect replication for the year, region, illness cell.

D.3 Issues with Disease Selection

As documented in the paper, the exclusion of campylobacter runs contrary to the public health consensus. We here detail other issues with J&L's disease selection.

Comparing the J&L and Simon disease sets, Table 5 highlights the disease selection issues. In spite of the fact that the two papers purport to implement the same research design (with the same principal authors), the two papers study radically different disease sets. Of 31 discharge codes, only 11 discharge codes are shared between the papers (e.g., salmonella, food poisoning unspecified). J&L studies 13 discharge codes (rows shaded light grey) that are excluded by Simon (e.g., cysticercosis, amoebiasis, taeniasis, and gastrointestinal anthrax). Conversely, Simon includes 7 discharge codes excluded by J&L. Likely because it was published in a public health journal, Simon coheres with the public health consensus by including campylobacter, foodborne botulism, listeria, vibrio, and yersinia. The raw discrepancy suggests that the J&L disease selection evinced serious issues, which we outline below.

E. Coli. J&L appears to include all E. Coli 008 ICD-9 subcodes, but there is no reason to expect restaurant grading to affect ETEC and EPEC E. Coli infections. ETEC is also known as "traveler's diarrhea" (Nataro and Kaper, 1998) and Scallan et al. (2011) estimate that 55% of hospitalizations in the United States for ETEC are acquired overseas. The FDA notes, "People in the U.S. usually don't get ETEC infections, unless they travel to areas of the world with poor sanitation" (Food and Drug Administration, 2012). EPEC infections most commonly occur in children less than 2 years old, and in countries outside the United States: "Recent estimates from the Centers for Disease Control and Prevention (CDC) on food-related illness in the United States listed only 4 hospitalizations as a result of EPEC infection; however, this pathogen continues to persist in other parts of the world and continues to be regarded as a serious threat to children under the age of 2" (Croxen et al., 2013).

Cysticercosis. Cysticercosis (a type of taeniasis) is principally travel-related. The FDA notes, "In the United States, the disease occurs primarily in individuals who have traveled or immigrated from endemic regions in Latin America, India, Asia, Eastern Europe, and Africa" (Food and Drug Administration, 2012). There is also a significant delay between the ingestion of Taenia solium eggs, and the onset of cysticercosis or taeniasis symptoms. For instance, it can take several years before neurocysticercosis symptoms are displayed (Food and Drug Administration, 2012).

Exclusion of Listeriosis. Listeriosis is recognized as "one of the leading causes of death from foodborne illness" (Food and Drug Administration, 2012), and is estimated to be food-borne in 99% of cases in the United States, and travel-related in just 3% of cases (Scallan et al., 2011).

Under-inclusion of Salmonella. Similarly, by focusing only on MDC 6 disorders, J&L may miss the most serious cases of salmonellosis that present non-gastroenteritic complications (e.g., salmonella septicemia, salmonella pneumonia, salmonella arthritis, and salmonella osteomyelitis). Such cases constitute roughly half of (non-typhoidal) salmonella hospitalizations in LA county from

1995-99. While some non-MDC 6 codes are properly excluded (e.g., salmonella meningitis largely affects babies), foodborne salmonellosis can manifest in non-gastroenteritic ways (Acheson and Hohmann, 2001).

Other Idiosyncratic Illnesses. Some diseases included by J&L appear entirely irrelevant from a public health understanding of foodborne disease in the United States (e.g., tularemia, amebiasis, gastrointestinal anthrax and taeniasis). These illnesses are absent from the major public health reviews (Mead et al., 1999; Scallan et al., 2011). For amebiasis, "water is the most common source of contamination", and like many of the other illnesses that J&L studies, it is often acquired overseas: "[Amebiasis is] not very common in the U.S., where it usually affects people who traveled here from a country with poor sanitation" (Food and Drug Administration, 2012). Gastrointestinal anthrax is also predominantly acquired overseas, by consuming meat from livestock infected with anthrax (Centers for Disease Control and Prevention, 2015).

Principal Diagnosis. The exclusive focus on principal diagnoses may also undercount foodborne hospitalizations. For instance, hemolytic-uremic syndrome (HUS) is a complication associated with E. coli, leading to life-threatening kidney failure (Gould et al., 2009). Due to the severity of HUS, it may be coded as the principal diagnosis (the reason for hospital admission), but a secondary diagnosis of E. Coli may be entered.[10]

[10] Lloyd and Rissing (1985) documents considerable discretion in entering discharge codes.

Appendix E Units of Analysis

In this Appendix, we explain why the geographic units of analysis used by J&L — three-digit ZIP codes – are inappropriate for the analysis undertaken. We also discuss why OSHPD data at the monthly, compared to quarterly, level are unlikely to provide any advantages for assessing the impact of restaurant grading on foodborne illness rates.

E.1 ZIP Codes as Geographic Units

It is widely documented that five-digit ZIP codes have serious limitations for public health and social science research (Krieger et al., 2002a,b; Grubesic, 2008). First, ZIP codes are administrative units created for efficient delivery of mail and are hence incompatible with jurisdictional lines relevant for understanding the impact of health laws. ZIP codes, for instance, do not correspond to municipal lines (Sater, 1994). Each ZIP code can, as a result, have radically different sizes and populations. Second, ZIP codes can change substantially over time. Between 1997 and 2001, for instance, the post office added 390 new ZIP codes and discontinued 120 (Krieger et al., 2002b, p. 1100). Mapping census data to ZIP codes is hence nontrivial. The 2000 census, for instance, created ZIP Code Tabulation Areas (ZCTAs) that are quite distinct from how ZIP codes were represented in the 1990 Census.

These limitations can have serious implications for research. Krieger et al. (2002a), for instance, shows that while census block and tract data detected demographic health disparities, ZIP code measures sometimes failed to detect such disparities or detected disparities inconsistent with census tract and block measures. As put succinctly in a presentation by a research scientists from the CA Department of Public Health: "Avoid using ZIP codes whenever possible" (Smorodinsky, 2010).

J&L's analysis compounds this problem by using three-digit ZIP codes. While five-digit ZIP codes generally accord with county lines, three-digit ZIP codes do not. Three digit ZIP codes correspond to a processing and distribution center of the US Postal Service ("Sectional Center

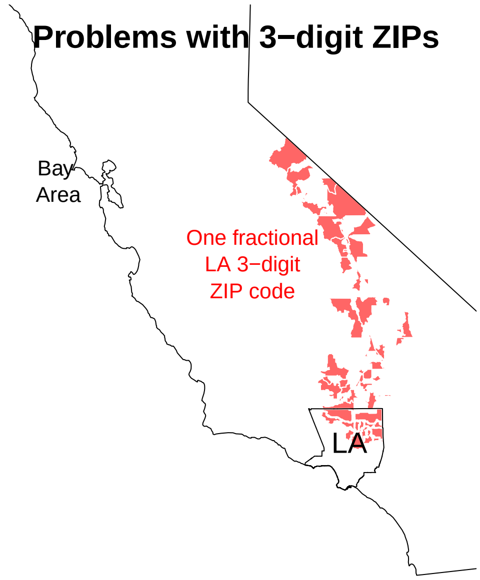

Figure 2: The primitive geographic unit of analysis for J&L is a three-digit ZIP code. These exacerbate well-known limitations to using five-digit ZIP codes. This figure plots the region for a single 3-digit border LA ZIP code (935), which spans much of the state of CA.

Facilities"). Out of 18 LA three-digit ZIP codes, five cross county lines,[11] requiring J&L to assign fractional treatments to those units, even though grading is adopted at the county and municipal level. To see how problematic that can be, Figure 2 plots the area for the 935 three-digit ZIP code. While that ZIP code contains much of northern LA, it spans vast portions of CA, reaching some 300 miles north to Yosemite Park and the Nevada border. Nor is it the case that three-digit ZIP codes escape realignment of ZIP codes. In 1999, for instance, realignment in Riverside county split the ZIP code 91719 into 92877, 92879, 92881, and 92883.[12]

Because our version of the OSHPD data contains information on five-digit ZIP codes, we can construct an alternative test of grading effects that is implicit in J&L's treatment of boundary ZIP codes. We focus on three-digit ZIP codes that were assigned fractional treatment by J&L. An observable implication of grading should be that if we disaggregate that unit into constituent five-digit ZIP codes, (a) hospitalization rates for units just inside and just outside of LA should be parallel pretreatment, but (b) rates for units inside LA should decrease around 1998 relative to

[11]We calculated this based on population, but J&L report six border ZIPs, potentially because of land area. Using land area is tricky, as ZIP code boundaries are poorly defined.

[12]Postal Bulletin, pp. 61-62. (1999, April 8).

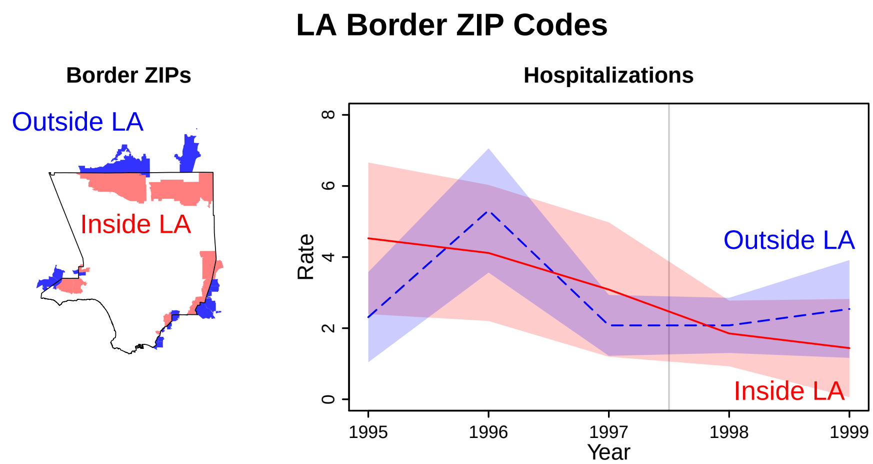

Figure 3: Analysis of border ZIP codes. The left panel plots areas for five-digit ZIP codes in boundary three-digit ZIP codes to which J&L assigned fractional treatment. The right panel plots time trends for five-digit ZIP codes just outside of LA and just inside of LA.

those outside of LA. Figure 3 conducts this border analysis, with the left panel displaying border (five-digit) ZIP codes just inside and just outside of LA. The right panel plots time trends, with pointwise 95% confidence intervals. This reveals no evidence consistent with grading effects. If anything, the drop after 1997 is as pronounced in units outside of LA, even if it stems from a spike around 1996.

E.2 Temporal Aggregation

In this subsection, we show that monthly OSHPD data is unlikely to provide any additional leverage over the research question. As mentioned in footnote 33, the principal reason is that foodborne hospitalizations are sparse. In 1999, the expected number of hospitalizations at the monthly level is 1.3 hospitalizations, compared to 4 at the quarterly level. While a general concern with higher levels of temporal aggregation is the loss of statistical power, Section IV shows that we can replicate statistically indistinguishable grading effects using J&L's model. In addition, because J&L models log(counts), it also appears to adjust for zero count observations, but such log transformations are known to be problematic (Buntin and Zaslavsky, 2004; O'hara and Kotze, 2010). Given the sparseness of foodborne hospitalizations, monthly data are likely to inflate zero count observations.

We flesh out one additional reason why monthly data — particularly given the tradeoff in terms of geographic aggregation to the three-digit ZIP code level — will not offer any significant benefits to assessing the effects of grading adoption. To understand this, we investigate the timing of how LA municipalities adopted grading. Figure 4 plots average values of m_{it} and v_{it} across LA ZIP codes from 1998-99, with dashed lines representing monthly data and solid lines representing quarterly data.[13] The quarterly and monthly values align well for m, as municipal adoption largely occurs by quarter. Mandatory grading became effective in the City of Los Angeles, home to 40% of the county population, on April 4, 1998, and in the City of Glendale, the third largest city in LA county, on March 24, 1998.

The right panel shows that the only substantial difference between monthly and quarterly adoption data occurs for v in the early months of 1998. This is because the county ordinance becomes effective January 16, decreasing v in January, but driving v up in February and March before subsequent municipal adoptions. (Recall that because ZIP codes include areas outside of LA county's public health jurisdiction, $v < 1$ after January 16 before any municipal adoption.) As a behavioral matter, it is highly implausible that voluntary grading effects should manifest themselves rapidly in the first two months of effectiveness. Restaurant inspections occur only one to three times per year (County of Los Angeles Public Health, 2017). It would hence take a considerable amount of time (1) for a sufficient number of restaurants to receive placards, (2) for consumers to infer that failure to post indicates a low sanitation score, and (3) for restaurants to improve sanitation practices due to this form of informational unraveling. If restaurants anticipate prospectively being graded, this also undermines the idea that monthly variation provides any substantial leverage. As Section IV shows, m and v are simply standing in for the years 1999 and 1998, respectively.

[13] See Equations 13 and 14 in Appendix L.

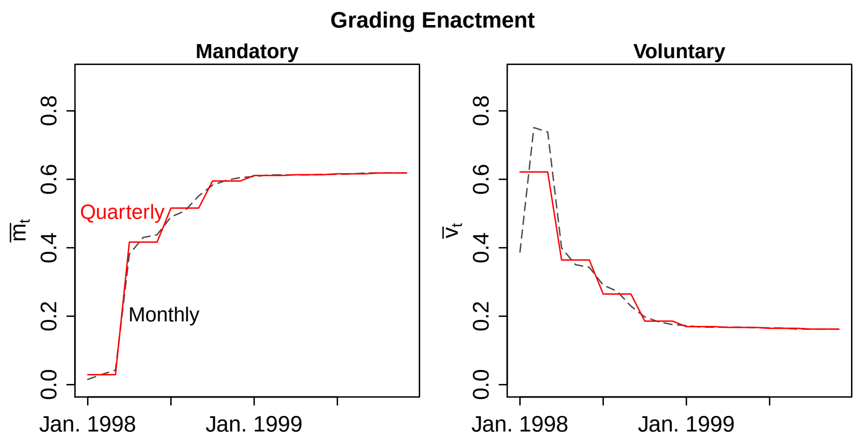

Figure 4: Timeline for grading enactment at monthly and quarterly level. The left panel plots the average value of m_{it} across LA 3-digit ZIP codes between January 1998 and December 1999. The dashed line shows the average value for m_{it} at the monthly level, while the solid line shows the average value for m_{it} at the quarterly level (where m_{it} values are calculated for 3-digit ZIP codes at the quarterly or monthly level according to equation 13 in Appendix L, and generalizing t to months for the monthly calculation). The right panel plots the average v values across LA 3-digit ZIP codes for the same time period.

Appendix F Triple Differences

F.1 Model specification

As described in Section I, J&L's analysis extends a DID analysis by adding a second control group of non-foodborne hospitalizations. Adding this control group could relax the DID assumption that there are no LA-specific time shocks coinciding with the introduction of grading in 1998. While the use of hospitalizations for all digestive system disorders as control illnesses may be questionable, it is worth considering what might comprise an ideal set of control illnesses. For example, we might be concerned that contemporaneous changes in LA's food supply, such as salmonella contamination in a local food supplier that serves both grocery stores and restaurants, could confound the impact of restaurant grading on foodborne hospitalizations.[14] To address this concern, we could split foodborne hospitalizations into two groups: those arising from restaurant meals, with an expected value across three-digit ZIP codes of $\bar{R}$, and those arising from home-cooked meals, with an expected value across three-digit ZIP codes of $\bar{H}$. We could then calculate two DID estimators in Equations 1 and 2. The triple differences estimator in equation 3 would adjust for the confounding effect of food supply safety changes in LA in 1998:

$$\delta_{\text{Restaurant}} = (\bar{R}^{\text{LA}}_{\text{After}} - \bar{R}^{\text{LA}}_{\text{Before}}) - (\bar{R}^{\text{CA}}_{\text{After}} - \bar{R}^{\text{CA}}_{\text{Before}}) \tag{1}$$

$$\delta_{\text{Home}} = (\bar{H}^{\text{LA}}_{\text{After}} - \bar{H}^{\text{LA}}_{\text{Before}}) - (\bar{H}^{\text{CA}}_{\text{After}} - \bar{H}^{\text{CA}}_{\text{Before}}) \tag{2}$$

$$\delta_{\text{Grading}} = \delta_{\text{Restaurant}} - \delta_{\text{Home}} \tag{3}$$

where superscripts LA and CA represent three-digit ZIP codes in LA and the rest of CA respectively, and the subscripts Before and After represent the time periods before and after grading was introduced in 1998 respectively. The triple differences estimator for δ_{Grading} requires four parameters corresponding to the four time-dependent differences that estimate $\delta_{Restaurant}$ and δ_{Home}. Equation 4 describes the model using fixed effects α_{ij} for (time-invariant) ZIP code-illness type

[14] This scenario recently occurred in the 2016 alfalfa sprout salmonella outbreak. Products from Sprouts Extraordinaire, which supplies both restaurants and grocery stores with alfalfa sprouts, were identified as a common source of reported illnesses (https://www.cdc.gov/salmonella/reading-08-16/index.html).

combinations, fixed effects τ_t for year-quarters, and three additional parameters for each remaining temporal comparison before and after grading:

$$\ln(a_{ijt}) = \alpha_{ij} + \tau_t + \beta_1(\text{LA}_i \times \text{After}_t) + \beta_2(\text{Restaurant}_j \times \text{After}_t) + \gamma_1(\text{LA}_i \times \text{After}_t \times \text{Restaurant}_j) \quad (4)$$

Table 8 describes how this parameterization relates to the expectations described in equations 1-3.

	Difference	Parameters
Restaurant	$(\bar{R}^{\text{LA}}_{\text{After}} - \bar{R}^{\text{LA}}_{\text{Before}})$	$\tau_{After} - \tau_{Before} + \beta_2 + \beta_1 + \gamma_1$
	$(\bar{R}^{\text{CA}}_{\text{After}} - \bar{R}^{\text{CA}}_{\text{Before}})$	$\tau_{After} - \tau_{Before} + \beta_2$
	$\delta_{\text{Restaurant}}$	$\beta_1 + \gamma_1$
Home	$(\bar{H}^{\text{LA}}_{\text{After}} - \bar{H}^{\text{LA}}_{\text{Before}})$	$\tau_{After} - \tau_{Before} + \beta_1$
	$(\bar{H}^{\text{CA}}_{\text{After}} - \bar{H}^{\text{CA}}_{\text{Before}})$	$\tau_{After} - \tau_{Before}$
	δ_{Home}	β_1
Net Treatment Effect	δ_{Grading}	γ_1

Table 8: Parameterization of triple differences estimate. The estimate is a difference between two DID's. The top panel is the DID for foodborne hospitalizations from restaurants. The second panel is the DID for foodborne hospitalizations from home meals. In principle, triple differences can adjust for LA-specific shocks (e.g., to the food supply).

J&L appears to pursue an identification strategy similar to the one outlined in Table 8, noting that "identification is based on time-series variation and cross-sectional variation provided by the presence of two control groups: California outside of Los Angeles and admissions for nonfood-related digestive disorders," the hospitalization control group analogous to our illnesses from home example (p. 439). However, J&L's specification only accounts for three of the four sources of temporal variation. We again rewrite m and v in the J&L specification as a single two-way interaction of LA (or proportion of ZIP population in LA) and post-1998:

$$\ln(a_{ijt}) = \alpha_{ij} + \tau_t + \beta_1(\text{LA}_i \times \text{After}_t) + \gamma_1(\text{LA}_i \times \text{After}_t \times \text{Food}_j) \quad (5)$$

We can now more clearly see that compared to Equation 4, β_2 is absent. Excluding β_2 results in a biased DID estimator for *both* foodborne and control hospitalizations. Equations 6-9 show the new estimators using J&L's control group of all other digestive disorder hospitalizations and assuming a balanced panel, where $\bar{F}$ represents the average over foodborne illness hospitalizations, $\bar{D}$ represents the average over all other digestive disorder hospitalizations, and $\overline{FD}$ represents the average over both foodborne illness and digestive disorder hospitalizations. Without a separate parameter for foodborne illness hospitalizations in CA after grading, the average trend over foodborne illness and digestive disorder hospitalizations in CA stands in for the CA component of each DID equation.

$$\overline{FD}^{\text{CA}}_{\text{After}} - \overline{FD}^{\text{CA}}_{\text{Before}} = \frac{1}{2} \times \left[(\bar{F}^{\text{CA}}_{\text{After}} - \bar{F}^{\text{CA}}_{\text{Before}}) + (\bar{D}^{\text{CA}}_{\text{After}} - \bar{D}^{\text{CA}}_{\text{Before}}) \right] \tag{6}$$

$$\delta_{\text{Foodborne}} = (\bar{F}^{\text{LA}}_{\text{After}} - \bar{F}^{\text{LA}}_{\text{Before}}) - (\overline{FD}^{\text{CA}}_{\text{After}} - \overline{FD}^{\text{CA}}_{\text{Before}}) \tag{7}$$

$$\delta_{\text{Digestive}} = (\bar{D}^{\text{LA}}_{\text{After}} - \bar{D}^{\text{LA}}_{\text{Before}}) - (\overline{FD}^{\text{CA}}_{\text{After}} - \overline{FD}^{\text{CA}}_{\text{Before}}) \tag{8}$$

$$\delta_{\text{Grading}} = \delta_{\text{Foodborne}} - \delta_{\text{Digestive}} \tag{9}$$

To provide the intuition of the resulting bias in $\delta_{Grading}$, Table 9 provides a stylized numerical example.[15] If both LA and CA drop by 10% in foodborne hospitalizations and increase by 10% in digestive disorder hospitalizations after 1998, the simple DID and the fully parameterized triple difference would correctly recover a 0% treatment effect (left and middle columns). By imposing trends to be identical in foodborne and digestive disorder trends in CA (but not LA), J&L's specification would estimate a foodborne DID of -10% and a digestive disorder DID of +10%, resulting in treatment effect estimate of -20% (right column). The J&L specification is the only one that would find a non-zero treatment effect.

The top two panels of figure 5 illustrate a stylized numerical example of the consequences of the J&L estimator under conditions similar to those observed in the OSHPD data. In the left panel, foodborne hospitalizations trend downwards from 1995-1999 for both LA and CA. In the right panel, digestive hospitalizations trend upwards for both LA and CA. Because the trends for CA's foodborne and digestive hospitalizations are of similar magnitude but opposite sign, constraining trends to be identical biases both towards zero, depicted in the dashed lines ($\overline{FD}^{\text{CA}}_{\text{After}} - \overline{FD}^{\text{CA}}_{\text{Before}}$).

[15] Here, we assume a balanced panel, as is the case for the LA data.

		Diff.-in-Diff.	Triple Diff.	J&L Spec.
Foodborne	LA	$\bar{F}^{\text{LA}}_{\text{After}} - \bar{F}^{\text{LA}}_{\text{Before}} =$ -10%	$\bar{F}^{\text{LA}}_{\text{After}} - \bar{F}^{\text{LA}}_{\text{Before}} =$ -10%	$\bar{F}^{\text{LA}}_{\text{After}} - \bar{F}^{\text{LA}}_{\text{Before}} =$ -10%
	CA	$\bar{F}^{\text{CA}}_{\text{After}} - \bar{F}^{\text{CA}}_{\text{Before}} =$ -10%	$\bar{F}^{\text{CA}}_{\text{After}} - \bar{F}^{\text{CA}}_{\text{Before}} =$ -10%	$\overline{FD}^{\text{CA}}_{\text{After}} - \overline{FD}^{\text{CA}}_{\text{Before}} =$ 0%
		$\delta_{\text{Foodborne}} =$ 0%	$\delta_{\text{Foodborne}} =$ 0%	$\delta_{\text{Foodborne}} =$ -10%
Digestive	LA		$\bar{D}^{\text{LA}}_{\text{After}} - \bar{D}^{\text{LA}}_{\text{Before}} =$ 10%	$\bar{D}^{\text{LA}}_{\text{After}} - \overline{D}^{\text{LA}}_{\text{Before}} =$ 10%
	CA		$\bar{D}^{\text{CA}}_{\text{After}} - \bar{D}^{\text{CA}}_{\text{Before}} =$ 10%	$\overline{FD}^{\text{CA}}_{\text{After}} - \overline{FD}^{\text{CA}}_{\text{Before}} =$ 0%
			$\delta_{\text{Digestive}} =$ 0%	$\delta_{\text{Digestive}} =$ 10%
Treatment Effect		$\delta_{\text{Grading}} =$ 0%	$\delta_{\text{Grading}} =$ 0%	$\delta_{\text{Grading}} =$ -20%

Table 9: Numerical illustration comparing DID, triple differences, and J&L specification. We use the notation of Equations 1-3 to refer to foodborne and digestive hospitalizations before or after 1998 across LA and CA. The J&L specification imposes homogeneous temporal effects for foodborne and digestive hospitalizations in CA: $(\overline{FD}^{\text{CA}}_{\text{After}} - \overline{FD}^{\text{CA}}_{\text{Before}})$. As a result, it is the only specification with a non-zero treatment effect in spite of the fact that LA and CA follow the the same trends for foodborne and digestive hospitalizations after 1998.

While a fully specified triple difference would find each DID estimate (between solid lines) to be zero, the J&L specification induces large DID estimates: the difference between the solid LA line and the dashed CA line widens both for foodborne and digestive disorders. Each of these differences magnifies the estimated treatment effect.

The J&L estimator can also fail to identify a true treatment effect by introducing bias of the opposite sign to the treatment effect. The bottom two panels of figure 5 illustrate this scenario. If we had indeed observed a larger drop in LA relative to CA for foodborne illness hospitalizations, and a larger drop in CA relative to LA for digestive disorder hospitalizations, we would estimate a large negative treatment effect with the correct triple differences specification. Yet, by replacing $(\overline{F}^{\text{CA}}_{\text{After}} - \overline{F}^{\text{CA}}_{\text{Before}})$ with $(\overline{FD}^{\text{CA}}_{\text{After}} - \overline{FD}^{\text{CA}}_{\text{Before}})$, the J&L estimator flips the sign of the foodborne DID to reflect a sharper decrease in CA rather than in LA. At the same time, by replacing $(\overline{D}^{\text{CA}}_{\text{After}} - \overline{D}^{\text{CA}}_{\text{Before}})$ with $(\overline{FD}^{\text{CA}}_{\text{After}} - \overline{FD}^{\text{CA}}_{\text{Before}})$, the J&L estimator attenuates the estimated drop in CA digestive disorder hospitalizations and underestimates the DID for digestive disorder hospitalizations. When subtracting these two biased DID estimates, now both small positive differences of equal magnitude, we obtain a treatment effect of zero rather than the large negative quantity we observe from the raw trends.

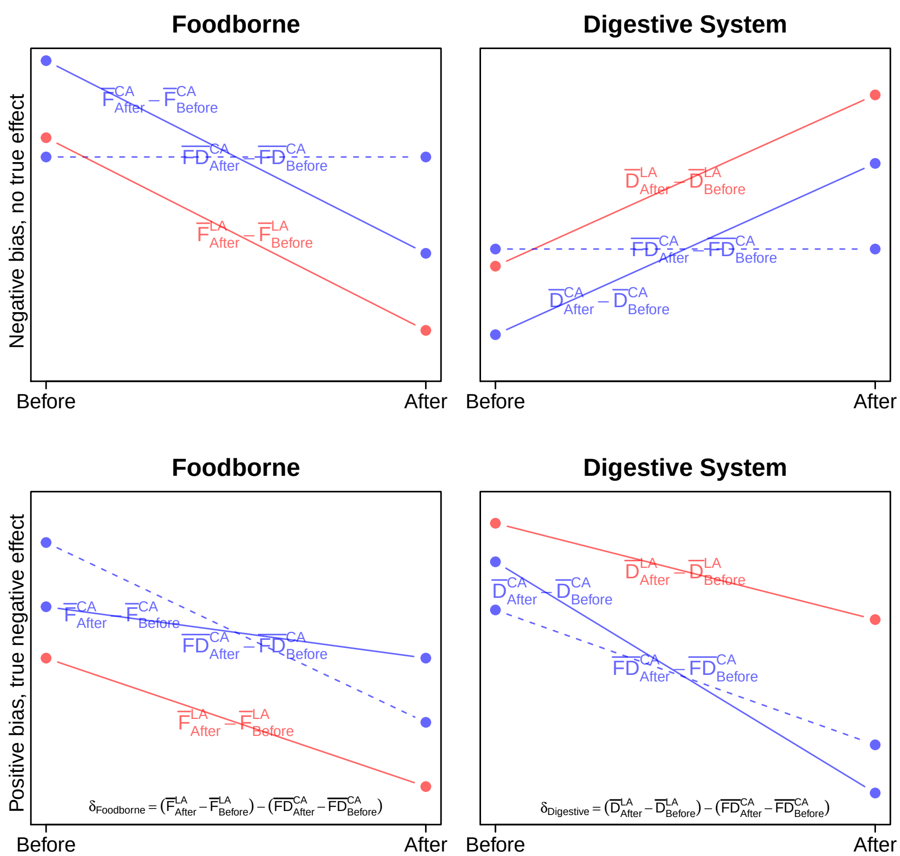

Figure 5: DID estimators for the J&L specification (dashed line) and the triple differences (solid line). J&L's specification replaces the control trend for foodborne illness hospitalizations in CA with the control trend for digestive disorder hospitalizations in CA. In the top two panels, because digestive disorder and foodborne illness hospitalizations go in opposite directions for CA over the observation window, constraining each trend to be equal to the average of the two biases the treatment effect upwards. In the bottom two panels, because CA digestive disorder hospitalizations drop more than any other series, the treatment effect is biased downwards.

Appendix G False Precision in Triple Differences Panel Designs

We now make the more general point that even a fully specified triple differences model introduces false precision into the treatment estimator with a panel as short as five years. A five-year observation window cannot properly characterize the cycle of region-and-time-specific outbreaks that drive temporal variation in foodborne hospitalizations. CDC data, for instance, indicate that the top 10% of outbreaks are responsible for 87% of hospitalizations for foodborne illnesses nationwide.[16] As a result, even a fully specified triple differences model with a short observation window has a high false positive rate.

		J&L Specification						Triple Differences					
Control	Treated	Type I Error			Power			Type I Error			Power		
group	ZIPs	5 yr	10 yr	15 yr	5 yr	10 yr	15 yr	5 yr	10 yr	15 yr	5 yr	10 yr	15 yr
CA	1-5	0.50	0.47	0.47	0.69	0.77	0.81	0.39	0.31	0.32	0.56	0.50	0.41
	6-10	0.19	0.24	0.35	0.68	0.93	0.94	0.11	0.11	0.10	0.43	0.47	0.40
	11-15	0.20	0.32	0.44	0.84	0.94	0.95	0.08	0.06	0.03	0.47	0.51	0.43
	16-20	0.22	0.37	0.51	0.93	0.97	0.98	0.07	0.05	0.02	0.48	0.42	0.29
Region	1-5	0.51	0.47	0.47	0.70	0.78	0.82	0.30	0.21	0.20	0.40	0.37	0.35
	6-10	0.19	0.24	0.36	0.68	0.93	0.94	0.10	0.07	0.07	0.35	0.41	0.42
	11-15	0.20	0.31	0.45	0.84	0.94	0.95	0.09	0.05	0.04	0.37	0.46	0.47
	16-20	0.22	0.36	0.51	0.93	0.98	0.98	0.08	0.03	0.01	0.42	0.47	0.46

Table 10: Type I error and power at a 20% effect size for randomly selected groups of one or more adjacent counties (excluding LA) after 4,000 random draws. Type I error is calculated as the percentage of placebo tests that reject the null hypothesis at a posited effect size of zero, and power is calculated as the percentage of tests that reject the null hypothesis at a posited effect size of 20%, for all treatment years between 1992 and 2004. Rates are broken out by the number of treated three-digit ZIPs in each posited treated county group (ZIPs), whether a regional control group was used (Region), the length of the observation window (5 yr, 10 yr, or 15 yr), and the model specification (J&L in Equation 6 or triple differences in Equation 7). All figures exclude San Bernardino for observation windows spanning 2004 or later, and Kern in observation windows spanning 2006 or later, to account for their subsequent restaurant grading adoptions. Minimum placebo test counts per cell are 50752, 34962, 13518, and 2878 for 1-5, 6-10, 10-15, and 16-20 treated three-digit ZIPs, respectively.

To illustrate this point, we conduct a series of Monte Carlo tests using the hospitalization data to calculate Type I error (given no effects) and power (given a 20% effect) for a variety of triple differences design choices. Table 10 presents Type I error and power at $\alpha = 0.05$ from 4,000 random draws of one or more adjacent counties (excluding LA) for which we posited treatment effects in the years between 1992 and 2004. Type I error is calculated as the percentage of placebo tests that

[16] This is based on CDC outbreak data from 1998-2015. Outbreaks are typically defined as illnesses by two or more persons due to eating a common food.

reject the null hypothesis at a posited effect size of zero. Power is calculated the percentage of tests that reject the null hypothesis at a posited effect size of 20%. We vary four design parameters: the number of treated three-digit ZIPs (rows),[17] the length of the observation window (columns),[18] the scope of the geographic control group (all of CA in top panel and region in the bottom panel),[19] and the model specification (J&L in left panel and triple differences in right panel).[20]

Our results confirm that the J&L model specification systematically underestimates the variance of the treatment effect, resulting in substantial Type I error. To understand why, recall that the J&L specification pools time fixed effects across foodborne and digestive hospitalizations for CA. Foodborne hospitalizations have much more temporal variability than digestive disorder hospitalizations, both because foodborne illnesses are outbreak-driven and the digestive category includes a much higher volume of cases. Pooling time fixed effects across the two categories hence underestimates the variance of the effect on foodborne hospitalizations.[21] To see the magnitude of this effect, at the smallest number of treated ZIPs, the shortest observation window, and CA as the control group, the J&L specification results in a Type I error rate of 50% (top left cell), when it should be 5%. Increasing the number of treated ZIPs reduces Type I error to 22%.[22] Increasing the observation period tends to *increase* Type I error, as outbreaks are folded into the point estimates without corresponding increase in standard errors due to pooling.[23]

Yet adding the missing interaction term to the J&L specification remains insufficient to provide a test with adequate size and power in the face of outbreaks. With 1-5 treated ZIPs, a five-year

[17] The median county in California has two three-digit ZIPs, and no county besides LA has more than five three-digit ZIPs. This skews the sampling distribution towards smaller numbers of treated ZIPs.

[18] For all tests, we keep the ratio of pretreatment to posttreatment years faithful to J&L's original design.

[19] We define regional control groups based on 1 degree latitude bands across the width of the state.

[20] As in Section V, we compare the J&L specification in Equation 6 or the fully specified triple differences in Equation 7.

[21] More formally, the within-cluster covariance of non-treated foodborne and digestive ZIP clusters 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 a N_g x 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 x 1 vector of OLS residuals for cluster g from Equation 6.

[22] This is because the asymptotics with the cluster-robust estimator are in N (Cameron and Miller, 2015; Conley and Taber, 2011).

[23] In this scenario, we would expect cluster-robust standard errors to inflate with more within-cluster time points due to outbreaks (see Hansen, 2007).

observation window, and CA as the control group, the triple differences model exhibits a Type I error rate of 39% (top left cell in triple differences panel). While increasing the number of treated ZIPs to the maximum reduces Type I error to 7%, power remains lower than 50% and decreases with a longer observation period. In contrast, using a regional control group at the maximum number of treated ZIPs reduces or preserves Type I error with no meaningful loss in power across observation periods (see bottom right row). Adding more time periods with a regional control group addresses some of the challenges of accounting for acute time- and region-specific outbreaks.

Yet as the 46% power with a 20% effect shows (bottom right cell), the approach remains underpowered to detect anything but large effects. Synthetic control methods (Abadie et al., 2010) provide another appropriately sized test to examine treatment effects on a single county, but Appendix K shows that these methods are similarly underpowered to detect moderate grading effects in foodborne hospitalizations.[24]

[24] Appendix K also finds no evidence to support grading effects in LA using synthetic control methods.

Appendix H Southern CA Spillover

In this Appendix, we address the concern that LA and Southern CA follow similar patterns around the adoption of grading because of spillover effects. In particular, we investigate a possible source of restaurant grading spillover effects between LA and Southern CA via the media coverage of LA's grading enactment. We obtained the coverage areas for KCBS-TV, the station that broadcasted the media exposé in November 1997.[25] Figure 6 shows the location of these towers along with their coverage contours. We calculated the percentage of each Southern California's three-digit ZIP code's population that was within the coverage area of these towers as a measurement of KCBS spillover.[26] If media spillover effects from the exposé drove the observed drop in foodborne illness across Southern California counties, then Southern California three-digit ZIP codes with more population within KCBS' coverage area should have experienced larger drops in foodborne illness compared to Southern California ZIP codes with little or no population within the coverage area.

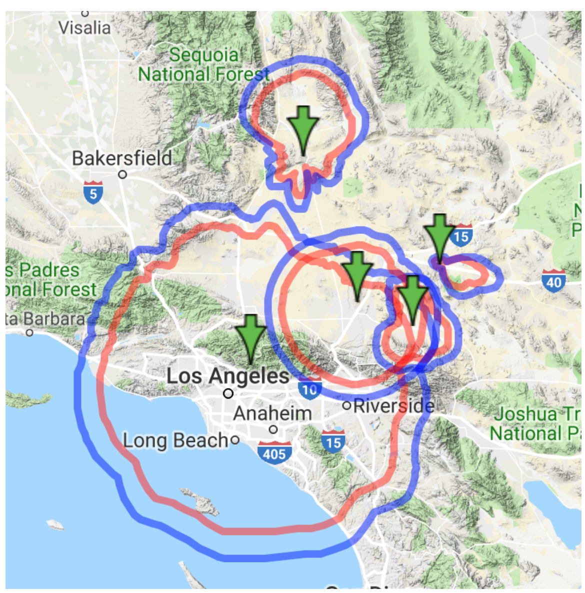

Figure 6: Coverage contour maps for television towers broadcasting KCBS. Green arrows represent the location of the KCBS-TV tower (closest to Los Angeles) and its four translator towers. Source: rabbitears.info.

[25]Source: https://rabbitears.info/market.php?request=station_search&callsign=9628

[26] As before, we exclude LA from Southern CA because it was actually treated with restaurant grading.

To test this, we estimated the model in Equation 10 on Southern California three-digit ZIP codes, where $\ln(a_{ijt})$ is the logged foodborne illness count in three-digit ZIP code i at year-quarter t for illness type j, α_{ij} are fixed effects for three-digit ZIP code illness type combinations, and τ_{tj} are fixed effects for year-quarter illness type combinations (equivalent to a full triple differences specification). The variable KCBS represents the proportion of the three-digit ZIP code's population within KCBS' coverage area, and the variable Post-Exposé is a dummy variable that is 1 in and after Q4 1997, and 0 otherwise. The coefficient γ_1 represents the effect on foodborne illness of being within the KCBS coverage area in Southern California after the exposé in Q4 1997.

$$\ln(a_{ijt}) = \alpha_{ij} + \tau_{tj} + \beta_1(\text{KCBS}_i \times \text{Post-Exposé}_t) + \gamma_1(\text{KCBS}_i \times \text{Post-Exposé}_t \times \text{Food}_j) \quad (10)$$

Table 11 shows that there was no difference in the change in foodborne illness rates after the KCBS media exposé between Southern California ZIP codes within and outside of KCBS' coverage area.

	1995-99	1993-2009
KCBS × Post-Exposé × Food	−0.05	−0.13
	(0.14)	(0.14)
KCBS × Post-Exposé	0.05	0.07
	(0.07)	(0.11)
N	800	2720
R^2	0.99	0.99

Table 11: The effect of KCBS coverage across Southern California ZIP codes. 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-quarter illness type combinations. We present results for the original observation period (1995-99) as well as an expanded observation period (1993-2009). The foodborne illness definition is consistent with J&L's original design. *p<0.10; **p<0.05; ***p<0.01

We also have reason to believe that the KCBS exposé was not widely covered outside of the KCBS network. We searched two different newspaper databases, ProQuest and NewsBank, for mention of the KCBS exposé between November 1997 and December 2009.[27] These databases jointly covered many of the major print news outlets in Southern California dating back to the

[27] Our search terms required the mention of KCBS and two variants of the exposé title, "Behind Kitchen Doors" and "Behind the Kitchen Door", case insensitive.

1990s. Table 12 presents the counts of mentions we found of the KCBS exposé within Southern California. Notably, we did not find any mention of the exposé in the San Diego Tribune, which has the third largest circulation in Southern California behind the Los Angeles Times and the Orange County Register.[28]

County	Publication	Article Count
Los Angeles	Los Angeles Times	6
Los Angeles	Long Beach Press-Telegram	2
Orange	Orange County Register	1

Table 12: Mentions of the KCBS exposé in major CA newspapers between November 1997 and December 2009.

[28] "Circulation numbers for the 25 largest newspapers." Associated Press (2012).

Appendix I Additional Robustness Checks

This Appendix assesses the robustness of our results to enactments of restaurant grading in other CA counties.

We researched each of CA's 58 counties by examining the county health department website and news reports where available. Based on these sources, we coded the enactment date of a restaurant grading system. We classified the type of system into one of two kinds: (1) a letter grading system, whereby letters are assigned based on inspection performance, as in LA; (2) a placarding system, where a typically color-coded placard indicates whether a restaurant passed its last inspection. Placarding systems typically also assign a "conditional pass," for low-scoring inspections. Because conditional passes often trigger return inspections, there is often little variability in observed grades. For instance, in November 2011, every single open restaurant in Santa Clara County had a green placard. Table 13 provides results. We can see a distributional difference between the two kinds of grading systems: early adopters were large jurisdictions (e.g., LA, San Diego, Riverside), which adopted letter grading systems. The most recent spate of adoptions, however, disproportionately adopted placarding systems. Two large counties, San Bernardino and Kern, which both border LA, adopted letter grading systems in 2004 and 2006, respectively.

We quantitatively examined the effect of restaurant letter grading in San Bernardino and Kern counties, as they are sufficiently populous and have over ten years of foodborne illness data post restaurant grading. These counties modeled their grading statutes after LA's,[29] resulting in nearly identical grading systems.[30] Figure 7 plots foodborne hospitalization and illness data for both of these counties against the rest of CA.[31] The left panels plot results for SB, which enacted grading in 2004. The variability in hospitalization rates (in the top panels) reinforces the importance of

[29]See Martin, Bakersfield Californian. (2006, February 9). Show which eateries make grade; Martin, H. (2004, June 9). S.B. county OKs plan for rating eateries. Los Angeles Times.

[30]Each of the three jurisdictions scores inspections on a 100-point scale. The cutoffs are identical for LA and San Bernardino, assigning an A for 90-100 points, B for 80-89 points, and C for 70-79 points. Kern uses the same cutoffs, except that 75-79 points correspond to a C. San Bernardino Ordinance § 33.1403; Kern County Public Health Services Department, Food Facility Grading Policy.

[31]Appendix I shows that none of the earlier results for LA are affected by the adoption of grading by SB and Kern counties.

County	Enactment year	Type	Population	County	Enactment year	Type	Population
Los Angeles	1998	L	9,818,605	Shasta			177,223
San Diego	1947	L	3,095,313	Imperial			174,528
Orange			3,010,232	Kings			152,982
Riverside	1963	L	2,189,641	Madera	2015	*L*[2]	150,865
San Bernardino	2004	L	2,035,210	Napa	2005	*L*[3]	136,484
Santa Clara	2014	P	1,781,642	Humboldt			134,623
Alameda	2012	P	1,510,271	Nevada			98,764
Sacramento	2007	P	1,418,788	Sutter			94,737
Contra Costa	2016	P	1,049,025	Mendocino			87,841
Fresno			930,450	Yuba	2017	P	72,155
Kern	2006	L	839,631	Lake			64,665
Ventura			823,318	Tehama			63,463
San Francisco	2007	P[1]	805,235	Tuolumne			55,365
San Mateo	2016	P	718,451	San Benito			55,269
San Joaquin			685,306	Calaveras			45,578
Stanislaus			514,453	Siskiyou			44,900
Sonoma	2016	P	483,878	Amador			38,091
Tulare			442,179	Lassen			34,895
Santa Barbara			423,895	Del Norte			28,610
Monterey			415,057	Glenn			28,122
Solano			413,344	Colusa			21,419
Placer	2016	P	348,432	Plumas			20,007
San Luis Obispo			269,637	Inyo			18,546
Santa Cruz			262,382	Mariposa			18,251
Merced			255,793	Mono			14,202
Marin	2014	P	252,409	Trinity			13,786
Butte	2014	P	220,000	Modoc			9,686
Yolo	2017	P	200,849	Sierra			3,240
El Dorado			181,058	Alpine			1,175

Table 13: All counties in CA with year of enactment of restaurant grading, if applicable, sorted by 2010 census population. Type is denoted by L for "letter grading system," where restaurants are assigned and required to post a letter grade, and P for "placarding system," where restaurants post a colored placard indicating whether the restaurant passed or conditionally passed the last health inspection. Grey highlighting indicates letter grading systems that are comparable to LA's.

[1] San Francisco's placarding system requires only that a symbol for scoring above 90% be posted, but that the inspection score card be available upon request. San Francisco Health Code, art. 8 §456.

[2] Madera County is a voluntary letter grade trial involving only three volunteer restaurants in 2015. Flanagan, K. (2015, February 14). Pilot food facility grade program continues in madera county. Sierra News.

[3] Napa Valley County assigns letter grades, but does not require establishments to post them. Goetting, J. (2006, March 23). Restaurant report availability and grades both improving in county. St. Helena Star.

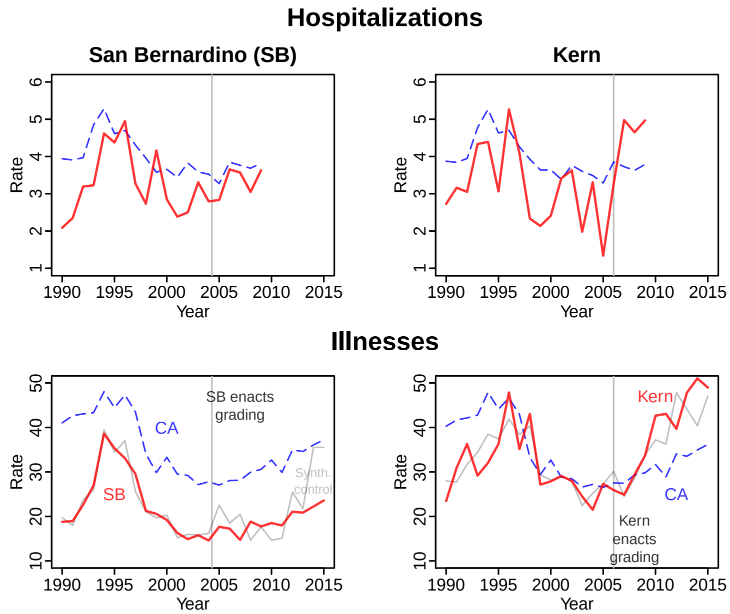

Figure 7: San Bernardino (SB) and Kern counties, compared to all other counties (CA), for foodborne hospitalizations (1993-2009) and illnesses (1990-2015). SB and Kern enacted restaurant grading in 2004 and 2006, respectively. Trends for CA exclude SB in the left panels and exclude Kern in the right panels. Light grey lines in illness panels depict synthetic control series. Although Kern experiences a sharp increase in its foodborne illness rate post-grading, a similar increase is observed in the other California counties comprising its synthetic control, likely due to adoption of culture-independent tests. p-values from permutation inference are 0.78 and 0.75 for San Bernardino and Kern, respectively.

examining reported illnesses (bottom panels). While SB and CA diverge in the 1990s, likely related to the salmonella outbreak, illness trends are parallel starting in 2000 with no evidence of any grading effect around 2004.[32] In fact, hospitalizations and illnesses appear to increase after Kern County's adoption in 2006.[33] Synthetic controls for illnesses (plotted in grey) confirm that there is no evidence of appreciable grading benefits in the most comparable CA counties that adopted grading since LA.[34]

We also show that the adoptions in San Bernardino and Kern do not affect any of the results for LA presented in the main body of the paper. We focus particularly on San Bernardino and Kern because they are the only other counties adopting letter grading systems in our observation period and both are otherwise included in the Southern CA group. First, Table 14 presents re-

[32]Results are the same plotting Southern California in lieu of the rest of CA.

[33]Similar jumps exist for other counties, including those in the synthetic control, with the most likely explanation being the increasing use of enzyme immunoassays and culture-independent tests for foodborne disease detection.

[34]p-values from permutation inference are 0.78 for SB and 0.75 for Kern.

sults by shortening the observation window to 1993-2003. This observation period excludes all subsequent adoptions of grading systems after LA. We observe comparable treatment effects using J&L's specification both for LA and for Southern CA. Second, using the full 1993-2009 observation window, we include treatment variables for San Bernardino and Kern counties corresponding to their enactment years. The left columns of Table 15 shows that the results for Southern CA remain comparable. Last, we remove San Bernardino and Kern county entirely, and again find comparable results in the right two columns of Table 15.

	Continuous treatment		**Binary treatment**	
	LA Treated	S. Cal. Treated	LA Treated	S. Cal. Treated
Foodborne × mandatory disclosure post-1998	-0.32^{***}	-0.34^{***}	-0.29^{***}	-0.35^{***}
	(0.05)	(0.09)	(0.05)	(0.08)
Foodborne × voluntary disclosure post-1998	-0.26^{***}	-0.38^{***}	-0.27^{***}	-0.32^{***}
	(0.09)	(0.12)	(0.06)	(0.09)
Mandatory disclosure post-1998 (Digestive)	0.05^{**}	0.17^{**}	0.06^{**}	0.17^{***}
	(0.03)	(0.07)	(0.03)	(0.06)
Voluntary disclosure post-1998 (Digestive)	0.09^{***}	0.07^{*}	0.04	0.07
	(0.03)	(0.04)	(0.03)	(0.05)
R^2	0.99	0.99	0.99	0.99
N	5,016	5,016	5,016	5,016

Table 14: Shortening observation window to 1993-2003 to assess sensitivity to San Bernardino and Kern enactments of restaurant grading in 2004 and 2006, respectively. 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. 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 1000 random draws from LA's observed m and v values, adjusting for boundary populations. For details, see Appendix L. Model (2) presents a representative model out of the 1,000 draws based on the lowest sum of squared distances from the median t-statistics for each parameter. In the Southern CA continuous models, the rejection rate for a significant m effect is 100% and for a significant v effect is 88.3%. *p<0.10; **p<0.05; ***p<0.01

Additionally, since we presented results from the 1993-2009 observation window for consistency across analyses that included and excluded campylobacter in Section V, we also test for sensitivity to using the longest observation window, 1983-2009. The validity of these results rests on the assumption that there was uniform adoption of the new campylobacter ICD code across hospitals between 1992 and 1993. The conclusions we draw from the longer observation period are consistent with the shorter observation windows: Southern California exhibits statistically indistinguishable treatment effects to LA's treatment effect (Table 16) because of the salmonella outbreak, and a fully specified triple difference estimates no statistically significant grading effects (Table 17).

	SB and Kern Dummies		**SB and Kern removed**	
	Continuous treatment	Binary treatment	Continuous treatment	Binary treatment
Foodborne × SC mandatory disclosure posttreatment	-0.40^{***}	-0.39^{***}	-0.34^{***}	-0.35^{***}
	(0.12)	(0.08)	(0.11)	(0.11)
Foodborne × SC voluntary disclosure posttreatment	-0.51^{***}	-0.32^{***}	-0.35^{**}	-0.30^{**}
	(0.13)	(0.09)	(0.17)	(0.12)
SC mandatory disclosure posttreatment (Digestive)	0.24^{**}	0.21^{***}	0.20^{**}	0.21^{**}
	(0.09)	(0.07)	(0.09)	(0.09)
SC voluntary disclosure posttreatment (Digestive)	0.06	0.07	0.09	0.06
	(0.05)	(0.05)	(0.07)	(0.07)
Foodborne × SB grading	−0.03	0.04		
	(0.13)	(0.13)		
Foodborne × Kern grading	0.31^{***}	0.39^{***}		
	(0.08)	(0.05)		
R^2	0.99	0.99	0.99	0.99
N	7,752	7,752	6,800	6,800

Table 15: Southern CA treatment models that add indicator variables for San Bernardino and Kern enactments of restaurant grading in 2004 and 2006 respectively (left columns), or drop all three-digit ZIP codes partially or fully in San Bernardino and Kern from the analysis (right columns). 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. 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 1000 random draws from LA's observed m and v values, adjusting for boundary populations. For details, see Appendix L. Models (1) and (3) 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. With the dummy variables for SB and Kern, the rejection rate for a significant m effect is 100% and for a significant v effect is 86% in the continuous model. With SB and Kern omitted, the rejection rate for a significant m effect is 100% and for a significant v effect is 53.4% in the continuous model. *p<0.10; **p<0.05; ***p<0.01

	Campylobacter Excluded		*Campylobacter Included*	
	J&L Spec.	**Triple Diff.**	**J&L Spec.**	**Triple Diff.**
	(1)	(2)	(3)	(4)
Foodborne × LA mandatory disclosure post-1998	-0.21^{***} (0.04)	-0.01 (0.07)	-0.14^{***} (0.04)	-0.03 (0.07)
Foodborne × LA voluntary disclosure post-1998	-0.19^{***} (0.07)	0.01 (0.09)	-0.17^{**} (0.08)	-0.06 (0.10)
Foodborne × CA post-1998		-0.18^{***} (0.05)		-0.10^{*} (0.05)
LA Mandatory disclosure post-1998 (Digestive)	-0.03 (0.04)	-0.13^{***} (0.05)	-0.08^{**} (0.03)	-0.13^{***} (0.05)
LA Voluntary disclosure post-1998 (Digestive)	0.01 (0.03)	-0.10^{**} (0.05)	-0.04 (0.03)	-0.09^{**} (0.05)
R^2	0.98	0.98	0.98	0.98
N	11,808	11,808	11,808	11,808

Table 16: Lengthening observation window to 1983-2009 to assess sensitivity to the observation period for omitting the two-way interaction term with and without campylobacter in the disease selection. 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. *p<0.10; **p<0.05; ***p<0.01

	Continuous treatment		**Binary treatment**	
	LA Treated	S. Cal. Treated	LA Treated	S. Cal. Treated
Foodborne × mandatory disclosure post-1998	-0.21^{***} (0.04)	-0.33^{**} (0.13)	-0.21^{***} (0.03)	-0.33^{***} (0.12)
Foodborne × voluntary disclosure post-1998	-0.19^{***} (0.07)	-0.31^{**} (0.15)	-0.17^{***} (0.05)	-0.28^{***} (0.11)
Mandatory disclosure post-1998 (Digestive)	-0.03 (0.04)	0.28^{**} (0.11)	-0.03 (0.03)	0.26^{***} (0.10)
Voluntary disclosure post-1998 (Digestive)	0.01 (0.03)	0.14 (0.11)	-0.03 (0.03)	0.12 (0.08)
R^2	0.98	0.98	0.98	0.98
N	11,808	11,808	11,808	11,808

Table 17: Lengthening observation window to 1983-2009 to assess sensitivity to the observation period for Southern California placebo analysis. 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. 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 1000 random draws from LA's observed m and v values, adjusting for boundary populations. Model (2) presents a representative model 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 Appendix L. In the Southern CA continuous models, the rejection rate for a significant m effect is 99.1% and for a significant v effect is 59.5%. *p<0.10; **p<0.05; ***p<0.01

Appendix J OSHPD and Notifiable Conditions Data

J.1 OSHPD Hospitalization Data

We employ public use versions of the Hospital Inpatient Discharge Data from CA's Office of Statewide Health Planning and Development (OSHPD). The data files include individual patient-level discharge records, including patient ZIP code, demographic information, diagnoses and treatments. OSHPD uses a set of masking rules to protect patient confidentiality.

We obtained annual files from 1983-2009 from the Stanford University library.[35] Data from 1983-1994 were stored on 8mm magnetic tape as fixed-width format ASCII files, so we built a computer system to recover and convert these files to proper format. From 1995 to 2009, files were available in properly delimited CSV format.

We aggregate the data into three versions.

1. County - year version. Although J&L's analysis uses three-digit ZIP codes as geographic units, as we illustrate in Appendix E, such units do not cleanly correspond to jurisdictional lines. We hence create a version of the dataset with county of residence based on five-digit ZIP code. Beginning in 1991, OSHPD matched the county of residence to each patient record based on five-digit ZIP code, using US Postal Service (USPS) ZIP code data. To obtain a record of county for 1983-1989, we applied the same process by imputing county of residence using historical five-digit ZIP code information. To do so, we obtained five-digit ZIP code to county mappings from 1980, 1990, and 2014.[36] For each five-digit ZIP code, we identified the earliest available mapping and assigned that county. A 1980 mapping, for instance, may not be available if a ZIP code was created after the 1980 census, in which case the 1990 mapping is used. Because five-digit ZIP codes can cross county lines, the earliest available county mapping included multiple counties for about 9% of five-digit ZIPs. To mimic OSHPD's process of using the primary county designated by USPS, we

[35] We could not locate the 1988 data files.

[36] The 1980 and 1990 mappings come from the Missouri Census Data Center (`http://mcdc.missouri.edu/applications/uexplore.shtml`). The 2014 mapping comes from the public version of the ZIP code database (`https://www.unitedstateszipcodes.org/zip-code-database/`).

selected the county that matched OSHPD's county mapping in 1995 for these border ZIPs. Any ZIP that appeared in the hospitalization data but not our county mapping (because of intercensal ZIP changes) was assigned the county of the hospital. Because OSHPD's data format transitioned in 1990, we have only three-digit ZIP codes available from 1990-1994. To match patient county of residence in 1990, we used the county of the hospital for every discharge. We use this version of the dataset to display rate time series for Figures 1, 2, 4, 9 in the main text and 1 of the Appendix.

2. Three-digit ZIP - quarter version. Our second version of aggregated OSHPD data corresponds most closely to that used in the analysis by J&L. Counts are aggregated to the three-digit ZIP code - quarter - illness type level. One complication that arises in this aggregation stems from OSHPD's masking rules. In instances where cell counts are low, to prevent individual identification, OSHPD uses a masking algorithm which can mean that some observations have years, but not quarters, identified. As a result a small number of cases are masked when using quarterly observations. Table 18 presents counts of foodborne and digestive system hospitalizations in the dataset at the yearly level (unmasked) and at the quarterly level (masked). The fourth and sixth columns present the percentage difference due to using quarterly vs. yearly data. The differences are slight, ranging from 2-2.7%. For comparability to J&L, this dataset is used for regression analyses in Tables 3 and 4 and Figure 7.

	Foodborne			**Digestive system**		
Year	Unmasked	Masked	% diff.	Unmasked	Masked	% diff.
1995	985	965	2.03	183,623	179,096	2.47
1996	1,091	1,061	2.75	188,648	183,989	2.47
1997	1,016	989	2.66	198,935	194,168	2.40
1998	992	969	2.32	205,899	201,117	2.32
1999	912	887	2.74	211,014	206,290	2.24

Table 18: Impact of OSHPD masking rules on counts by temporal unit of aggregation. The unmasked columns present counts for the year and the masked columns present counts where quarter is coded. % diff. indicates the proportion of observations lost due to masking.

3. County - quarter version. Our third version of the hospitalization data is aggregated at the county - quarter - illness level. For comparability to J&L, we use this dataset for the synthetic controls analysis of Section K.1. Because of the masking rules described above, this results in some cases being dropped.

Lastly, J&L also applies a filter to include only cases of patients admitted from home and that were unscheduled. Our analyses above do not apply this filter for three reasons. First, it is not substantively obvious why this hospitalization filter is appropriate for the research design. An admission from an emergency room at a different hospital due to complications from foodborne illness, for instance, would be excluded by the home filter. Second, this filter is not recorded in the same fashion prior to 1995 by OSHPD, making it impossible to extend the time series with the filter in place. It does not appear to be applied by Simon et al. (2005), as that version uses a 1993-2000 observation window. Third, the filter does not affect results. For instance, Figure 8 plots the ZIP code data from 1995-2009 with the filter on the x-axis against the data without the filter on the y-axis. The correlation coefficient is 0.997. Table 19 shows that the regression results

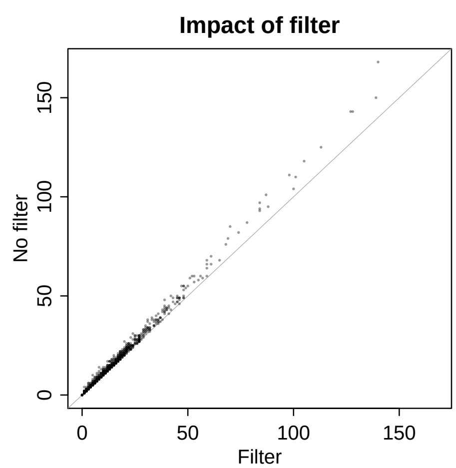

Figure 8: Impact of hospitalization filter. Each data point represents the count of foodborne hospitalizations for a ZIP-year cell. The x-axis applies the filter for unscheduled admissions from home and the y-axis includes all admissions. Counts are highly correlated.

are also comparable with and without the admissions filter.

J.2 Notifiable Conditions Data

Our data source for notifiable conditions (illness data) are state reports of communicable diseases. The reporting format changes somewhat across the years:

	Admission filter	No filter
Foodborne × LA mandatory disclosure post-1998	−0.32***	−0.31***
	(0.08)	(0.07)
Food-related × LA voluntary disclosure post-1998	−0.22***	−0.27***
	(0.08)	(0.08)
LA mandatory disclosure post-1998 (Digestive)	0.04	0.04*
	(0.03)	(0.03)
LA voluntary disclosure post-1998 (Digestive)	0.07	0.08**
	(0.05)	(0.04)
R^2	0.99	0.99
N	2,280	2,280

Table 19: J&L's filter of admissions from home as part of an unscheduled visit does not significantly change the coefficients in J&L's model. 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. *p<0.10; **p<0.05; ***p<0.01

Years	Title	Obs. period
1964-1980	Communicable Diseases reports	annual
1981-1989	California Morbidity reports	weekly
1990	California Morbidity report	annual
1991-2015	Communicable Diseases reports	annual

For 1981-89, we aggregated salmonellosis counts from weekly reports. In 1989, we estimate salmonella counts, as the reporting changed from weekly to bi-weekly in the middle of the year. For 1990, illnesses were aggregated for Humboldt and Del Norte counties, so we assigned counts proportionally based on the disease-specific 1991 allocation. To validate the data, we manually checked years where reported disease counts / rates appeared to be outliers. For instance, the LA spike in 2008 in Figure 4 corresponds to a salmonella javiana outbreak where contaminated fruit in a multisite preschool program affected 594 individuals. The only instance we were not able to confirm is a severe outlier for Riverside in 1974, where salmonella cases jump from 59 in 1973 to 248 in 1984 and back to 53 in 1975. The major outbreak appeared to occur in December 21, where 107 cases were reported, but no local newspapers reported in the outbreak. We exclude this anomaly as a likely data entry error, particularly because the Riverside outbreak in 1964 was widely reported.

For the 1990-2015 period, we collect information on salmonella, campylobacter, listeria, and vibrio. While it would be desirable to study a more comprehensive set of foodborne diseases, it is more important to exclude reporting effects. E. Coli O157, for instance, became mandatorily reportable in CA in 1996 (Belshé et al., 2003). While it was voluntarily reportable from 1993 (Bissell and Sebesta, 1995), there are likely to be substantial differences in reporting practices across areas over time. In LA, the E. Coli O157 rate in the three years prior to 1996 was 9.67, but this rose to 23 in the three years after mandatory reporting came into place. San Diego, on the other hand, reported relatively stable E. Coli rates.

For similar reasons, yersiniosis, staphylococcal food poisoning, and foodborne botulism are not included in the illness data. Yersiniosis data is available only after 2000 (compare, for example, Belshé et al. (2003) to Dooley and Smith (2015)). Staphylococcus aureus data is available only after 2007 (Dooley and Smith, 2015), and even then, reported data are limited to fatal or ICU cases. Botulism is excluded, because the early data does not distinguish between foodborne botulism and infant or wound botulism (see, e.g., Hastings et al. (1991)). Counts associated with these illnesses are small (e.g., CA had 81 cases of yersiniosis in 2001, compared to 4141 for salmonellosis and CA had 2 cases of foodborne botulism in 1992 compared to 5705 for salmonellosis), and therefore unlikely to affect findings.

Appendix K Synthetic Control Methods

This Appendix presents synthetic control methods as an alternative to traditional panel methods to estimate the effect of restaurant grading on foodborne illness. Subsection K.1 shows that the synthetic controls analysis finds no effect of restaurant grading in LA county. Subsection K.2 presents simulation evidence to show that permutation inference controls the false rejection rate, with statistical power to detect large effects. Subsection K.3 provides county weights and covariate weights used for the illness and hospitalization models. Subsection K.4 shows that the findings are robust to a wide range of disease selections and observation periods.

K.1 Synthetic control analysis

Synthetic control methods generalize DID approaches to potentially account for unobserved, time-specific confounding (Abadie et al., 2010, p. 495-96). Using the implementation of Abadie et al. (2011), we construct a linear combination of control counties (the synthetic control region) that is otherwise similar in pretreatment outcome time trends to LA (Abadie and Gardeazabal, 2003; Abadie et al., 2010). More formally, let $t \in \{1, \ldots, T\}$ index time, $j \in \{1, \ldots, J\}$ index control counties, and define weights $W = \{w_1, ..., w_J\}$, such that $0 \leq w_j \leq 1$ and $\sum_{j=1}^{J} w_j = 1$. Let $R^1 = \{r_1^1, \ldots, r_T^1\}$ be the vector for foodborne illness (or hospitalization) rates in LA and R^0 be a $T \times J$ matrix of rates for J control counties in T time periods. The synthetic control (with foodborne illness rate $r_t^0 = \sum_{j=1}^{J} R_{t,j}^0 w_j$ at time t) is constructed by selecting weights to minimize the pretreatment mean squared prediction error:

$$\text{MSPE}_{\text{pre}} = \frac{1}{n_{pre}} \sum_{t<1998} \left(r_t^1 - r_t^0\right)^2. \tag{11}$$

where n_{pre} is the number of pretreatment years.

The left panels of Figure 9 depict the time series of LA (solid) and the synthetic control region (dashed) for hospitalizations in the top and illnesses in the bottom left panels. We obtain reasonable balance in the pretreatment period. In the top left panel, the synthetic control exhibits the peaks

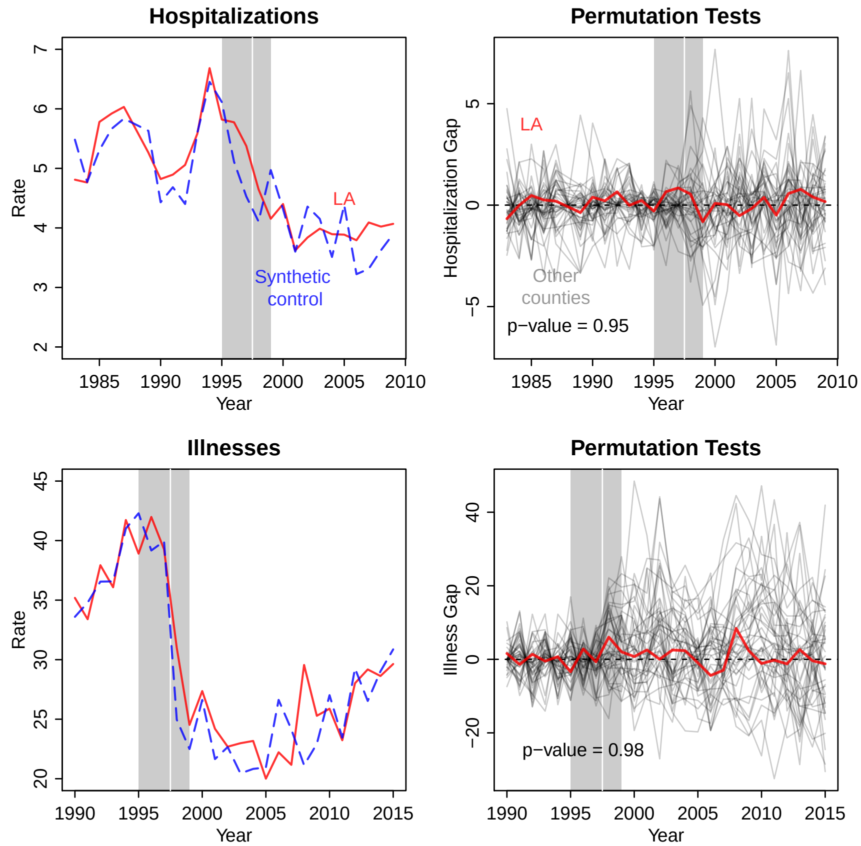

Figure 9: Synthetic control analysis. The top panels present analyses for hospitalization data from 1983-2009. The bottom panels present analyses for illness data from 1990-2015. The left panels present the LA time series (solid lines) against the time series of the synthetic control (dashed lines). The right panel present the year-by-year difference in rates between the treated and control units, with bold (red) line indicating the difference between LA and its synthetic control and thin grey lines indicating those for all other permutations of treatment across counties. There is no evidence that difference in illnesses or hospitalizations in LA exceeds that of the reference distribution (p-values are 0.95 and 0.98 for hospitalizations and illnesses, respectively).

from 1985-87 and 1993-94 salmonella outbreaks. Similarly, the synthetic control region in the bottom left panel for illnesses is largely comprised of San Diego and Orange counties, which makes substantive sense given the regional salmonella outbreak. After 1998, synthetic control region trends stay remarkably close to those of LA, exhibiting no evidence consistent with the 20% grading effect. For hospitalizations, LA's rate, if anything, trends above that of the synthetic control region. For illnesses, the time series do not exhibit any divergence.

To construct a test, we use permutation inference with a DID test statistic τ:[37]

$$\tau = \frac{1}{n_{post}} \sum_{t \geq 1998} \left(r_t^1 - r_t^0 \right) - \frac{1}{n_{pre}} \sum_{t < 1998} \left(r_t^1 - r_t^0 \right) \tag{12}$$

where n_{post} is the number of posttreatment years.

As a reference null distribution, we construct synthetic controls for every control county, calculating τ. As in Abadie et al. (2010), we exclude placebo counties from the reference distribution if MSPE_{pre} exceeds a threshold.[38] The right panels of Figure 9 plot (year-by-year) differences between LA and the synthetic control in bold line and placebo units in thin lines. The LA difference is close to zero for the entire posttreatment observation period, and deviations are substantially smaller than those for placebo counties. We fail to reject the null for illnesses ($p = 0.98$) and hospitalizations ($p = 0.95$).

K.2 Power

One concern with permutation inference may be that statistical power is low. As noted by Abadie et al. (2010), permutation inference does not address uncertainty from *sampling*, but rather uncertainty about the *counterfactual* outcome, and is closely related to randomization inference (Ho and Imai, 2006). To understand the power of the test, Table 20 presents rejection rates under effect sizes ranging from 0 to 100% in the hospitalization data for three estimators: a conventional DID estimator (with cluster-robust standard errors), permutation inference with a DID test statistic, and permutation inference with synthetic control methods. In the first three columns, we study power with an intervention in 1998 for hospitalization data from 1983 to 2009 with all CA counties. The first row shows the poor properties of conventional DID estimators with low numbers of treated units consistent with Conley and Taber (2011): the false rejection rate is 49%. While the likelihood of rejecting the null increases in the effect size, specificity is low. Permutation inference alone performs

[37] Permutation inference is better-suited when few groups adopt the treatment (see Abadie et al., 2010, p. 497; Cameron and Miller, 2015, pp. 349-50; Conley and Taber, 2011).

[38] We present results excluding counties where MSPE_{pre} exceeds 20 times that of LA, resulting in a reference distribution of 44 counties for illnesses and 41 counties for hospitalizations. Results are insensitive to the MSPE exclusion threshold.

poorly. Permutation inference with synthetic controls, however, performs reasonably well: the false rejection rate with an effect size of zero is around the α level, and the rejection rate increases with the effect size.

The right columns perform similar tests using only Southern CA counties. Because the reference distribution for a single intervention year is small, we use a five-year moving window from 1993-2009 to construct the reference distribution for permutation inference.[39] The parametric DID has similar properties, but permutation inference with a DID estimator performs relatively well. When the effect size goes from 0 to -0.5, the rejection rates increase more rapidly for permutation DID (from 0.04 to 0.60) than for parametric DID (0.38 to 0.84). Interestingly, permutation inference with synthetic controls appears to perform worse than permutation DID conditioning on Southern CA.

These results yield several takeaways. First, permutation inference both has the power to detect sizable effects, while also controlling the false rejection rate. Second, while synthetic controls may construct reasonable control groups, prior knowledge can improve a test. Our substantive knowledge of the salmonella outbreak (and other commonalities of Southern CA counties in terms of the food supply) may lead one to prefer the control group constructed with prior knowledge. Indeed, that is likely why permutation DID outperforms synthetic control methods in the Southern CA models. We caution, however, that the nature of foodborne outbreaks limits the ability to detect moderately sized effects.

K.3 Weights

We use the Synth software package in R to implement synthetic control matching (Abadie et al., 2011). The specific optimization is over two parameters: weights on K covariates ω_k, based on the predictive power of covariates on outcomes in the pre-treatment period, and weights on J control counties W, minimizing the difference between true covariate values and covariate values for the

[39] We note that five-year window may not provide a long enough pretreatment time series to construct a credible synthetic control group. We do so here to understand the application of permutation inference to the original observation period by J&L.

Effect size	California			Southern California		
	Parametric DID	Permutation DID	Permutation Synthetic	Parametric DID	Permutation DID	Permutation Synthetic
0.00	0.49	0.03	0.06	0.38	0.04	0.01
-0.10	0.49	0.03	0.08	0.37	0.10	0.05
-0.20	0.69	0.06	0.08	0.45	0.15	0.08
-0.30	0.80	0.09	0.27	0.60	0.27	0.15
-0.40	0.89	0.11	0.35	0.74	0.42	0.28
-0.50	0.97	0.14	0.53	0.84	0.60	0.32
-0.60	1.00	0.17	0.65	0.89	0.75	0.47
-0.70	1.00	0.17	0.78	0.90	0.92	0.61
-0.80	1.00	0.20	0.80	0.96	0.98	0.67
-0.90	1.00	0.20	0.82	0.98	0.98	0.78
-1.00	1.00	0.20	0.88	0.98	0.99	0.82

Table 20: Rejection rates for a parametric DID estimator (with cluster robust standard errors, ZIP fixed effects, and quarter fixed effects), permutation inference using the DID coefficient as the test statistic, and permutation inference using synthetic controls. Effect size indicates the posited effect of and each cell indicates the rejection rate at two-tailed $\alpha = 0.05$. The left three columns conduct the power analysis with all counties in CA (except LA), using 1998 as the treatment year and hospitalization data from 1983-2009, including campylobacter. The right columns conduct a power analysis with Southern CA counties only (excluding LA), using a 5-year moving window on hospitalization data from 1993-2009, including campylobacter.

synthetic control. Because J&L did not control for any covariates other than fixed effects, we focus on obtaining balance in the pretreatment outcome time series.

Table 21 provides weights that minimize the loss functions for the illness model in the left column and the hospitalization model in the right column. Due to changes in mandatory reporting of diseases, we use 1990-2015 as the observation period for the illness model. The principal foodborne diseases that were mandatorily reported during this period are salmonella, campylobacter, listeria, and vibrio. (E. Coli does not become reportable until 1996, and we re-run models that include E. Coli in Subsection K.4.) We can see that weights construct a highly plausible control group, comprised principally of Orange County and San Diego County, which have very similar pretreatment salmonella trends, as seen in Figure 4. The illness model therefore appropriately adjusts for the salmonella outbreak in Southern CA.

The hospitalization model includes the disease set studied by J&L, as well as campylobacter, from 1983-2009. (The International Classification of Diseases (ICD) revised its treatment of campylobacter in 1992, and we show in Subsection K.4 that the results are the same starting the observation period in 1993.) In contrast to the illness model, the hospitalization model places the

		Illness Model	Hospitalization Model
County Weights (w_j)	Orange	0.74	
	San Diego	0.12	
	Sacramento	0.06	
	Alpine	0.05	
	San Francisco	0.03	0.13
	Monterey		0.19
	Riverside		0.18
	Mendocino		0.18
	Butte		0.12
	Northwestern		0.11
	Imperial		0.05
	Yuba		0.03
	San Benito		0.01
Pretreatment Outcome Weights (ω_k)	1983-1984		0.08
	1985-86		0.08
	1987-89		0.24
	1990	0.03	
	1990-91		0.08
	1991-92	0.33	
	1992		0.02
	1993-94	0.16	0.13
	1994-95		0.16
	1995-96	0.42	0.17
	1996-97	0.06	
	1997		0.06

Table 21: Weights for synthetic control methods. The left column presents weights for the illness model and the right column presents weights for the hospitalization model. County weights (w_j) that construct the synthetic control unit are presented in the top panel and pretreatment outcome weights (ω_k) are presented in the bottom panel. OSHPD masks smaller counties, so Northwestern refers to Colusa, Glenn and Trinity counties combined.

highest weights on Monterey, Riverside, and Mendocino counties. As can be seen from the outcome weights in Table 21, this is largely driven by hospitalizations occurring around 1985-87. Each of these counties exhibited hospitalization increases around that time. Recall that the increase in LA at that time was caused by a drug-resistant strain of salmonella newport. The drug-resistance of that salmonella strand explains why the rate increase around 1985-87 is more prominent in the hospitalization data than in the illness data – compare, for instance, the dark and grey lines for LA in Figure 4. The salmonella newport outbreak was caused by a slaughterhouse that distributed meat both to LA and Northern CA (Spika et al., 1987, p. 567), which potentially makes the control group plausible. Intuitively, the synthetic control methods should track the geography of slaughterhouse/meatpacking distribution networks. These results reinforce the point that outbreaks can make identification challenging with foodborne illnesses and the importance of examining robustness to both illness and hospitalization data.

K.4 Robustness

Here we demonstrate that the synthetic control results are invariant to (a) the observation window, (b) disease selection, and (c) geographic distance. Across all results, we fail to reject the null hypothesis of no grading effects. Rows A of Table 22 present the baseline models presented in above, which use the 1983-2009 observation window for hospitalizations and the 1990-2015 observation window for illnesses. The table denotes which diseases are included, the two control counties with the highest weights in the LA synthetic control, and the test statistic τ and p-value.

Observation Window. Rows B of Table 22 use the original 1995-97 pretreatment window of J&L. The major downside to this short pretreatment window is that balance may not be particularly credible: an ideal DID would show a long, comparable pretreatment time series. Yuba County, for instance, is heavily weighted in these models because hospitalization rates are close to those of LA in 1995 and 1997. As the fifth smallest county in CA, it may not be a particularly credible control unit, which is why we focus on a model with a longer pretreatment time trend in Section K.1. We nevertheless examine this observation period because it allows us to apply a hospitalization filter used by J&L, which limits cases to patients admitted from home and unscheduled.[40] Row C presents hospitalization results using 1992 as the beginning of the observation period, as 1992 is the year that a separate ICD code (00843) was established for campylobacter (Healthcare Cost and Utilization Project, HCUP, 2016). Each of these models fails to reject the null hypothesis.

Disease Selection. We also investigate robustness of these results to different disease selections. First, for the 1995-2009 models, we examine the four combinations of selecting campylobacter and E. Coli. The reason for examining the impact of E. Coli separately is that mandatory reporting

[40]OSHPD data does not record these fields prior to 1995.

	Years	Campylobacter	E. Coli	Salmonella only	Principal County	Secondary County	τ	p-value	Geo. Model p-value
Hospitalizations									
A	1983-2009	Yes	Yes	No	Monterey (0.19)	Riverside (0.18)	-0.10	0.95	0.98
B	1995-2009	No	Yes	No	Yuba (0.30)	Del Norte (0.25)	0.06	0.97	0.48
		No	No	No	Yuba (0.60)	Kings (0.23)	-0.04	1.00	0.67
		Yes	No	No	Yuba (0.48)	Orange (0.35)	-0.09	0.86	0.27
		Yes	Yes	No	Orange (0.64)	Yuba (0.22)	-0.25	0.79	0.65
C	1992-2009	Yes	Yes	No	Northwestern (0.48)	Kings (0.33)	-0.30	0.78	0.61
D	1983-2009	No	No	Yes	San Diego (0.46)	Orange (0.44)	-0.07	0.84	0.93
Illnesses									
A	1990-2015	Yes	No	No	Orange (0.74)	San Diego (0.12)	0.89	0.98	0.29
B	1995-2015	No	Yes	No	Orange (0.65)	San Diego (0.30)	-0.43	0.95	-
		No	No	No	Orange (0.70)	San Diego (0.23)	0.54	0.96	-
		Yes	No	No	Orange (0.10)	Alpine (0.03)	-3.00	0.80	-
		Yes	Yes	No	Orange (0.87)	Alpine (0.08)	4.20	0.52	0.60
E	1990-2015	Yes	Yes	No	Orange (0.74)	San Diego (0.12)	-1.30	0.83	0.26
D	1990-2015	No	No	Yes	Orange (0.4)	San Diego (0.4)	-1.20	0.70	0.19

Table 22: Invariance of synthetic control results to observation window, disease selection, and geographic distance. The top panel presents models for hospitalization discharge data. The bottom panel presents models for illness data. Years indicate the observation window. Campylobacter and E. Coli indicate whether the disease is included in outcomes. Salmonella only indicates that the model is fit exclusively with salmonella outcomes. Principal and secondary counties indicate the control counties receiving the highest weights, with weights in parentheses. τ is the DID test statistic in Equation 12. The p-value is calculated via permutation inference, with placebo treatments for all control counties. Geo. Model refers to model where counties in synthetic control group must be within 250 miles of treated (or placebo treated) unit. For geographic models, the principal and secondary counties and τ are different than those reported for the main model. For reference, rows are labeled A (baseline), B (observation window starting in 1995, corresponding with J&L's observation period), C (observation window starting with 1992, when campylobacter was assigned a new ICD code), D (salmonella only) and E (full time period for illnesses data including E. Coli). E. Coli includes only E. Coli O157 for illness data, as that is the serotype subject to the earliest mandatory reporting. Models in rows B apply the J&L hospitalization filter to examine only unscheduled admissions from home. For rows D, we used a pretreatment MSPE cutoff of 22, as the MSPE for LA's synthetic control is quite low.

for E. Coli illnesses began statewide only in 1996 (Belshé et al., 2003). Voluntary reporting existed in the state since 1993 (Bissell and Sebesta, 1995), but because we find evidence of divergent reporting practices, our main illness model in Section K.1 includes only the four foodborne illnesses that were subject to mandatory reporting from 1990-2015 (salmonella, campylobacter, listeria, and vibrio). Row E adds E. Coli to the baseline illness model. In addition, we fit models in rows D that examines only salmonella, since salmonella is the predominant pathogen resulting in hospitalizations and reported illnesses. We fail to reject the null across all of these models.

Geographic Distance. As the Southern CA salmonella outbreak shows, there may be distinct reasons to construct a control group that is graphically proximate to the treated unit. Proximate counties may share similar food supply chains, climate (which can affect food handling practices), and restaurants, and hence be subject to similar shocks in foodborne risk. The principal and secondary counties in Table 22 suggest that illness data provides a more credible control group based on geographic proximity. Orange County, which borders LA, consistently receives the highest weight. Here we hence investigate the sensitivity to geographic proximity. We do so by restricting the donor pool of control units to be within 250 miles of the treated (or placebo treated) unit. The downside to this geographic model is that there are insufficient number of well-matched placebo counties when the pretreatment observation period begins in 1995. The last column of Table 22 presents p-values from these permutation tests, showing nearly identical results.

Appendix L Calculation for m and v

L.1 Los Angeles

Here we describe how we calculated m and v, the measures for the proportion of the population in each three-digit ZIP code subject to mandatory and voluntary enactment of restaurant grading respectively.

To compile information about municipal enactment, we began with the list of adoptees provided by LA.[41] For any municipality either not listed in the table or listed as not having enacted grading, we examined the city ordinance and contacted city officials to clarify the status. Particularly for smaller municipalities, we encountered some degree of confusion. For instance, while La Habra Heights adopted the LA County Public Health Code after the grading ordinance was passed, it does not appear to have a functioning grading system. As the cities of Long Beach, Pasadena, and Vernon operate health inspections independently from the county (and did not adopt LA's ordinance), they are coded as having enacted neither voluntary nor mandatory grading.

Calculating m and v requires splitting the population of a three-digit ZIP code into the components subject to each grading regime. We used our 1994 postal service dataset at the five-digit ZIP code level to map each five-digit ZIP code to every intersecting municipal area (incorporated or unincorporated) and county (Blodgett, 2017).[42] While useful for geographic mapping, this dataset does not specify the population that resides in each ZIP-municipality-county combination. To estimate this for the start of J&L's observation window, we obtained the 1995 population of all incorporated and unincorporated areas in each county in CA from the CA Department of Finance. We allocated the 1990 Census population of each five-digit ZIP code to municipalities or unincorporated areas according to the percentage of the county's population that resided in each municipality or unincorporated part of the county.[43] This method assumes that the municipal population distri-

[41] `http://publichealth.lacounty.gov/eh/misc/cityord.htm`

[42] For each ZIP-municipality pair, the dataset assigns one primary county and, if necessary, a secondary county. Only 136 ZIP-municipality pairs have a secondary county assigned. We use the primary county for the purposes of calculating m and v.

[43] For the purposes of this analysis, we treat individual unincorporated municipalities as a single population entity because they are subject to same grading regimes at the same times.

bution at the county level is a good proxy for the municipal population at the five-digit ZIP code level, but it is unclear how J&L divides the population of five-digit ZIP codes into municipalities without making a similar assumption.

To each incorporated ZIP-municipality population segment under the jurisdiction of the LA County Department of Public Health, we assigned a voluntary grading effectiveness date of January 16, 1998, and a mandatory effectiveness date as described above. To each unincorporated segment in LA county, we assigned a mandatory grading date of January 16, 1998, since unincorporated areas of LA county were subject to mandatory grade-card posting when the county ordinance went into effect. With these dates, we could calculate the number of days that each ZIP-municipality-county population segment was subject to each grading regime.

We then calculated m and v for ZIP code i and quarter t according to the following formula:

$$m_{it} = \frac{\sum_{k=1}^{M_i} \text{pop}_{ik}\ \text{days}_{kt}^{\text{mandatory}}}{\text{pop}_i\ \text{days}_t} \tag{13}$$

$$v_{it} = \frac{\sum_{k=1}^{M_i} \text{pop}_{ik}\ \text{days}_{kt}^{\text{voluntary}}}{\text{pop}_i\ \text{days}_t} \tag{14}$$

where M_i is the number of distinct jurisdictions (municipalities and unincorporated area) in ZIP code i, pop_{ik} is the population in jurisdiction k and ZIP code i (i.e., the ZIP code's population in the unincorporated county or an enacting municipality), and pop_i is the total population in ZIP code i. Furthermore, $\text{days}_{kt}^{\text{mandatory}}$ is the number of days that municipality k has mandatory grading in quarter t (which will be 0 for all jurisdictions before 1998, and will continue to be 0 for jurisdictions outside of LA after 1998), $\text{days}_{kt}^{\text{voluntary}}$ is the number of days that municipality k has voluntary grading in quarter t, and days_t is the number of days in quarter t. For three-digit ZIP codes entirely within LA county, m and v sum to one. For three-digit ZIP codes that straddle the border of LA county, m and v sum to the proportion of the population that resides within LA county boundaries, as these ZIP codes contain jurisdictions that are outside of LA county and are therefore subject to neither mandatory nor voluntary enactment.

L.2 Southern California

In order to assign placebo treatments to Southern CA, we could not simply randomly sample m and v for LA, because five of the 18 LA three digit ZIP codes cross county lines. For problems with these units of analysis see Appendix E. It would make no sense to assign $m = 0.7$ to Southern CA, when that LA ZIP code is coded as 0.7 only because 30% of the ZIP code is outside of LA.

We hence adjusted m and v constraining k to include only jurisdictions fully within LA county. This assured that m and v vectors always summed to one for each three-digit ZIP code in LA county within each quarter, hence retaining the crucial information about timing and population distribution of municipal enactment, but removing the influence of county borders. For ease of exposition, we call these adjusted m and v vectors m^{LA} and v^{LA}, respectively.

Border ZIP codes also exist for Southern CA, so our placebo tests adjust each simulated placebo treatment m^{LA} and v^{LA} for the proportion of the three-digit ZIP code i that is in Southern CA. We calculate the proportion as:

$$\text{prop socal}_i = \frac{\sum_{n=1}^{S} \text{pop}_{ni}}{\sum_{n=1}^{N} \text{pop}_{ni}} \tag{15}$$

where S is the number of five-digit ZIP codes within i that fall within Southern CA placebo county borders according to our 1994 postal service dataset, N is the total number of five-digit ZIP codes within i, and pop_{ni} is the 1990 Census population at the five-digit ZIP code level.

We perform the Southern CA placebo tests by assigning placebo treatments 1,000 times via the following process:

1. For each three-digit ZIP code i partially or fully within Southern CA placebo counties, randomly sample an LA three-digit ZIP code j with replacement and assign its m_j^{LA} and v_j^{LA} vectors to three-digit ZIP i;

2. Multiply m_i^{LA} and v_i^{LA} by the population proportion of three-digit ZIP i that resides within any of the Southern CA placebo counties (from Equation 15);

3. Set m^{LA} and v^{LA} to zero for all three-digit ZIP codes fully outside of Southern CA placebo counties.

4. Estimate the J&L specification using m^{LA} and v^{LA} in place of m and v.

Out of these 1,000 simulations, we report a representative model based on the median t-statistic for γ_1 in Table 3 and report the rejection rate for all 1,000 models in the caption. The process spelled out above avoids confounding LA borders with Southern CA placebo county borders.

Appendix M Population Data

In this Appendix, we describe the data sources and process that we used to augment the OSHPD hospitalization and the reported illness data with county time-varying population estimates. This process is straightforward for the illness dataset at the county level, but presents considerable difficulties when ZIP codes are the primitive units of analysis (for discussion about why ZIP codes are poor units of analysis, see Appendix E).

M.1 County Time-Varying Population Data

We obtained CA county population data for 1964-2015. We used census estimates, but for the years 1964-1969, we obtained our population estimates from the CA Department of Finance (State of California, Department of Finance, 2017). For the years between 1970 and 2015, we used intercensal estimates from the National Bureau of Economic Research (Roth, 2007).

M.2 ZIP Code Population Data

In order to calculate m and v, we need to know the proportion of a ZIP code's population that is in LA County. We obtained a ZIP code to county mapping in 1994, but population estimates for ZIP codes added since 1990 are missing. We describe here the process to estimate population for those units.[44]

Types of ZIP Codes. To create our population dataset, our goal is to represent geographic units of residence. The Postal Services ZIP codes, however, can include "point ZIPs," which do not represent a geographical area (e.g., PO boxes, ZIP codes for businesses), as well as physical ZIPs, which represent a geographical area. Because the first three digits represent the postal office processing center, we treat a point ZIP in the OSHPD data as located within the same three-digit physical ZIP code.

[44] It is unclear how J&L addressed these dynamic changes in ZIP codes. One possibility is that J&L calculated the proportion in LA county at the three-digit ZIP code level based exclusively on 1990 Census and ZIP code information. This would miss dynamic changes from ZIP code realignments.

Sources. Our base dataset was 1990 census population data for 1523 CA five-digit ZIP codes, from the CDC website (Bureau, 1992). Population estimates provided through this portal correspond to the 1990 census tape STF3B.[45] Because J&L spans 1995-1999, and ZIP code boundaries change over time (US Postal Bulletins Consortium, 2014), this dataset did not include 1990 population for all ZIP codes in the patient discharge data. We hence obtained a 1994 postal service dataset of ZIP codes from the University of Missouri's Dexter (Data Extractor) (Blodgett, 2017). In order to establish if a ZIP code came into existence between 1990 and 1994, we used the Digitized US Postal Bulletins database (US Postal Bulletins Consortium, 2014).

Missing ZIP Codes. We identified 126 physical ZIP codes in existence in 1994 that were absent from the 1990 census population data. This absence could be due to two reasons. First, the ZIP code may have existed in 1990, but was pooled with another ZIP code when reporting population estimates.[46] For the 1990 census, the Census Bureau worked with private vendors, but when census blocks crossed ZIP codes, which happened frequently, the entire block was assigned to a single ZIP code (Missouri Census Data Center, 2010). The second possibility is that the ZIP code came into existence between 1990 and 1994, in which case this should be documented in the Postal Bulletins. We found this to be the case for 11 out of 126 ZIP codes. Because both scenarios mean that population was imputed to a different ZIP code in 1990, we developed a method to redistribute population counts to 1994 ZIP codes.

ZIP Code Blocks. We reconstructed the mapping that the Census Bureau had used to group ZIP codes when reporting 1990 population estimates. Specifically, we assigned all 1994 ZIP codes to "blocks" that consisted of:

1. <u>Identical ZIP Codes Blocks.</u> The 1994 five-digit ZIP code alone, if

[45] We used this tape as it contains ZIP code information, whereas STF3A is at the county, census tract and census block level.

[46] The 1990 census was the last decennial census to be conducted before the introduction of ZCTAs (ZIP Code Tabulation Areas) (Missouri Census Data Center, 2010), which standardized the reporting of census data at an approximate ZIP code level.

(a) that ZIP code was included in both the 1990 census and the 1994 USPS list, and

(b) no adjacent five-digit ZIP codes from the same three-digit ZIP code were missing from the 1990 census data;

2. Blocks with Newly Created ZIP Codes. The 1994 five-digit ZIP code and its corresponding 1990 three-digit ZIP code, if the ZIP code came into existence between 1990 and 1994 according to the Postal Bulletins;

3. Blocks with Pooled ZIP Codes. The 1994 five-digit ZIP code plus adjacent five-digit ZIP codes that existed in 1990 in the same three-digit ZIP code[47], if the adjacent ZIP code was listed as created in the postal bulletin for the five-digit ZIP code (so that it was likely pooled with neighboring ZIP codes in the 1990 process).

If multiple 1994 five-digit ZIP codes missing from the 1990 census shared adjacent five-digit ZIP codes, we pooled the entire set of 1994 ZIP codes together, provided they were in the same three-digit ZIP code. This resulted in 295 ZIP codes from the 1622 physical 1994 ZIP codes being assigned to ZIP code blocks larger than the ZIP code by itself. Assuming that the census correctly captured the true population of CA in 1990, we assigned population to a ZIP code block by summing any non-zero 1990 population estimates for ZIP codes in the ZIP code block.

Population Imputation. The resulting dataset contains 1649 observations, representing 1523 ZIP codes from the 1990 census and 126 previously missing 1994 ZIP codes. For the former, we can readily use 1990 population estimates. For the previously missing ZIP codes, we needed to estimate the population based on the known 1990 ZIP block population, pop_{bl}, where bl indexes blocks. To do so, we use the known 1995 total hospitalization counts at the five-digit ZIP code level, hosp.1995_i, where i indexes ZIP codes. For the 1523 ZIP codes where hosp.1995 and pop are observed, we regress pop against hosp.1995. Figure 10 shows that this simple linear fit works well.

[47] We used the 2010 USPS ZIP code map to determine geographic boundaries.

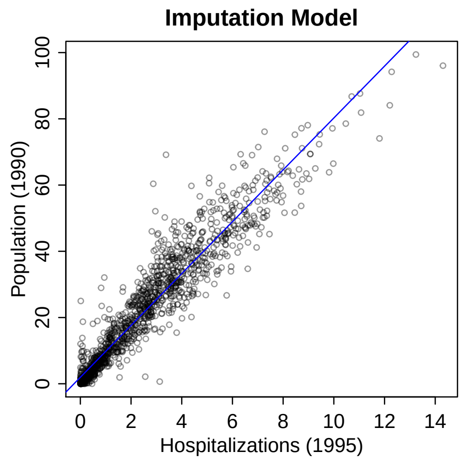

Figure 10: Population imputation model for missing ZIP code population. The x-axis presents hospitalizations (1995) in thousands, available for all five-digit ZIP codes. The y-axis presents population estimates (in thousands) from the 1990 Census, available for 1523 ZIP codes. We used this model to infer population for newly created ZIP codes or previously pooled ZIP codes.

We use this model to predict pop for the 126 previously missing ZIP codes, $\widehat{\text{pop}}_i$. To ensure consistency in block totals, we then allocate the fraction of pop_{bl} to each ZIP code i using the predicted ratio:

$$\frac{\widehat{\text{pop}}_i}{\widehat{\text{pop}}_{bl}} \text{pop}_{bl}, \tag{16}$$

subject to:

$$\widehat{\text{pop}}_{bl} = \sum_{i \in \text{bl}} \widehat{\text{pop}}_i. \tag{17}$$

References for Appendix B

Jurisdiction	Source
State of Florida	Torres, MN (2007, July 30). Could your restaurant earn an A? Consumer advocates push simpler inspection ratings. South Florida Sun - Sentinel.
Los Angeles County, CA	Rong-Gong Lin (2010, September 14). On the A-B-C bandwagon: food trucks may soon be graded with letter placards. Los Angeles Times.
New York City, NY	Collins, G. (2010, Mar 17). City restaurants required to post cleanliness grades. New York Times.
	Kristof, K. (2010, Jul 29). Ballpark update: stadiums must tell about toxic food. CBS News.
State of Connecticut	Pytka, E. (2005). Publicly posted health inspection grade cards. Connecticut General Assembly Office of Legislative Research Report No. 2005-R-0403.
Orange County, CA	Orange County Grand Jury Report. (2007-2008). Restaurant inspections – what no one is telling you.
	Johnson, B. A. (2014, January 8). A major failing. Orange County Register.
San Bernardino County, CA	Martin, H. (2004, June 9). S.B. County Oks Plan for Rating Eateries. Los Angeles Times.
King County, WA	Phuong Cat, L. P. (2004, July 9). Restaurant inspections skipped, fines for infractions infrequent; post check results so more can be learned about violators, critics say. Seattle Post - Intelligencer.
Santa Clara County, CA	Susko, J., Putnam, J., and Villareal M. (2013, February 6). Silicon Valley restaurants: no grades, no accountability. NBC Bay Area News.
Sacramento County, CA	MS Enkoji Bee, S. W. (2005, September 20). Keeping score on cleanliness reports of unsanitary conditions spurred a new disclosure law for restaurants in los angeles county; diners depend on the letter grades, and now sacramento county is considering a similar system. Sacramento Bee.
Cuyahoga County, OH	CantonRep.com. (2010, February 7). Cleveland eyes restaurant grading system.
	Spector, K. (2010, February 7). Grading system for Cuyahoga County restaurants under discussion. Cleveland Plain Dealer.
Allegheny County, PA	Allegheny County Board of Health. (2014, March 3). Update on restaurant grading system. Approved Meeting Minutes, Allegheny County.
	Sabatini, P. (2008, September 7). Dining Dangers Consumer Watchdog Wants Safety Inspectors to Post Letter Grades in Restaurant Windows. Pittsburgh Post - Gazette.
Contra Costa County, CA	Richards, S. (2016, April 19). Contra Costa's new color-coded food inspection grading placards. Mercury News.
	Felsenfeld, P. (2005, February 4). Health inspection info may be posted. Contra Costa Times.
Pima County, AZ	Stauffer, T. (2008, August 18). Force eateries to post inspection grades? Tucson Citizen.
Kern County, CA	The Bakersfield Californian. (2006, February 9). Show which eateries make the grade.
	Price, R. (2006, February 10). Restaurant inspection system needs upgrade. Bakersfield Californian.
Ventura County, CA	Ventura County Grand Jury. (2008-2009). Is your favorite restaurant clean?
San Francisco, CA	Mayor Ed Lee announces open data partnership with Yelp to offer restaurant health inspection scores to improve public health, transparency. (2013, January 17). PR Newswire.

San Mateo County, CA	San Mateo County Civil Grand Jury Report. (2003-2004). Food inspection in San Mateo County.
Boston, MA	Rocheleau, M. (2015, November 9). Boston to assign restaurants letter grades. Boston Globe.
Stanislaus County, CA	Milbourn, T. (2004, July 28). Modesto, Calif., health inspector has appetite for food safety. Knight Ridder Tribune Business News.
Minneapolis, MN	Roper, E. (2016, April 13). Here are the A, B, Cs of why Minneapolis inspectors don't grade restaurants. Minneapolis Star Tribune.
Marin County, CA	O'Malley, M. (2006, January 18). Better restaurant monitoring needed. Marin Independent Journal.
Muskegon County, MI	McVicar, B. (2011, August 1). Some states grade restaurants based on health inspections. Mlive.
United Kingdom	Food Standard Agency. (2008, May 20). UK-wide scores on the doors scheme on hygiene standards in food businesses.
New South Wales, Australia	NSW Food Authority. (2011). "Scores on Doors" Pilot Evaluation Report. New South Wales, Australia.
Hong Kong, China	Legislative Council Secretariat. (2008). Food hygiene information system in selected places. Hong Kong, China.
New Zealand	Filion, K. and Powell, D. (2011). "Designing a national restaurant inspection disclosure system for New Zealand." Journal of Food Protection 74(11):1869-1874.
Hamilton, Ontario, Canada	Vallance-Jones, F. (2007, April 23). Expert: Hamilton signs 'misleading'; one year after los angeles introduced its restaurant grading system, hospitalizations for food-borne illnesses dropped by 20 per cent. Spectator.

Table 23: References for J&L citations described in Table 3

Institution	Area	Source
Federal Trade Commission (FTC)	Consumer protection	Federal Trade Commission (2016, September 15). Putting disclosures to the test.
Environmental Protection Agency (EPA)	Environment	Environmental Protection Agency (2011, January 18). Benefits of environmental information disclosure proceedings.
Federal Communications Commission (FCC)	Communications	Federal Communications Commission (2013, September 27). Improving the resiliency of mobile wireless communications networks.
Board of Governors of the Federal Reserve System	Consumer finance	Kroszner, Randall S. (2007, May 23). Creating more effective consumer disclosures.
Organisation for Economic Co-operation and Development (OECD)	Regulatory enforcement	Blanc, Florentin (2013). Inspection reforms: Why, how, and with what results.
CA Department of Toxic Substances Control	Product safety	Kahn, Mathew E. and DeShazo, J.R. (2010, September 8). Economic analysis of California's green chemistry regulations for safer consumer products.
CA State Assembly Committee on Health	Health	California Healthcare Foundation (2009, February 17). What is transparency in health care and why does it matter?

Table 24: References for wider policy discussions citing J&L

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.

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2011. "Synth: An R package for synthetic control methods in comparative case studies." *Journal of Statistical Software* 42(13):1–17.

Abadie, Alberto, and Javier Gardeazabal. 2003. "The Economic Costs of Conflict: A Case Study of the Basque Country." *The American Economic Review* 93(1):113–132.

Acheson, David, and Elizabeth L. Hohmann. 2001. "Nontyphoidal salmonellosis." *Clinical Infectious Diseases* 32(2):263–269.

Anding, Jenna D., Chris Boleman, and Britta Thompson. 2007. "Self-Reported Changes in Food Safety Behaviors among Foodservice Employees: Impact of a Retail Food Safety Education Program." *Journal of Food Science Education* 6(4):72–76.

Belshé, S. Kimberley, Diana M. Bontá, Mark Starr, Duc Vugia, and Stanley R. Bissell. 2003. "Communicable Diseases in California 1999-2000." Technical report, State of California - Department of Health Services.

Bissell, Stanley, and Shu Sebesta. 1995. "Communicable Diseases in California 1993."

Blodgett, John. 2017. "Dexter – Data Extracter Tool."

Boehnke, Richard H., and Catherine Graham. 2000. "International survey on public posting of restaurant inspection reports, and/or grade card posting schemes based upon health inspections." *Region of Ottawa-Carleton Health Department ed., Ottawa, Canada* .

Buntin, Melinda Beeuwkes, and Alan M. Zaslavsky. 2004. "Too much ado about two-part models and transformation?: Comparing methods of modeling Medicare expenditures." *Journal of Health Economics* 23(3):525–542.

Bureau, US Census. 1992. "CDC. 1990 Census Data on Housing and Population." Technical report, US Census Bureau.

Cameron, A. Colin, and Douglas L. Miller. 2015. "A practitioner's guide to cluster-robust inference." *Journal of Human Resources* 50(2):317–372.

Centers for Disease Control and Prevention. 2015. "How People Are Infected | Anthrax | CDC." Technical report, Centers for Disease Control and Prevention.

Conley, Timothy G., and Christopher R. Taber. 2011. "Inference with 'difference in differences' with a small number of policy changes." *The Review of Economics and Statistics* 93(1):113–125.

Cotterchio, Michelle, Julia Gunn, Thomas Coffill, Patricia Tormey, and M. Anita Barry. 1998. "Effect of a manager training program on sanitary conditions in restaurants." *Public Health Reports* 113(4):353.

County of Los Angeles Public Health. 2017. "LA County Public Health - Environmental Health." Technical report, County of Los Angeles Public Health.

Croxen, Matthew A., Robyn J. Law, Roland Scholz, Kristie M. Keeney, Marta Wlodarska, and B. Brett Finlay. 2013. "Recent advances in understanding enteric pathogenic Escherichia coli." *Clinical Microbiology Reviews* 26(4):822–880.

Cruz, Miguel A., Dolores J. Katz, and Juan A. Suarez. 2001. "An assessment of the ability of routine restaurant inspections to predict food-borne outbreaks in Miami-Dade County, Florida." *American Journal of Public Health* 91(5):821.

Dooley, Diana S., and Karen L. Smith. 2015. "Yearly Summaries of Selected General Communicable Diseases in California, 2001-2010." Technical report, California Department of Public Health.

Egan, M. B., M. M. Raats, S. M. Grubb, A. Eves, M. L. Lumbers, M. S. Dean, and M. R. Adams. 2007. "A review of food safety and food hygiene training studies in the commercial sector." *Food Control* 18(10):1180–1190.

Food and Drug Administration. 2012. *Bad bug book, foodborne pathogenic microorganisms and natural toxins*. FDA, second edition.

Goldman, Lynn R., P. H. Daniel F. Smith, Raymond R. Neutra, L. Duncan Saunders, Esther M. Pond, James Stratton, Kim Waller, Richard J. Jackson, and Kenneth W. Kizer. 1990. "Pesticide Food Poisoning from Contaminated Watermelons in California, 1985." *Archives of Environmental Health: An International Journal* 45(4):229–236.

Gould, L. Hannah, Linda Demma, Timothy F. Jones, Sharon Hurd, Duc J. Vugia, Kirk Smith, Beletshachew Shiferaw, Suzanne Segler, Amanda Palmer, Shelley Zanksky, Patricia M. Griffin, and Emerging Infections Program FoodNet Working Group. 2009. "Hemolytic Uremic Syndrome and Death in Persons with Escherichia coli O157:H7 Infection, Foodborne Diseases Active Surveillance Network Sites, 2000–2006." *Clinical Infectious Diseases* 49:1480–1485.

Grubesic, Tony H. 2008. "Zip codes and spatial analysis: Problems and prospects." *Socio-Economic Planning Sciences* 42(2):129–149.

Hansen, Christian B. 2007. "Asymptotic properties of a robust variance matrix estimator for panel data when T is large." *Journal of Econometrics* 141:597–620.

Hastings, John, Shu Sebesta, Terry Thompson, and John Williamson. 1991. "California Morbidity 1990." Technical report, State of California - Department of Health Services.

Healthcare Cost and Utilization Project, HCUP. 2016. "I9_Formats_Txt." Technical report, Agency for Healthcare Research and Quality, US Department of Health and Human Services.

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

Ho, Daniel E., and Kosuke Imai. 2006. "Randomization inference with natural experiments: An analysis of ballot effects in the 2003 California recall election." *Journal of the American Statistical Association* 101(475):888–900.

Irwin, Kathleen, Jane Ballard, John Grendon, and John Kobayashi. 1989. "Results of routine restaurant inspections can predict outbreaks of foodborne illness: the Seattle-King County experience." *American Journal of Public Health* 79(5):586–590.

Jones, Timothy F., Boris I. Pavlin, Bonnie J. LaFleur, L. Amanda Ingram, and William Schaffner. 2004. "Restaurant inspection scores and foodborne disease." *Emerging Infectious Diseases* 10(4):688–692.

Krieger, Nancy, Jarvis T. Chen, Pamela D. Waterman, Mah-Jabeen Soobader, S. V. Subramanian, and Rosa Carson. 2002a. "Geocoding and monitoring of US socioeconomic inequalities in mortality and cancer incidence: does the choice of area-based measure and geographic level matter? the Public Health Disparities Geocoding Project." *American Journal of Epidemiology* 156(5):471–482.

Krieger, Nancy, Pamela Waterman, Jarvis T. Chen, Mah-Jabeen Soobader, S. V. Subramanian, and Rosa Carson. 2002b. "Zip code caveat: Bias due to spatiotemporal mismatches between zip codes and us census–defined geographic areas—The public health disparities geocoding project." *American Journal of Public Health* 92(7):1100–1102.

Lloyd, Susan S., and J. Peter Rissing. 1985. "Physician and coding errors in patient records." *JAMA* 254(10):1330–1336.

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.

Missouri Census Data Center. 2010. "ZIP Codes Resources, 2000 Edition." Technical report, Missouri Census Data Center.

Nataro, James P., and James B. Kaper. 1998. "Diarrheagenic escherichia coli." *Clinical Microbiology Reviews* 11(1):142–201.

NYCDHMH. 2012. "Restaurant Grading in New York City at 18 Months." Technical report, New York City Department of Health and Mental Hygiene.

O'hara, Robert B., and D. Johan Kotze. 2010. "Do not log-transform count data." *Methods in Ecology and Evolution* 1(2):118–122.

Puzo, Daniel P. 1986. "Growth Enhancers Under Suspicion as Cause of Salmonella Outbreak." *Los Angeles Times* .

Roth, Jean. 2007. "Census U.S. Intercensal County Population Data, 1970-2014." Technical report, The National Bureau of Economic Research.

Sater, Douglas K. 1994. "Geographic coding of administrative records–current research in ZIP/sector-to-county coding process." In *Annual Meeting of the Population Association of America*.

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.

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.

Serapiglia, Tino, Erin Kennedy, Sylvanus Thompson, and Ron de Burger. 2007. "Association of Food Premises Inspection and Disclosure Program with Retail-Acquired Foodborne Illness and Operator Non-compliance in Toronto." *Journal of Environmental Health* 70(1):54–59.

Simon, Paul A., Phillip Leslie, Grace Run, Ginger Zhe Jin, Reporter Roshan, 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.

Smorodinsky, Svetlana. 2010. "Everything You Wanted To Know About Zip Code Use in Public Health and Were Afraid To Ask." Western Regional Epidemiology Network.

Spika, J. S., S. H. Waterman, G. W. Hoo, M. E. St Louis, R. E. Pacer, S. M. James, M. L. Bissett, L. W. Mayer, J. Y. Chiu, and B. Hall. 1987. "Chloramphenicol-resistant Salmonella newport traced through hamburger to dairy farms. A major persisting source of human salmonellosis in California." *The New England Journal of Medicine* 316(10):565–570.

State of California, Department of Finance. 2017. "Demographic Estimates." Technical report, State of California, Department of Finance.

US Postal Bulletins Consortium. 2014. "US Postal Bulletins." Technical report, US Postal Bulletins Consortium.

Wiant, Chris J. 1999. "Scores, grades, and communicating about food safety." *Journal of Environmental Health* 61(9):37–39.