AJAE Appendix for "Shaming, stringency, and shirking: Evidence from food-safety inspections"

John Bovay*

May 9, 2024

*Department of Agricultural and Applied Economics, Virginia Tech. bovay@vt.edu.

Note: The material contained herein is supplementary to the article named in the title and published in the *American Journal of Agricultural Economics.*

Appendix A: Details on data-cleaning procedure

The data set I obtained from FSIS does not include any indication of the sample-set groupings that were used to determine regulatory compliance and category designation over 1999–2015, and FSIS did not provide further guidance on this issue. Inspection of the data reveals clear patterns of 51 samples being collected over a short period, followed by a gap (often, approximately one year) before another set of 51 samples. However, it is clear that inspectors often collected slightly more and occasionally slightly fewer than 51 samples. FSIS personnel confirmed that inspectors were supposed to collect samples until *results* from 51 tests were available, which explains the frequent appearance of 52 to 56 samples over a brief period, followed by a gap. FSIS also sometimes terminated collection before reaching 51 samples, if a threshold was certain to be exceeded. After some preliminary data cleaning to eliminate duplicate observations, I assign observations into sample sets by identifying lengthy gaps between observations while maximizing the number of sample sets with 51 observations. Specifically, I identify the start of a new sample set as occurring when the gap between observations was at least x times as long as the average gap over the previous 51 observations, where x is chosen for each policy period as the integer that maximizes the number of sample sets with 51 observations. This method generates sample sets with lengths reasonably close to the expected length: at least 80% of all sample sets in each of the regulatory periods have 50 to 56 observations. To eliminate noise that would be generated through mis-assigning observations to sample sets, for the main analysis of sections 5 and 6, I only include observations from sample sets of length $[n, ..., N]$, where n and N are the minimum and maximum sample-set lengths such that at least 1% of sample sets have lengths n and N. Note again that the 51-sample sets were eliminated effective May 6, 2015.

Appendix B: Robustness and placebo tests

This appendix provides the results of various robustness and placebo tests described in the text.

Effects of known categorization: Placebo tests

Table B1 presents the results of RD models that use placebo cutoffs near the $c = 0$ and $c = 1$ cutoffs that yield significant estimates in table 3. As discussed in the main body of the paper, only two of the 36 cutoffs in table B1 have statistically significant coefficients with $p < 0.1$ and with the correct, positive sign; the distribution of p-values is approximately uniform. In conclusion, the placebo cutoffs do not raise concerns about the validity of the main results.

Effects of policy changes: robustness tests

This appendix subsection presents the results of robustness tests relevant to the RDiT design discussed in section 7. For RDiT approaches to analysis of policy changes, Hausman and Rapson (2018) recommend a few additional robustness tests.[1] First, as recommended by Cattaneo, Idrobo, and Titiunik (2020) for RD designs where the data have many "mass points", I collapse the data set and use the daily share of samples positive, across all establishments, as my dependent variable. The results, in panel A of table B2, essentially conform with the results in panel B of table 5: the introduction of public disclosure in 2008 led to a 4.5 percentage point decrease in the share of samples positive, while the tightening of standards in 2011 led to a 6.3 percentage point increase. In this specification, the 2006 introduction of the categorization system is also estimated to have led to a statistically significant 3.4 percentage point increase in the share of samples positive. The result for 2006 is of the same sign as the insignificant result shown for that year in table 5, but is of larger magnitude.

Second, I employ a "donut" approach as recommended by Barreca et al. (2011) to ensure that *Salmonella* sampling dates were not subject to manipulation around the dates of the policy changes,

[1] All results described in the rest of this appendix use the same data set as panel B of table 5, dropping all establishments that ever exited.

which might have occurred if sampling dates were misreported or establishments briefly shut down before or after policy changes. These results are again similar to the main results in table 5. The donut specifications, removing all observations within 1 to 7 days on both sides of policy changes, yield somewhat larger estimated effects of the 2008 policy change (a 4.9 to 5.8 percentage point decrease in the share of samples positive) and somewhat smaller estimated effects of the 2011 policy change (a 5.5 to 6.7 percentage point increase) than the main specification. Panel B of table B2 shows results for the RDiT regression with all observations within 7 days of the policy changes removed. In all of the donut specifications, the 2006 policy change is estimated to have insignificant effects on the share of samples positive.

Third, I drop all observations belonging to sample sets that span two policy periods. Under the policy regimes in place through 2015, category status was assigned on the basis of sample sets as they were completed; incomplete sample sets were not reset at the time of the policy changes. When I drop observations from sample sets that span policy periods, the estimated RDiT effects change somewhat: the introduction of disclosure in 2008 resulted in a 2.9 percentage point decrease in the share of samples positive (though not statistically significant), while the 2011 tightening of standards led to an 11.1 percentage point increase ($p = 0.002$). The 2006 policy change had an insignificant effect.

While the various specifications yield somewhat different point estimates, the sign and magnitude of the estimates are fairly consistent. The introduction of mandatory disclosure in 2008 resulted in a significant improvement in average *Salmonella* test results, roughly a 55 percent reduction in the share of samples positive. Perversely, though, the tightening of standards in 2011 resulted in a significant worsening of test results, more than doubling the share of samples positive.

As another robustness test, I use several sets of placebo dates of policy changes. Each policy change was preceded by an announcement in the Federal Register about the scheduled policy change. In Panel A of table B3, I use the dates of the relevant Federal Register announcements as the cutoffs. I find that *Salmonella* test results did not change discontinuously at the dates of the announcements. In Panels B through E of table B3, I use placebo dates 120, 240, 360, and 480

days before the actual policy changes. Under the null hypothesis, with 12 placebo cutoff values, one placebo would be expected to have $p \leq 0.083$. In table B3, the lowest p-value is 0.094. We can therefore conclude that the placebo effects are the consequence of random variation and that the estimated effects of the policy changes in table 5 are valid.

Table B1: Placebo effects of known categorization on *Salmonella* outcomes

	(1)	(2)	(3)	(4)	(5)	(6)
Placebo RD cutoff (c)	-0.15	-0.1	-0.05	0.05	0.1	0.15
Panel A: 1999 to 2006, $\kappa = 12$ *positive samples*						
$D_0^p = 1$	-0.021	0.016	0.076	-0.063	-0.053	0.001
Robust p-value	0.554	0.427	0.107	0.009	0.010	0.844
Panel B: 2006 to 2008, $\kappa = 6$ *positive samples*						
$D_0^p = 1$	0.024	-0.030	0.001	-0.016	0.004	-0.009
Robust p-value	0.437	0.195	0.597	0.317	0.866	0.235
Panel C: 2006 to 2008, $\kappa = 12$ *positive samples*						
$D_0^p = 1$	0.070	-0.070	0.018	-0.096	-0.158	-0.071
Robust p-value	0.221	0.194	0.924	0.096	0.041	0.262
Panel D: 2011 to 2015, $\kappa = 2$ *positive samples*						
$D_0^p = 1$	0.059	0.011	-0.018	0.004	-0.013	-0.023
Robust p-value	0.086	0.807	0.110	0.345	0.002	0.000
Panel E: 2011 to 2015, $\kappa = 5$ *positive samples*						
$D_0^p = 1$	-0.056	-0.022	0.102	0.001	0.035	-0.011
Robust p-value	0.181	0.331	0.430	0.796	0.000	0.019
Placebo RD cutoff (c)	0.8	0.85	0.9	1.05	1.1	1.15
Panel F: 2006 to 2008, $\kappa = 12$ *positive samples*						
$D_1^p = 1$	0.007	-0.014	-0.008	-0.056	-0.010	0.011
Robust p-value	0.827	0.517	0.761	0.042	0.832	0.528

Notes: This table presents results of regressions paralleling those in table 3 with statistically significant results but for placebo cutoffs not associated with any change in disclosure status. D_0^p and D_1^p are analogous to D_0 and D_1 in table 3 but use the placebo cutoffs indicated at the top of the columns. Panels A through E report results for three placebo cutoffs on either side of the actual cutoff ($c = 0$) according to $c \pm 0.05n$, where $n = \{1, 2, 3\}$. Panel F uses the nearest placebo cutoffs to the actual cutoff ($c = 1$) that are multiples of 0.05, for which there are enough observations on either side of the placebo cutoffs to estimate the optimal bandwidths. Each panel represents regressions using observations from the policy regimes beginning and ending in the indicated years. For sample sets that span the dates of policy change, observations are included in the later period if the samples were taken after the Federal Register announcement that preceded the policy change. All regressions are local linear RD regressions with triangular kernels, using $leeway\kappa$ as the running variable. Bandwidths, robust p-values, and confidence intervals are calculated using the `rdms` command in Stata (Cattaneo, Titiunik, and Vazquez-Bare, 2020), although bandwidths and confidence intervals are suppressed in this table. Bandwidths are chosen to minimize mean squared error on either side of each cutoff.

Table B2: Effects of policy changes on average *Salmonella* test outcomes: Robustness tests

Policy introduced	Categorization (private)	Public disclosure	Public disclosure w/ tighter standards
Date of implementation (c)	5/30/2006	3/28/2008	7/1/2011
	(1)	(2)	(3)
Panel A: Observations collapsed by sample collection date			
$t \geq c$	0.034	−0.045	0.063
Robust p-value	0.046	0.015	0.001
95% CI (lower limit)	0.00	−0.07	0.03
(upper limit)	0.08	−0.01	0.12
Observations	381	326	380
Left bandwidth	372	275	284
Right bandwidth	175	200	260
Panel B: "Donut" approach: Drop all observations within 7 days of policy changes			
$t \geq c$	0.024	−0.057	0.055
Robust p-value	0.294	0.039	0.088
95% CI (lower limit)	−0.03	−0.11	−0.01
(upper limit)	0.10	−0.00	0.14
Observations	15236	7183	5414
Left bandwidth	366	204	199
Right bandwidth	220	237	233
Panel C: Drop all observations belonging to sample sets that span policy periods			
$t \geq c$	0.034	−0.029	0.111
Robust p-value	0.248	0.385	0.002
95% CI (lower limit)	−0.03	−0.11	0.05
(upper limit)	0.13	0.04	0.21
Observations	11540	3625	3692
Left bandwidth	345	148	177
Right bandwidth	212	199	258

Notes: See notes to table 5.

Table B3: Effects of policy changes on average *Salmonella* test outcomes: Placebo cutoff dates

Policy introduced	Categorization (private)	Public disclosure	Public disclosure w/ tighter standards
	(1)	(2)	(3)
Panel A: Cutoffs c = Federal Register announcement dates			
$t \geq c$	−0.038	−0.031	0.015
Robust p-value	0.640	0.883	0.269
95% CI (lower limit)	−0.15	−0.10	−0.01
(upper limit)	0.09	0.08	0.05
Observations	9621	3289	5747
Left bandwidth	356	172	165
Right bandwidth	89	60	139
Panel B: Cutoffs c = 120 days before policy changes			
$t \geq c$	−0.015	0.022	−0.008
Robust p-value	0.877	0.334	0.915
95% CI (lower limit)	−0.07	−0.04	−0.04
(upper limit)	0.08	0.12	0.05
Observations	11627	2664	2944
Left bandwidth	426	144	174
Right bandwidth	117	38	120
Panel C: Cutoffs c = 240 days before policy changes			
$t \geq c$	−0.018	0.053	0.002
Robust p-value	0.187	0.236	0.925
95% CI (lower limit)	−0.07	−0.04	−0.09
(upper limit)	0.01	0.16	0.08
Observations	27891	4213	7277
Left bandwidth	1190	96	294
Right bandwidth	237	104	113
Panel D: Cutoffs c = 360 days before policy changes			
$t \geq c$	0.019	0.029	0.030
Robust p-value	0.573	0.094	0.111
95% CI (lower limit)	−0.08	−0.01	−0.01
(upper limit)	0.15	0.08	0.08
Observations	15410	4474	6121
Left bandwidth	523	68	235
Right bandwidth	146	76	112
Panel E: Cutoffs c = 480 days before policy changes			
$t \geq c$	−0.026	0.032	−0.006
Robust p-value	0.226	0.204	0.988
95% CI (lower limit)	−0.11	−0.02	−0.06
(upper limit)	0.03	0.08	0.06
Observations	13538	5298	6117
Left bandwidth	545	66	238
Right bandwidth	160	130	108

Notes: For additional details on the regression specifications, see notes to table 5.

Appendix C: Analysis of additional policy regimes in place over 2015–2017

For clarity and ease of exposition, the body of the paper analyzes *Salmonella* test outcomes and shirking only for the four policy periods in place from January 4, 1999 until May 5, 2015. The data set I obtained from FSIS by FOIA request covers two additional policy regimes. This appendix describes those policy regimes and analysis of shirking or moral hazard over these periods.

Effective May 6, 2015, the 51-sample-set framework was replaced with a system of categorization based on aggregated results over rolling 52-week windows. Under the new system, categories were defined using the same shares: an establishment with more than 9.8% of samples positive (i.e., 5/51) during any window of the windows ending the previous month would be placed on the Category 3 list and would remain on that list for a three-month period. The rolling-window system was introduced because FSIS officials recognized that under the sample-set system, establishment operators might increase efforts related to *Salmonella* control during the weeks that establishments were under scrutiny but shirk during all other weeks of the year.[2] Moreover, the rolling-window system seemed it would be an effective way to mitigate shirking: each week, a new rolling window began, so the end-of-sample-set incentives to shirk might be countered by incentives to obtain good categorization in the coming year.

Shortly after the rolling-window system was introduced, FSIS began using a new chemical solution (neutralizing buffered peptone water) as part of the test procedure.[3] After this change, which was implemented on July 1, 2016, the share of positive test results rapidly rose, and on November 20, 2016, FSIS suspended public disclosure of *Salmonella* category information for chicken-slaughter establishments but continued to sample carcasses for *Salmonella*. No date was given for the resumption of disclosure; on December 15, 2017, FSIS announced that disclosure would resume the following month. Thus, during the final period analyzed, there were no immediate consequences for poor test outcomes. Establishment operators may have anticipated that the tests might ultimately be incorporated into their categorization, but they would not have known

[2]See https://www.federalregister.gov/documents/2015/01/26/2015-01323/changes-to-the-salmonella-and-campylobacter-verification-testing-program-proposed-performance, page 3945.

[3]See https://www.govinfo.gov/content/pkg/FR-2018-11-09/pdf/2018-24540.pdf.

this for certain.

Tables C1, C3, and C4 in this appendix present the results of regression models equivalent to those in tables 3, 4 and 5, covering the periods 2015–16 (rolling windows) and 2016–17 (disclosure hiatus). Figures C1 and C2 present RD plots that correspond to tables C1 and C4. Tables C2, C5, and C6 present robustness and placebo tests equivalent to those in appendix B, covering the periods 2015–16 and 2016–17.

Effects of known categorization on Salmonella test outcomes, 2015–17

RD plots for the 2015–16 and 2016–17 periods are shown in figure C1, and regression results equivalent to those shown in table 3 are shown in table C1. During the 2015–16 period, sample sets were no longer used and establishments with more than 9.8 percent of samples positive during any 52-week window ending within the last three months were listed as Category 3 on the FSIS website. Similar to the 2008–11 period, establishment operators apparently exerted effort to meet the Category 1 standard but then reduced effort once exceeding the threshold. Establishments were 4.1 percentage points more likely to have positive samples after failing to meet the Category 1 standard for the soonest-ending window (table C1, panel A, column 1). In addition, during this period, establishments appear to have reduced effort after good performance ensured they would meet the Category 2 standard and therefore avoid information disclosure. Establishments were 5.2 percentage points more likely to have positive samples after meeting the Category 2 standard for the soonest-ending window during 2015–16 (table C1, Panel B, column 2).

Under the hiatus in disclosure (2016–17), crossing thresholds associated with any of the categories had statistically insignificant effects on *Salmonella* test outcomes.

Table C2 presents results for regressions parallel to those in table C1 using placebo cutoff values for the running variables ($leeway\kappa$). Similar to table B1, the thresholds shown here are placebo cutoffs near the statistically significant estimates from table C1. Specifically, the placebo cutoff values are three multiples of 0.05 in either direction from $c = 0$; and the nearest multiples of 0.05 to $c = 1$ for which optimal bandwidths (in the sense of minimizing mean squared errors) could be

computed using the `rdms` command in Stata (Cattaneo et al., 2020c). In table C2, three of the 12 RD coefficients are statistically significant with $p < 0.1$, but none of these have the correct (i.e., positive) sign. Moreover, when considering the results in table C2 together with those in table B1, only three of the 48 coefficients have the correct sign and $p < 0.1$. In conclusion, the placebo tests do not raise significant concerns about the conclusions drawn from table C1.

Distance from thresholds and Salmonella test outcomes, 2015–17

Table C3 presents results of regressions that demonstrate the positive correlations between *leeway*2 (*leeway*5) for the soonest-ending window and the likelihood of positive *Salmonella* test results. The regressions are similar to those in table 4, except that FSIS did not use sample sets during these periods. So, instead of using sample sets to calculate the values of the running variable *leeway* and the regressor for share of samples positive, these regressions use the soonest-ending window. Also, they use establishment–month–year fixed effects instead of establishment–sample-set fixed effects. In 2015–16, when the *leeway*2 value was 10 percentage points higher, the probability of a positive *Salmonella* test result was 2.00 percentage points higher ($p = 0.012$; elasticity = 0.92; panel A, column 2). Under the disclosure hiatus, there was no statistically significant relationship between *leeway*2 and *Salmonella* test results in the preferred specification, which controls for the share of samples positive in the soonest-ending window. When the *leeway*5 value was 10 percentage points higher, the probability of a positive *Salmonella* test result was 3.03 percentage points higher in 2015–16 ($p < 0.001$; elasticity = 3.36; panel B, column 2) and 4.16 percentage points higher in 2016–17 ($p = 0.018$; elasticity = 0.55; panel B, column 4).

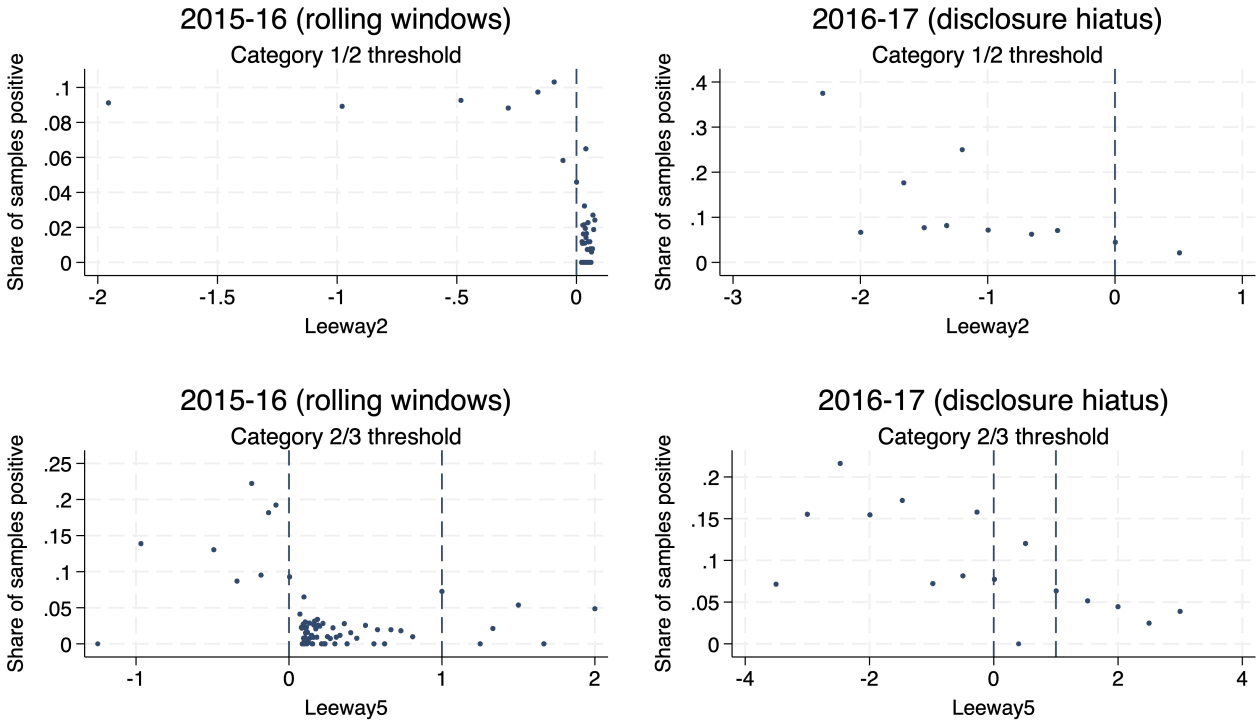
Although the correlation between *leeway*2 and positive *Salmonella* test results lessened under the disclosure hiatus from 2016 to 2017, the correlation between *leeway*5 and test results increased in this period, relative to 2011–15 and 2015–16. In other words, establishment operators appear to have relaxed efforts around *Salmonella* control when they had more leeway with respect to the Category 2/3 threshold, and did so more in 2016–17 than during the earlier periods when the same threshold applied.

Effects of policy changes, 2015 and 2016

Table C4 presents results of RDiT regressions for the policy changes in 2015 and 2016. Figure C2 presents a corresponding RD plot. These policy changes had insignificant effects on average *Salmonella* test outcomes under the main specifications.[4] When collapsing the data set and using the daily share of samples positive as the dependent variable (rather than carcass-level test results), the 2015 introduction of rolling windows is estimated to have decreased the share of samples positive by 2.6 percentage points, evidence of the effectiveness of the rolling-windows system (table C5, panel A, column 1). The additional robustness tests and placebo tests presented in tables C5 and C6 do not raise concerns about the validity of the main result. In conclusion, the 2015 introduction of rolling windows may have improved average test results, but the estimated effects are not as robust as those presented in table 5, which shows that the introduction of public disclosure in 2008 reduced the share of samples positive by about 55 percent.

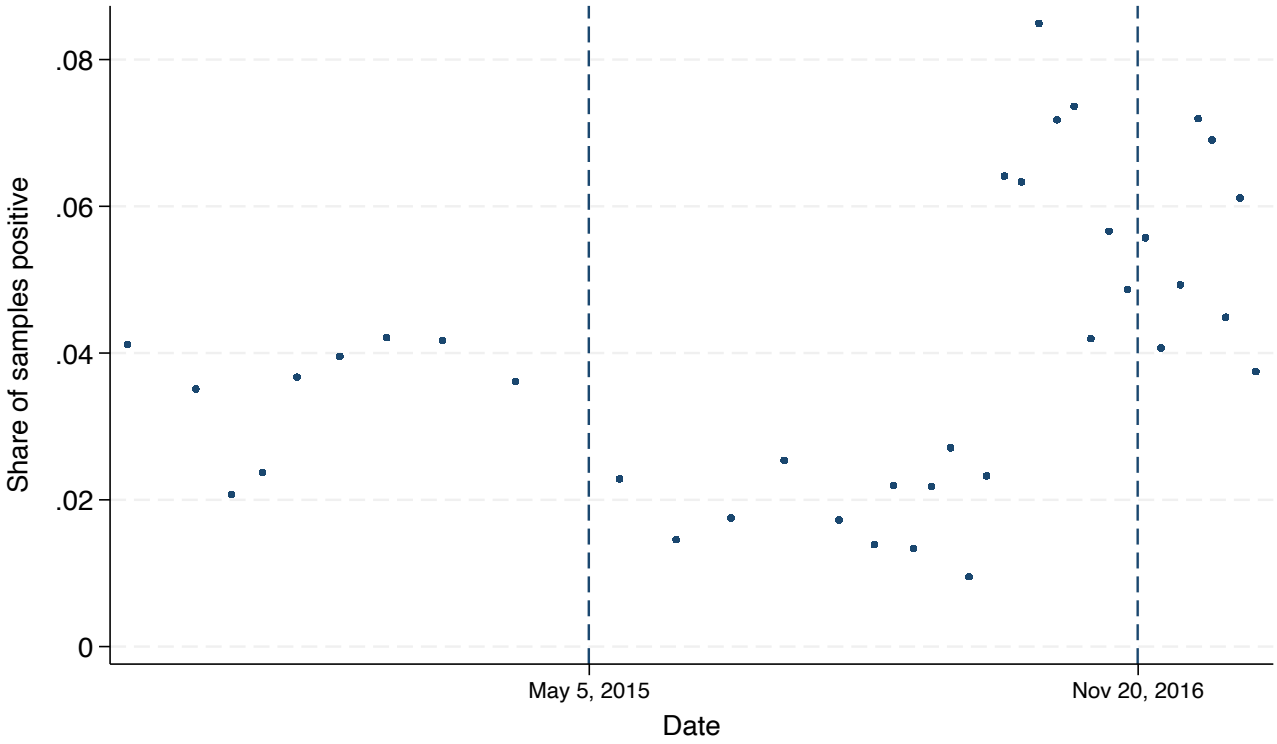
[4]In figure C2, a large rise in the share of samples positive starting July 1, 2016, can be seen. As mentioned above, this can be attributed to a change in the FSIS test procedure. The bandwidths used in estimating the effects of the 2015 and 2016 policy changes do not include July 1, 2016, when the test procedure changed.

Figure C1: RD plots: effects of known categorization on *Salmonella* outcomes



Notes: These RD plots provide graphical evidence corresponding with table C1, within the ranges of the running variables that correspond with the MSE-optimal bandwidths used in table C1. As in figures 4 and 5, quantile-spaced bins are generated using integrated MSE-optimal spacings estimators (Calonico, Cattaneo, and Titiunik, 2014, 2015), and fit lines are not included because they tend to increase the type I error rate of visual inference (Korting et al., 2023). The $leeway2 = 1$ threshold is not shown in the upper two plots because $leeway2$ takes on only two different values at and above the cutoff so the RD models cannot be estimated around this threshold.

Figure C2: RD in time plot



Notes: This RD plot provides graphical evidence corresponding with panel B of table C4, within the temporal range that corresponds with the MSE-optimal bandwidths used in table C4. As in figure 6, quantile-spaced bins are generated using integrated MSE-optimal spacings estimators (Calonico, Cattaneo, and Titiunik, 2014, 2015), and fit lines are not included because they tend to increase the type I error rate of visual inference (Korting et al., 2023).

Table C1: Effects of known categorization on *Salmonella* outcomes, 2015–17

Panel A: Cutoffs not associated with disclosure				
Policy regime	Rolling windows		Disclosure hiatus	
Years	2015 to 2016		2016 to 2017	
Threshold	$D_0 = 1$		$D_0 = 1$	
Implication	Cat. 2 or 3		Cat. 2 or 3	
Max. # pos. samples (κ)	2		2	
	(1)		(2)	
Known categorization ($D_0 = 1$)	0.041		0.029	
Robust p-value	0.047		0.125	
95% CI (lower limit)	0.00		−0.01	
(upper limit)	0.08		0.07	
Observations	7467		4390	
Left bandwidth	2.13		2.49	
Right bandwidth	0.08		0.67	
Panel B: Cutoffs associated with disclosure				
Policy regime	Rolling windows		Disclosure hiatus	
Years	2015 to 2016		2016 to 2017	
Threshold	$D_0 = 1$	$D_1 = 1$	$D_0 = 1$	$D_1 = 1$
Implication	Cat. 3	Cat. 1 or 2	Cat. 3	Cat. 1 or 2
Max. # pos. samples (κ)	5	5	5	5
	(1)	(2)	(3)	(4)
Known categorization ($D_0 = 1$ or $D_1 = 1$)	0.125	0.052	−0.004	−0.107
Robust p-value	0.217	0.008	0.860	0.587
95% CI (lower limit)	−0.03	0.01	−0.06	−0.60
(upper limit)	0.15	0.09	0.05	1.05
Observations	7435	8494	2365	4145
Left bandwidth	1.27	1.00	3.72	1.00
Right bandwidth	0.62	1.48	0.67	2.02

Notes: Each regression uses observations from the policy regimes beginning and ending in the indicated years. All regressions are local linear RD regressions with triangular kernels, using $leeway\kappa$ as the running variable, as described in the text. $D_0 = 1$ if $leeway\kappa < 0$ and $D_1 = 1$ if $leeway\kappa \geq 1$; each of these conditions are equivalent to the known categorization outcomes reflected by the Implications rows in the table. Bandwidths, robust p-values, and confidence intervals are calculated using the `rdms` command in Stata (Cattaneo, Titiunik, and Vazquez-Bare, 2020), clustering on establishment using nearest-neighbor estimation for the variance-covariance estimator. Bandwidths are chosen to minimize mean squared error on either side of each cutoff. The RD models cannot be estimated with $D_1 = 1$ and $\kappa = 2$ because the running variable takes on too few different values at and above the cutoff $c = 1$ (namely, the only possible values are 1 and 2).

Table C2: Placebo effects of known categorization on *Salmonella* outcomes

	(1)	(2)	(3)	(4)	(5)	(6)
Placebo RD cutoff (c)	−0.15	−0.1	−0.05	0.05	0.1	0.15
Panel A: 2015 to 2016, $\kappa = 2$ positive samples						
$D_0^p = 1$	−0.015	0.011	0.032	−0.002	−0.003	−0.010
Robust p-value	0.295	0.715	0.589	0.756	0.799	0.039
Placebo RD cutoff (c)	0.7	0.75	0.8	1.25	1.3	1.5
Panel B: 2015 to 2016, $\kappa = 5$ positive samples						
$D_1^p = 1$	0.035	0.032	0.039	−0.048	−0.036	−0.015
Robust p-value	0.102	0.268	0.130	0.001	0.020	0.878

Notes: This table presents results of regressions paralleling those in table C1 with statistically significant results but for placebo cutoffs not associated with any change in disclosure status. Each panel uses the nearest placebo cutoffs to the actual cutoff ($c = 0$ in panel A; $c = 1$ in panel B) that are multiples of 0.05, for which there are enough observations on either side of the placebo cutoffs to estimate the optimal bandwidths around c. Each panel represents regressions using observations from the policy regimes beginning and ending in the indicated years. As in table B1, D_0^p and D_1^p are analogous to D_0 and D_1 in table 3 but use the placebo cutoffs indicated at the top of the columns. All regressions are local linear RD regressions with triangular kernels, using $leeway\kappa$ as the running variable. Bandwidths, robust p-values, and confidence intervals are calculated using the `rdms` command in Stata (Cattaneo, Titiunik, and Vazquez-Bare, 2020), although bandwidths and confidence intervals are suppressed in this table. Bandwidths are chosen to minimize mean squared error on either side of each cutoff.

Table C3: Effects of distance from category thresholds on *Salmonella* test outcomes, 2015–17

	Rolling windows, 2015 to 2016		Disclosure hiatus, 2016 to 2017	
	(1)	(2)	(3)	(4)
Panel A				
Distance from Category 1 threshold (soonest-ending window)	0.258	0.200	0.261	−0.121
	(0.08)	(0.08)	(0.11)	(0.15)
Share of samples positive, soonest-ending window		−2.942		−14.065
		(1.02)		(6.50)
Observations	7787	7604	1863	1863
Elasticity	1.12	0.92	0.72	−0.34
Panel B				
Distance from Category 2 threshold (soonest-ending window)	0.368	0.303	0.660	0.416
	(0.07)	(0.06)	(0.13)	(0.17)
Share of samples positive, soonest-ending window		−2.581		−8.067
		(0.48)		(5.30)
Observations	7117	6934	971	971
Elasticity	3.84	3.36	0.88	0.55

Notes: This table represents the results of similar regressions to those shown in table 4, for the 2015–16 policy period during which sample sets were replaced with overlapping sampling windows, and the 2016–17 hiatus in public disclosure. Panel A demonstrates the effects of distance from the Category 1 threshold ($leeway2$) on *Salmonella* test outcomes; Panel B the effects of distance from the Category 2 threshold ($leeway5$). The main variables of interest are $leeway2$ and $leeway5$ for the soonest-ending window, but the even-numbered columns also control for the share of samples positive in the soonest-ending window. All regressions use establishment–month–year fixed effects. Standard errors, clustered by establishment, are given in parentheses. Elasticities reported are the elasticities of the share of samples positive with respect to $leeway\kappa$, calculated using the mean share of samples positive and the mean value of $leeway\kappa$. Observations are included only if $leeway\kappa \in [0, 1)$.

Table C4: Effects of policy changes on average *Salmonella* test outcomes

Policy introduced	Rolling windows	Disclosure hiatus
Date of implementation (c)	5/6/2015	11/20/2016
	(1)	(2)
Panel A: All establishments included		
$t \geq c$	−0.015	0.005
Robust p-value	0.388	0.819
95% CI (lower limit)	−0.05	−0.03
(upper limit)	0.02	0.04
Observations	11935	5734
Left bandwidth	392	98
Right bandwidth	165	128
Panel B: Establishments that ever exited excluded		
$t \geq c$	−0.015	0.005
Robust p-value	0.393	0.803
95% CI (lower limit)	−0.05	−0.03
(upper limit)	0.02	0.04
Observations	13650	5795
Left bandwidth	512	99
Right bandwidth	167	129

Notes: This table reports the results of RD in time regressions that use the dates of policy implementation as the cutoffs (c). See additional notes to table 5.

Table C5: Effects of policy changes on average *Salmonella* test outcomes: Robustness tests

Policy introduced	Rolling windows	Disclosure hiatus
Date of implementation (c)	5/6/2015	11/20/2016
	(1)	(2)
Panel A: Observations collapsed by sample collection date		
$t \geq c$	−0.026	0.145
Robust p-value	0.066	0.115
95% CI (lower limit)	−0.06	−0.04
(upper limit)	0.00	0.38
Observations	444	144
Left bandwidth	390	81
Right bandwidth	183	98
Panel B: "Donut" approach: Drop all observations within 7 days of policy changes		
$t \geq c$	−0.024	0.006
Robust p-value	0.143	0.874
95% CI (lower limit)	−0.06	−0.04
(upper limit)	0.01	0.05
Observations	14017	4631
Left bandwidth	513	83
Right bandwidth	186	108
Panel C: Drop all observations belonging to sample sets that span policy periods		
$t \geq c$	−0.012	0.005
Robust p-value	0.475	0.785
95% CI (lower limit)	−0.04	−0.03
(upper limit)	0.02	0.04
Observations	13245	5790
Left bandwidth	496	98
Right bandwidth	170	130

Notes: For additional details on the regression specifications, see notes to table 5.

Table C6: Effects of policy changes on average *Salmonella* test outcomes: Placebo cutoff dates

	Rolling windows	Disclosure hiatus
	(1)	(2)
Panel A: Cutoffs c = Federal Register announcement dates		
$t \geq c$	−0.015	
Robust p-value	0.813	
95% CI (lower limit)	−0.05	
(upper limit)	0.04	
Observations	12262	
Left bandwidth	502	
Right bandwidth	99	
Panel B: Cutoffs c = 120 days before policy changes		
$t \geq c$	−0.013	0.016
Robust p-value	0.894	0.435
95% CI (lower limit)	−0.05	−0.02
(upper limit)	0.04	0.05
Observations	9699	5456
Left bandwidth	357	72
Right bandwidth	119	150
Panel C: Cutoffs c = 240 days before policy changes		
$t \geq c$	−0.020	−0.007
Robust p-value	0.069	0.573
95% CI (lower limit)	−0.07	−0.03
(upper limit)	0.00	0.02
Observations	8811	4492
Left bandwidth	295	72
Right bandwidth	83	122
Panel D: Cutoffs c = 360 days before policy changes		
$t \geq c$	−0.006	0.004
Robust p-value	0.547	0.589
95% CI (lower limit)	−0.03	−0.03
(upper limit)	0.02	0.04
Observations	8271	3253
Left bandwidth	232	46
Right bandwidth	115	102
Panel E: Cutoffs c = 480 days before policy changes		
$t \geq c$	−0.008	−0.012
Robust p-value	0.726	0.157
95% CI (lower limit)	−0.04	−0.04
(upper limit)	0.03	0.01
Observations	12195	5227
Left bandwidth	316	84
Right bandwidth	161	164

Notes: Panel A does not include column (2) because the hiatus in disclosure was not preceded by a Federal Register announcement. For additional details on the regression specifications, see notes to table 5.

References

Barreca, A.I., M. Guldi, J.M. Lindo, and G.R. Waddell. 2011. "Saving babies? Revisiting the effect of very low birth weight classification." *The Quarterly Journal of Economics* 126:2117–2123.

Calonico, S., M.D. Cattaneo, and R. Titiunik. 2015. "Optimal data-driven regression discontinuity plots." *Journal of the American Statistical Association* 110:1753–1769.

—. 2014. "Robust data-driven inference in the regression-discontinuity design." *The Stata Journal* 14:909–946.

Cattaneo, M.D., N. Idrobo, and R. Titiunik. 2020. *A Practical Introduction to Regression Discontinuity Designs: Extensions*. Cambridge: Cambridge University Press.

Cattaneo, M.D., R. Titiunik, and G. Vazquez-Bare. 2020. "Analysis of regression discontinuity designs with multiple cutoffs or multiple scores." *The Stata Journal* 20:866–891.

Hausman, C., and D.S. Rapson. 2018. "Regression discontinuity in time: Considerations for empirical applications." *Annual Review of Resource Economics* 10:533–552.

Korting, C., C. Lieberman, J. Matsudaira, Z. Pei, and Y. Shen. 2023. "Visual inference and graphical representation in regression discontinuity designs." *The Quarterly Journal of Economics* 138:1977–2019.