

ONLINE APPENDIX

Improving Regulatory Effectiveness through Better Targeting: Evidence from OSHA

Matthew S. Johnson

Sanford School of Public Policy, Duke University

David I. Levine

Haas School of Business, University of California

Michael W. Toffel

Harvard Business School

Online Appendix

- Appendix A SST Program Details
- Appendix B Appendix Tables and Figures
- Appendix C Sample Attrition
- Appendix D Validity of ODI-reported Injury Rates
- Appendix E Estimating the Social Cost of Serious Injuries during Our Sample Period
- Appendix F Pre-specification Plan
- Appendix G Predictors in the Machine Learning Analyses
- Appendix H Assessing Generalizability: Can Results from the Randomized Sample be Applied to the Nonrandomized Sample?
- Appendix I Robustness Checks on Estimate of Intention-to-treat Effect on Injuries
- Appendix J Assessing the Lucas Critique: Can We Estimate Effects of New Targeting Rules Using Historical Data?
- Appendix K Did Inspections of Establishments near the SST Cutoffs Reduce Injuries?

Appendix A SST Program Details

Each year from 1996 to 2011, the OSHA Data Initiative (ODI) surveyed between 60,000 and 80,000 workplaces with at least 40 employees in a set of hazardous industries OSHA identified based on US Bureau of Labor Statistics (BLS) data.^{1,2} OSHA sent the ODI survey mid-year and establishments reported summary data on their injuries, illnesses, and employment from the previous calendar year, based on logs that OSHA required them to keep of every work-related injury and illness.³

As part of the SST program that OSHA operated from 1999 to 2014, each year—between April and August—OSHA created a primary and a secondary SST target list based on the prior year’s ODI survey data. The primary list consisted of the roughly 3,500 establishments whose prior-year ODI survey indicated the highest rates of injuries causing days away from work, restricted work, or a transfer (DART injuries) or the highest rates of injuries causing days away from work (DAFW injuries). Both DART rate and DAFW rate were measured as the number of such injuries reported over a year per 100 full-time employees working 40-hour weeks that year. The secondary list contained the roughly 7,000 establishments with the next-highest rates. In the early years of our sample period, these cutoffs were based exclusively on DART rates; starting in 2005, they were a function of both DART *and* DAFW rates. For example, in 2008, the primary list included establishments whose 2007 ODI data indicated that in 2006 they experienced DART rates of at least 11 or DAFW rates of at least 9 and the secondary list included establishments reporting DART rates between 7 and 11 or DAFW rates between 5 and 9. The specific cutoffs for the primary and secondary lists changed each year and, beginning in 2009, varied by industry. We restrict our analysis to the 2001–2010 target lists because those are the only years for which we could obtain primary and secondary lists from OSHA. Establishments on the primary and secondary lists in those years reported average DART rates of 12.8 and 7.0,

¹ Specifically, the US Bureau of Labor Statistics (BLS) Survey of Occupational Injuries and Illnesses gathered data each year from a sample of approximately 200,000 establishments drawn from all private-sector industry establishments. OSHA selected for the SST program a subset of industries that BLS classified as “high hazard industries.” OSHA used BLS annual “high hazard industries” lists until 2003, when BLS stopped updating them, and from then on used the 2003 edition.

² OSHA intended to survey each establishment meeting these criteria—that is, with at least 40 employees and in the specified hazardous industries—at least once every three years.

³ OSHA Form 300, which OSHA standard 29 CFR 1904 requires employers to complete, is available at <https://www.osha.gov/recordkeeping/RKform300pkg-fillable-enabled.pdf>.

respectively, both of which are several times the average DART rate of 2.3 for all private-sector establishments over this period (US Bureau of Labor Statistics 2015).

After constructing the year's primary and secondary lists, OSHA notified each of its 81 area offices of all those establishments on the primary list that were located in that office's region. If an area office did not anticipate having sufficient resources to inspect its entire primary list, a "cycle" ensued whereby the office entered the number it anticipated being able to inspect into OSHA software. The software then randomly assigned the subset of establishments from the primary list that the area office was to inspect. If the area office inspected them all before OSHA headquarters issued the next year's list, another cycle ensued whereby the area office estimated how many additional inspections it could conduct and the software generated a new random set of establishments from the remainder of its primary list. When an area office had at least attempted inspections at all of its primary list, it repeated this process with the secondary list (for details, see US Occupational Safety and Health Administration 2008b). Thus, most area offices were assigned to inspect a *random* subset of either their primary or secondary list (but never both).

When an OSHA inspector arrived to conduct an SST inspection, he or she explained that the establishment was being inspected because it had a relatively high injury rate.⁴ The inspection itself was similar to other types of OSHA inspections: the inspector walked through the establishment to assess hazards that could lead to injuries or illnesses, then conducted a closing conference with representatives from management and (sometimes) the employees. The inspector typically discussed any violations and also "the strengths and weaknesses of the employer's occupational safety and health system and any other applicable programs, and advises the employer of the benefits of an effective program and provides information, such as OSHA's website, describing program elements" (US Occupational Safety and Health Administration 2016: 3–20). If the inspector discovered violations of OSHA regulations, a few weeks later OSHA would issue a citation and typically assess a fine. Establishments could appeal fines and OSHA often reduced them if the violation was remediated immediately.

⁴ OSHA did not inform establishments that they were on the SST target list until an inspector showed up unannounced for an inspection.

References

- US Bureau of Labor Statistics.** 2015. “Industry Injury and Illness Data.”
http://www.bls.gov/iif/oshsum.htm#10Supplemental_News_Release_Tables.
- US Occupational Safety and Health Administration.** 2008b. “Site-Specific Targeting 2008 Directive 08-03 (CPL 02), Effective Date May 19, 2008.”
<https://www.osha.gov/enforcement/directives/08-07-cpl-02>, accessed March 2019.
- US Occupational Safety and Health Administration.** 2016. “OSHA Field Operations Manual, Directive CPL-02-00-160, Effective Date August 2, 2016.”
https://www.osha.gov/OshDoc/Directive_pdf/CPL_02-00-160.pdf, accessed March 2018.

Appendix B Appendix Tables and Figures

Table B1: Pipeline from Full SST Target Lists to the Randomized Sample

	Primary list				Secondary list				Primary & secondary lists			
	Assigned to inspection		Not assigned		Assigned to inspection		Not assigned		Assigned to inspection		Not assigned	
	#	% of total	#	% of total	#	% of total	#	% of total	#	% of total	#	% of total
Number of establishment-directives ^a on 2010 SST target lists and...												
In states under federal OSHA jurisdiction ^b	20,708	100.0	13,314	100.0	7,455	100.0	50,349	100.0	28,163	100.0	63,663	100.0
<i>Restrict to area-office-directives that randomized lists</i>												
On a primary list that was started but not exhausted ^c	8,084	39.0	10,063	75.6				
Primary list exhausted, and on a secondary started but not exhausted ^c	5,150	69.1	7,472	14.8				
Located in an area-office-directive that randomized (overall)	13,234	47.0	17,535	27.5
<i>Restrict to establishments eligible for SST inspection</i>												
Not subject to deletion criteria ^d	6,589	31.8	7,747	58.2	4,145	55.6	5,234	10.4	10,734	38.1	12,981	20.4
<i>Drop establishments targeted for concerns with ODI reporting quality</i>												
Has non-missing ODI data in directive year ^e	6,363	30.7	7,474	56.1	4,032	54.1	5,094	10.1	10,395	36.9	12,568	19.7
DART and DAFW rates meet selection criteria for corresponding list ^e	5,935	28.7	7,099	53.3	3,923	52.6	5,003	9.9	9,858	35.0	12,102	19.0
<i>Cross-checks that establishment exists</i>												
Found in NETS	5,813	28.1	6,941	52.1	3,850	51.6	4,913	9.8	9,663	34.3	11,854	18.6
Alive in year $t - 2$ [NETS] ^f	5,496	26.5	6,556	49.2	3,684	49.4	4,721	9.4	9,180	32.6	11,277	17.7
Alive in year t [NETS] ^g	5,316	25.7	6,341	47.6	3,593	48.2	4,585	9.1	8,909	31.6	10,926	17.2
<i>Final steps for analysis sample</i>												
Not a nursing home in 2002 directive ^h	5,151	24.9	5,980	44.9	3,398	45.6	4,079	8.1	8,549	30.4	10,059	15.8
SST cycle is opened ⁱ	4,806	23.2	5,980	44.9	3,308	44.4	4,079	8.1	8,114	28.8	10,059	15.8
DART rate, employment, hours (all from $t - 2$) in common support	4,805	23.2	5,968	44.8	3,305	44.3	4,073	8.1	8,110	28.8	10,041	15.8
Area-office-directive has ≥ 1 assigned and not-assigned meeting restrictions	4,279	20.7	5,253	39.5	2,698	36.2	3,911	7.8	6,977	24.8	9,164	14.4

^a An establishment-directive refers to a specific instance of an establishment being placed on a particular year's SST target list.

^b Restricts to the 29 states under federal OSHA jurisdiction. While a few of the 21 states with state-run OSHA offices participated in SST, they were not subject to oversight from the federal office. See Figure A.1 for details.

^c Restricts to establishments on (a) the primary list and belonging to an area-office-directive that assigned strictly between 5 and 95% of its primary list to inspection, or (b) the secondary list and belonging to an area-office-directive that assigned 100% of its primary list to inspection and assigned strictly between 5 and 95% of its secondary list to inspection. This is the subset of the target lists that was randomized.

^d An establishment is subject to the deletion criteria if, within 2 years prior to the directive start date—or 3 years, beginning with the 2009 SST directive start date—it had an inspection in IMIS coded as a comprehensive safety inspection or as a records-only inspection or if it is a nursing home and had a focused inspection.

^e A subset of establishments that either do not respond to the ODI survey or report very low injury rates are placed on the target list each year to assess the reliability of their reported data. Because these establishments are targeted precisely because of concerns over the accuracy of their data, we remove them from our sample.

^f Drops establishments that, according to NETS, were not in operation two calendar years before the directive year.

^g Drops establishments not alive at the start of the directive year, as such establishments were ineligible for SST inspection.

^h The 2002 SST directive said nursing homes were to be excluded from the 2002 target list, due to OSHA's concurrent National Emphases Program on nursing homes. We therefore drop such establishments from the sample.

ⁱ An area office could assign subsets of its target list to inspection in cycles. Once it inspected each establishment in a cycle, it could create another one. In some cases, if an Area Office had created a cycle but had not yet opened it (i.e., begin inspecting it), it was told by OSHA headquarters to move on (e.g. to the next year's target list). We therefore drop such unopened cycles from our sample. We identify unopened SST cycles as those in which less than 5% of eligible establishments in the cycle show up in IMIS with an SST inspection in the directive year.

Table B2: Industry Tabulation for SST Target Lists and Randomized Sample Subset

	(1)	(2)	(3)	(4)
	All SST target lists		Randomized sample	
	Count	% of total	Count	% of total
Agriculture, forestry, fishing	764	0.8%	150	0.9%
Mining	59	< .01%	4	< .01%
Construction	38	< .01%	4	< .01%
Manufacturing	50,349	54.8%	9,148	56.7%
Wholesale trade	10,436	11.4%	1,770	11.0%
Retail trade	4,827	5.3%	863	5.3%
Transportation, warehousing	14,894	16.2%	2,291	14.2%
Other services	1,173	1.3%	221	1.4%
Nursing homes	9,291	10.1%	1,690	10.5%
Number of establishment-directives	91,831		16,141	

An establishment-directive corresponds to a unique instance of an establishment being included on an annual SST target list from 2001 to 2010 (some establishments are included in multiple years' target lists). The sample in Columns 1 and 2 includes the entire 2001–2010 SST target lists in states under federal OSHA jurisdiction. The subsample in Columns 3 and 4 includes the subset of establishment-directives on the SST target lists that are included in our randomized sample, as described in Table A.1.

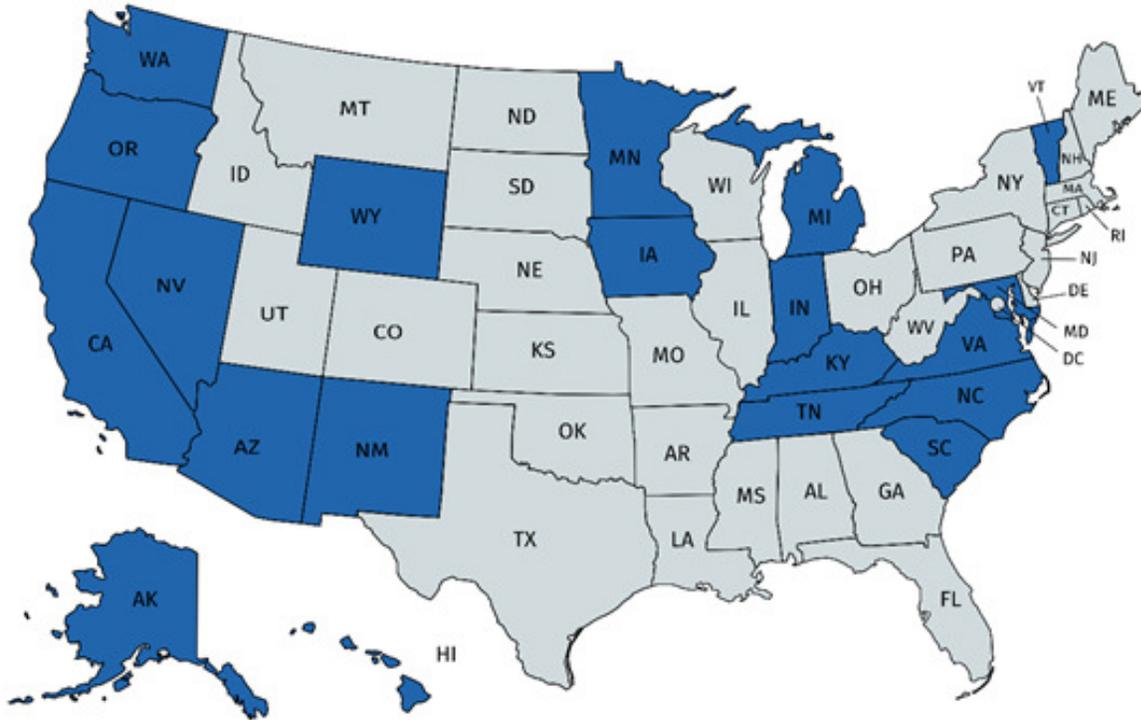
Table B3: Targeting on Predicted Injuries or Estimated Treatment Effects Yields Similar Benefits Whether via Machine Learning Metrics Based a) Only on Data OSHA Would Have Had Available or b) on All Data

Sample = <i>Predicted injuries</i> and <i>Treatment effects</i> estimated on:	2007–2010 target lists			
	2001–2006 sample		2001–2010 sample	
	(1)	(2)	(3)	(4)
Assigned to inspection if targeting on highest <i>predicted injuries</i>	-0.792 (0.367)*		-0.967 (0.292)***	
Assigned to inspection if targeting on highest <i>estimated treatment effects</i>		-0.589 (0.323)+		-0.664 (0.286)**
# observations	4,785	4,785	4,785	4,785

The dependent variable in each column is equal to the average number of injuries an establishment experienced over the 5-year period comprising the directive year and 4 subsequent years. The sample is establishment-directives in the randomized sample on the 2007–2010 SST target lists. *Would be assigned in benchmark policy* is a dummy equal to 1 if an establishment’s *estimated treatment effect* or *predicted injuries* is high enough to be assigned to inspection in this policy. See Section 3.4.2 for details. In Columns 1–2, *estimated treatment effects* and *predicted injuries* are estimated from a causal forest (Super Learner) run on the 2001–2010 target lists, using the method described in Section 2.3. The regression estimates correspond to the median $\hat{\gamma}_1$ from Equation 5 across 250 sample splits, estimated on the 2007–2010 sample. In Columns 3–4, the models underlying *estimated treatment effects* and *predicted injuries* are trained on the 2001–2006 samples, then applied out of sample to the 2007–2010 samples; the estimates correspond to $\hat{\gamma}_1$ from Equation 5 estimated once on the 2007–2010 sample.

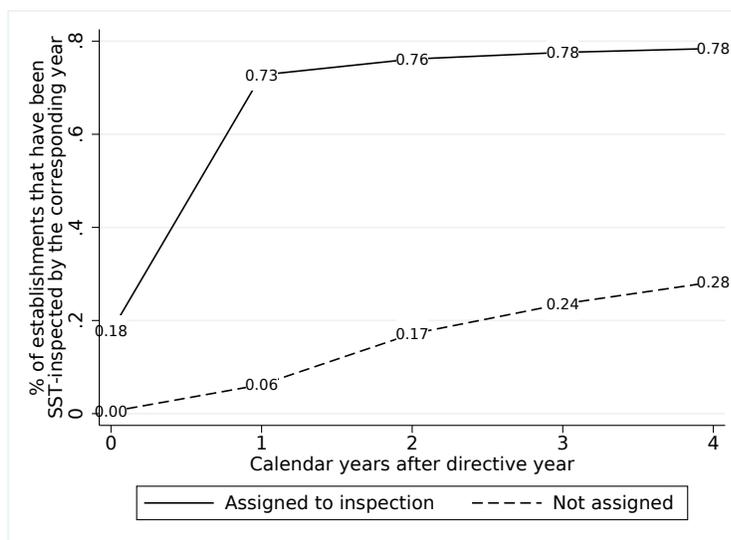
Robust standard errors in parentheses. +p<.1, *p<.05, **p<.01.

Figure B1: The 29 States That OSHA Directly Regulates (in Gray)



Private-sector establishments in the 29 states in gray are under federal OSHA jurisdiction. Source: www.osha.gov/dcsp/osp/. Map created with <https://mapchart.net/usa.html>.

Figure B2: SST Inspection Rates by Year Relative to Directive Year



The figure shows the percent of establishments in our randomized sample with at least one completed SST inspection by the end of each calendar year relative to the directive year, separately for those assigned and not assigned to inspection.

Appendix C Sample Attrition

As discussed in Section 3.1, we do not observe ODI-reported injury data in any year of the post-period (the directive year and four following calendar years) for 15 percent of our randomized sample (though most have post-period data for other outcomes we examine). Such attrition might be a concern if, for example, it is correlated with assignment to SST inspection. Here we discuss the sources of our sample attrition and assess its relationship with assignment to inspection.

Table C.1 illustrates factors leading to sample attrition. While the overall attrition rate is 14.9 percent, this falls to 10.9 percent among those establishments on the 2001–2007 SST target lists. Because the ODI survey ended in 2011, establishments on the 2008–2010 lists had fewer opportunities to be surveyed. When we further restrict the sample to those establishments that never change industry and whose employment never drops below 40 in the post-period, the attrition rate drops slightly from 10.9 percent to 9.8 percent. This suggests that factors that would render establishments ineligible for the ODI survey are not a major source of sample attrition.

The final row of Table C.1 assesses the role of establishment survival: one clear way to exit the sample is to shut down. Indeed, further restricting the sample to those establishments alive during the entire post-period reduces the attrition rate to 5.9 percent. Thus, over 60 percent of the sample attrition can be explained by straightforward observable characteristics.

Table C.1: Sources of ODI Attrition

	(1)	(2)	(3)
	Number of establishment-directives in the randomized sample that...		
	...lack ODI data in the post-period	...have ODI data in the post-period	% with no ODI data: (1)/[(1)+(2)]
Analysis sample	2,405	13,736	14.9%
...and in 2001–2007 target lists	1,269	10,353	10.9%
...and employment [NETS] remained above 40	1,064	9,674	9.9%
...and never changed industry	916	8,457	9.8%
...and remained alive during sample period	491	7,864	5.9%

More pressing than why sample attrition occurs is whether it is correlated with assignment to SST inspection. In Table C.2, we report the coefficients from a series of regressions that predict an indicator variable equal to 1 if an establishment has ODI-reported data

in any of the post-period years, with the key explanatory variable being *assigned to SST inspection* and controlling for directive-year fixed effects. The columns report estimates of this model on each of the sample restrictions in Table C.1.

Reassuringly, in all columns, the coefficient on *assigned to SST inspection* is tiny and statistically indistinguishable from zero, implying that the attrition in ODI-reported data is unlikely to bias our estimates of the effects of SST inspections on ODI-reported outcomes.

Table C.2. Does Assignment to Inspection Predict ODI Attrition?

Sample =	(1) Randomized sample	(2) 2001– 2007 target lists	(3) ...and employment [NETS] remained above 40	(4) ...and never changed industry	(5) ...and remained alive
Assigned to SST inspection	0.0022 (0.0055)	0.0077 (0.0058)	0.0086 (0.0058)	0.0087 (0.0062)	0.0062 (0.0052)
Directive-year fixed effects	Y	Y	Y	Y	Y
Observations	16,141	11,622	10,738	9,373	8,355
R-squared	0.041	0.003	0.003	0.003	0.003
Dependent variable sample mean	0.851	0.891	0.901	0.902	0.941

Each column reports estimates from a separate OLS regression, in which the dependent variable is an indicator of whether an establishment has ODI-reported data in any of the five years made up of the directive year and the four following years. Robust standard errors in parentheses. +p<.1, *p<.05, **p<.01.

Appendix D Validity of ODI-reported Injury Rates

Because our analysis relies on injury data that establishments self-report to OSHA as part of the OSHA Data Initiative (ODI), data accuracy could be a concern.

First, we note that measurement error in ODI-reported injuries—at least for those injuries employees report to their employer—might not be a significant concern in practice. Messiou and Zaidman (2005) compared establishment-level workers' compensation data to ODI-reported data in 2003 and—while finding differences—found no systematic underreporting of injuries to ODI. Moreover, OSHA routinely audits a random sample of ODI respondents to verify the accuracy of their ODI responses by comparing them to the establishment's OSHA log forms, assessing large fines if the ODI response is found to be inaccurate. The threat of such audits provides employers incentives to report accurately to ODI and OSHA's prior audits have found low rates of misreporting (ERG and National Opinion Research Center 2009; ERG 2013).

Still, one may be concerned about measurement error affecting our estimates. There is very likely classical measurement error (that is, pure noise) in injuries reported to ODI. In addition, there is evidence in other contexts that injuries reported to government surveys are often an undercount. Many factors could explain this divergence, but two primary ones are that (a) some employees might not report some injuries to their employers and (b) some employers might not report some injuries to OSHA (for a thorough discussion of these factors, see Azaroff, Levenstein, and Wegman 2002). Younger employees are less likely to report their injuries, as are employees who suffer less-serious injuries and those who work in states that offer less-generous workers' compensation benefits (Biddle and Roberts 2003). Smaller employers are less likely to report injuries to OSHA (Oleinick, Gluck, and Guire 1995; Dong et al. 2011) and all employers are less likely to report less-serious injuries to OSHA (Boden, Nestoriak, and Pierce 2010).⁵

As long as these sources of measurement error in injury reporting are unaffected by OSHA inspections, they will increase the standard errors of our estimates but not bias the coefficients. More worrisome is the potential for inspections to affect the accuracy of self-

⁵ A few studies have compared the US Bureau of Labor Statistics' Survey on Occupational Injuries and Illnesses (SOII) to workers' compensation data to estimate the reliability of data collected by the SOII. While the SOII is distinct from ODI, its format is quite similar and both rely on employers' logs of OSHA-recordable injuries; thus, lessons from these studies probably apply to ODI. A consistent finding is that injuries which are more acute and easier to diagnose (such as amputations) are reported quite accurately in the BLS survey, whereas chronic injuries (such as carpal tunnel syndrome), injuries that are more difficult to diagnose (such as those that result in hearing loss), and occupational illnesses are more likely to be misreported (Ruser 2008; Nestoriak and Pierce 2009).

reported injuries. On the one hand, inspections could *increase* reported injuries; for example, if OSHA issues recordkeeping violations that motivate employers to keep more complete injury records. In this case, even if inspections truly lead to lower injuries, this effect would bias regression estimates towards inspections *increasing* injuries reported to the ODI. On the other hand, inspections could *decrease* the reporting of subsequent injuries by leading establishments to perceive that this would reduce the likelihood of future inspections. In this case, inspections could lead to fewer *reported* injuries while having no effect on their actual occurrence.

While the extent of such bias is unobservable, we nonetheless address this concern in several ways. First, our primary outcome is the most serious class of injuries—those causing days away from work—for which the potential for measurement error is smaller than for total injuries (Boden, Nestoriak, and Pierce 2010). Second, given evidence that underreporting injuries is more common among smaller establishments, the extent of underreporting should be mitigated by the fact that ODI only surveyed establishments with at least 40 employees until 2009 and with at least 20 employees thereafter. Third, we conducted a robustness check which excluded from our analysis any establishment that OSHA had ever cited for a recordkeeping violation, which in our sample constitutes roughly 7 percent of establishments assigned to inspection and 4 percent of establishments not assigned to inspection.

References

- Azaroff, Lenore S., Charles Levenstein, and David H. Wegman.** 2002. “Occupational Injury and Illness Surveillance: Conceptual Filters Explain Underreporting.” *American Journal of Public Health* 92 (9): 1421–1429.
- Biddle, Jeff, and Karen Roberts.** 2003. “Claiming Behavior in Workers’ Compensation.” *Journal of Risk and Insurance* 70 (4): 759–780.
- Boden, Leslie I., Nicole Nestoriak, and Brooks Pierce.** 2010. “Using Capture-recapture Analysis to Identify Factors Associated with Differential Reporting of Workplace Injuries and Illnesses.” *2010 JSM Proceedings, Statistical Computing Section*. Alexandria, VA: American Statistical Association.
- Dong, Xiuwen S., Alissa Fujimoto, Knut Ringen, Erich Stafford, James W. Platner, Janie L. Gittleman, and Xuanwen Wang.** 2011. “Injury Underreporting among Small Establishments in the Construction Industry.” *American Journal of Industrial Medicine* 54 (5): 339–349.

- ERG.** 2013. “Analysis of OSHA’s National Emphasis Program on Injury and Illness Recordkeeping (RK NEP).” Prepared for the US Occupational Safety and Health Administration’s Office of Statistical Analysis [contract J-099-F-2-8441], November 1, 2013.
- ERG and National Opinion Research Center.** 2009. “OSHA Data Initiative Collection Quality Control: Analysis of Audits on CY 2006 Employer Injury and Illness Recordkeeping.” Prepared for the US Occupational Safety and Health Administration’s Office of Statistical Analysis [contract no. J-099-F-2-8441], November 25, 2009, <https://www.reginfo.gov/public/do/DownloadDocument?objectID=14894901>.
- Messiou, Eleni, and Brian Zaidman.** 2005. “Comparing Workers’ Compensation Claims and OSHA Data Initiative Cases.” St. Paul, MN: Minnesota Department of Labor and Industry.
- Nestoriak, Nicole, and Brooks Pierce.** 2009. “Comparing Workers’ Compensation Claims with Establishments’ Responses to the SOII.” *Monthly Labor Review* 132 (5): 57–64.
- Oleinick, Arthur, Jeremy V. Gluck, and Kenneth E. Guire.** 1995. “Establishment Size and Risk of Occupational Injury.” *American Journal of Industrial Medicine* 28 (1): 1–21.
- Ruser, John W.** 2008. “Examining Evidence on Whether BLS Undercounts Workplace Injuries and Illnesses.” *Monthly Labor Review* 131 (8): 20–32.

Appendix E Estimating the Social Cost of Serious Injuries during Our Sample Period

Waehrer et al. (2007) estimate that the social cost of a serious injury—the combined costs to employers, workers, and the rest of society from injuries causing days away from work—in 2002 was \$37,016. Our goal is to estimate the cost in 2005—the median year of our sample—but in 2018 dollars.

According to Leigh (2011), medical costs make up roughly a quarter of the social costs of injuries, with the remaining three-quarters made up of indirect costs such as foregone wages and loss to home production (e.g. child care, cooking and cleaning). Because medical care spending rose 24 percent from 2002 to 2005,⁶ we scale up 25 percent of the \$37,016 figure (\$9,254) by 24 percent to estimate that the medical cost portion was \$11,475 in 2005. We assume that indirect costs grew at the rate of inflation, noting that, according to the US Bureau of Labor Statistics, the consumer price index (CPI) rose by 8.6 percent from 2002 to 2005.⁷ We therefore scale up the indirect portion (75 percent of \$37,016) by 8.6 percent to estimate it as \$30,150 in 2005. Thus, we estimate a serious injury in 2005 to cost \$41,625 (\$11,475 + \$30,150) in 2005 dollars.

To convert this to 2018 dollars, we note that the US Bureau of Labor Statistics reports that the CPI rose by 28.6 percent from 2005 to 2018.⁸ Thus, the social cost of a 2005 serious injury in 2018 dollars is $\$41,625 * 1.286 = \$53,529$.

References

Leigh, J. Paul. 2011. “Economic Burden of Occupational Injury and Illness in the United States.” *Milbank Quarterly* 89(4): 728–772.

US Bureau of Labor Statistics. 2015. “Industry Injury and Illness Data.” http://www.bls.gov/iif/oshsum.htm#10Supplemental_News_Release_Tables.

⁶ Peterson-Kaiser Health Systems Tracker, “US Health Expenditures 1960–2015,” <http://www.healthsystemtracker.org/interactive/?display=US%2520%2524%2520Billions&service=Hospitals%2520CPhysicians%2520%2526%2520Clinics%2520CPrescription%2520Drug>, accessed July 2017.

⁷ Analysis based on Consumer Price Index Databases, US Bureau of Labor Statistics, available at <https://www.bls.gov/cpi/data.htm>. Accessed August 2022.

⁸ Analysis based on Consumer Price Index Databases, US Bureau of Labor Statistics, available at <https://www.bls.gov/cpi/data.htm>. Accessed August 2022.

Wahrer, Geetha M., Xiuwen S. Dong, Ted Miller, Elizabeth Haile, and Yurong Men. 2007.
“Costs of Occupational Injuries in Construction in the United States.” *Accident Analysis & Prevention* 39 (6): 1258–1266.

Appendix F Pre-specification Plan

We pre-specified our design and posted our subsequent pre-analysis plan to the Open Science Framework at <https://osf.io/2snka/>.

The first version of our pre-analysis plan, posted in July 2015, provided the basic outline of our study and described our primary outcome variables and our planned empirical specifications to estimate the overall effects of inspections. We also uploaded the Stata code we would use to estimate our regressions.

To ensure that estimates are not vulnerable to large outliers or threats to identification, we conducted a series of tests to pre-specify our regression specification. We first blinded ourselves to each establishment's assignment status. In each of 500 simulation runs, we created an *assigned to placebo inspection* dummy that randomly assigned 0 or 1 to each establishment-directive in our randomized sample, with a probability corresponding to the proportion of establishments that OSHA historically assigned to inspection. We then estimated the effect of *assigned to placebo inspection* on various outcomes using several functional forms and approaches to handle outliers.⁹ Our objective was to identify which specifications most often yielded a precisely estimated zero coefficient on the *assigned to placebo inspection* dummies across all simulations. This procedure led us to choose the specification described in Section 2.1.2.¹⁰

After posting this plan, we found several minor glitches in our pre-specified design, which we therefore updated over the next months. For example, because we initially believed a large share of establishments assigned to control in one year would become assigned to treatment (that is, assigned to inspection) in later years, we originally planned to estimate the effects of inspections using outcomes within a window of three years before and after the focal year. However, while creating our analysis sample, we learned that this “crossover” of controls was not as large as we thought and that our power would increase if we estimated outcomes using a window of four years before and after the focal year. As another example, we pre-specified that one specification would control for “employment” but we had intended “ln(employment).”

⁹ For example, we considered OLS specifications in regressions with injury rate (both level and log) as the dependent variable.

¹⁰ We posted our pre-analysis plan to the Open Science Framework on July 10, 2015 at <https://osf.io/2snka/>.

After specifying our randomized sample in the original pre-analysis plan, we learned of some unique features of the SST program in 2002 and 2003 that we deemed important to incorporate into our analysis. We also made some improvements to our fuzzy linking between the SST target lists and IMIS, which slightly changed our analysis sample.

We incorporated these changes in an updated version of our pre-analysis plan, which we uploaded to the Open Science Framework in January 2016.

Our initial and updated pre-analysis plans included plans to assess the effects of OSHA inspections on business outcomes, but we decided not to include those analyses in this paper.

Appendix G Predictors in the Machine Learning Analyses

This appendix lists the variables we included in the two machine learning exercises:

- 1) Using causal forest to construct establishments' *estimated treatment effects*
- 2) Using Super Learner to construct establishments' *predicted injuries*

When any establishment was missing a variable, we replaced it with that variable's sample mean.

Location and year variables

- Dummies for 10 OSHA regions
- Dummy if the establishment is in a large metro area
- Number of days after a work-related injury until the injured worker can receive workers' compensation, as determined by the establishment's state
- State leave-one-out mean¹¹ annual serious injury rate,¹² lagged 2 years
- Dummies for directive year

Industry and size variables

- Establishment's annual total working hours, lagged 1 and 2 years
- Establishment's annual log employment reported in NETS, lagged 1 year
- 4-digit SIC leave-one-out mean annual serious injury rate, lagged 2 years
- 3-digit SIC leave-one-out mean annual penalties assessed at OSHA inspections, lagged 1 year
- Dummy for manufacturing sector
- Dummy for nursing home sector

Compliance-related variables

- Dummy if establishment had any OSHA inspection prior to directive year
- Dummy if establishment had any OSHA complaint inspection from $t-1$ to $t-3$

¹¹ A leave-one-out mean is the mean of the variable, excluding the focal establishment.

¹² Serious injury rate refers to the annual number of injuries causing days away from work per 100 full-time workers.

Other establishment characteristics

- Establishment age reported in NETS, lagged 2 years
- Dummy for standalone firm, lagged 1 year
- Establishment's minimum monthly PAYDEX score, lagged 2 years

Variables related to injuries and to ODI

- Number of years establishment was previously on an SST target list
- Establishment's average annual number of serious injuries, $t-1$ to $t-4$
- Establishment's annual serious injury rate, lagged 1, 2, and 3 years
- Establishment's annual transfer/restriction injury rate, lagged 2 years
- Establishment's annual other recordable injury rate, lagged 2 years
- Establishment's annual serious injury rate, squared, lagged 2 years
- Establishment's total annual number of days away from work due to injuries, lagged 2 years
- Dummy for "has ODI data in $t-1$ "
- Dummy for "has ODI data in $t-3$ "
- Establishment's annual serious injury rate from $t-2$, interacted with dummies for 4 employment quartiles (from NETS) in $t-2$.

The causal forest to estimate treatment effects also included the percentage of establishments selected for inspection in the establishment's area-office-directive primary or secondary list.

Appendix H Assessing Generalizability: Can Results from the Randomized Sample be Applied to the Nonrandomized Sample?

Our estimation of the number of injuries OSHA could avert under alternative targeting policies relies on using half of the randomized sample to (a) generate *predicted injuries* that establishments would experience if not assigned to inspection (that is, $Y(0)$), and (b) generate *estimated treatment effects* that establishments would experience if they were assigned to inspection (that is, $Y(1) - Y(0)$). We then use these estimates, which were trained and validated on the randomized sample, to estimate the effects of counterfactual policies that target establishments from the overall historical SST target lists, including establishments beyond our randomized sample (the “nonrandomized sample”). Establishments in the nonrandomized sample include those that (a) were on an area office’s target list in a directive year in which fewer than 5 percent or more than 95 percent of the establishments listed were assigned to inspection, (b) had been inspected under SST in the prior two years, and (c) met other exclusion criteria described in Table B.1 in Appendix B. In this appendix, we first briefly describe our approach to generalizing from the randomized sample. We then discuss and empirically assess threats to the validity of our approach.

Generalizing from the Randomized Sample

We adapt the Chernozhukov et al. (2020) procedure to generate estimates that pertain to the entire historical SST target lists as follows. Each time we partition the randomized sample into training and holdout samples, we also keep a random 50-percent partition of the nonrandomized sample. In the training sample, we use the causal forest to estimate treatment effects and the Super Learner to estimate predicted injuries if not assigned to inspection. We then apply these models to construct *predicted injuries* and *estimated treatment effects* for all establishments in the holdout sample *and* the random 50-percent subset of the establishments in the nonrandomized sample that we kept. We use the combined holdout sample and nonrandomized subsample to construct the groupings that correspond to each targeting policy. Next, we estimate regressions of Equation 5 on the holdout sample to generate the coefficients ($\hat{\gamma}$ s) that correspond to the mean number of injuries averted among the establishments in each group. We estimate the total number of injuries averted under a counterfactual policy by

multiplying the $\hat{\gamma}$ s by the number of establishments in each group (in both the randomized and nonrandomized samples) that OSHA assigns to inspection in the policy.

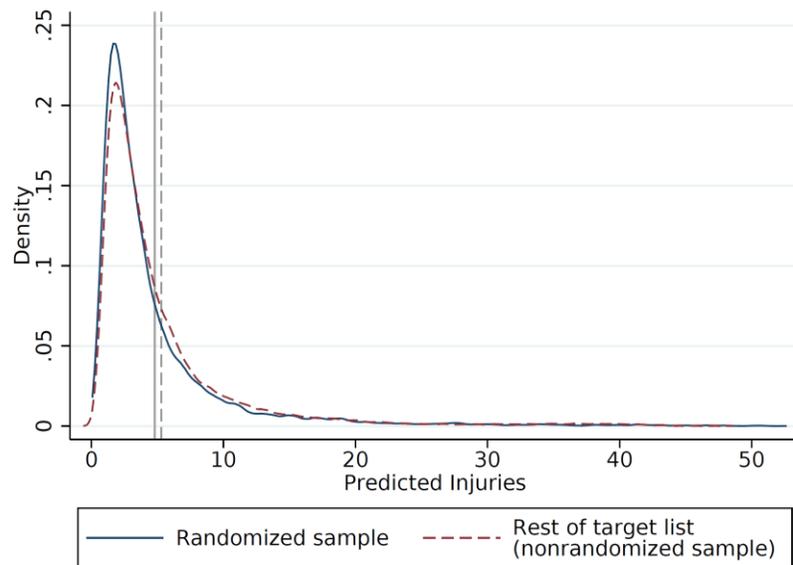
Addressing Potential Threats to Validity

There are two reasons that the estimates from our procedure might not generalize to the entire historical SST target lists and thus might result in misleading estimates of targeting policy counterfactuals. First, if establishments in the randomized and nonrandomized samples have different distributions of observable characteristics (different Z s), then our estimate of the average treatment effect of establishments that would be targeted in a particular policy (estimated on the randomized sample only) would be biased. Second, establishments in the randomized and nonrandomized sample could have different *unobservable* characteristics, in which case a causal forest model estimated on the randomized sample would poorly predict treatment effects for the nonrandomized sample (even if the two groups' distributions of observables were indistinguishable). We address these concerns below.

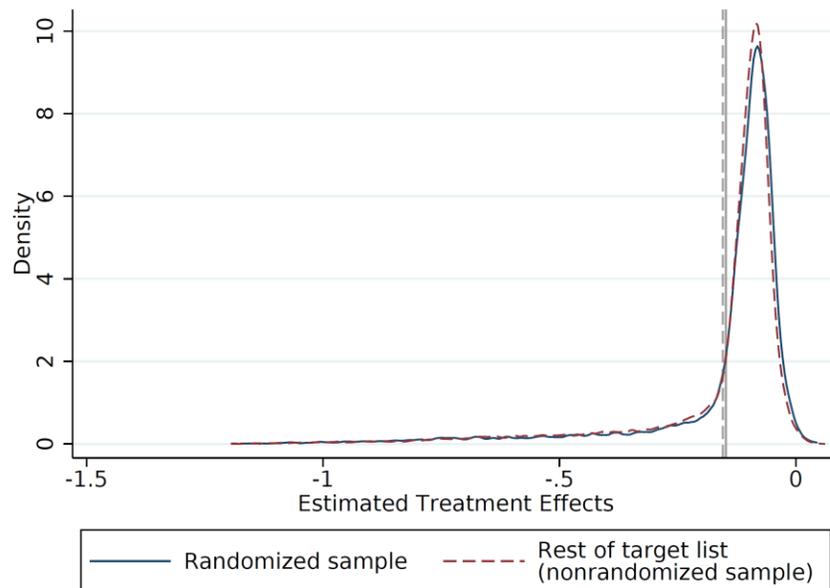
Do Establishments in the Two Samples Have Different Observable Characteristics?

If establishments in the randomized and nonrandomized samples have different distributions of observable characteristics, then the two groups might well have different distributions of underlying *treatment effects* and *expected injuries*. However, this potential concern is alleviated in our context because we find that the distributions of our estimates of these two metrics in the randomized sample are very similar to their distributions in the nonrandomized sample, as is apparent in Panels (a) and (b) of Figure H.1. Panel (a) depicts the distribution of *predicted injuries* generated from applying a Super Learner that was trained on those establishments in the randomized sample that were not assigned to inspection, and then applied to the rest of the randomized sample and to the nonrandomized sample. Panel (b) depicts the distribution of estimates of *treatment effects* generated from applying a causal forest on all establishments in the randomized sample (including those assigned to inspection) and then applied to the nonrandomized sample. Because the distributions of both metrics are very similar across the two samples, it is unlikely that any differences in *observable* characteristics between the randomized and nonrandomized samples render our estimates of counterfactual policies inconsistent.

Figure H.1. Distribution of *predicted injuries* and *estimated treatment effects* for Establishments on the SST Historical Target List in the Randomized versus Nonrandomized Sample
 Panel A: *Predicted injuries*



Panel B: *Estimated treatment effects*



The sample for these kernel density plots is the establishments on the 2001–2010 historical SST target lists. *Predicted injuries* are estimated using a Super Learner to predict \bar{y}_{it}^{post} among all establishments not assigned to inspection in the randomized sample and extrapolating the estimates to the rest of the target list. *Estimated treatment effects* are estimated using a causal forest to predict treatment effects for all establishments in the randomized sample and extrapolating the estimates to the rest of the target list.

Do Establishments in the Two Samples Have Different Unobservable Characteristics?

Our machine learning estimates of the models underlying an establishment's treatment effect of assignment to inspection and its *expected injuries* if not assigned to inspection are estimated on the randomized sample. These models could have poor out-of-sample predictive power for the nonrandomized sample if establishments in the two samples had different *unobservable* characteristics.

Unfortunately, we cannot test the out-of-sample predictive power of our model to estimate treatment effects because an establishment's underlying *treatment effect* ($Y(1) - Y(0)$) is unobservable. Fortunately, we *can* test whether the model that predicts *expected injuries* ($Y(0)$) generates predictions that are less accurate in the nonrandomized sample than in the randomized sample because uninspected establishments' injuries are (eventually) observable. We can do the same for an estimate of the model determining the number of injuries an establishment would experience if assigned to inspection ($Y(1)$).

Recall that in the Chernozhukov et al. (2020) algorithm, we estimate *expected injuries* ($Y(0)$) using a Super Learner to predict $\overline{y_{it}^{post}}$ (the number of serious injuries an establishment experiences averaged over the assignment year and four subsequent years) for the establishments assigned to control in the training sample. We then use the results to generate $\widehat{Y(0)}$ —our prediction of *expected injuries*—for those in the holdout sample and the nonrandomized sample. We repeat this process 250 times.

To test whether the predictions of *expected injuries* are more accurate or less accurate for the holdout subset of the randomized sample compared to the nonrandomized sample, we conduct the following exercise. For each of our 250 iterations, among establishments not assigned to inspection, we regress realized $\overline{y_{it}^{post}}$ on $\widehat{Y(0)}$ separately for those in the holdout sample and those in the nonrandomized sample. Among those establishments that OSHA did not assign to inspection, $\overline{y_{it}^{post}}$ equates to $Y(0)$. Thus, we are regressing establishments' realized *expected injuries* ($Y(0)$) on its predicted value, $\widehat{Y(0)}$. We save the median coefficient and standard error on $\widehat{Y(0)}$, as well as the median R^2 , across the 250 iterations.

We find that for these establishments not assigned to inspection, both in the holdout sample and in the nonrandomized sample, the median coefficients on $\widehat{Y(0)}$ are close to 1 and the estimates are statistically indistinguishable from each other (Table H.1, Columns 1 and 2). The

R^2 s are high and nearly identical and the root mean squared error (MSE) are also nearly identical. In other words, our model to predict *expected injuries*, which was estimated on half of the randomized sample, has equal predictive power with respect to realized injuries for the other half of the randomized sample and for the nonrandomized sample.

We go one step further to perform a similar exercise to assess predictions of the number of injuries an establishment would experience if assigned to inspection; that is, $E[Y(1)|Z]$, or $Y(1)$. In each of our 250 iterations, we estimate $Y(1)$ using a Super Learner to predict injuries among those in the training sample that were assigned to inspection. We then use this model to generate $\widehat{Y(1)}$ for both the holdout and nonrandomized samples. Analogous to what we did before, in each iteration we regress realized y_{it}^{post} on its predicted value for establishments assigned to inspection, first in the holdout sample and then in the nonrandomized sample.

As reported in Columns 3 and 4 of Table H.1, we find that $\widehat{Y(1)}$ has essentially equal predictive power for the holdout sample and the nonrandomized sample.

Overall, this exercise bolsters confidence that our estimates derived from applying the Chernozhukov et al. (2020) procedure to the randomized sample generate unbiased estimates of the number of injuries OSHA could avert by re-targeting the *entire* SST target list.

References

Chernozhukov, Victor, Mert Demirer, Esther Duflo, and Ivan Fernandez-Val. 2020. "Generic Machine Learning Inference on Heterogenous Treatment Effects in Randomized Experiments." NBER Working Paper No. 24678.

Table H.1. Comparing the Predictive Power of Establishments’ Estimated Injuries If Not Treated, $B(Z)$, for the Randomized Controls and for the Nonrandomized Samples

Dep var =	Average annual # of serious injuries during post-period			
	(1)	(2)	(3)	(4)
Sample =	Establishments <i>not assigned</i> to inspection in the...		Establishments <i>assigned</i> to inspection in the...	
	Holdout sample (subset of the randomized sample)	Nonrandomized sample	Holdout sample (subset of the randomized sample)	Nonrandomized sample
Expected injuries if not assigned ($\widehat{Y(0)}$)	1.028 (0.033)	1.063 (0.008)		
Expected injuries if assigned ($\widehat{Y(1)}$)			1.025 (0.036)	1.086 (0.013)
Root MSE	3.82	4.03	3.84	4.20
R ²	0.68	0.70	0.68	0.71

The table assesses whether our machine-learning–based estimates of the annual number of serious injuries an establishment would experience if not assigned to inspection ($Y(0)$) and the number it would experience if it were assigned to inspection ($Y(1)$) have differential predictive power for establishments in the randomized sample vs. the nonrandomized sample. For each of 250 splits of the data, we train the Super Learner algorithm on establishments not assigned to inspection in the training sample (a random half of the randomized sample) to train the model to predict $Y(0)$ and then apply the model to construct *predicted injuries* ($\widehat{Y(0)}$) in the holdout sample (the other half of the randomized sample) and the nonrandomized sample. In each split, focusing on establishments that were not historically assigned to inspection, we regress establishments’ realized mean annual serious injury count on their $\widehat{Y(0)}$, separately for establishments in the holdout sample (Column 1) and the nonrandomized sample (Column 2). We use an analogous procedure in Columns 3 and 4 to assess the predictive power of $\widehat{Y(1)}$, restricting to establishments that were assigned to inspection in the historical policy. The table reports the median coefficient and SE on $\widehat{Y(0)}$ and $\widehat{Y(1)}$ and the median R², across the 250 splits.

Appendix I Robustness Checks on Estimate of Intention-to-treat Effect on Injuries

We consider several other specifications as robustness checks for our ANCOVA model (Equation 1).

In a pre-specified robustness check, we estimated the effect of assignment to inspection using a difference-in-differences specification.¹³

We also sought to minimize the chances that our estimates were contaminated by inspections leading to more complete reporting of injuries. Therefore, we reestimated the intention-to-treat specification corresponding to Column 1 of Table 3 but excluded any establishment that OSHA had ever cited for recordkeeping violations during our sample period (7 percent and 4 percent of establishments assigned and not assigned to inspection, respectively).¹⁴

We also averaged the outcomes during the directive year and four following years into one observation per establishment-directive to construct a new dependent variable, $\overline{y_{it}^{post}}$. We reestimate the ANCOVA model (omitting the τ -year fixed effects, θ_τ) on this outcome, using OLS rather than Poisson because the outcome includes non-integer values.

As a final check, we report the average intention-to-treat estimate from the Chernozhukov et al. (2020) procedure. Chernozhukov et al. (2020) also lets us estimate the overall average effect of assignment to inspection, which is β_I from the following regression model estimated on the holdout sample:

$$Y = \alpha_1 + \alpha_2 B(Z) + \beta_1 (D - p(Z)) + \beta_2 (D - p(Z)) * (S - ES) + \epsilon. \quad (\text{I. 1})$$

See Chernozhukov et al. (2020) for details on notation. We estimate $\widehat{\beta}_1$ from the holdout sample. As noted in Section 2.3.1, we conduct 250 iterations of this process on 250 sample splits, then reverse the roles of the training and holdout samples and repeat the same process, obtaining another set of coefficients for each iteration pertaining to the other half of the sample split. For

¹³ We estimate the following difference-in-differences regression model for this robustness check:

$$y_{ijt\tau} = \alpha_1 \text{Assigned}_{it} * \mathbb{1}(\tau \geq 0) + \mu_{jt} * \mathbb{1}(\tau \geq 0) + \lambda_{it} + \theta_\tau + \epsilon_{jit\tau}.$$

All variables here are defined as in Equations 1 and 2. We control for P_τ separately to account for changes in injuries (and other outcomes) following the directive year that would have occurred even without an SST inspection.

¹⁴ As one other robustness check, we drop establishment-year observations for which the establishment, according to the NETS database, is no longer in operation. We lose only a few hundred observations and obtain essentially identical estimates (results not shown).

each iteration we take the average of these two coefficients. We use the median point estimate as our point estimate of the average treatment effect and use the 2.5th and 97.5th percentiles as the range of our 90-percent confidence interval.

We report results from these specification checks in Table I.1. The difference-in-differences estimate (Column 1: $\beta = -0.033$, $SE = 0.017$) is nearly identical to the estimate from the ANCOVA specification ($\beta = -0.035$, $SE = 0.017$). Excluding establishments with any recordkeeping violation during our sample period (Column 2) yields a coefficient ($\beta = -0.040$, $SE=0.018$) slightly larger in magnitude than that from our baseline specification, which is consistent with the idea that establishments cited for recordkeeping violations are subsequently less likely to underreport injuries. Our OLS estimate on the collapsed outcome variable (Column 3) yields a point estimate of $\beta = -0.178$ ($SE = 0.081$), which, as a percent of the control mean ($-0.178 / 4.62 = 3.8$ percent), is essentially identical to the estimate from the Poisson model. Finally, the point estimate using Chernozhukov et al. (2020) ($\beta = -0.168$, 90% CI = $[-0.193, -0.140]$) is quite similar to our OLS estimate in Column 3 but is actually more precise. In short, our results are robust to these several specification checks.

References

Chernozhukov, Victor, Mert Demirer, Esther Duflo, and Ivan Fernandez-Val. 2020. "Generic Machine Learning Inference on Heterogenous Treatment Effects in Randomized Experiments." NBER Working Paper No. 24678.

Table I.1: Intention-to-treat Effects of SST Inspection on Serious Injuries: Robustness and Alternative Specification

	(1)	(2)	(3)	(4)
	Dep var = Number of serious injuries			
	Fixed effects	Drop record-keeping violators	Collapse to mean over the post-period	[Chernozhukov et al. (2020)]
	[Poisson]	[Poisson]	[OLS]	
Assigned to inspection	-0.033 (0.017)*	-0.040 (0.018)*	-0.178 (0.081)*	-0.168 [-0.193, -0.140]**
# observations	89,509	38,171	13,736	13,736
# establishment-directive years	15,715	12,909	13,736	13,736
# establishments	12,630	10,509	11,083	11,083
# area-office-directives	383	383	383	383
Mean dep var, estabs not assigned, post-period	5.37	5.26	4.62	4.62

The table shows the results of regressions with coefficient on an *Assigned to inspection* dummy, SEs in parentheses in Columns 1–4 and 90% confidence interval in brackets in Column 5. *Assigned to inspection* is equal to 1 in years beginning with the directive year, for establishments selected for SST inspection in the directive year. OLS and Poisson coefficients are estimates of the level change and percent change, respectively, in the dependent variable associated with *Assigned to inspection*.

Poisson drops establishments with only one observation or with constant values across observations.

Columns 2 and 3 report results from an ANCOVA regression. Column 1 reports results from a specification with establishment fixed effects. If an establishment appears more than once, a separate establishment fixed effect is included for each directive year. Standard errors in Columns 1–3 clustered by establishment. **p<.01., *p<.05, †p<.1.

Column 4 reports the estimate of the average ITT from the procedure in Chernozhukov et al. (2020), described in the text.

Each regression is restricted to the randomized sample, described in Table A.1, and to a window of four years before and after the directive year.

Appendix J Assessing the Lucas Critique: Can We Estimate Effects of New Targeting Rules Using Historical Data?

Our approach to estimating the effects of alternative targeting strategies does not consider the potential behavioral effects on *uninspected* establishments. Because alternative targeting strategies change the *threat* of inspection, such behavioral changes could render our estimates misleading. In Section 3.4.1, we provided evidence that such effects are, in fact, unlikely to be important in our setting. This appendix provides more details and evidence to support that conclusion. In particular, we use a regression discontinuity (RD) design that yields no evidence that establishments experienced a change in injuries in response to discrete changes in the threat of inspection in the historical SST program. These non-results bolster confidence that our main analyses yield accurate estimates of the effects of alternative targeting regimes despite our not considering their potential effects on *uninspected* establishments.

Quantifying the Change in the Threat of Inspection at the Primary and Secondary Cutoffs

To determine establishments' distance from the primary and secondary list cutoffs, we reviewed the annual SST directives over 2001–2010 to identify the relevant injury rate cutoffs for each year's directive (which in all cases were based on injury rates two years prior). As described in Appendix A, in the early years of this range, these cutoffs were based exclusively on DART rates. For example, the 2002 primary list comprised establishments whose DART rates exceeded 14 in the year 2000. In the later years, the cutoffs were a function of DART *and* DAFW rates. For example, the 2006 primary list comprised establishments whose 2004 DART rates exceeded 12 and establishments whose 2004 DAFW rates exceeded 9. For each directive year t , we calculate the difference between establishment i 's relevant injury rate from $t-2$ and the relevant cutoff (which we designate c). For example, $DA RT_{i,t-2} - c_{pri,dart,t}$ is this quantity for an establishment's DART rate relative to the DART element of the primary list cutoff. The distance metric to determine an establishment's eligibility for the primary list in year t is calculated as:

$$Distance_{it}^{pri} = \max\{(DA RT_{i,t-2} - c_{pri,dart,t}), (DA FW_{i,t-2} - c_{pri,dafw,t})\}. \quad (J.1)$$

An establishment is eligible for the primary list in year t if this distance metric is zero or positive; larger values denote establishments with injury rates further away (increasingly worse) than the cutoff. We analogously calculate this distance metric for the secondary list.

Our sample restrictions for this analysis are similar to those in our main analysis: we drop establishments (a) whose recent inspection histories make them ineligible for SST inspection, (b) whose ODI data quality were flagged as questionable, (c) that were not alive in year $t-2$ according to NETS, or (d) that reported fewer than 40 employees to ODI in year $t-2$. See Table B.1 for details about these criteria.¹⁵ The regression discontinuity design does not require randomization, so we include all area offices, not just the subset of those that randomized their target lists.

Figures 3a and 3b and the associated discussion in Section 3.4.1 provided graphical evidence that establishments just above the primary or secondary cutoffs faced a higher threat of inspection than those just below. To quantify the magnitude of this change in threat level, we estimate the following regression model on the sample described in the previous paragraph:

$$y_{it\tau} = \alpha + \gamma * \mathbb{1}(Distance_{it}^{pri} \geq 0) + f(Distance_{it}^{pri}) + \beta X_{it} + \epsilon_{it\tau}. \quad (J.2)$$

The outcome variable y is an indicator coded 1 if establishment i was SST-inspected in year t or $t+1$. $\mathbb{1}(\cdot)$ is an indicator function that equals 1 for establishments eligible for the primary list—that is, when $Distance_{it}^{pri} \geq 0$ —and 0 for those ineligible. $f(\cdot)$ is a polynomial in $Distance_{it}^{pri}$; we use the approach developed by Calonico et al. (2019) to select the degree of the polynomial. To improve precision on our estimate of γ , we include the following controls in X_{it} : year fixed effects, OSHA-region fixed effects, manufacturing and nursing-home-industry dummies, and the log of the establishment’s total working hours in year $t-2$. We use an analogous specification—but with $Distance_{it}^{sec}$ —for the secondary list cutoff. To establish the sample for these regressions, we use the approach developed by Calonico et al. (2019) to select the bandwidth around the cutoffs—that is, the range of $Distance_{it}^{pri}$ and $Distance_{it}^{sec}$ above and below zero included in our regressions—that minimizes mean squared error (MSE) and which we refer to as the MSE-optimal bandwidth.

Table J.1 reports the results. Corroborating the visual evidence in Figure 3b, Column 1 reveals that just barely making it above the primary cutoff yields a 31-percentage-point increase

¹⁵ We also drop nursing homes for all years except 2009–2010 for this analysis. Nursing homes were excluded from the secondary list in all years prior to 2009. While they were included on the primary list in earlier years, their inclusion was not based on the same cutoff rule as other industries.

in the probability of receiving an SST inspection in that or the following year ($p < 0.01$). Column 2 reveals an essentially identical increased threat of inspection at the secondary cutoff.

Quantifying the Change in Injuries Associated with the Change in Threat of Inspection at the Primary and Secondary Cutoffs

Having established that the threat of inspection changes substantially at the primary and secondary list thresholds, we use a regression discontinuity design to examine whether establishments above these thresholds, which faced a greater threat of inspection, had better safety outcomes, which would suggest that establishments facing a higher inspection threat improved their safety management effort. Because a regression discontinuity design is valid only if being just above or just below the cutoff is random, we first conduct a test of whether establishments manipulated their injury rates around the cutoffs, in which case the randomization would be violated.

Do Establishments Manipulate Reported Injuries to Reduce Threat of Inspection?

As shown above, among establishments with injury rates near the cutoff for primary or secondary list, those whose injury rates two years prior were just above a cutoff face discontinuously higher threat (risk) of inspection than those whose historical injury rates were just below a cutoff. If the threat effect were important and managers knew OSHA's targeting rule, establishments would have an incentive to underreport their injuries to keep below the cutoffs. We look for evidence of this by assessing whether there is bunching of reported injury rates just below the cutoff levels for the primary and secondary lists.

Conceptually, such bunching is unlikely for two reasons. First, it is difficult for establishments to misreport their injuries to ODI, as explained in Appendix D. Second, the cutoffs for the SST directive in a given year were based on injury rates from two years prior and the SST cutoffs usually changed from year to year, making them difficult to predict.

To test for the presence of bunching, Figure J.1 reports the density of $Distance_{it}^{pri}$ and $Distance_{it}^{sec}$. If establishments are strategic and able to misreport their injuries, then we should see bunching of reported injuries just below the cutoffs for the primary and secondary lists. However, the density appears smooth in both cases and a formal manipulation test based on

Cattaneo, Jansson, and Ma (2018) fails to reject that the density is not continuous.¹⁶ Thus, we find no evidence that establishments strategically misreport injuries to reduce their threat of inspection.

Do Establishments Facing a Higher Risk of Inspection Subsequently Experience Fewer Injuries?

Comparing injury rates of establishments just above these cutoffs to those of establishments just below them risks confounding two factors that might improve safety: the threat of being inspected and the experience of being inspected. The latter contaminates our desired estimate of the threat factor because establishments above these thresholds *are actually inspected* at higher rates than those below these thresholds.

To isolate the threat effect on subsequent injury rates, we leverage the fact that in each year, some area offices inspected only a subset of establishments on their primary list and that other area offices completed their primary list and inspected only a subset of establishments on their secondary list. Thus, many establishments on primary and secondary lists were not inspected; we focus on a subset of these.

Specifically, we use a regression discontinuity approach to estimate Equation J.1 for the following subsamples for which γ should reflect purely an estimate of increased threat of inspection. We first focus on establishments eligible for SST inspection that met the following three criteria: they (a) were in the jurisdiction of an area office that began its primary list, (b) had injury rates just above or just below the primary list cutoff (defining “just above” and “just below” using the MSE-optimal bandwidth approach described above), and (c) were *not* selected for SST inspection. We then focus on a second set of establishments eligible for inspection that met the following three analogous criteria pertaining to the secondary list: they (a) were in the jurisdiction of an area office that completed its primary list and began its secondary list, (b) had injury rates just above or just below the *secondary* list cutoff, and (c) were *not* selected for SST inspection. Establishments in these two subsamples faced a discontinuous threat of inspection at the cutoff but, due to OSHA’s resource constraints, were not actually inspected.

Table J.2 reports our results. In Column 1, the dependent variable is an indicator coded 1 when an establishment was subjected to an SST inspection in year t or $t+1$. Even though the

¹⁶ This evidence corroborates Li and Singleton (2019), who also find no evidence of bunching around the SST cutoffs.

estimate in Column 1 of Table J.1 indicated that the *overall* threat of inspection changed substantially at the primary cutoff, the actual inspection rates for this subsample barely changed across the cutoff. In Column 2, the dependent variable is $\log(\text{serious injuries})$ in year $t+1$. Corroborating the visual evidence in Figure 3d, the estimate indicates no difference in subsequent injuries between those just above and just below the primary cutoff ($\beta = 0.00$, $SE = 0.06$). Columns 3 and 4 indicate a similar null threat effect for establishments just above the secondary cutoff.

We conduct another test to assess the effect of heightened threat on injury rates. Columns 5 and 6 focus on all establishments that (a) were in the jurisdiction of an area office that did not begin its secondary list and (b) had injury rates just above or below the secondary cutoff (again using the MSE-optimal bandwidth). While those just above the cutoff were at greater threat of inspection, none of the establishments in this sample were actually subjected to an SST inspection. As expected, Column 5 of Table J.2 shows that the change in the *realized* probability of inspection among these establishments near the cutoff is essentially zero. Corroborating the estimates in Columns 2 and 4, Column 6 shows no difference in subsequent injuries at the secondary cutoff for this subsample ($\beta=0.01$; $SE=0.03$).

To summarize, the estimates in this section provide no evidence that establishments' safety changed in response to a change in their threat of inspection. First, we showed that reported injuries were not strategically bunched just below the primary or secondary list cutoffs, which would have suggested strategic misreporting to reduce the threat of inspection. Second, our regression results indicate that uninspected establishments' injuries were unaffected by a change in their threat of inspection.

This unresponsiveness to changes in the threat of inspection implies that such potential behavioral changes by establishments in response to a change in targeting strategy would be second-order relative to the specific deterrence effects of *realized* inspections. Thus, these results give us confidence that our main analysis estimates the overall effects of alternative targeting strategies on injuries, and that effects on *uninspected* establishments are minimal and thus not a serious confounding factor in our context.

References

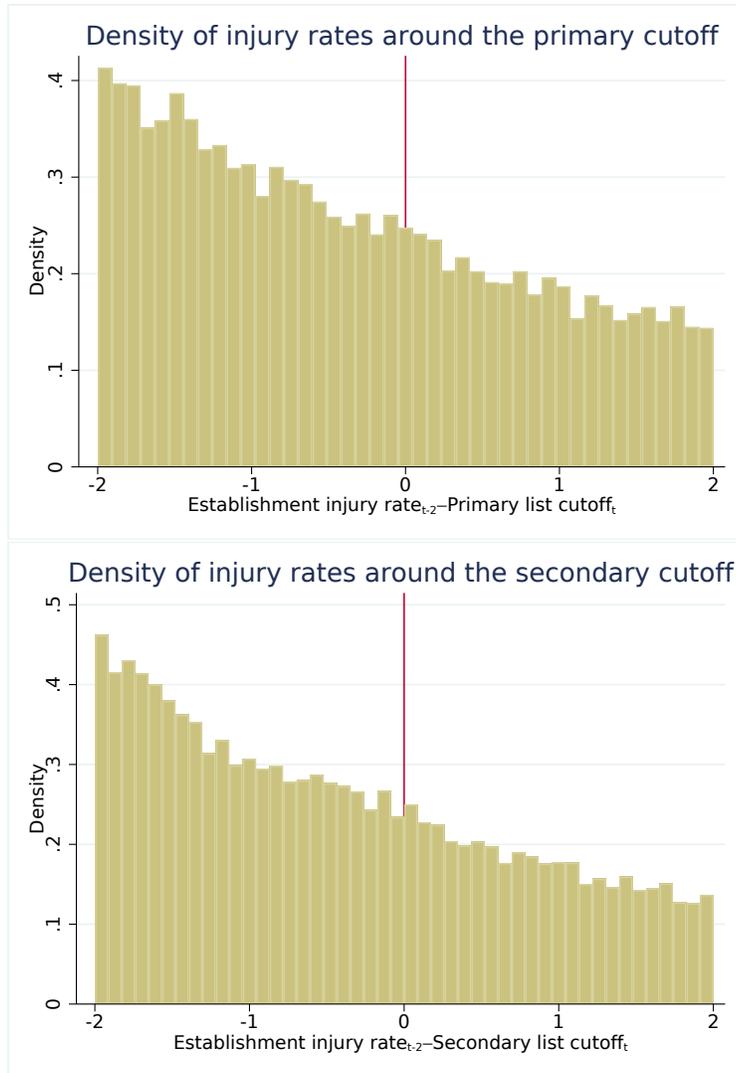
- Calonico, Sebastian, Matias D. Cattaneo, Max H. Farrell, and Rocio Titiunik. 2019.** “Regression Discontinuity Designs Using Covariates.” *Review of Economics and Statistics* 101 (3): 442–451.
- Cattaneo, Matias D., Michael Jansson, and Xinwei Ma. 2018.** “Manipulation Testing Based on Density Discontinuity.” *Stata Journal* 18 (1): 234–261.
- Li, Ling, and Perry Singleton. 2019.** “The Effect of Workplace Inspections on Worker Safety.” *ILR Review* 72 (3): 718–748.

Table J1: An Injury Rate Just Above the Primary or Secondary List Cutoff Raised Establishments' Risk of Inspection

Dep var =	(1)	(2)
Sample=	SST-inspected in t or $t + 1$	SST-inspected in t or $t + 1$
	Establishments in boundary of area office that... began primary list (c = primary cutoff)	Establishments in boundary of area office that... began secondary list (c = secondary cutoff)
Injury rate \geq cutoff	0.31 (0.01)**	0.30 (0.02)**
Robust p-value	0.000	0.000
# observations	22,054	8,170
Left bandwidth	2.30	1.15
Right bandwidth	1.53	1.70

The running variable in Column 1, defined in Equation H.1, determines whether an establishment's injury rates from two years prior render it eligible for the SST primary list. The running variable in Column 2 is defined analogously for eligibility for the secondary list. In each column, # observations refers to the number of establishments with injury rates in year $t - 2$ within the left and right bandwidths around the cutoff (also reported in each panel). The MSE-optimal bandwidths are selected, based on the method in Cattaneo, Jansson, and Ma (2018), separately for each regression. All regressions include controls for region fixed effects and year fixed effects, a manufacturing dummy, and the log of the establishment's working hours from $t - 2$.

Figure J1: Density of Establishments' Injury Rates Relative to the Primary and Secondary List Cutoffs



Note: These figures display the density of $Distance_{it}^{pri}$ and $Distance_{it}^{sec}$ —defined in Equation H.1—in the top and bottom panels, respectively. $Distance_{it}^{pri}$ and $Distance_{it}^{sec}$ are the “running variables” that determine whether an establishment’s injury rate from year $t - 2$ is high enough to make it eligible for the SST primary or secondary list, respectively, in year t . The sample includes those establishments that a) were considered for the 2001–2010 SST Target Lists (except nursing homes for all years prior to 2009), b) had not received an inspection in the prior two years, and c) reported injury rates that resulted in the absolute value of their $Distance_{it}^{pri}$ or $Distance_{it}^{sec}$ being less than 2.0.

Table J2: Establishments Facing a Higher Threat of Inspection Did Not Experience a Change in Injuries

Sample= Dep var =	(1)	(2)	(3)	(4)	(5)	(6)
	Establishments not assigned to inspection and in boundary of area office that... began primary		began secondary		did not begin secondary	
	SST-inspected in t or $t + 1$	log injuries, $t + 1$	SST-inspected in t or $t + 1$	log injuries, $t + 1$	SST-inspected in t or $t + 1$	log injuries, $t + 1$
Injury rate \geq cutoff	0.02 (0.01)	0.00 (0.06)	0.02 (0.01)*	-0.03 (0.05)	-0.00 (0.00)	0.02 (0.03)
Robust p-value	0.143	0.985	0.029	0.565	0.801	0.546
# observations	14,263	4,776	8,166	4,072	23,345	12,187
Left bandwidth	1.86	1.20	1.28	1.34	1.73	1.52
Right bandwidth	1.75	1.01	1.23	1.23	0.92	1.77

The running variable in Columns 1–2, defined in Equation H.1, determines whether an establishment’s injury rates from two years prior render it eligible for the SST primary list. The running variable in Columns 3–6 is defined analogously for eligibility for the secondary list. In each column, # observations refers to the number of establishments with injury rates in year $t - 2$ within the left and right bandwidths around the cutoff (also reported in each panel). The MSE-optimal bandwidths are selected, based on the method in Cattaneo, Jansson, and Ma (2018), separately for each regression. All regressions include controls for region fixed effects and year fixed effects, a manufacturing dummy, and the log of the establishment’s working hours from $t - 2$.

Appendix K Did Inspections of Establishments near the SST Cutoffs Reduce Injuries?

The analysis in Section 3.4.1 and Appendix J showed that the increased *threat* of inspection near the SST cutoff did not affect establishments' subsequent injury rates. Using a slightly different sample, we can also use this regression discontinuity approach to estimate the local average treatment effect of (realized) SST inspections on establishments' subsequent injuries. This approach complements our main approach outlined in Section 2.3.1, which leverages the subset of randomized inspections to estimate the average effects of SST inspections.

In Panel A of Figure K.1, we replicate Figure 3c, but use a broader sample of establishments eligible for an SST inspection whose lagged DART rates were just above or below the primary list cutoff that year, including not only establishments meeting these criteria that were *not* inspected (as in Figure 3c) but also those that *were* inspected. With this broader sample, Panel A of Figure K.1 depicts the sum of the direct and threat effects of inspections. Because Figure 3c already showed no threat effect at the cutoff, any change in injuries at the cutoff in Figure K.1 represents the direct effect of being inspected. Panel A of this figure reveals that establishments just barely above the primary list cutoff and thus placed on the primary list subsequently exhibited a discontinuous downward shift in serious injuries the following year. Panel B illustrates a similar discontinuity at the secondary cutoff.

We quantify this graphical effect by applying the same regression discontinuity design discussed in Appendix J to samples akin to those used in Appendix J, except here we also include establishments that were actually assigned to inspection. We report results in Table K.1. Given this sample, $\hat{\gamma}$ represents the combined effects on injuries of a heightened threat of inspection and of an actual inspection, but because the results in Appendix J reveal essentially no threat effect, $\hat{\gamma}$ effectively represents an intention-to-treat estimate of the effect of being inspected. Corroborating the visual evidence in Figure K.1, the point estimate in Column 1 indicates that establishments just above the primary cutoff subsequently experience 7 percent ($p=0.08$) fewer serious injuries than those just below.¹⁷ Column 2 indicates a similar reduction in

¹⁷ This evidence corroborates Li and Singleton (2019), who estimate the effects of SST inspections using a similar research design (using the primary list cutoff only) and find similar magnitudes.

injuries for establishments just above the secondary cutoff ($\hat{\beta} = -0.08, p = 0.09$). In Column 3, we combine the two samples from Columns 1–2 to generate a single intention-to-treat estimate of the effect of receiving an SST inspection, based on crossing either the primary or secondary cutoff. The point estimate is quite similar to that from both Columns 1 and 2 and, consistent with the larger sample size, more precise ($\hat{\beta} = -0.07, p = 0.02$).

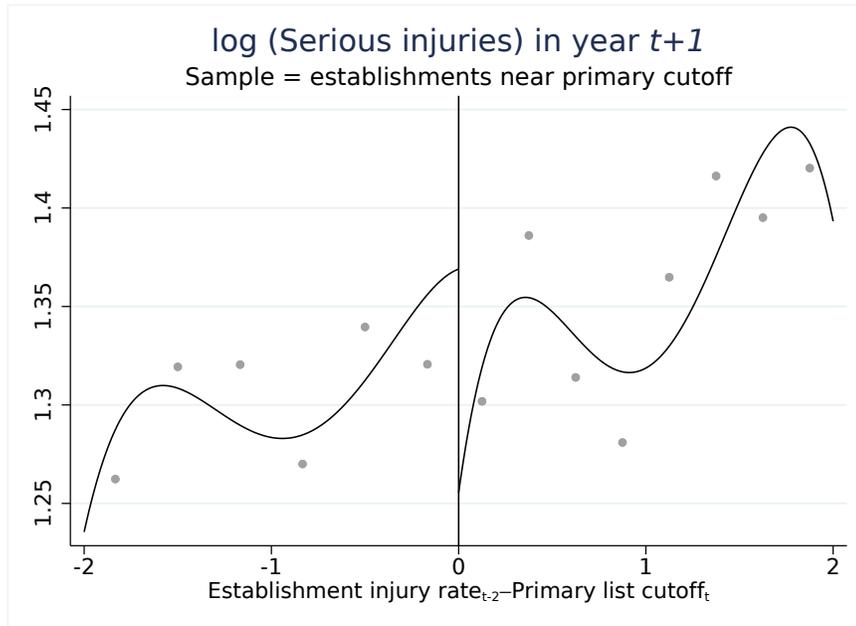
These results bolster the evidence in the main text that realized inspections improve safety.¹⁸

References

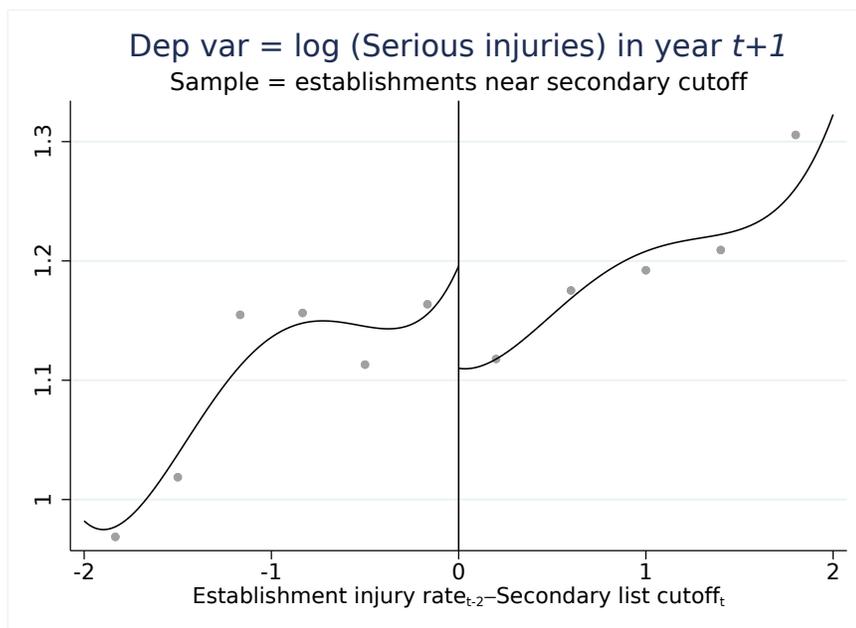
- Cattaneo, Matias D., Michael Jansson, and Xinwei Ma.** 2018. “Manipulation Testing Based on Density Discontinuity.” *Stata Journal* 18 (1): 234–261.
- Li, Ling, and Perry Singleton.** 2019. “The Effect of Workplace Inspections on Worker Safety.” *ILR Review* 72 (3): 718–748.

¹⁸ It is instructive to compare the magnitude of the average treatment effect of SST inspections implied by our main estimates in Section 3.1 to the effect size implied in Table J.1. Specifically, the estimate reported in Column 1 in Table J.1 implies an “intent to treat” effect of SST inspection on serious injuries of 6.7 percent ($\hat{\beta} = -0.07, \exp(-0.07) = -0.067$). Scaled by the “first-stage” coefficient in Column 1 (the extent to which being just above the primary cutoff raised the risk of inspection) implies a local average treatment effect of SST inspections of $-0.067/0.31 = 21.7$ percent. This estimate is twice as large as our main estimate of the average treatment effect of SST inspections in Section 3.1, which is 9 percent. The implied magnitude is similar for the secondary cutoff. There are two reasons that these estimates might differ. First, our main approach estimates an average treatment effect, whereas the result in Table K.1 is a local average treatment effect pertaining only to the establishments with injury rates near the cutoff; if treatment effects are heterogeneous, these two effects could differ. Second, the results in Table K.1 only consider injuries one year after the directive year, whereas our main estimates consider the effects on injuries up to four years after the directive year.

Figure K1: Further Evidence That SST Inspections Led to Fewer Injuries: Intention-to-treat Estimates from a Fuzzy Regression Discontinuity Design



(a)



(b)

Each figure displays a binned scatterplot, in which the x-variable is the “running variable” that determines whether an establishment’s injury rate from year $t - 2$ is high enough to make it eligible for the SST primary list (top panel) or secondary list (bottom panel) in year t . For each figure, the y-variable is provided in the title. The sample includes those establishments that were considered for the 2001–2010 SST target lists (except nursing homes for all years prior to 2009) and were eligible for an SST inspection.

Table K1: Further Evidence that SST Inspections Led to Fewer Injuries: Intention-to-treat Estimates from a Fuzzy Regression Discontinuity Design

Sample=	(1) Establishments in boundary of area office that began... primary list	(2) secondary list	(3) either primary or secondary
Dep var =	log injuries, $t + 1$		
Injury rate \geq cutoff	-0.07 (0.04)+	-0.08 (0.04)+	-0.07 (0.03)*
Robust p-value	0.083	0.091	0.021
# observations	7,447	4,664	
Left bandwidth	1.02	1.57	1.17
Right bandwidth	1.74	1.48	1.56

The running variable in Column 1, defined in Equation J.1, determines whether an establishment's injury rates from two years prior render it eligible for the SST primary list. The running variable in Column 2 is defined analogously for eligibility for the secondary list. In each column, # observations refers to the number of establishments with injury rates in year $t - 2$ within the left and right bandwidths around the cutoff (also reported in each panel). The MSE-optimal bandwidths are selected, based on the method in Cattaneo, Jansson, and Ma (2018), separately for each regression. All regressions include controls for region fixed effects and year fixed effects, a manufacturing dummy, and the log of the establishment's working hours from $t - 2$.