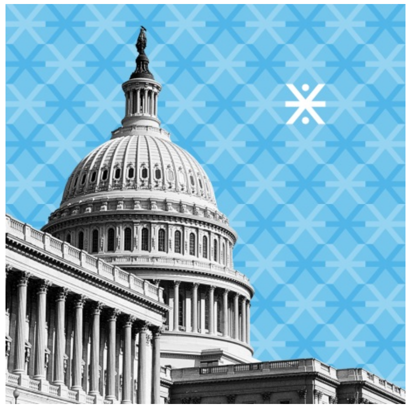
summit

complexity simplified.



Evaluation of the Occupational Safety and Health Administration's Site-Specific Targeting Program Final Report

Summit Consulting

Submitted to
Chief Evaluation Office
U.S. Department of Labor
Project Officer: Molly Irwin
Contract Number: DOLU139634581

CONTACT:

Ed Dieterle, Ed.D.

Director of Program Evaluation

202.407.8300 Tel

ed.dieterle@summitllc.us

AUTHORS:

Balint Peto

Laura Hoesly

George Cave, Ph.D.

David Kretch

Ed Dieterle, Ed.D.

June 2016

Table of Contents

List of Tables

List of Figures

Acknowledgements

Summit would like to acknowledge technical working group members Dr. Wayne Gray (Clark University), Dr. Randall Juras (Abt Associates), Jacob Klerman (Abt Associates), and Dr. John Mendeloff (University of Pittsburgh) for their helpful feedback on this report. Summit would also like to acknowledge Marios Michaelides and colleagues from IMPAQ International for their initial design and data analysis work.

Summit would like to thank staff members at Department of Labor's Occupational Safety and Health Administration (OSHA) and Chief Evaluation Office (CEO) for their continuous and tireless contribution and guidance throughout the project.

EXECUTIVE SUMMARY

Assuring that all workers in the United States have safe and healthful working conditions is the mission of the U.S. Department of Labor's (DOL) Occupational Safety and Health Administration (OSHA).[1] The Site-Specific Targeting (SST) program, a planned inspection program managed by OSHA, aims to improve health and safety of workplaces under OHSA's jurisdiction by targeting enforcement actions on establishments with historically high injury and illness rates.[2] The enforcement actions of the SST program include: (1) high-rate letters, which warn workplaces about their high injury-illness rates, and (2) inspections of worksites for compliance with safety and health regulations. In 2013, OSHA targeted 9,400 worksites nationwide using these enforcement actions.[3]

To determine whether the SST program has an impact on improving regulatory compliance and workplace health and safety, the DOL's Chief Evaluation Office (CEO) and OSHA contracted Summit Consulting LLC (Summit) to evaluate the program. This evaluation, started by IMPAQ International in 2010 and taken over by Summit in late 2013, assesses the impacts of SST on two main outcomes of interest:

- **Regulatory Compliance**, measured by the probability of OSHA citing a worksite for a violation during a follow-up inspection; and
- **Health and Safety**, measured by the follow-up injury/illness rate [i.e., the DART (days away, restricted, transferred) rate].

The evaluation assesses the direct impact of receiving letters or inspections. It also assesses the indirect impact of *being assigned* by OSHA to these enforcement actions, which allows measuring potential deterrent effects that fear of inspections may have on regulatory compliance and health and safety.

We apply both a randomized controlled trial (RCT) design and a regression discontinuity design (RDD) to assess impacts:

- **RCT:** An experimental design in which worksites were randomly assigned to one of two treatments (i.e., worksites receiving a high-rate letter or receiving both a high-rate letter and an SST inspection) or to a control group. This design allows us to attribute all significant observed differences in outcomes between the control group and either of the treatment groups to the respective treatment. 2,520 worksites were included in this experimental study. The treatments occurred in 2011 and outcomes were observed in 2012-2015.
- **RDD:** A quasi-experimental design, which capitalizes on OSHA's assignment of worksites to various categories based on injury-illness rate cutoffs. With this design, we can attribute observed differences in outcomes between groups on either side of the cutoffs to whether OSHA listed the worksites on treatment lists or to whether they received a specific treatment. The RDD relies mainly on the OSHA Data Initiative (ODI) survey of 7,045 worksites.

The original research design and analysis plan included only the RCT. Summit supplemented the experimental study with a quasi-experimental design to offset implementation and data limitations of the RCT and to capitalize on data OSHA had already collected.

[1] OSHA's jurisdiction covers almost all U.S. employees. Exceptions include miners, the self-employed, some transportation workers, and many public employees.
[2] SST applies to all non-Federal establishments in non-construction industries that are not in State Plan States.
[3] https://www.osha.gov/as/opa/foia/highrate_2013.html

Overall, neither the RCT nor the RDD study found statistically significant impacts. However, based on the results of this study, we can rule out impacts of very large magnitudes. Specifically, across the RDD and the RCT studies, we can say high-rate letters did not decrease worksites DART rates by more than 1.5 case-rates (or 22% of the control group mean), and that inspections did not decrease DART rates by more than 2.9 case-rates (or 40%) in the three years following the treatment. Similarly, we can rule out that high-rate letters, or inspections, decreased the probability of a post-treatment violation by more than 5.7% and 17.1% points, respectively.

We did find promising results in a few of the treatment-outcome combinations. Namely, we estimated that inspections decreased post-treatment DART-rates for Primary worksites across all designs and subgroups, although none of these estimates were statistically significant. In other treatment-outcome combinations and subsets of the sample, we found the direction of the impacts to be sometimes negative as expected, and at other times positive, which was unexpected, but again, none were statistically significant.

Previous literature suggests that letters can have a 1.6-1.7% and inspections a 3.3-3.4% impact on improving workplace injuries and illnesses, expressed as percentages of sample means. [4] It also suggests that inspections can have a 6.3% impact on improving regulatory compliance. In this study, we could neither confirm nor refute findings from the literature.

The lack of statistically significant impacts from our RCT and RDD studies does not mean that OSHA's enforcement actions had no beneficial impact on workplace health and safety. The results of this study only imply that if the enforcement actions had any impact on worksites' regulatory compliance and health and safety, that impact is unlikely to be larger than what the estimated confidence intervals indicate.

Even if this study can only rule out the existence of some very large impacts, the knowledge and techniques developed through our data preparation may help enhance OSHA's research infrastructure and could be used to facilitate and support future evaluation and analytic work with OSHA data. [5]

[4] Previous literature includes: Inspections' impact on workplace safety/health: (Levine, Toffel, and Johnson 2012), (Gray and Mendeloff 2002), and (Ruser and Smith 1991). Inspections' impact on regulatory compliance is (Weil 2001). Letters' impact on workplace safety/health is (Eastern Research Group 2004).

[5] See above for exact numbers.

I. BACKGROUND AND CONTEXT

Summit Consulting, LLC (Summit), under contract to the Chief Evaluation Office (CEO) in the Department of Labor (DOL), has completed an evaluation of the Occupational Safety and Health Administration's (OSHA) Site-Specific Targeting (SST) Program. IMPAQ International originally designed the impact evaluation in 2010. Summit received the materials from IMPAQ in late 2013-early 2014.

The goal of the evaluation is to assess the effectiveness of two enforcement actions in improving workplace health and safety outcomes. SST's two enforcement actions are:

1. High-rate letters, which inform employers that they have especially high rates of injury or illness for their industry, and suggest ways to improve the safety of their worksites.[6]
2. Programmed inspections, in which OSHA inspectors examine worksites for compliance with OSHA standards, and which can potentially result in citations, penalties, and hazard abatement orders.

1. Occupational Safety and Health Administration

OSHA's mission is to promote and to assure workplace safety and health. It is also charged with reducing workplace fatalities, injuries and illnesses for working men and women. Toward fulfillment of this mission, OSHA sets and enforces standards, conducts inspections, and provides training, outreach, education, and assistance.[7]

Specifically, OSHA conducts two classes of inspections of worksites to enforce compliance with OSHA requirements: unprogrammed and programmed inspections. In FY2014, OSHA conducted 36,163 total inspections: 19,222 programmed inspections and 16,941 unprogrammed inspections.

OSHA's unprogrammed inspections are in response to complaints, referrals, and incidents, while programmed inspections are aimed at specific industries or worksites that have experienced high rates of injuries or illnesses.[8] OSHA provides outreach and education to employers and workers about how to reduce on-the-job hazards.[9]

[6] A copy of the letter is presented in Appendix A
[7] See: https://www.osha.gov/about.html
[8] https://www.osha.gov/dep/2014_enforcement_summary.html
[9] For example, OSHA's free and confidential On-site Consultation program for firms with fewer than 250 workers, with information available at https://www.osha.gov/dcsp/smallbusiness/consult.html.

2. The SST Program

Until 2014, SST was one of OSHA's programmed inspection programs. SST focuses enforcement actions on the worksites with the highest rates of injury or illness. OSHA sets thresholds to determine which worksites will be subject to SST. The thresholds are set using industry-specific DART ("days away, restricted, or transferred") case rates and DAFWII ("days away from work injury and illness") case rates.[10]

SST uses inspections and high-rate letters to enforce OSHA standards and lower workplace injuries and illnesses. SST inspections are comprehensive examinations of all potentially hazardous areas of a worksite. Inspections can take as little as three hours or up to four weeks, depending on the size of the worksite and the complexity of the inspection. Because of the time and expense of inspections, OSHA is limited in the number of inspections it can conduct. To extend the reach of the SST program beyond inspections, OSHA uses high-rate letters. These letters are much less expensive and time-consuming than inspections. Letters inform employers that they have high rates of injury or illness and suggest ways to improve the safety of their worksites.[11]

Based on these thresholds, OSHA classifies worksites into four groups as follows:

1. Worksites with injury/illness rates at or above the first tier of injury/illness rate thresholds are selected to be part of the 'Primary List' of worksites to be subject to SST inspection.
2. Worksites with rates above the second tier of thresholds but below the Primary List thresholds are selected to be part of the 'Secondary List' and are inspected after all of the Primary List worksites are inspected, given available resources.
3. Worksites above the last, lowest threshold, including all worksites in the Primary and Secondary Lists, are sent a high-rate letter.
4. Worksites below the lowest do not receive a high-rate letter and are not on the SST inspection list but could still potentially receive an inspection for another reason.

We illustrate this in Table 1.

Table 1. OSHA enforcement groups and actions

Group			Enforcement Action		
Number	Name	Definition	Receives Letter?	On SST Inspection List?	Probability of Inspection
#1	Primary List	Above "Primary" threshold	Yes	Yes	High
#2	Secondary List	Above "Secondary" but below "Primary" threshold	Yes	Yes	Medium-high
#3	Letter List	Above "letter" but below "Secondary" threshold	Yes	No	Low, not zero
#4	No letter group	Below "letter" threshold	No	No	Low, not zero

[10] DART rates represent the number of cases with days away from work, or job transfer, or restriction due to injury and illness. DAFWII rates represent the number of cases with days away from work due to injury and illness. The rates are calculated as the equivalence of incidents per 100 employees working 40 hours per week, 50 weeks per year.
[11] An example of a high-rate letter can be found in Appendix A.

3. Evaluation overview

This evaluation uses a randomized control trial (RCT) design and a regression discontinuity design (RDD).

The RCT design assesses the impacts of two treatments:

- high-rate letter only (letter) and
- high-rate letter and inspection (letter plus inspection).

The RCT study assesses the impact on regulatory compliance outcomes (whether or not the worksite is cited for a violation) and health and safety outcomes (injury/illness rate measured by the DART case rate).

The regression discontinuity design (RDD) study capitalizes on OSHA's assignment of worksites to enforcement groups based on injury and illness thresholds. It separately assesses the effect of:

- being on the high-rate letter list,
- being on the primary list, and
- being on the secondary list.

The RDD study also assesses the effect of actually receiving a programmed inspection (this is an estimate of the impact of receiving an inspection for worksites that know they are at risk of receiving one). The RDD study assesses the impact on health and safety outcomes (injury/illness rate measured by the DART case rate).

4. Evaluation history

Throughout the history of this evaluation, changes occurred in the data environment, the design, and the implementation. The study was re-designed in 2012 to account for the cancellation of the ODI survey which would have provided the outcome data for health and safety outcomes. As a result of this unforeseeable change, outcome measures had to come from the Form 300A data. Form 300A data are recorded at the worksite and collected during inspections. However, some worksites fail to keep records or inspectors fail to collect the data during the inspection.

Table 2 summarizes the major steps in the history of the SST program evaluation.

Table 2. Evaluation history

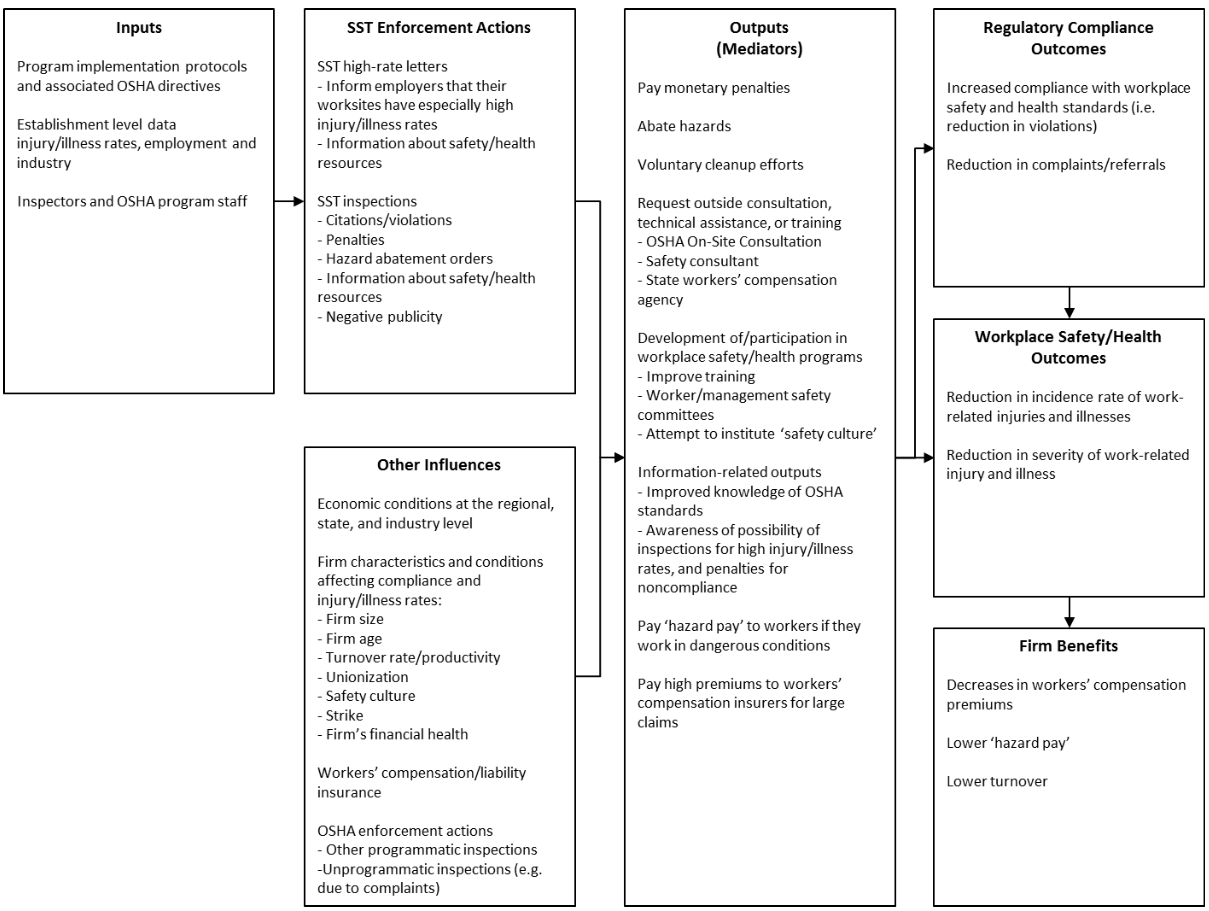
Year	Event	Comment
2010	IMPAQ designs impact evaluation. Experimental design: Two treatment groups and one control group (per SST list) Treatment to be administered in 2011. All follow-up inspections originally planned for 2013 to collect regulatory compliance outcomes Planned to use 2013 ODI survey for health/safety outcomes	
2011-2012	OSHA selects evaluation sites, administers treatments.	SST year 2011 effective: 9/9/2011 – 1/4/2013
2012	IMPAQ submits evaluation design report. ODI survey cancelled. Follow-up inspections which collect Form 300A data for health and safety outcomes replace ODI. OSHA moves follow-up inspections for letter + inspection treatment sites to 2014 due to rule prohibiting multiple programmed inspections within a two-year period. OSHA splits control groups into two; follow-up in 2013 *or* 2014.	Unforeseeable major change in design – outcome source affected. Foreseeable change in design – increased implementation risk.
2013	Summit begins work on SST project	
2015	Regression Discontinuity Design added to offset implementation and data limitations of RCT	

Due to the rule prohibiting multiple programmed inspections within a two-year period, the original timing of follow-up inspections was not possible. The follow-up inspections of worksites in the letter plus inspection treatment group had to be delayed by a year. This change caused additional administrative burden on OSHA. They needed to keep track of the multiple control and treatment groups and the associated timing of inspections in addition to their daily activities.

5. Theories of Change and Compliance with the SST Program

The conceptual framework for this evaluation is based in part on the Becker-Stigler regulatory compliance model.[12] The Becker-Stigler enforcement model posits that employers who operate worksites are profit-maximizers. They compare the expected cost of various levels of compliance (including no compliance) with health and safety regulations—including the direct benefits of improving health and safety and avoiding OSHA penalties— and choose the level of compliance that maximizes profit. [13] SST's targeted enforcement actions are posited to increase compliance and thus lower rates of injury and illness, or both by lowering employers' expected costs of compliance relative to their expected costs of noncompliance. To this model, we add elements of behavioral economics[14] and the theory of repeated games[15] to account for phenomena that do not fit neatly into the Becker-Stigler model.

Figure 1. SST program logic model



[12] (Becker and Stigler 1974)
[13] (Viscusi, Harrington, and Vernon 2005)
[14] (Camerer, Loewenstein, and Rabin 2011) & (Diamond, Vartiainen, and Yrjö Jahnssonin säätiö. 2007)
[15] (Myerson 1991)

Theories of Compliance

Noncompliance as well as compliance entails added costs for many employers. These costs include, but are not limited to:

- Penalties for violations,
- Costs of compensating injured/ill workers,
- Reduced productivity due to injured workers being replaced with new, less experienced workers,
- Reduced productivity due to high turnovers rates, and
- Demands for higher wagers due to unsafe conditions.[16]

The employer takes all of these factors into consideration in assessing its expected cost of current hazards and associated injury/illness rates. The expected cost of compliance, on the other hand, depends on the cost of abating hazards. If the expected cost of compliance is lower than that of noncompliance, an employer will become more compliant. The influence of letters and inspections on worksites is described below.

Letters

OSHA sends high-rate letters to selected worksites, and these letters have the following means of impact:

- Inform unaware worksites of their particularly high injury/illness rates,
- Inform about OSHA's free and confidential On-site Consultation Program and a list of other organizations that the employer can contact,
- Warn worksites of possible impending inspections,
- Note the benefits of and offer resources for improving workplace safety and health conditions,
- Warn about the costs of being noncompliant.

In addition to all of these direct impacts from OSHA, worksites may receive additional safety-related messages from safety consultants, equipment vendors, industry associations, or unions. These third-party groups acquire the list of high-rate letter recipients from OSHA's website.

Inspections

OSHA conducts inspections at a selected list of worksites. OSHA does the following during and after inspections:

- Imposes penalties for violations and fines for failing to abate hazards,
- Informs worksites of conditions that violate safety and health standards,
- Posts notices where hazards are found in the worksite,
- Releases the results of inspections online, and
- Publicizes enforcement actions and hazards through press releases.

We included a detailed description of the theories of compliance in Appendix B.

[16] (Viscusi 1993)

6. Research Questions

Each of the two designs used in this study has accompanying research questions to guide the analysis. The questions are as follows:

RCT

1. What are the impacts on regulatory compliance and health and safety outcomes of being on the SST high-rate-letter list compared to not being on either the SST high-rate letter list or the SST inspection list?
2. What are the impacts on regulatory compliance and health and safety outcomes of being on the SST high-rate-letter and SST inspection lists compared to not being on either list?

RDD

1. What impact does being on the high-rate letter list have on injury/illness rates?

2a. What impact does receiving a programmed inspection have on injury/illness rates?

2b. What impact does being on the SST inspection list have on injury/illness rates?

II. RANDOMIZED CONTROLLED TRIAL

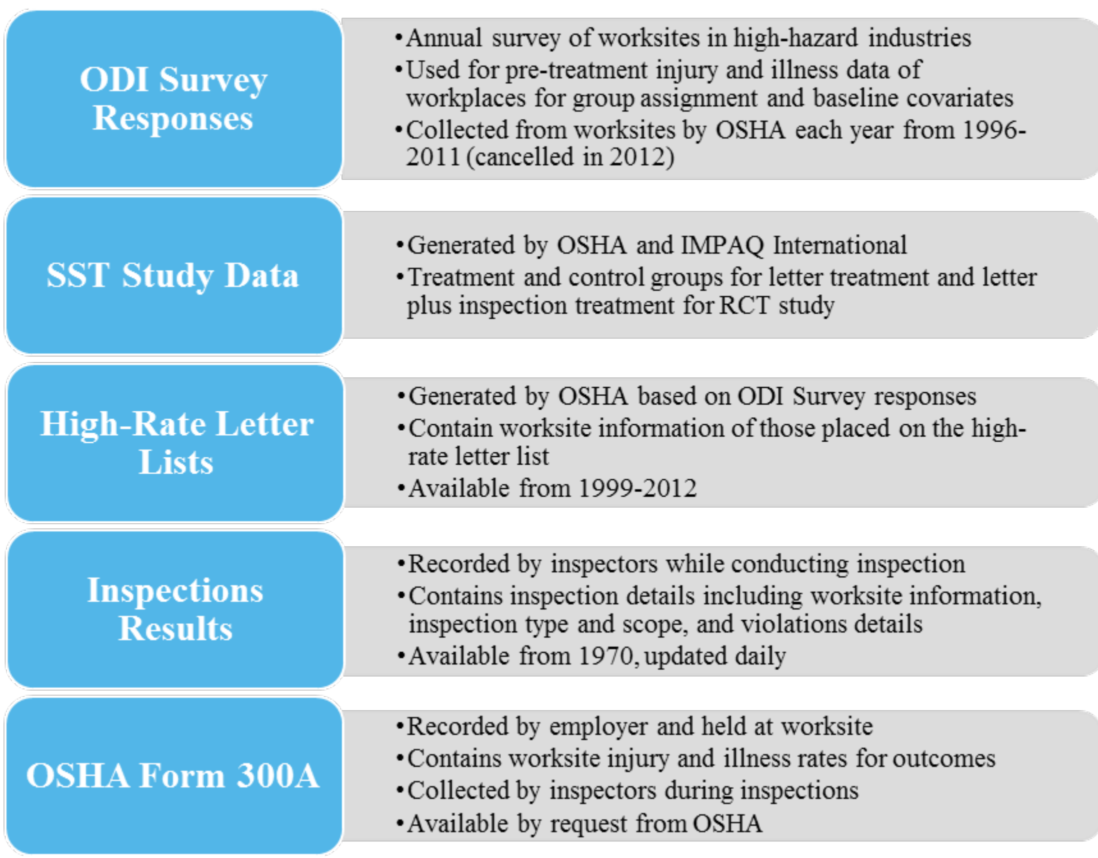
Based on the original study design, we applied the Randomized Controlled Trial (RCT) design to assess the impacts of enforcement actions on the outcomes.

1. Data

The data we used for the RCT analysis come from the following sources: OSHA Data Initiatives (ODI) survey responses, inspection results, high-rate letter lists, and OSHA Form 300A data. These data sources are summarized below in Figure 2.

The RCT used injury and illness rates from the 2010 ODI data to determine which worksites belong to the primary and secondary study groups. We used 2010 ODI data for values of some of the baseline covariates. The study group assignments were stored in an SST RCT Study database that was made specifically for the RCT study. We used the high-rate-letter lists to determine which worksites we placed on the OSHA's high-rate letter lists (to check treatment implementation). Further, the RCT uses inspections data from 2011 (the year of the inspection treatments), 2013, and 2014 (the years of the follow-up inspections). The injury/illness rate outcomes for the RCT come from Form 300A data collected during follow-up inspections. (Note that the original study design planned to use ODI data for injury/illness outcome measures but the ODI survey was cancelled for the year 2012 and beyond.)

Figure 2. RCT Data sources



Key variables

Table 3 lists the assignment and outcome variables used in the analysis along with the source for each of the variables.

Table 4 lists the covariates used in the RCT analysis.

Table 3. Assignment and outcome variables, RCT

Study	RCT-Letter	RCT-Letter + Insp	
	Variable	Variable	Source
Variable used to determine primary / secondary list assignment	DART 2009		2010 ODI Survey/SST Study Data
Regulatory Compliance Outcome	Violation After 2011	Violation After 2012	OSHA Inspections Data
Health/Safety Outcome	DART 2011	DART 2012	Form 300A Data

Table 4. Covariates, RCT

RCT-Letter, Letter Plus Inspection	
Covariate	**Source**
Industry	2010 ODI Survey
Aggregated Regions	2010 ODI Survey/SST Study Data
Number of Employees	2010 ODI Survey/SST Study Data
Pre-treatment inspection, penalty	OSHA Inspections Data
DART Rate 2009	2010 ODI Survey/SST Study Data
High-Rate Letters Received before 2011	High-Rate Letter Lists

The data used in the RCT are summarized in Appendix G.

Basic statistics of the RCT dataset

Table 5 presents the number of observations (treatment and control groups combined) broken down by some of the covariates in the RCT study. The RCT datasets contain about 1,200 sites for each treatment tested (“letter” and “letter plus inspection”).

Table 5. Basic descriptive statistics of the RCT study

Worksite Characteristic	Letter		Letter + Insp	
Industry	**N**	**%**	**N**	**%**
Agriculture, Forestry, and Fishing	14	1%	18	1%
Mining	1	<1%	2	<1%
Manufacturing	675	54%	729	58%
Transportation and Communication	249	20%	213	17%
Wholesale Trade	136	11%	135	11%
Retail Trade	75	6%	65	5%
Services	105	8%	103	8%
Region	**N**	**%**	**N**	**%**
Region 1	191	15%	152	12%
Region 2	150	12%	141	11%
Region 3	180	14%	184	15%
Region 4	178	14%	185	15%
Region 5	263	21%	278	22%
Region 6	161	13%	173	14%
Region 7	67	5%	72	6%
Region 8	52	4%	65	5%
Region 10	13	1%	15	1%
Number of Employees	**N**	**%**	**N**	**%**
0-99 Employees	833	66%	875	69%
100+ Employees	422	34%	390	31%
Inspection/Penalty Combinations	**N**	**%**	**N**	**%**
No inspection, no penalty	612	49%	610	48%
Inspection, but no penalty	152	12%	149	12%
Inspection and penalty	491	39%	506	40%
DART Rate	**N**	**%**	**N**	**%**
0.00-9.99	796	63%	835	66%
10-19.99	399	32%	375	30%
20+	60	5%	55	4%
High-Rate Letters Received before treatment	**N**	**%**	**N**	**%**
None	375	30%	396	31%
1-5 Letters	671	53%	694	55%
6+ Letters	209	17%	175	14%
Total	**1,255**	**100%**	**1,265**	**100%**

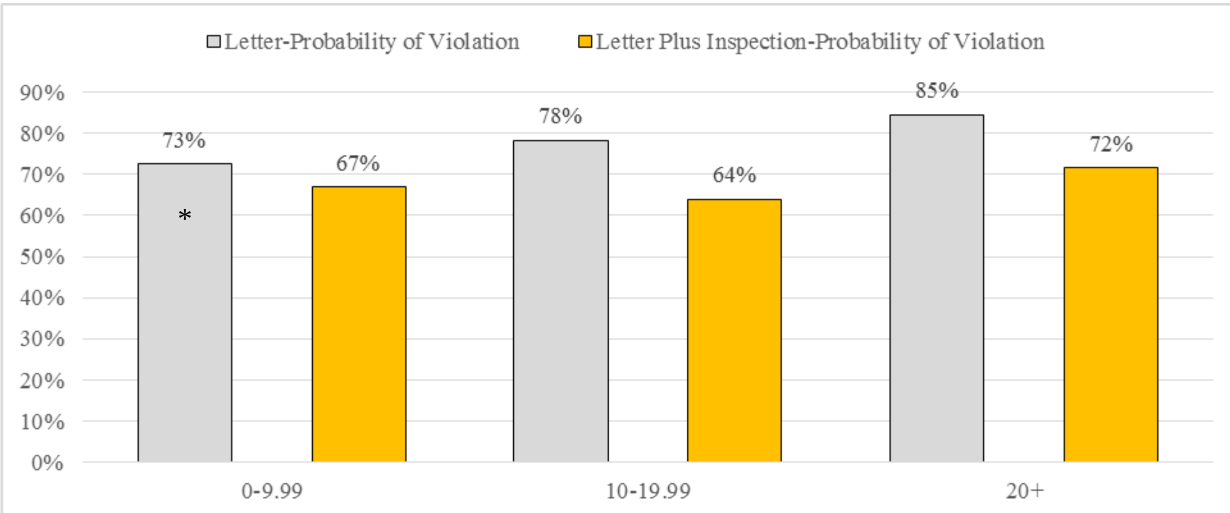
As Table 5 shows, the allocation of worksites across the industry, regional, and other groups is approximately even across study groups.

Injury illness rates and violation probability

As an additional analysis we examined the relationship between pre-treatment injury/illness rates and how likely a worksite is to be cited for a violation after controlling for observable covariates. Some aspects of injury/illness rates we assume to be time-invariant, and therefore we presume that worksites with higher injury/illness rates prior to receiving an inspection would have a higher probability of receiving a violation and worksites with lower rates would have a lower probability.

The results shown in Figure 3 suggest that overall, there is variation in probability of violation across DART rate groupings.

Figure 3. Worksite probability of violation, by pre-treatment DART rate, letter and letter plus inspection study, RCT



Note: DART groups marked with an asterisk have probabilities of violation that are statistically significantly different from the grand mean. The grand mean was 74.97% for letter sites and 66.32% for letter plus inspection sites.

What this analysis tells us is that for letter treatment sites, the lowest bracket of pre-treatment injury-illness rates is associated with a significantly lower post-treatment probability of violation. The figure also suggests that the probability of violation generally increases slightly as the pre-treatment injury-illness rate increases, but the measured probabilities are not significant at the conventional 5% level.

2. Study Design and Methodology

Design overview and implementation timeline

The RCT study includes 2,520 worksites. Half of the worksites were assigned to the letter treatment group, and the other half of the sites were assigned to the letter plus inspection treatment group. Each of those groups is further subset into two lists, a Primary List and a Secondary List. Primary and Secondary worksites are differentiated by injury/illness rates (see Table 1). Although the study was originally designed to assess impacts separately for Primary and Secondary worksites, we combined the two to increase power.

Given that the Primary and Secondary control and treatment groups were selected independently from non-overlapping populations (worksites on the Primary inspection list and worksites on the Secondary inspection list), combining the treatment and control groups resulted in treatment and control groups that allow for drawing statistically valid a meaningful conclusions about the combined population.[17] The key advantage of combining is that it increases the sample size thereby increasing statistical power and the ability of the design to detect an impact if there is one. Its disadvantage is that it masks subgroup results, eg. smaller impacts on Secondary sites water down potentially larger impacts on Primary sites. Therefore, we report the results of both the combined and the subgroup analyses.

For each treatment, the combined treatment group has 840 sites, and the control group has 415 and 425 sites (letter treatment and letter plus inspection treatment respectively). For each subgroup (Primary and Secondary lists), the treatment group has 420 sites, and the control group has about 210 sites. The study groups and sample sizes are summarized in Table 6.

Table 6. Summary of RCT impact evaluation study groups

List	Treatment/ Control	Letter	Letter Plus Inspection	Total
All worksites	**Treatment group**	**840**	**840**	**1,680**
	Control group	**415**	**425**	**840**
Primary sites	**Treatment group**	420	420	**840**
	Control group	204	216	**420**
Secondary sites	**Treatment group**	420	420	**840**
	Control Group	211	209	**420**
	Total	**1,255**	**1,265**	**2,520**

The timing and enforcement actions are shown in Table 7. OSHA sent the high-rate letters in calendar year 2011. OSHA conducted inspections in the 2011 SST year (between 9/9/2011 and 1/4/2013). Although control group worksites were not supposed to get treatment at the baseline, their names appeared on the DOL OSHA website as high-rate letter recipients.

Unfortunately, we have no information on which worksites actually received high-rate letters, and therefore we do not know the proportion of control vs. treatment worksites treated with letters, we only

[17] When combined, Primary and Secondary lists act like strata in a stratified simple random sampling design in that the two lists have different proportional contributions to the overall estimate. Therefore, we applied formulas for estimating means and variances from stratified samples as described in (Thompson, 2002, p. 120).

know which worksites were published on the list of high-rate letter recipients and which received an SST inspection as treatment. We indicated this in the table below.

Table 7. RCT implementation timeline

Study	Study Group	Treatment Letter Inspection	Follow-Up Inspection	N	Number and % of sites treated
Letter	**Treatment**	2011	2013	840	N/A**
	Control	2011*	2013	415	N/A**
Letter Plus Inspection	**Treatment**	2011	2014	840	616 (73.3%)
	Control	2011*	2014	425	29 (6.8%)

**Although, based on verbal information from OSHA, control group worksites did not receive high-rate letters, their names were published on the DOL OSHA website in 2011 as high-rate letter recipients.*
***We do not have information on whether or not worksites actually received the physical letters.*

Both treatment and control group worksites were envisioned to receive follow-up inspections to measure the outcomes of both groups for post-treatment comparisons. Due to the 2-year inspection restriction rule, the letter plus inspection treatment group's follow-up inspections were delayed mostly to 2014, while the letter treatment group's follow-up inspections mainly took place in 2013. For us to minimize sample attrition, we took into account all inspections in the 2013-2015 for the letter plus inspection group and 2012-2015 year-range for the letter group.

Randomization for the RCT

Random assignment for the RCT was a multi-step process. Worksites were taken from the Primary and Secondary Lists and further allocated into one of four main groups: letter treatment group, letter + inspection treatment group, and two control groups, resulting in a total of eight study groups.

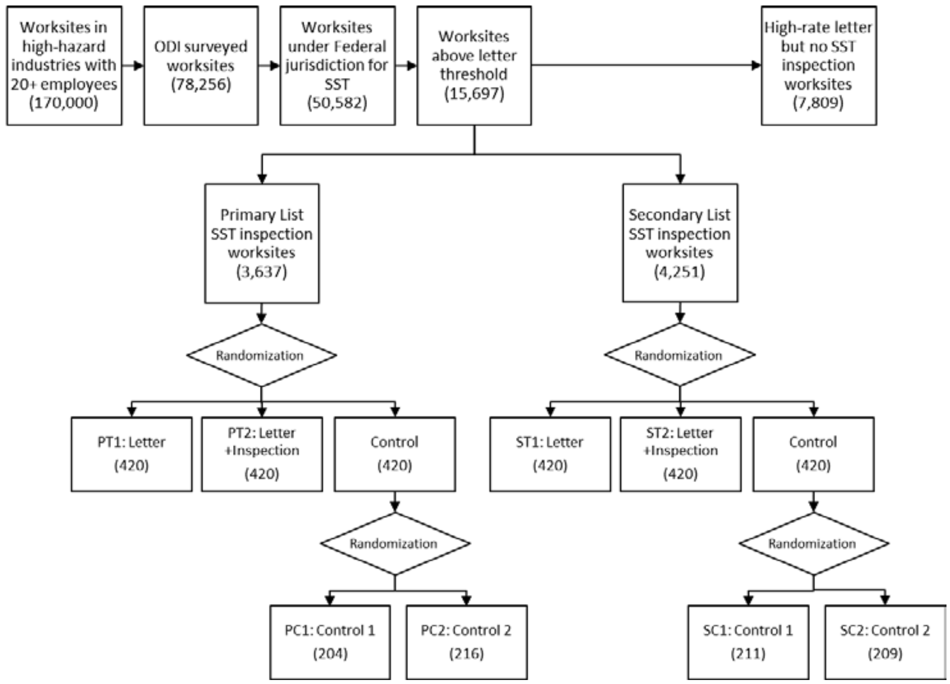
1. Study sites were selected using a systematic sampling procedure wherein sites are sorted by 2-digit SIC code, OSHA region, and worksite size, then every *k*th site is selected.[18]
2. Study sites were randomly assigned to letter treatment group, letter + inspection treatment group, or control group using the same systematic sampling procedure.
3. Control group sites were randomly assigned to one of two control groups using simple random sampling.

A diagram of the random assignment process, along with the assigned study group code names and the number of worksites in each group, is shown in Figure 4.[19]Naturally, the number of treatment and control group sites randomly selected for the study does not add up to the number of total worksites on the two lists in the population they were selected from. Furthermore, because not all worksites above the letter threshold are also above the Primary or Secondary thresholds, the total numbers of sites in the primary or

[18] (National Opinion Research Center 1996)
[19] Study sites were selected using a systematic sampling procedure wherein sites are sorted by 2-digit SIC code, OSHA region, and worksite size, then every *k*th site is selected. Sites were assigned to treatment or control group using the same systematic sampling procedure. Control groups were assigned to one of two control groups using simple random sampling. This process was done separately for Primary and Secondary List sites.

secondary lists 3,637 and 4,251 respectively) do not add up to the total number of worksites above the letter threshold (15,697).

Figure 4: RCT Impact Evaluation Random Assignment Process



Although the randomization process handled Primary and Secondary sites separately, we also conducted the analysis combining the two groups to be able to assess outcomes overall, using a larger number of observations with the appropriate stratum weights.

To check if the random assignment process was effective in ensuring measured impacts cannot be attributed to baseline differences in covariates, we compared pre-treatment values of all covariates between treatment and control groups. We found the randomization process to be effective. See Appendix F for details.

Model specification

We used weighted least squares (WLS) linear regression to estimate the program's impact on the rate of injuries and illnesses (as measured by the DART rate) and also to estimate the program's impact on sites' probability of violation.[20] We estimated impacts separately for Primary and Secondary List worksites as well as for the combined set. The weights come from the cell-weight adjustment for sample attrition as described below in "Sample attrition" section on page 23.

The equations below show the regression specifications for the RCT. Because regulatory compliance outcomes are binary, this report presents the coefficients of linear probability model regressions. The regression specification for regulatory compliance outcomes is thus:

$$\mathrm{Prob}(Z_i = 1 | T_{1i}, X_i) = \alpha_1 + \delta_1 T_{1i} + \sum_m \beta_m X_{mi} + \varepsilon_i \qquad (1)$$

[20] When the outcome of interest is binary, this model is known as a linear probability model (LPM).

The regression specification for continuous outcomes (DART rates in the case of health and safety outcomes) is:

$$Y_i = \alpha_1 + \delta_1 T_{1i} + w_i \sum_m \beta_m X_{mi} + \varepsilon_i \qquad (2)$$

where

- Z_i regulatory compliance outcomes: whether or not worksite i was cited for a violation after treatment (after 2011 in case of the letter treatment sites, after 2012 for the letter + inspections sites).
- Y_i health and safety outcomes: DART rate observed in worksite i after treatment (For the letter treatment sites, we used the earliest available DART rate for each worksite after 2011. The default is DART rate of 2012, but if that was missing, we used DART rate of 2013, and so on. Similarly for the letter + inspections treatment, we used the earliest available DART rate for each worksite after 2012.)
- T_{1i} indicator variable that takes on the value of 1 if worksite i has been randomized into the treatment group and 0 otherwise,
- w_i cell weight – adjusted for sample attrition for worksite i
- X_{mi} baseline characteristic m in worksite i (as described in Table 3)
- ε_i is a normally distributed error term,
- α_1 is the intercept – the estimated value of the outcome at the base levels of all covariates

$\hat{\delta}_1$, the coefficient estimates for T_1, give the estimated impact on the outcome of being assigned to the treatment group (letter or letter + inspection) relative to being assigned to the control group.

3. Results

This RCT study examined the effect of (1) being on the SST high-rate-letter list and (2) being on the SST high-rate-letter and SST inspection lists on two outcomes of interest: (A) a worksite's probability of being cited for an OSHA violation during their follow-up inspection and (B) post-program injury/illness rates. The counterfactuals (the bases for comparison) were control group worksites that were not supposed to have received either an SST high-rate letter or an SST inspection, but were still exposed to perceived threats of other programmed and unprogrammed enforcement actions.

Table 8, Table 9, and Table 10 present the results of the RCT study of all treatment-outcome combinations, for the Primary and the Secondary inspection list worksites.[21] The tables present robust standard errors to account for potential heteroscedasticity.

Interpretations of the key impact coefficients in Table 8, Table 9, and Table 10 below are as follows:

The treatment coefficients represent the mean difference between the treatment and control group worksites after being scheduled for a "letter" treatment or a "letter plus inspection" treatment. These estimates hold observable baseline covariates constant, therefore accounting for any initial differences that may have remained after randomizations between the two groups. See Appendix F for details of comparisons in baseline values of covariates between treatment and control group worksites.

[21] The results presented in the table are the results of weighted least squares (WLS) regressions run on the RCT dataset with covariates after having performed the cell-weight adjustment for nonresponse as described on page 32.

The coefficients allow us to observe whether worksites in the treatment group were more or less likely to be cited for a violation after treatment as compared to the control group. We can also observe whether, on average, treatment group sites have higher or lower injury/illness rates after treatment as compared to control group sites.

The theory of change (see discussion starting on page 7) suggests that worksites that were part of the treatment groups should have a lower probability of receiving a violation during the follow-up inspection and lower follow-up injury/illness rates. Therefore the expected sign of all coefficients is negative.

Table 8. Estimated impacts on key outcomes, all worksites, RCT[22]

All worksites	Injury/Illness Rate (Health and Safety)		Probability of Violation (Regulatory Compliance)	
Treatment	**Letter**	**Letter + Insp**	**Letter**	**Letter + Insp**
Difference in Means (Estimated Impact)	**-0.05** CI [-1.16; 1.05]	**-0.47** CI [-1.58; 0.63]	**1.5%** CI [-4.2%; 7.2%]	**0.5%** CI [-7.4%; 8.4%]
Standard Error	0.56	0.56	2.9%	4.0%
Control Group Mean	6.52	6.39	73.5%	65.1%
Treatment Group Mean	6.46	5.92	75.1%	65.6%
Observations	703	572	998	804
R-Sq.	0.09	0.21	0.05	0.05
F-Stat	6.26	10.74	4.41	3.78

* p<0.05

Table 9. Estimated impacts on key outcomes, Primary worksites, RCT

All worksites	Injury/Illness Rate (Health and Safety)		Probability of Violation (Regulatory Compliance)	
Treatment	**Letter**	**Letter + Insp**	**Letter**	**Letter + Insp**
Difference in Means (Estimated Impact)	**0.30** CI [-1.50; 2.10]	**-0.99** CI [-2.86; 0.88]	**2.5%** CI [-5.4%; 10.3%]	**-7.5%** CI [-17.1%; 2.2%]
Standard Error	0.92	0.95	4.0%	4.9%
Control Group Mean	6.75	7.01	76.4%	77.3%
Treatment Group Mean	7.05	6.02	78.8%	69.8%
Observations	345	275	498	409
R-Sq.	0.12	0.24	0.07	0.06
F-Stat	5.94	9.24	3.81	3.99

* p<0.05

[22] Note that the combined estimates are not necessarily in between the Primary and Secondary estimates due to differences in the non-response cells and the resulting non-response weights identified by the reciprocal cell response weighting method. (See discussion on Page 25.)

Table 10. Estimated impacts on key outcomes, Secondary worksites, RCT

All worksites	Injury/Illness Rate (Health and Safety)		Probability of Violation (Regulatory Compliance)	
Treatment	Letter	Letter + Insp	Letter	Letter + Insp
Difference in Means (Estimated Impact)	**-0.18** CI [-1.32; 0.96]	**-0.48** CI [-2.36; 1.39]	**2.6%** CI [-5.7%; 10.9%]	**5.4%** CI [-6.7%; 17.5%]
Standard Error	0.58	0.95	4.2%	6.2%
Control Group Mean	6.17	6.34	69.5%	56.3%
Treatment Group Mean	5.98	5.85	72.2%	61.7%
Observations	358	297	500	395
R-Sq.	0.074	0.273	0.053	0.045
F-Stat	2.291	4.784	2.907	1.73

* $p<0.05$

Letter treatment findings

We first discuss the findings of letter treatment analysis in the RCT design for probability of receiving a violation and injury/illness rate outcomes for each of the study groups.

Research Questions and Answers

Questions:

> *What are the impacts on regulatory compliance and health and safety outcomes of being on the SST high-rate-letter list compared to not being on either the SST high-rate letter list or the SST inspection list?*
>
> *What are the impacts on regulatory compliance and health and safety outcomes of being on the SST high-rate-letter and SST inspection lists compared to not being on either list?*

Answers:

Health and Safety:

The estimated impacts, that is the differences between treatment and control worksites in the post-treatment period injury/illness rates are -0.05 (or -0.8% of the control group mean) for the combined set of all worksites, and are 0.30 (4.4%) and -0.18 (-2.9%) for Primary and Secondary sites, respectively[22]. The results for the combined set of sites and for Secondary sites were in the expected direction, but neither of the estimated impacts on health and safety outcomes were statistically significant at the 5% level.

The confidence intervals for each of the estimates are also shown in the tables. These intervals show the likely range of the impact. Therefore, the impact was unlikely to be extremely large, in the positive or negative direction. We can conclude that receiving a high-rate letter is unlikely to have reduced worksites' injury and illness rates by more than 1.16 (DART case rate measure, -17.8% of the control group mean), and it is unlikely to have increased it by more than 1.05 (16.1%).

Based on our results, we can make similar conclusions about Primary and Secondary list worksites, the confidence intervals are [-1.50; 2.10] and [-1.32; 0.96], respectively, indicating impacts in the ranges of [-22.2%, 31.1%] and [-21.4%, 15.6%] as expressed in the percentage of control group means.

Regulatory Compliance:

As for regulatory compliance outcomes, we can see in Table 8 that the estimated impact (or the difference between the treatment and control means) among all sites for the "letter" treatment is a 1.5 percentage-point increase in the probability of a violation after the program period (measured in probability percentages in years 2013 to 2015). Treatment sites showed a 75.1% average probability, compared to 73.5% for control sites[23]. The sign of the impact (an increase) is in the unexpected direction, but it is not statistically significant at the 5% level.

The estimated impact of the letter treatment among Primary sites only is 2.5 percentage points – treatment sites have a 78.8% probability of violation after the program period, compared to 76.4% for control sites. The difference is not statistically significant, and is positive (not in the expected direction). See Table 9.

Among Secondary sites (Table 10), we estimated the treatment group to be 2.6 percentage points more likely to be cited for a violation, on average. This result was also not statistically significant and it is also positive (not in the expected direction).We can conclude that high-rate letters are unlikely to decrease the probability of a violation by more than 4.2 percentage points, or increase it by more than 7.1 percentage points.

We compare our findings to results in previous literature in the discussion starting on page 29.

Letter plus inspection treatment findings

In this section, we discuss the findings of letter plus inspection treatment in the RCT design for regulatory compliance and health and safety outcomes.

Research Questions and answers

Questions:

What are the impacts on regulatory compliance and health and safety outcomes of being on the SST high-rate-letter list compared to not being on either the SST high-rate letter list or the SST inspection list?

What are the impacts on regulatory compliance and health and safety outcomes of being on the SST high-rate-letter and SST inspection lists compared to not being on either list?

Answers:

Health and Safety:

The differences in the post-treatment period injury/illness rates (measured in DART case rates after the treatment period in years 2013 to 2015) are -0.47 (-7.4% of the control group mean), -0.99 (-14.1%) and -0.48 (-7.6%) for all worksites, Primary sites, and Secondary sites, respectively[22]. The results in each case are in the right direction, and they can be considered relatively large for Primary list worksites (almost 1 DART case rate or 14.1% of the control group mean), but none of the estimated impacts on health and safety outcomes were statistically significant at the 5% level.

[23] The impact is not precisely equal to the difference between treatment and control due to rounding.

As with the letter treatment, if we look at the lower and upper bounds of the confidence intervals, we can conclude that if receiving a high-rate letter and an inspection had an impact on worksites' health and safety, that impact is unlikely to be extremely large. We can conclude that receiving a high-rate letter and an inspection is unlikely to have reduced worksites' injury and illness rates by more than 1.58 (DART case rate measure, -24.7% of the control group mean), and it is unlikely to have increased it by more than 0.63 (9.9%).

Based on our results, we can make similar conclusions about Primary and Secondary list worksites, the confidence intervals are [-1.50; 2.10] and [-1.32; 0.96], respectively, indicating impacts in the ranges of [-40.8%, 12.6%] and [-37.2%, 21.9%] as expressed in the percentage of control group means..

Regulatory Compliance:

We can see in Table 8 that among all sites the "letter plus inspection" treatment has an estimated 0.5 percentage-point impact on the probability of post-treatment violation (in post-treatment probability of a violation). This estimated impact on regulatory compliance is in the unexpected direction, and is not statistically significant at the 5% level.

As shown in Table 9, the "letter plus inspection" treatment had an estimated -7.5 percentage-point impact on post-treatment violation probability among Primary sites. While not statistically significant, the difference is relatively large and negative (in the expected direction). Among Secondary sites (Table 10), the treatment group was, on average, 5.4 percentage points more likely to be cited for a violation. This result was in the unexpected direction but also not statistically significant.

Based on the lower and upper bounds of the confidence intervals, we can conclude that if a high-rate letter and an inspection had an impact on worksites' regulatory compliance, that impact is unlikely to be extremely large. We can conclude that receiving a high-rate letter and an inspection is unlikely to decrease the probability of a violation by more than 7.4 percentage points, or increase it by more than 8.4 percentage points. We can make similar conclusions about Primary and Secondary list worksites, the confidence intervals are [-17.1%, 2.2%] and [-6.7%; 17.5%], respectively.

Exploratory Analyses - Impacts on Additional Outcomes

In the planning phases of our analyses, regulatory compliance and health and safety were selected as primary outcomes. We pre-specified primary outcomes to avoid the appearance of "cherry picking" results based on findings. The pre-selected key outcomes are the probability of violation after the treatment period (regulatory compliance) and DART rates after the treatment period (health and safety).

The analysis presented below includes additional outcomes of interest. However, since these were not pre-specified, these analyses should be considered exploratory. Due to the fact that the nominal level of significance (the probability of rejecting a true null hypothesis) of unplanned outcomes is too high, analysis of these additional outcomes provide a sensitivity test for our primary outcomes.

We find the results on these alternative outcomes are similar to the results on our key or primary outcome measures. We found no significant impacts for the combined set of worksites or the Secondary list worksites, and found some significant impacts of the letter plus inspections treatment for Primary list worksites. This is in line with the findings of our primary analysis, where we found consistently relatively large (but statistically not significant) impacts in the right direction for the letter plus inspections treatment on Primary worksites.

However we emphasize that these results should be considered exploratory and despite the fact that some of these findings show nominal statistical significance, they should not be used to make conclusions.

The tables below present the results of these analyses on additional outcomes. [24]

Table 11. Effect of treatment on additional outcomes, all worksites, RCT

All Sites	Letter			Letter + Insp		
	Estimates and CI	S.E.	Control/ Trt. Means	Estimates and CI	S.E.	Control/ Trt. Means
Number of Violations After Treatment	0.26 [-0.47,0.99]	0.37	C: 4.32 T: 4.57	-0.68 [-1.51,0.15]	0.42	C: 3.61 T: 2.92
Number of Other-than-Serious Violations After Treatment	0.11 [-0.10,0.32]	0.11	C: 1.12 T: 1.23	-0.16 [-0.42,0.1]	0.13	C: 1.01 T: 0.85
Other-than-Serious Violation After Treatment (0/1)	0.02 [-0.05,0.09]	0.03	C: 0.51 T: 0.53	0.00 [-0.08,0.08]	0.04	C: 0.41 T: 0.41
Number of Serious Violations After Treatment	0.15 [-0.48,0.77]	0.32	C: 3.19 T: 3.34	-0.52 [-1.25,0.20]	0.37	C: 2.60 T: 2.07
Serious Violation After Treatment (0/1)	0.02 [-0.04,0.08]	0.03	C: 0.62 T: 0.64	0.01 [-0.07,0.09]	0.04	C: 0.52 T: 0.53
Penalty Amount After Treatment	$132.48 [-2050.64,2315.60]	$1,112.49	C: 7086.61 T: 7219.08	-$2,957.96 [-6231.94,316.03]	$1,667.88	C: 8147.45 T: 5189.50
Penalty Received After Treatment (0/1)	0.03 [-0.03,0.09]	0.03	C: 0.67 T: 0.70	0.01 [-0.08,0.09]	0.04	C: 0.58 T: 0.58
DAFWII Rate	0.42 [-0.15,0.99]	0.29	C: 2.95 T: 3.37	0.22 [-0.43,0.87]	0.33	C: 2.93 T: 3.15
Rate of Days Away from Work	32.47 * [5.11,59.82]	13.93	C: 83.94 T: 116.41	-7.48 * [-36.14,21.18]	14.59	C: 101.58 T: 94.09

* p<0.05

[24] Much like with our primary analysis, the combined estimates are not necessarily in between the Primary and Secondary estimates due to differences in the non-response cells and the resulting non-response weights identified by the reciprocal cell response weighting method.

4. Limitations and Power

This study had a number of limitations, which may help to explain why the results of the RCT impact estimations are inconclusive.

- **Small sample size/Sample attrition:** The study had a small sample size and a larger-than-expected sample attrition (see Table 12 and Table 13), which reduced the power of the study: it became less likely that the study would be able to detect an impact if there was one. The sample attrition issue was more pronounced for "letter plus inspection" worksites than for "letter" sites, and this shows in the power calculations. Comparing estimated impacts in previous research to the updated power calculations with the known response rates, the study was not likely to detect impacts in either of the treatment-outcome combinations[25]. (See Table 19).

  Additionally, the proportion of sites in the "letter plus inspection" treatment group that had outcomes was statistically significantly different from the control group, in other words, there was differential sample attrition between treatment and control sites. We adjusted for differential sample attrition using cell based inverse probability weighting, which should provide valid results, but the validity of this adjustment is not testable, and the strength of random assignment is weakened. Unobserved differences between treatment and control sites that actually responded could also be responsible for differences in measured outcomes.

- **Implementation Imperfections:** Two primary imperfections with the implementation of this study may have contributed to the inconclusive results.
  - **Letters:** Both treatment and control worksites were inadvertently published online at the DOL website as high-rate letter recipients. This removed a potential avenue for improving compliance, receiving third-party contacts about being on the list, since both treatment and control group sites were exposed to this avenue of the high-rate letter impact. (See "Theories of Change and Compliance with the SST Program" chapter.)
  - **Inspections:** Some worksites on the letter plus inspection treatment list did not receive an SST inspection in 2011 or 2012, and some control group sites that should not have, did. The analysis compares the outcomes of those assigned to treatment to those assigned to control and does not account for imperfect implementation. The treatment impact is therefore watered down.

Sample attrition

Sample attrition (unavailable outcome information at follow-up on some study participants) was higher than previously anticipated in each of the study groups. As we show below in the "Power calculations - RCT study" section on page 29, the RCT study was unlikely to detect impact even with the original sample sizes, and high sample attrition further reduced the effective sample size and hence the probability of the study detecting an impact.

For regulatory compliance outcomes, we define sample attrition as "no post-treatment inspections" (post-2011 for letter treatment and post-2012 for letter plus inspection treatment sites). For health and safety outcomes, we define sample attrition as no post-2011 DART rates (letter), and no post-2012 DART rates (letter plus inspection). We analyzed sample attrition carefully before conducting the regression analyses.

[25] There are 4 treatment-outcome combinations for each pair of two treatments ("letter" and "letter plus inspection") and two outcomes ("regulatory compliance" and "health and safety")

Table 12 and Table 13 present the distribution of non-response in the control and in the treatment groups for all worksites and separately for the primary and secondary sites, for all treatment-outcome combinations.

When planning the RCT analysis, we expected sample attrition to be low (10%). Excluding sites going out of business, all study sites were expected to have received a follow-up inspection. As Table 12 and Table 13 show however, the level of sample attrition turned out to be between 16% and 73%, which is much higher than anticipated.

Table 12 shows the number of study sites without follow-up inspections (sample attrition for regulatory compliance outcomes), while Table 13 shows the number of study sites without follow-up DART rates (sample attrition for health and safety outcomes). A p-value of less than 0.05 indicates that there is a significant difference in sample attrition between treatment and control sites. For example, there are highly significant differences between the number of sites without a follow-up inspection in the control group (55%) and the treatment group (27%) in the letter plus inspection study.

Table 12. Sample Attrition for Regulatory Compliance - Study Sites without Follow-Up Inspections

Treatment	SST List	Group	Total Number of Sites	Sites w/o Follow-Up Inspections		p-value of the Difference
				N	%	
Letter	**Total**	**Control**	**415**	**70**	**17%**	**0.03**
		Treatment	**840**	**187**	**22%**	
	Primary	Control	204	32	16%	0.05
		Treatment	420	94	22%	
	Secondary	Control	211	38	18%	0.23
		Treatment	420	93	22%	
Letter + Insp	**Total**	**Control**	**425**	**235**	**55%**	**0.00**
		Treatment	**840**	**226**	**27%**	
	Primary	Control	216	112	52%	0.00
		Treatment	420	115	27%	
	Secondary	Control	209	123	59%	0.00
		Treatment	420	111	26%	

Note: Outcomes are counted as missing if they are not present in 2012-2015 for the letter treatment study and in 2013-2015 for the letter plus inspection treatment study. A p-value of less than 0.05 indicates that there is a statistically significant difference between treatment and control sites in sample attrition rates.

Table 13. Sample Attrition for Health and Safety - Study Sites without Follow-Up DART Rates

Treatment	SST List	Group	Total Number of Sites	Sites w/o Follow-up DART Rates		p-value of the Difference
				N	%	
Letter	**Total**	**Control**	**415**	**169**	**41%**	**0.04**
		Treatment	**840**	**394**	**47%**	
	Primary	Control	204	80	39%	0.02
		Treatment	420	208	50%	
	Secondary	Control	211	89	42%	0.61
		Treatment	420	186	44%	
Letter + Insp	**Total**	**Control**	**425**	**300**	**71%**	**0.00**
		Treatment	**840**	**393**	**47%**	
	Primary	Control	216	149	69%	0.00
		Treatment	420	212	51%	
	Secondary	Control	209	151	72%	0.00
		Treatment	420	181	43%	

Note: Outcomes are counted as missing if they are not present in 2012-2015 for the letter treatment study and in 2013-2015 for the letter plus inspection treatment study. A p-value of less than 0.05 indicates that there is a statistically significant difference between treatment and control sites in non-response rates.

According to IMPAQ International's 2010 study design, health and safety outcomes were supposed to be acquired through the ODI survey. There was good reason to believe that the study worksites would respond to a follow-up survey given that they responded to the original survey. Further, the ODI was a high quality survey that was conducted annually for 15 years. However, OSHA canceled the ODI in 2012, and thus Form 300A data replaced it as the source of health and safety outcomes.

Substituting Form 300A data gathered at follow-up inspections led to significantly greater sample attrition than anticipated. OSHA operated the ODI independently of inspections. Since many scheduled follow-up inspections did not occur, these higher sample attrition risks were realized in the RCT outcome data.

The higher-than-anticipated sample attrition led to higher minimum detectable impacts (MDIs), which reducing the chance of the study capturing an impact.

We provide further details of the possible reasons for sample attrition for regulatory compliance outcomes in Appendix B.

Correcting for sample attrition

Given that on a number of occasions sample attrition is different between the treatment and the control group (see Table 12 and Table 13), the key advantage of randomly assigning worksites to treatment and control groups is at risk. Observed differences between treatment and control groups may be a result of the differences between how sample attrition affected those two groups and not only the difference in treatment anymore.

Because of differential sample attrition, and because of the generally high sample attrition, we used Little's method of weighting by reciprocal cell response rates to correct for sample attrition separately for

Primary sites, Secondary sites, and the combined set (Little 1986). This method weights observations that are similar to lost sample members to account for unobserved members.

For example, if a certain industry faces high sample attrition, observed worksites from this industry are weighted appropriately such that observed control group sites are treated as if the entire control group had been observed. This approach should provide unbiased estimates of the mean control group outcome and the impact of the treatment assuming that sample attrition is unrelated to unobservable characteristics. We dropped worksites without outcomes from the dataset for all subsequent calculations, and we weighted the remaining sites by the calculated weights.

Imperfect implementation

All treatment worksites and none of the control group worksites should have received an SST inspection as treatment (in 2011 and 2012)[26]. As shown in Table 14, only 73.3% of the treatment worksites received treatment (an SST inspection in 2011 or 2012). About 6-7% of the control group worksites also received treatment.

Table 14. RCT, Letter Plus Inspection Treatment, Receipt of SST Inspection during Treatment Period, by Group and List

Group	List	Did Not Receive SST Inspection in 2011 or 2012		Received SST Inspection in 2011 or 2012		Total
		N	%	N	%	
Control	**Total**	**396**	**93%**	**29**	**7%**	**425**
	Primary	202	94%	14	6%	216
	Secondary	194	93%	15	7%	209
Treatment	**Total**	**224**	**27%**	**616**	**73%**	**840**
	Primary	112	27%	308	73%	420
	Secondary	112	27%	308	73%	420

If we look at all inspections, rather than just SST inspections, we see that almost 20% of control sites and 80% of the treatment sites were inspected during the treatment period. If we assume that non-SST inspections generally have similar impact on regulatory compliance and health and safety as SST inspections (the actual treatment), we can say that the differences in treatment were about 66percentage points (83% - 17%) while they should ideally be 100 percentage points.

[26] Note that treatment inspections are different from follow-up inspections. The former is the treatment we are measuring the impact of and therefore should be different between treatment and control (in 2011 and 2012). On the other hand, follow-up inspections are the means for measuring outcomes, and thus treatment and control group worksites alike should all receive them after the treatment (in 2013-2015).

Table 15: RCT, Letter Plus Inspection Treatment, Receipt of ANY Inspection during Treatment Period, by Group and List

Group	List	Sites that did not receive any Inspection in 2011 or 2012		Sites that did receive any Inspection in 2011 or 2012		Total
		N	%	N	%	
Control	**Total**	**351**	**83%**	**74**	**17%**	**425**
	Primary	178	82%	38	18%	216
	Secondary	173	83%	36	17%	209
Treatment	**Total**	**179**	**21%**	**661**	**79%**	**840**
	Primary	91	22%	329	78%	420
	Secondary	88	21%	332	79%	420

Control group worksites published as high-rate letter recipients

During the 2011 implementation of study treatments, control group worksites' names were inadvertently listed as high-rate letter recipients on the DOL website.

Re-examining the theory of change helps us understand how this error reduced the chance for capturing the impact of high-rate letters. Not only does the actual receipt of a high-rate letter encourage worksites to improve compliance, a worksite could also be affected by contact from third parties.

Several sources post articles online when the names of high-rate letter recipients are published; some include advice or suggest sources for advice to letter recipients. These sources include safety journals and newsletters, such as the EHS Journal: Practical Solutions for Environmental, Health and Safety Professionals[27] and J.J. Keller's OSHA Safety Training Newsletter.[28] Safety consultants and equipment vendors can access this list as a source for leads for sales. These consultants and vendors may contact the worksites with marketing messages and sales calls, reiterating OSHA's message to the worksite to increase compliance.

As seen in Table 16 and Table 17 below, between 85% and 88% of the treatment group worksites were published online as high-rate letter recipients, in line with the study design. However, between 80% and 91% of the control group worksites in both the letter only and the letter plus inspection treatment groups were also published online as high-rate letter recipients.

[27] Stacey Lucas. "U.S. OSHA Targets 15,000 Facilities with High Incident Rates." *EHS Journal*. April 4 2010.
[28] Judie Smithers. "OSHA notifies employers with high injury rates." *J.J. Keller's OSHA Safety Training Newsletter*. Vol. 19/No. 6, June 2012.

Table 16: Study Sites Listed on the High-Rate Letter List Online, by Treatment and Group

Treatment	Group	Total Number of Sites	Sites that were on the high-rate letter list in 2011		Sites that were not on the high-rate letter list in 2011	
			N	%	N	%
Letter	Control	415	347	84%	68	16%
	Treatment	840	726	86%	114	14%
Letter + Insp	Control	425	371	87%	54	13%
	Treatment	840	722	86%	118	14%

Table 17: Study Sites Listed on the High-Rate Letter List Online, by Treatment, List and Group

Treatment	List	Group	Total Number of Sites	Sites that were on the high-rate letter list in 2011		Sites that were not on the high-rate letter list in 2011	
				N	%	N	%
Letter	Primary	Control	204	163	80%	41	20%
		Treatment	420	356	85%	64	15%
	Secondary	Control	211	184	87%	27	13%
		Treatment	420	370	88%	50	12%
Letter + Insp	Primary	Control	216	196	91%	20	9%
		Treatment	420	363	86%	57	14%
	Secondary	Control	209	175	84%	34	16%
		Treatment	420	359	85%	61	15%

Therefore the RCT study experienced reduced ability to isolate one of the key avenues of effect of the letter treatment, and therefore the probability of detecting a difference between treatment and control groups became radically smaller than if control group worksites' names had not been posted. The only difference between treatment and control groups for the letter treatment study was reduced to physically receiving the letter.

Follow-up inspections

When we examine inspections received by letter plus inspection treatment worksites in the RCT study, we find that the proportion of sites that received a follow-up inspection is different for the treatment group than the control group. As it turned out, this difference was also caused by an implementation error.

According to the RCT study design, follow-up inspections were supposed to provide data to measure the outcomes in both the treatment and the control groups so that the difference in outcomes can be assessed. During follow-up inspections, inspectors were to collect violations data for regulatory compliance outcomes and the Form 300A which reports injury and illness data for health and safety outcomes. The impacts of the treatment (differences in outcomes) cannot be calculated if we lack outcomes for the treatment and control groups.

As seen in Table 12, in the letter plus inspections treatment study, about 27% of worksites in the treatment group while 55% of control group sites did not receive follow-up inspections. All of the worksites should have received follow-up inspections so there should not be a difference in how many did not between the treatment and control group. However, we see that these sites did receive inspections at a statistically significantly different rate. As Table 18 shows, 215 worksites, 109 in the Primary and 106 in the Secondary group, were not assigned to follow-up inspections. All of them were control group sites.

A few of the worksites that were not assigned to follow-up inspections were still inspected by non-SST inspections (20 in the Primary and 13 in the secondary Group, 33 altogether), and as a result, we did have access to their injury-illness records after the treatment.

We applied logistic regressions with the covariates used for the impact estimations to determine if there are significant differences between assigned and unassigned control group sites after applying the non-response weights. We found no significant differences between the two groups.

Still, this error in the RCT implementation added to the problem of high sample attrition rates.

Table 18. RCT, Proportion of worksites unassigned to a follow-up inspection (letter plus inspection treatment)

		Not assigned to Follow-Up Inspection List			Assigned to Follow-Up Inspection List		
		Not inspected	Inspected	Total	Not inspected	Inspected	Total
All worksites	Control	182 (84.7%)	33 (15.4%)	215 (100.0%)	53 (25.2%)	157 (74.8%)	210 (100.0%)
	Treatment	0 (0.0%)	0 (0.0%)	0 (0.0%)	115 (26.9%)	305 (73.1%)	420 (100.0%)
	Total	**182 (84.65%)**	**33 (15.35%)**	**215 (100.00%)**	**168 (26.67%)**	**462 (73.33%)**	**630 (100.00%)**
Primary	Control	89 (81.7%)	20 (18.3%)	109 (100.0%)	23 (21.5%)	84 (78.5%)	107 (100.0%)
	Treatment	0 (0.0%)	0 (0.0%)	0 (0.0%)	115 (27.4%)	305 (72.6%)	420 (100.0%)
	Total	**89 (81.7%)**	**20 (18.3%)**	**109 (100.0%)**	**138 (26.2%)**	**389 (73.8%)**	**527 (100.0%)**
Secondary	Control	93 (87.7%)	13 (12.3%)	106 (0.0%)	30 (29.1%)	73 (70.9%)	103 (100.0%)
	Treatment	0 (0.0%)	0 (0.0%)	0 (0.0%)	111 (26.4%)	309 (73.6%)	420 (100.0%)
	Total	**93 (87.7%)**	**13 (12.3%)**	**106 (0.0%)**	**141 (27.0%)**	**382 (73.0%)**	**523 (100.0%)**

Power calculations - RCT study

Statistical power is the ability of a test to detect an effect, if the effect actually exists. Impact evaluation designs need a minimum sample size to be able to determine with a given certainty if the there is an impact. In this study, the sample sizes are given, and therefore we are able to determine the smallest effect that could be detected given the sample size. We call this the minimum detectable impact (MDI). It essentially inverts the question; it does not ask what the sample size has to be to detect this impact, it asks what size the impact needs to be to be detectable with this given sample.

A small minimum detectable impact means that the study can detect even small impacts at a given level of statistical significance, whereas large minimum detectable impacts mean that the study would not be able to detect impacts unless they were in reality very large.

Before compiling the data and conducting the analysis, we calculated the MDIs for the RCT design assuming a 10% non-response rate. The actual sample attrition rate was larger than anticipated (between 16% and 72%, depending on the study group), further increasing the MDI and decreasing the probability of detecting an impact. The primary group MDI increased more for the letter plus inspection treatment worksites due to implementation error: no follow-up inspections conducted for some of the control group worksites. The MDI change as a results of sample attrition for the letter plus inspections estimated treatment impact on DART rates ranged between 68% and 75% compared to 32%-33% for letter treatment sites. The MDI change for the estimated impacts on the probability of post-treatment violation was smaller, but still over 10% for the letter sites and over 30% for the letter plus inspection sites.

The MDIs and the percentage increases are shown in Table 19.

Table 19. RCT – minimum detectable impacts after sample attrition

Treat-ment	SST List	Probability of post-treatment violation				DART rate			
		Number of Sites with Outcome	MDI as % of sample mean before sample attrition	MDI as % of sample mean after sample attrition	MDI % change as a result of sample attrition	Number of Sites with Outcomes	MDI as % of sample mean before sample attrition	MDI as % of sample mean after sample attrition	MDI % change as a result of sample attrition
Letter	**All**	**998**	**8.6%**	**9.5%**	**11%**	**692**	**13.5%**	**17.8%**	**32%**
	Pri	498	10.0%	11.0%	10%	336	21.1%	28.0%	33%
	Sec	500	11.8%	13.1%	11%	356	15.9%	21.0%	32%
Letter Plus Insp	**All**	**804**	**10.6%**	**14.7%**	**39%**	**567**	**12.5%**	**21.3%**	**71%**
	Pri	409	11.7%	15.8%	35%	274	18.8%	31.5%	68%
	Sec	395	15.1%	21.7%	44%	293	16.3%	28.6%	75%

Table 20. Impact estimates from previous research[29]

Treatment	Outcome	Impact Estimates (1-Year Impact)	Average 1-Year impact Estimate
Inspections	Workplace safety/health	9.4%, 4.0% to 4.6%, 0.0%, 0.9%, 4.4%, 4.2%, 0%	3.3% to 3.4%
Inspections	Regulatory compliance	6.3%	6.3%
Letter	Workplace safety/health	1.6% to 1.7%	1.6% to 1.7%

The first row collects all the impact estimates for inspections (not including impact estimates that include only inspections with penalties). The average of these is a reduction in lost workday injury/illness rate between 3.3% and 3.4%. Given that the RCT study measured impacts over two years, if these are the true annual impact, then two years' impact would be 6.6% to 6.8% if change per year is constant into the second year. The most recent study's estimate of 9.4% is potentially more relevant than an average of all estimated impacts covering over thirty years.

This study's minimum detectable impacts for DART case rate are 18.8% for Primary List worksites, 16.3% for Secondary List worksites, and 12.5% overall. DART case rate is similar to the lost workday injury/illness rate used in this previous research.[30] These MDIs, are both above the 6.6% estimate for two years' impact, so the study was likely to fail to detect impacts that do exist. If the 9.4% annual improvement is taken as the true impact, with 18.8% impact over two years, then the study could have been expected to detect impact, but sample attrition increased MDIs to levels where the impact became even more unlikely to be detected.

Note that no previous estimate exists in the literature for the letter treatment's impact on regulatory compliance, but the one previous estimate for the impact of letters on health and safety outcomes (Eastern Research Group 2004).[31] The ERG study estimated a 1.6% to 1.7% annual improvement in lost workday injury/illness rate, which is well below the MDIs for either of the studies. The MDIs for the impact of letters on regulatory compliance outcomes in the RCT were originally estimated to be 10.0% and 11.8% for Primary List and Secondary List sites, respectively, and 8.56% overall, and was further increased by sample attrition. Although the previous literature estimated the impact on a different outcome, in order for us to be able to make the comparison, we have to assume the magnitude of the impact to be similar. For health and safety outcomes, the original MDIs were much higher and were more affected by sample attrition, therefore the study was not expected to detect impacts.

At the same time, there is one previous estimate for inspections' impact on regulatory compliance, the value of the estimated change in compliance is 6.3% (Weil 2001). Because of the methodology of that study, which measured changes in compliance in between repeated inspections, one cannot necessarily interpret the 6.3% improvement as an annual figure. However, if the 6.3% estimate is interpreted as an annual improvement in compliance, and we disregard the very likely scenario of the impacts diminishing over time, the two-year improvement estimate is 12.6%, which is slightly higher than the originally calculated MDI: 11.7% for Primary, 15.07% for secondary, and 10.58% overall.

[29] Previous literature includes: Inspections' impact on workplace safety/health: (Levine, Toffel, and Johnson 2012), (Gray and Mendeloff 2002), and (Ruser and Smith 1991). Inspections' impact on regulatory compliance is (Weil 2001). Letters' impact on workplace safety/health is (Eastern Research Group 2004).
[30] DART case rate differs by including cases that led to restricted work or transfer to alternative work.
[31] Eastern Research Group (ERG). "Evaluation of Osha's Impact on Workplace Injuries and Illnesses in Manufacturing Using Establishment-Specific Targeting of Interventions: Programmed Inspections and High Hazard Notification Letters. Final Report.". Lexington, MA, 2004.

As Table 19 shows, the RCT MDIs, which were large to begin with, increased substantially with non-response. Therefore, comparing the MDIs re-calculated with known response rates to the results of the very few relevant previous studies (one assessing the impact of the same treatment on a different outcome, and one assessing the impact of a more effective treatment on the same outcome), the RCT study was unlikely to have a sufficient number of respondents to detect the impact of either of the treatments on either of the outcomes.

In general, the impacts of the treatment found by this study and presented in Table 8, Table 9, and Table 10 are much lower than the minimum detectable impacts calculated and presented in Table 19.

III. REGRESSION DISCONTINUITY DESIGN

Beyond the originally planned RCT, we also applied a regression discontinuity design (RDD) to strengthen the findings and offset the limitations of the RCT design. The RDD capitalizes on the injury-illness rate cutoffs OSHA uses to assign worksites to various treatments.

1. Data

The data we used for the Regression Discontinuity Design (RDD) analysis come from the following sources: OSHA Data Initiative (ODI) survey responses, inspection results, and high-rate letter lists.

We separated the RDD sites into "treatment" and "comparison" groups using the 2007 ODI data reporting 2006 injury-illness rates, and the corresponding thresholds for OSHA enforcement actions in those years to mimic the 2008 SST program. Sites over the threshold are "treatment" sites, those under the threshold are "comparison" sites. We used injury/illness rate outcomes from 2011-2012 ODI years reporting 2010-2011 injury-illness data. We excluded from the RDD analysis worksites that do not belong under the SST program's jurisdiction, including construction industry worksites, worksites in State Plan States, and worksites with fewer than 40 employees as reported in their ODI survey data.[32] The RDD does not examine regulatory compliance outcomes (ie. the probability of a violation), because that would require all worksites (treatment and comparison alike) to be inspected.

Figure 5. Data sources of the RDD study

ODI Survey Responses	• Annual survey of ~80,000 worksites in high-hazard industries • Includes injury and illness data from workplaces • Collected from worksites by OSHA each year from 1996-2011 (cancelled in 2012)
Inspections Results	• Recorded by inspectors while conducting inspections • Contains inspection details including worksite information, inspection type and scope, and violations details • Available from 1970, updated daily
High-Rate Letter Lists	• Generated by OSHA based on ODI Survey responses • Contain worksite information of those placed on the high-rate letter list • Available from 1999

[32] The 2008 SST Directive applies to the date when the RDD treatment occurred. The 2008 SST Directive applies to worksites with 40 or more employees.

Key Variables

Table 3 lists the assignment and outcome variables used in the analysis along with the source for each of the variables.

Table 4 lists the covariates used.

Table 21. Assignment and outcome variables, RDD

Study	RDD	
	Variable	Source
Assignment Variable	DART 2006/DAFWII 2006	ODI Survey
Regulatory Compliance Outcome	N/A	N/A
Health/Safety Outcome	DART Rate 2010/2011	ODI Survey

Note: Regulatory compliance outcomes were not examined through the RDD because outcomes are not available for uninspected worksites.

Table 22. Covariates, RDD

RDD	
Covariate	Source
Industry	ODI Survey
Region	ODI Survey
Number of Employees	ODI Survey
Penalties Received Before 2006	OSHA Inspections Data
High-Rate Letters Received before 2006	High-Rate Letter Lists

The data used in the RDD are summarized in Appendix G.

Basic statistics of RDD dataset

Table 5 presents the number of observations broken down by some of the covariates for all worksites above and below the thresholds but within the bandwidths jointly for primary, secondary, and letter treatment sites.

Table 23. Basic descriptive statistics of the RDD study

Worksite Characteristic	RDD	
Industry	**N**	**%**
Agriculture, Forestry, and Fishing	36	1%
Mining	1	<1%
Manufacturing	3,378	48%
Transportation and Communication	646	9%
Wholesale Trade	591	8%
Retail Trade	941	13%
Services	1,438	20%
Missing	14	<1%
Region	**N**	**%**
Region 1	763	11%
Region 2	792	11%
Region 3	834	12%
Region 4	1,074	15%
Region 5	1,630	23%
Region 6	1,041	15%
Region 7	571	8%
Region 8	273	4%
Region 10	67	1%
Number of Employees	**N**	**%**
0-99 Employees	2,996	43%
100+ Employees	3,942	56%
Missing	107	2%
Inspection/Penalty Combinations	**N**	**%**
No inspection, no penalty	4,151	59%
Inspection, but no penalty	623	9%
Inspection and penalty	2,271	32%
DART Rate	**N**	**%**
0.00-9.99	6,122	87%
10-19.99	911	13%
20+	12	0%
High-Rate Letters Received before 2006	**N**	**%**
None	4,022	57%
1-5 Letters	2,667	38%
6+ Letters	356	5%
Total	**7,045**	**100%**

As Table 5 shows, the allocation of worksites across the industry, regional, and other groups is similar to the RCT allocation. The DART rates are somewhat lower in the RDD data, but that is expected, since the worksites around the lower letter threshold decrease the average DART rate, whereas the RCT study dataset only contains sites above the Primary/Secondary thresholds.[33]

[33] The number of observations in the RDD datasets was calculated only for worksites within the calculated bandwidth around the thresholds.

The “Number of Employees” cutoff was higher in the RDD dataset because the 2008 SST Directive (and therefore the RDD study) only targeted sites with more than 40 employees, while the RCT dataset included worksites over 20 employees.

2. Study design and methodology

In general, regression discontinuity designs can be used to estimate impacts of treatments when there is a cutoff or threshold used to determine which subjects will receive a treatment (see for example Angrist and Pischke 2008). In this study, we capitalize on the fact that OSHA categorizes worksites into groups based on injury and illness rate as measured by the DART and DAFWII rates reported in the ODI survey.

We refer to DART and DAFWII rates as the “forcing”, “assignment”, or “running” variables in the context of the RDD. The analysis took advantage of the differences in probability of receiving a letter or inspection at the injury/illness rate thresholds to estimate the impact of each enforcement action. See Chapter 2 “The SST Program” for details about how OSHA uses thresholds to assign worksites to inspection groups.

Design overview and implementation timeline

We assessed the impact of the following three treatments on injury/illness rates:

- Being assigned to the high-rate letter list
- Being assigned to an SST inspection list (Primary/Secondary)
- Receiving a programmed inspection.

We assessed the impact of being assigned to enforcement lists, not necessarily receiving it, by intention-to-treat (ITT) analyses. Worksites may be assigned to an SST inspection but may not actually be inspected. Therefore the ITT analysis takes into account the additional general deterrence effect of fear of being inspected, besides estimating the specific deterrence of inspections.

We assessed the impact of receiving a programmed inspection by a treatment-on-the-treated (TOT) analysis (or a “fuzzy” RDD). The TOT analysis estimates the impact of actually receiving an inspection as opposed to not receiving an inspection regardless of whether or not the worksite was assigned for inspection, thereby ignoring the general deterrent effects of fear of inspections. The TOT analysis uses the instrumental variable approach, a two-stage least squares method to disentangle the specific deterrence effect of inspections.

Starting in 2005, OSHA used both the DART and DAFWII rates to determine assignment to the Primary and the Secondary list. To account for both running variables, we use the “centering approach” recommended by Wong, Steiner, and Cook (2013) to calculate a combined value of the running variable.

We used unweighted linear models, estimated upon worksites whose running variables are within a bandwidth of the thresholds. We calculated the bandwidths around the thresholds using the optimal bandwidth selection method (Imbens and Kalyanaraman 2011).[34] The table below presents the number and proportion of sites within the calculated bandwidth for each study group.

[34] Summit, with recommendations from the TWG, considered several alternatives for linear estimation techniques vs. higher order polynomials, using kernel-weight adjustments, and selecting bandwidth before selecting the final methodology.

Table 24. Number and proportion of sites within the calculated bandwidth

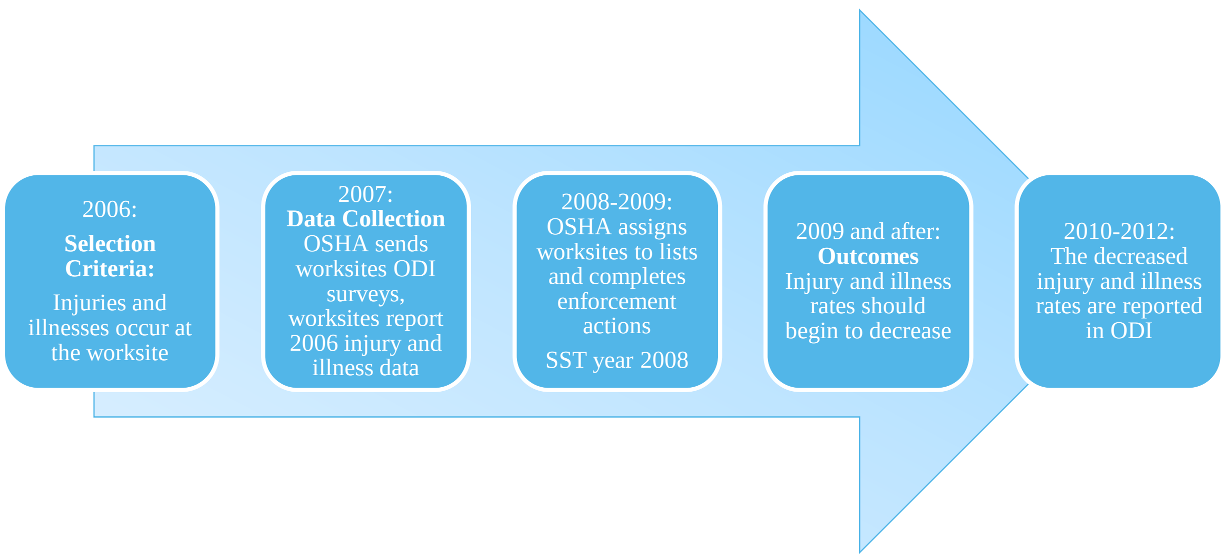
Treatment	Letter	Primary	Secondary
Running variable	**DART Rate 2006**	**Centered Injury Rate 2006**	**Centered Injury Rate 2006**
Threshold*	8	0	0
Bandwidth*	4.18	4.10	2.09
Total sites	16,166	16,166	16,166
Sites within bandwidth	5,023	1,426	4,025
Percent within bandwidth	31.07%	8.82%	24.90%

**DART case rate units (Letter column) and Centered Injury Rate units (Primary and Secondary columns)*

Figure 6 illustrates the timeline of how the injuries and illnesses are translated into enforcement actions and impacts. This timeline also explains how we designed the RDD study.

2006: Injuries and illnesses occur at the worksite.
2007: OSHA sends ODI surveys to worksites, worksites submit injury/illness data from 2006
2008: OSHA selects the sites with high injury and illness rates (DART and/or DAFWII), and begins applying the enforcement actions (sends letters, begins SST inspections)
2009: OSHA completes enforcement actions (finishes SST inspections)
2009 and after: The injury and illness rates begin to decrease.
2010 and after: Decreased rates are reported in the ODI data

Figure 6. RDD implementation timeline



Therefore, to assess the impact of the 2008 SST program using an RDD, we first used the 2007 ODI data to select sites in the various enforcement lists. (The 2008 SST program assigns treatment to worksites using thresholds on the 2007 ODI data.) Second, the RDD takes inspections data from 2008 to assess the probability of an SST inspection below and above the thresholds. At this point we also took data from the high-rate letter lists for the letter treatment analysis. Finally, we used outcome data from the 2010-2012 ODI data to estimate the coefficients of the treatment on health and safety outcomes.

ITT analysis of the letter treatment

The ITT analysis of the letter treatment answers the research question "What impact does being on the high-rate letter list have on injury/illness rates?"

To answer this research question, we use an RDD around the letter injury/illness rate threshold, in which we estimated the impact of being over the letter threshold on the DART rate. We used a linear regression and restricted the analysis to worksites within the bandwidth. The linear regression also allows for controlling for observable covariates, such as industry or area office.

The local linear regression model used in this RDD analysis takes the following form:

$$Y_i = \beta_0 + \beta_1 R_i + \delta S_i + \sum_m \gamma_m z_{mi} + \varepsilon_i \quad (3)$$

Where

- Y_i is the outcome variable of interest (DART and DAFWII rates in the year after receiving the letter)
- S_i is our variable of interest, it is the dichotomous indicator variable of whether the ith worksite has a DART rate (LWDII rate) over the threshold in the base year. $S_i = 1$ if $R_i \geq R_0$, and $S_i = 0$ if $R_i < R_0$
- R_i is the value of the assignment variable (LWDII rate before 2001, and DART rate after 2001) for the ith worksite
- R_0 is the threshold for being assigned to the high-rate letter list
- z_{mi} is the baseline characteristic m for the ith worksite (eg. number of employees, hours per employee, OSHA region, OSHA area office, industry)

The estimate of interest is δ, the coefficient of the assignment indicator variable. The value of δ is interpreted as the average change in the DART or DAFWII rate in the year after being assigned to the high-rate letter list.

TOT analysis of the inspection treatment

The TOT analysis answers the question: "What impact does receiving a programmed inspection have on injury/illness rates?"

To answer this research question, we use two-stage least squares (2SLS) regression (also known as a "fuzzy" RDD). In the first stage, we estimate whether being over the threshold influences the likelihood of receiving a programmed inspection. In the second stage, we use the predicted probabilities of receiving a programmed inspection from the first stage as an explanatory variable to estimate its impact on the outcomes. We control for different sets of covariates in the first and the second stages.[35]

It must be noted that the design uses programmed inspections, not SST inspections. This is because some of the worksites that received an SST inspection possibly would have received another programmed inspection in lieu of the SST program. To the extent that comparison worksites received non-SST programmed inspections, estimated ITT impacts of SST inspections would underestimate the real impact of programmed inspections.

The two-stage least square model's first stage is the following:

$$T_i = \mu_0 + \mu_1 R_i + \pi S_i + \sum_m \gamma_m z_{mi} + \varepsilon_i \quad (4)$$

In this model, S_i, is the instrumental variable for T_i, the probability of inspection. The second stage of this model is then:

[35] We do not use the probability of receiving a programmed inspection, rather than receiving an SST inspection, because SST inspections and other programmed inspections may be close substitutes (compare (Heckman et al. 2000).

$$Y_i = \beta_0 + \beta_1 R_i + \delta \tilde{T}_i + \sum_k \delta_k v_{ki} + \eta_i$$

(5)

Where

T_i	is the dichotomous treatment variable, equal to 1 if the *i*th worksite received a programmed inspection, 0 if not.
$\bar{T}_i$	is the predicted probability of worksite *i* receiving a programmed inspection
S_i	is the dichotomous indicator variable of whether the *i*th worksite is over the threshold $S_i = 1$ if $x_i \geq x_0$, and $S_i = 0$ if $R_i < R_0$.
R_i	is the value of the assignment or "running" variable (DART / DAFWII / LWDII rates) for the *i*th worksite.
R_0	is the DART/DAFWII/LWDII-threshold for being on the Primary/Secondary list of sites
π	is the first-stage effect of S_i – the effect of being assigned to the treatment group on the probability of inspection
z_{mi}	is the baseline characteristic m for the *i*th worksite (eg. number of employees, hours per employee, OSHA region, OSHA area office, industry, and the pre-treatment compliance variables: number of violations, serious violations)
v_{ki}	is the baseline characteristic k for the *i*th worksite

The estimate of interest is δ, the coefficient of the predicted probability of worksite *i* receiving a programmed inspection. It is interpreted as: the average change in the outcome variable of interest (DART and DAFWII rates) as a result of *receiving* a programmed inspection (TOT).

ITT analysis of the inspection treatment

Research question 2b is: "What impact does being on the SST inspection list have on injury/illness rates?"

The estimation technique to answer this research question is similar to the letter treatment: we use a sharp RDD around the inspections list thresholds to estimate the coefficient of the binary variable of being over the threshold in a local linear regression on the outcomes. The linear regression also allows for controlling for observable covariates, such as industry or area office.

The local linear regression estimation uses the following equation:

(6)

$$Y_i = \beta_0 + \beta_1 R_i + \delta S_i + \sum_m \gamma_m z_{mi} + \varepsilon_i$$

Where

Y_i	is the outcome variable of interest (DART and DAFWII rates)
S_i	is our variable of interest, it is the dichotomous indicator variable of whether the *i*th worksite is over the threshold in the base year. $S_i = 1$ if $R_i \geq R_0$, and $S_i = 0$ if $R_i < R_0$
R_i	is the value of the assignment variable (LWDII rate before 2001, and DART rate or the centered value of the DART and DAFWII rates combined after 2001) for the *i*th worksite
R_0	is the threshold for being assigned to the high-rate letter list
z_{mi}	is the baseline characteristic m for the *i*th worksite (eg. number of employees, hours per employee, OSHA region, OSHA area office, industry)

The estimate of interest is δ, the coefficient of the assignment indicator variable. The value of δ is interpreted as the average change in the outcome in the year after being assigned to the SST inspection list. Note that the primary and secondary thresholds are handled in one dataset after pooling them together, and we make no distinction between the impacts around the primary and secondary thresholds. See below for a discussion about pooling the two datasets together.

Multiple Assignment Variables

Starting with SST year 2005, OSHA applied a combination of DART and DAFWII rates to assign worksites to the Primary and Secondary inspections lists. Since SST year 2005, the Secondary and Primary thresholds are defined as "a DART rate over *X* or a DAFWII rate over *Y* (only one of these criteria must be met)". (Note that *Y* is always smaller than or equal to *X*.) This means there are two assignment variables for the primary and secondary inspection lists.

As recommended by Wong, Steiner, and Cook (2013), we used the "centering approach" to deal with the multiple assignment variable problem. We selected this approach because of its straightforward implementation and interpretation. Reardon and Robinson (2012) and Wong, Steiner, and Cook (2013) list options for dealing with the multiple assignment variable problem in RDD.

Therefore, for the 2006-2011 cohort and the Primary and Secondary list worksite, we generated one single assignment variable using the two original assignment variables as follows:

a. First, for each worksite *i*, we centered DART and DAFWII rates to their respective cutoffs, that is $dart_i - dart_c$ and $dafwii_i - dafwii_c$
b. Second, we chose the maximum centered value $z_i = \max(dart_i - dart_c, dafwii_i - dafwii_c)$ as the worksite's sole assignment score.[36] This ensures that the value of z_i is larger than zero when either DART or DAFWII or both are over their respective cutoffs, and it is negative only when both are below the cutoffs.

The way we generated the centered injury-illness rates assumes equal variances of the DART and the DAFWII case rates.

RDD Bandwidth Selection

The bandwidth around the threshold in each case is calculated using the optimal bandwidth selection method that Imbens and Kalyanaraman (2011) describe. This approach uses the following information to calculate the optimal bandwidth:

- The assignment or "running" variable and its distribution;
- The outcome variable and its distribution;
- The running variable threshold used for estimation;
- The kernel-weighting method specification.

(Imbens and Kalyanaraman 2011) define the optimal bandwidth selection for single-stage RDD design, like we use for the ITT letter and ITT inspection analyses. The authors go on to advocate that in the case of a fuzzy regression discontinuity design, it is acceptable to use the same bandwidth for both stages, which is calculated by using the same procedure to find the optimal bandwidth calculated for the outcome variable. We used Stata's RDrobust package to calculate the optimal bandwidth for every outcome-running-variable-threshold combination, dropped observations outside the calculated optimal bandwidth and used the remaining observations to conduct all analyses (Calonico, Cattaneo, and Titiunik 2014).

[36] Note that Wong, Steiner, and Cook (2013) recommend using the min. function, but the assignment is different here, the relationship between the two variables is an OR relationship, while in the Wong, Steiner, and Cook (2013) study it is an AND relationship. Also note that this calculation assumes equal variances of the DART and the DAFWII case rates.

3. Results

The Regression Discontinuity Design (RDD) analysis examines the effect of being over the various enforcement list thresholds (letter threshold, Primary threshold, and Secondary threshold) on post-treatment injury/illness rates as opposed to being under the threshold. When OSHA set the thresholds to determine which sites would be put on the high-rate letter, Primary, and Secondary lists, the characteristics of the worksites were not considered. Therefore, because we include the running variable as a covariate, as we approach the cutoff, any differences in injury/illness between the worksites that are above and below these thresholds can be attributed to differences in treatment scheduled for these sites (this is more and more true as we approach the thresholds). Table 25 presents the number of worksites within the calculated bandwidth for each of those groups, and reports the number of sites below and above the threshold separately. We restrict all analyses to these bandwidths to avoid biasing the estimates, and as a result, RDD findings can only be generalized to the vicinity of the thresholds.

Table 25. Total number of RDD worksites on either side of the threshold, SST 2008

	Within Bandwidth					
	Letter		**Primary**		**Secondary**	
Under the threshold	3,004	59.8%	888	62.3%	2,545	63.2%
At or over the threshold	2,019	40.2%	538	37.7%	1,480	36.8%
Total	**5,023**	**100.0%**	**1,426**	**100.0%**	**4,025**	**100.0%**

Intention to Treat (ITT) Analysis

Table 26 presents the results of the RDD ITT analysis separately for the Letter, the Primary and the Secondary set of worksites within the bandwidths around their respective thresholds. The RDD ITT analysis measures the effect of being above these thresholds. We measured the effect as the difference between the injury/illness rates of sites above the threshold as compared to those below. The first column in Table 26 presents the effect of being above the high-rate letter threshold. The second and third columns present the effect of being over the Primary and Secondary threshold, respectively.

Table 26: RDD regression results, impact of being over the high-rate letter/primary/secondary threshold (with covariates), ITT

All worksites	RDD Letter	RDD Primary	RDD Secondary
Difference in Mean DART case rate (Estimated Impact)	**0.13** CI [-0.18, 0.44]	**-0.01** CI[-0.66, 0.64]	**0.31** CI[-0.06, 0.68]
Standard Error	0.16	0.33	0.19
Below Threshold Mean (comparison group)	3.54	4.63	3.73
Above Threshold Mean (treatment group)	3.68	4.62	4.04
Observations	4,925	1,425	3,983
R-Sq.	0.09	0.07	0.09
F-Stat	29.02	7.36	26.53

* p<0.05. Note: Covariates were used to calculate these estimates. The list of covariates is presented in Table 22. The number of observations in this table may differ from the number of observations presented in Table 25 if there are observations with missing covariates that were not included in the results calculations.

RDD – Letter treatment findings (ITT)

We first discuss the findings of letter treatment analysis in the RDD design for health and safety outcomes for each of the study groups. These results show the impacts of being over the high-rate letter threshold on worksite health and safety (measured by the 2010-2011 DART rate).

Research Questions and Answers

Question:

What impact does being on the high-rate letter list have on injury/illness rates?

Answer:

As shown in Table 26, the estimated impact of being on the high-rate letter list on DART rates, that is the difference between post-treatment DART rates in worksites above and below the letter threshold 0.13 (or 3.7% of the comparison group mean). This result is not statistically significant at the 5% level.

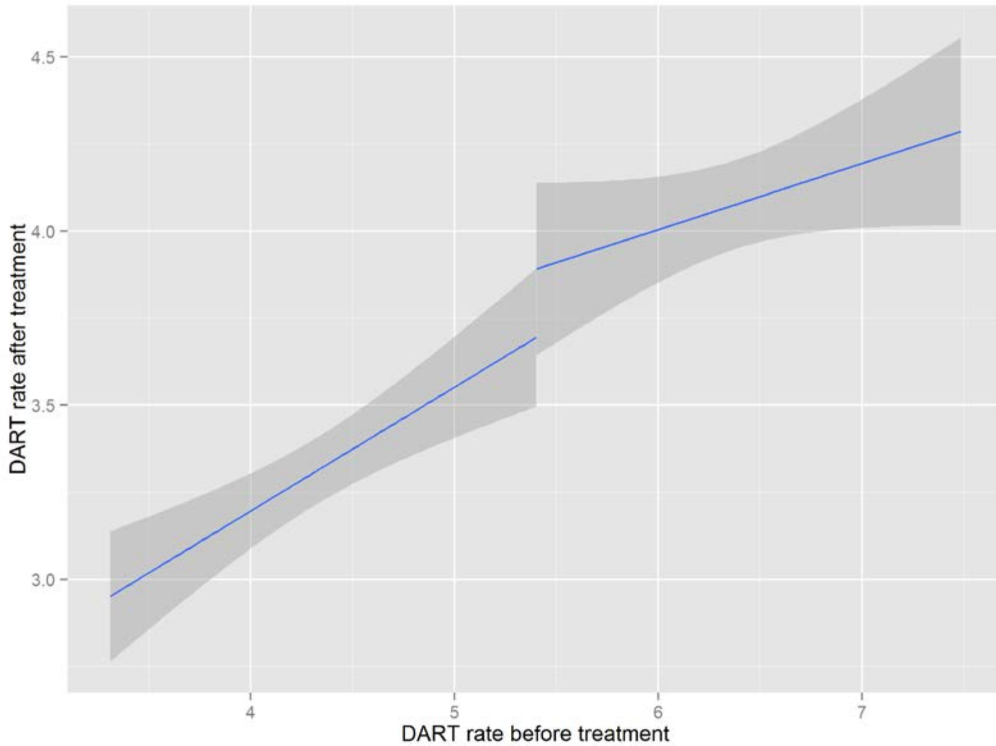
Similar to the RCT analysis results, being above the letter threshold was expected to decrease worksites' injury/illness rates; the sign of the estimated impact is unexpected. But the results are not statistically significant at the 5% level.

The confidence interval around the estimate is also shown in Table 26. They show the likely range of the impact: it was unlikely to be very large in either direction. We can conclude that being over the high-rate letter threshold is unlikely to have reduced worksites' injury and illness rates by more than 0.18 (DART case rate measure, -5.1% of the comparison group mean) and it is unlikely to have increased it by more than 0.44 (12.4%); the lower and upper bounds of the confidence intervals, respectively.

These results are also presented on the graph below (Figure 7). The graph presents the levels of post-treatment injury-illness rates (DART rates) by the running variable (DART-rate before treatment); it fits two linear regressions on the data: one for worksites under the threshold and one for those over it. Note that these graphs do not control for covariates as the regressions do, but we can still see the slight increase in injury-illness rates right at the threshold, just like we saw in the regression results in Table 26.

The graph also presents confidence intervals around the regression lines. For the model to be able to detect an impact, there should be a much larger difference between post-treatment injury/illness rates at the letter threshold. The jump at the threshold should be larger so that the confidence intervals would not overlap at the threshold.

Figure 7. DART rate after treatment plotted by the running variable (Letter sites)

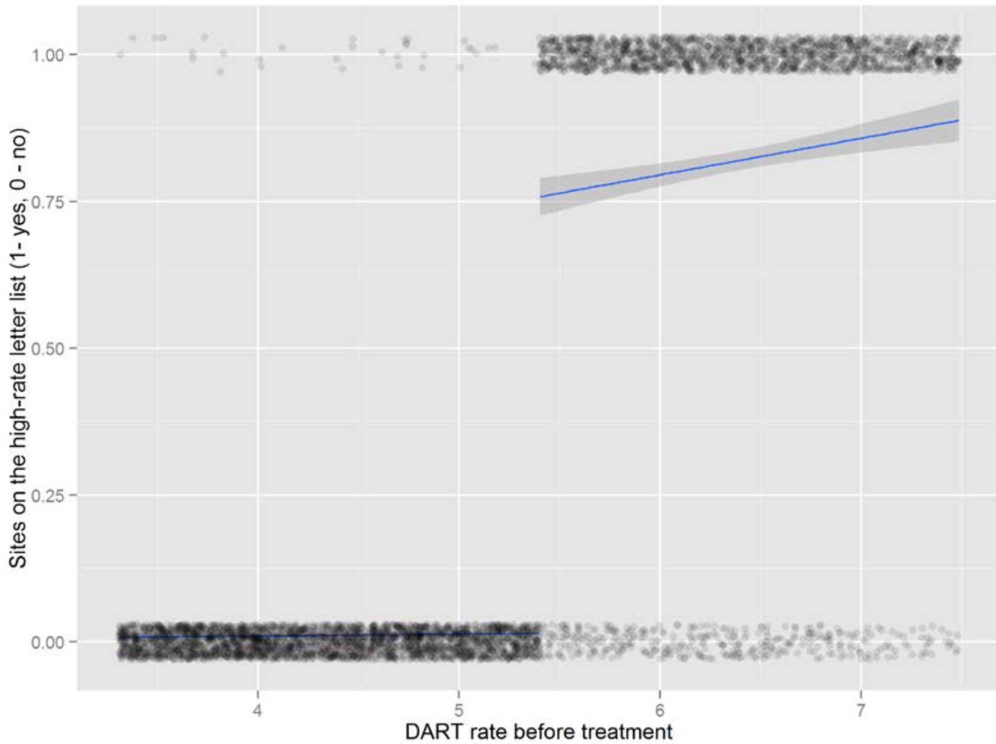


We also examined the implementation of the letter treatment, and found that not all worksites above the threshold actually appeared on the high-rate letter lists.

Figure 8 below presents the probability of a worksite being on the high-rate letter list by the running variable (the pre-treatment DART-rate). The graph shows a large discontinuity at the DART rate threshold (5.4) in the probability of a worksite being on the high-rate letter list, but the probability of being on the high-rate letter list only jumps to about 80%, not 100%. We would expect all worksites[37] above the high-rate letter threshold to be placed on the high-rate letter list. In addition, a few worksites under the letter threshold also appeared on the high-rate letter lists. We can also see from the graph that very few worksites under the thresholds received a high-rate letter.

[37] All worksites in the dataset: that is worksites with more than 40 employees in eligible industries and states.

Figure 8. The probability of being on the high-rate letter list plotted by the running variable (Letter sites)



RDD – Inspection treatment ITT findings

The RDD ITT analyzes the impact on injury/illness rates of being on the Primary and Secondary inspections list in each of the study groups.

Research Questions and Answers

Question:

What impact does being on the SST inspection list have on injury/illness rates?

Answer:

As shown in Table 26, the estimated impact of inspections on DART rates is -0.01 (-0.2% of the comparison group mean) and 0.31 (8.3%), for Primary and Secondary list worksites respectively.

The inspections treatment was expected to result in lower injury/illness rates in the treatment group as compared to the comparison group (that the coefficients of the treatment would be negative). The coefficient of the treatment is negative as expected for Primary sites, but positive for secondary sites, and not statistically significant in either case at the 5% level.

Based on the confidence intervals for each of the estimates shown in Table 26, the impact was unlikely to be very large in either direction for Primary or Secondary sites. If being over the Primary and Secondary thresholds had an impact on worksite injury and illness, it was less than 5% likely to reduce worksites' injury and illness rates by more than 0.66 (-14.3%) and 0.06 (-1.6%), respectively, and it was unlikely to have increased it by more than 0.64 (13.8%) and 0.68 (18.2%), respectively (DART case rate measures).

Figure 9 and Figure 10 present the levels of post-treatment DART rates by the running variable (DART-rate before treatment); it fits two linear regressions on the data: one for worksites under the threshold and one for those over it. Because in this case we are working with the centered injury-illness rates, the threshold is zero in each case.

Consistently with Table 26 results, we see a slight negative shift in the regression lines at the threshold for Primary worksites, and a positive shift for Secondary worksites. If we had a significant treatment effect, we would see a much larger discontinuity in the graph at the threshold. In Figure 9, we barely see any discontinuity at the threshold. In Figure 10, we do see some discontinuity in the injury/illness rate after treatment for sites above the threshold: an increase, a statistically not significant impact in the unexpected direction.

Figure 9. DART rate after treatment plotted by the running variable (Primary Sites)

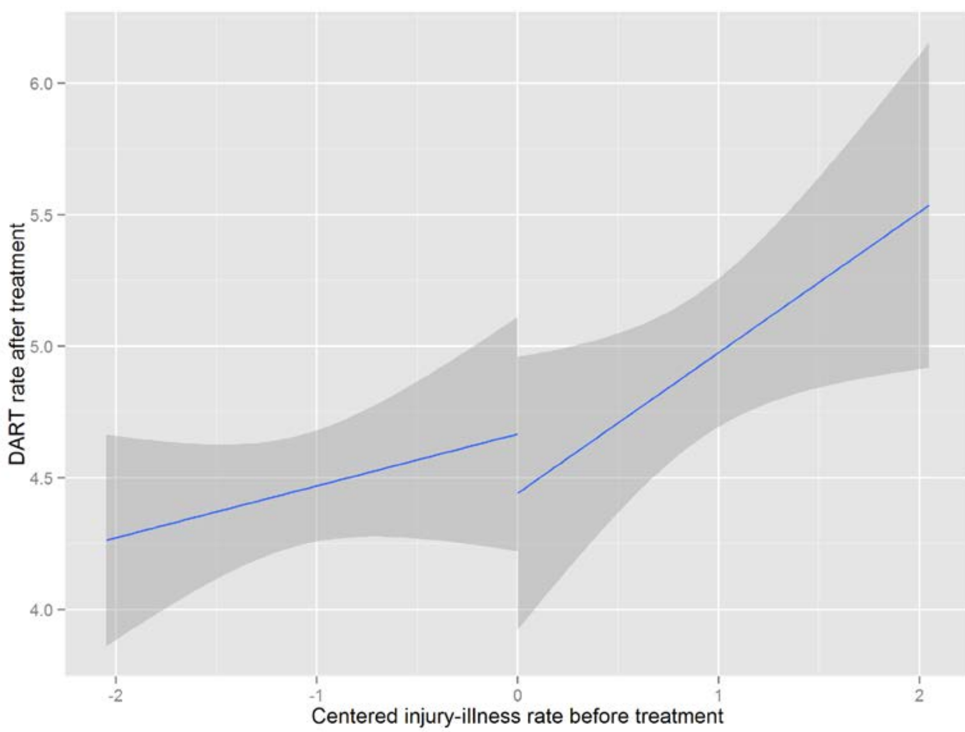
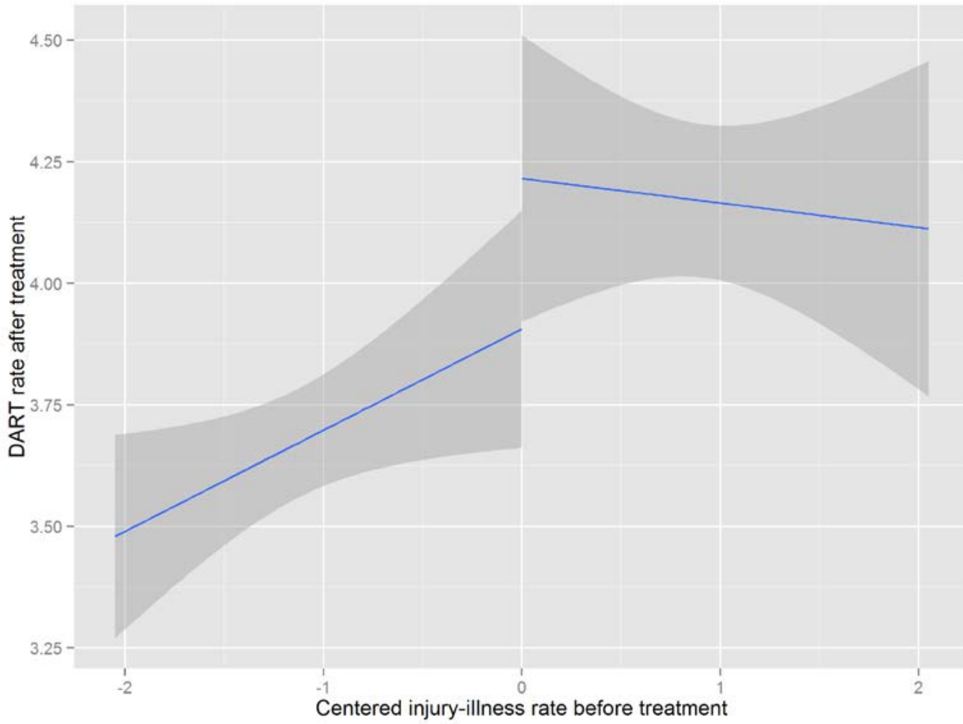


Figure 10. DART rate after treatment plotted by the running variable (Secondary Sites)



Treatment on the Treated (TOT) Analysis

In addition to assessing the impact of being on either side of the set thresholds, we also assessed the impact of actually receiving an inspection on injury/illness rates. We could not verify receipt of a high-rate letter, so we only conduct this analysis (the TOT analysis) on the Primary and Secondary sites. This analysis employs a two-stage design to assess this impact. The first stage estimates how likely a worksite is to receive an inspection. The second stage takes into account the likelihood of receiving an inspection to estimate the difference in injury/illness rates between the worksites that did and did not receive inspections.[38] It is important to note that TOT results assume that health and safety outcomes are only impacted by inspections that were actually conducted. TOT therefore ignores the additional impact of fear of inspections that occurs at worksites that do not actually receive an inspection. If fear of inspections has an impact, then TOT estimates using the instrumental variable method will provide upward-biased estimates of the effect of the inspection policy.

Interpreting the coefficients in Table 27 below, we see that first stage coefficients of being over the threshold tell us the difference in the probability of receiving a programmed inspection between worksites above the threshold and those below the threshold after controlling for covariates. This tells us the story of implementation: *is there a difference in the treatment between the treatment and the comparison group?* If the treatment and comparison groups were not treated differently, we cannot reliably estimate the differences in outcomes between the two groups. The second stage coefficients show the actual impact of receiving a programmed inspection on injury-illness rates. They report the difference in injury/illness rates after treatment between worksites that received a programmed inspection and those that did not.

[38] In this analysis, we consider receiving any programmed inspection as treated, rather than only receiving an SST inspection. Control worksites may receive a programmed inspection under another program and therefore are considered to be treated.

Table 27. RDD - TOT regression results of inspections

First Stage Regressions	The effect of instruments (being over the threshold) on receiving an inspection	
Study Group	**Primary**	**Secondary**
Outcome	**Received a programmed inspection in 2008**	**Received a programmed inspection in 2008**
Over Primary/Secondary Threshold 2006	**16.6%*** CI [7.5%, 25.6%]	**-0.7%** CI [-4.1%, 27.5%]
Standard Error	0.05	0.02
Observations	1,425	3,983
F-Stat	12.02	7.65
Prob > F	0.000	0.000
Second Stage Regressions	**The effect of receiving an inspection on post-inspection DART Rate**	
Study Group	**Primary**	**Secondary**
Outcome	**Injury/Illness Rate-2010/2011**	**Injury/Illness Rate-2010/2011**
Predicted probability of a Programmed Inspection in 2008	**-0.075** CI [-3.96, 3.81]	**-46.073** CI [-286.77, 194.62]
Standard Error	1.98	122.80
Observations	1,425	3,983
Centered R-Squared	0.07	-18.95
F-Stat	7.34	1.64
Prob > F	0.000	0.052

* p<0.05. Note: Covariates, including the running variable were used to calculate these estimates. The list of covariates is presented in Table 22. The number of observations in this table may differ from the number of observations presented in Table 25 if there are observations with missing covariates that were not included in the results calculations.

RDD – Inspection treatment TOT findings

The RDD TOT results show the impact of receiving a programmed inspection on worksite injury/illness rates (measured by the post-treatment DART rate).

Research Questions and Answers

Question:

What impact does receiving a programmed inspection have on injury and illness rates?

Answer:

The estimated impacts of programmed inspections on post-treatment DART rates are a decrease of 0.075 (-1.62% of the comparison group mean) and 46.07 (>1,000%) for Primary and Secondary sites respectively. Neither of those results are statistically significant at 5% level. We expected programmed inspections to decrease worksites' injury/illness rates, and the signs of the estimated impacts are negative as expected.

Based on the confidence intervals for each of the estimates shown in Table 27, the impact of receiving a programmed inspection was unlikely to be very large in either direction for Primary or Secondary sites. If receiving a programmed inspection had an impact on Primary and Secondary worksites' injury and illness, they were less than 5% likely to reduce worksites' injury and illness rates by more than 3.963 (-85.53%) and 286.766 (>1,000%), respectively for Primary and Secondary worksites, and they were unlikely to have increased it by more than 3.813 (82.29%) and 194.620 (>1,000%), respectively (DART case rate measures).

The very large impact and the very wide confidence interval found for Secondary sites on the second stage is an unreliable result as we did not find a significant first-stage impact for Secondary sites. In other words, we are facing the weak instrument problem discussed on page 51. We found no difference in the probability of receiving a programmed inspection between treatment and comparison sites. If no difference in treatment occurs, we should not expect differences in outcomes.

On the other hand, we did find a significant positive difference of 16.6% between Primary worksites above and those under the threshold in their probability of receiving a programmed inspection. This means that there actually was a difference in treatment between the treatment and the comparison group.

4. Study limitations and power

The following factors have reduced the chances of the regression discontinuity design (RDD) study to find a statistically significant impact:

- **Worksites that did not receive an SST inspection may have received some other programmed inspection instead.** As we noted, we do not use the probability of receiving an SST inspection for this estimation but instead the probability of receiving a programmed inspection. We made this design decision knowing that SST inspections and other programmed inspections may be close substitutes (Heckman et al. 2000). To the extent that comparison worksites received non-SST programmed inspections, estimated ITT impacts of SST inspections would underestimate the real impact of programmed inspections.

  Worksites receiving other programmed inspections dilutes the impact that the design can estimate. Worksites over the threshold are more likely to receive an SST inspection. However, there is a smaller difference between receipt of programmed inspections between worksites above and below the threshold. While worksites below the threshold may not receive an SST inspection, they are likely to receive another type of programmed inspection. Therefore there will not be as big of a difference between resulting outcomes as those are affected by both SST and non-SST programmed inspections.

  As Table 28 shows, the ratio of SST inspections among all programmed inspections is different between Primary treatment and comparison worksites (95.6% vs. 87.6%), but less different between Secondary treatment and comparison worksites (70.6% vs. 67.8%). This means that non-SST programmed inspections more often replaced SST inspections for Secondary than for Primary worksites.

Table 28. Programmed vs. SST inspections

Inspection type	Primary worksites		Secondary worksites		Total	
	SST	Progr.	SST	Progr.	SST	Progr.
Treatment	173	181	89	126	262	307
	173/181 = 95.6%		89/126 = 70.6%		262/307 = 85.3%	
Comparison	120	137	141	208	261	345
	120 / 137 = 87.6%		141/208 = 67.8%		261/345 =75.7%	
Total	**293**	**318**	**230**	**334**	**523**	**652**
	293/318 = 92.1%		**230/334 = 68.9%**		**523/652 = 80.2%**	

- **Treatment-on-the treated analysis for the secondary list suffered from a weak first-stage instrument.** In the treatment-on-the treated analysis, first stage results indicate the difference in the probability of receiving a programmed inspection between sites below and above the threshold. A large enough difference means that the impact found on the second stage is more likely to be statistically significant.

  We expected the first stage coefficients to be highly positive. A positive first-stage coefficient indicates a difference in treatment between treatment and comparison groups (worksites above and below the threshold): a positive first-stage coefficient means that worksites above the threshold were more likely to receive programmed inspections than those below the threshold. As expected, we can see a highly statistically significant positive first-stage coefficient for Primary worksites, and an insignificant first-stage coefficient for Secondary worksites in the unexpected direction. We can partly attribute this phenomenon to the fact that non-SST programmed inspections were more likely to replace SST inspections for Secondary than for Primary sites (see Table 28.)

Figure 11 and Figure 12 show the first-stage results of the probabilities of worksites receiving a programmed inspection by the running variable (the centered injury – illness rate) (note that in these graphs we did not control for covariates as we did in the regressions). The graphs fit a linear regression line separately on worksites above and below the threshold to show the change in the probability of receiving a programmed inspection at the threshold.

A discontinuity in the probability occurs for Primary sites receiving a programmed inspection at the threshold centered to 0, but does not occur for Secondary sites.

Figure 11. The probability of programmed inspections plotted by the running variable (Primary Sites)

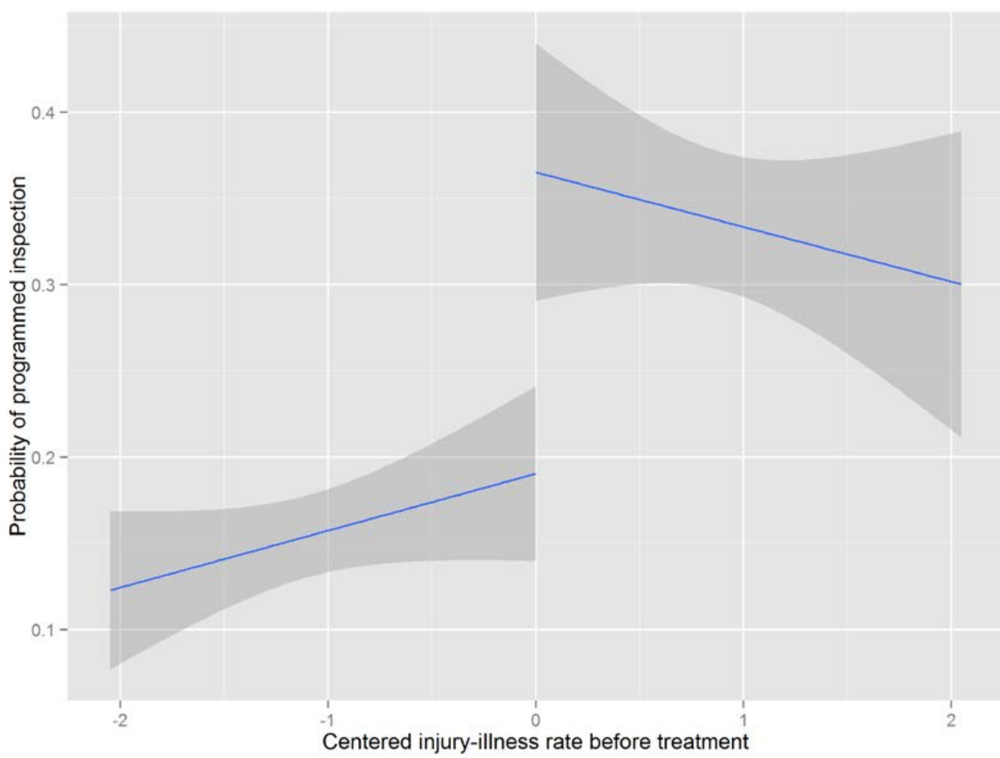
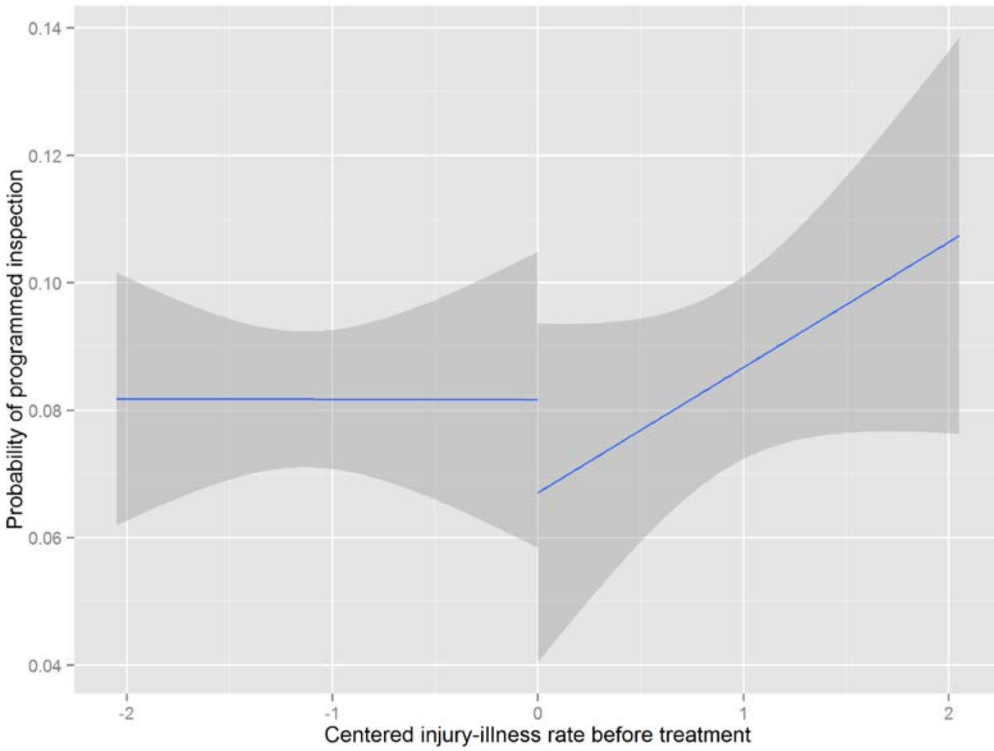


Figure 12. The probability of programmed inspections plotted by the running variable (Secondary Sites)



The discussion of TOT estimates requires introducing the concept of “weak instrument”: instruments that are poor predictors of the endogenous variable in the first-stage of the two-stage least squares.

Stock and Watson (2015) give the following rule of thumb for checking for weak instruments: “The first-stage F-statistic is the F-statistic testing the hypothesis that the coefficients on the instruments […] equal zero in the first stage of two stage least squares. When there is a single endogenous regressor, a first-stage F-statistic less than 10 indicates that the instruments are weak, in which case the TSLS estimator is biased (even in large samples) and TSLS t-statistics and confidence intervals are unreliable.”

For primary worksites, the first-stage results (see Table 27) are statistically significant, with an F-statistic of about 12. This F-statistic suggests that the instrument in the 2-stage least squares regression is slightly above the rule-of-thumb threshold for weak instruments. The probability for a primary worksite to receive a programmed inspection jumps from about 15-20% to 30-35% if the worksite was above the Primary inspection list threshold. This is a noticeable jump, however, with that size of a first-stage impact, we would need a very large second-stage impact (the impact of programmed inspections on outcomes) to be detectable.

For secondary worksites, in addition to the instrument’s weakness, the direction of the relationship is unexpected: worksites are *less* likely to be inspected if they are above the cutoff. The F-statistic of the first stage for secondary worksites is F = 7.65, lower than the rule-of-thumb threshold of 10, meaning therefore that the second stage estimates are unreliable.

Similar to the ITT case, programmed inspections take the place of SST inspections and water down their effects, causing instrument weakness.

As a result of the above the power of the RDD study decreased.

Power calculations - RDD study

We calculated the minimum detectable impacts (MDIs) for the RDD design using the ITT and TOT ANCOVA sample size formula that is presented in detail in Appendix B.

The Power of the ITT analysis

Table 29. RDD - ITT minimum detectable impacts

	Letter	Primary	Secondary
DART Case Rate	0.443	0.949	0.530
Sample Mean	3.617	4.630	4.015
% of Sample Mean	12.25%	20.49%	13.21%

The RDD study only assesses workplace safety and health, not regulatory compliance. The minimum detectable impacts of being over the threshold on DART case rate are 20.49% for the Primary List of worksites, 13.21% for the Secondary List worksites, and 12.25% for the letter list (percentage of the sample mean). The Primary and Secondary MDIs are above the 6.6% estimate for two years' impact (see Table 20), and the Letter MDI is over the 3.4% impact estimate over two years, meaning the RDD study does not have the power to detect impacts that do exist.

If the 9.4% annual improvement is taken as the true impact, with 18.8% impact over two years (respectively), then the study could be expected to detect impact for Secondary worksites. Note that we are comparing TOT estimates from previous research to calculated ITT MDIs, and thus these results are only approximate indications of the relative likelihood of detecting an impact.

When we compare ITT MDIs to the actual effect sizes found by the RDD ITT analysis (as presented in Table 26), we see that the minimum detectable impacts on DART case rate (0.443, 0.949, and 0.530 in Table 29) are disproportionately larger than the impact sizes found (-0.01, 0.31, and 0.13 in Table 26) respectively for Primary, Secondary, and Letter sites.

The Power of the TOT analysis

Table 30. RDD - TOT minimum detectable impacts

	Primary	Secondary
DART Case Rate	0.688	1.407
Sample Mean	4.630	4.015
% of Sample Mean	14.86%	35.05%

When we compare TOT MDIs of the RDD study as presented above (Table 30), we find that the minimum detectable impacts of receiving an inspection on DART case rate are 14.86% for the Primary List of worksites and 35.05% for the Secondary List worksites (as a percentage of the sample mean). These are both well above the 6.6% estimate for two years' impact (see Table 20). The RDD study was likely to fail to detect impacts if they exist.

The actual impact sizes found and presented in Table 27, the Primary impact found (-0.075) is well under the MDI of 0.688; although the Secondary impact size is larger than its respective MDI, the estimate is unreliable due to the weak first-stage result.

IV. CONCLUSIONS

1. Comparative Discussion of the RCT and the RDD study

In this section, we compare the designs and findings of the two causal studies: the RCT and the RDD study. Although subtle differences in the research questions asked and the interpretation of findings, both assess the impacts of letters and inspections on health and safety, and we find no contradictions between the findings.

When we compare results from the two studies we find that some of the estimated impacts on injury and illness rates[39] were in the expected direction in both the RDD and the RCT, others were in unexpected directions in both, and yet others showed differences in the direction of the impacts between the two methods. But none of those results were significant at the 5% level.

While an RDD design allows for retrospective application as it does not require changes to day-to-day procedures, the RCT design required OSHA to handle RCT sites differently than it would have ordinarily. Regardless, an RCT implementation more faithful to RCT plans could have provided a much stronger contrast between research groups than did the RDD.

The RDD study and the RCT study are different primarily due to the design through which causality is inferred. Even though both aim to assess the impact of OSHA's enforcement actions on outcomes, differences between the two studies occur in (1) implementation, (2) research questions, (3) the study years, and (4) in the numbers and characteristics of worksites. These comparisons are summarized below in Table 32 and discussed in the remainder of this section.

Table 31. Comparison of RCT and RDD designs

	RCT	RDD
Implementation	Required active participation from OSHA	Based on existing administrative data
Research questions	- Impact of randomization - Widely generalizable - Assesses regulatory compliance	-Impact of policy -Locally valid -Does not assess regulatory compliance
Timing	Treatment in 2011	Treatment in 2006
Number of sites	Relatively small	Large

Implementation

While the RCT study required active and constant attention from OSHA, the RDD study did not. For a successful RCT, worksites assigned to treatment groups would receive treatment and comparison group

[39] The RDD did not assess regulatory compliance.

member sites would not. Additionally, study sites needed to be assigned to follow-up inspections at the time designed by the study, and data needed to be collected appropriately.

The RDD study's use of existing administrative data protected the study from the risk of human errors in implementation. However, the uncontrolled nature of the quasi-experiment came at a price:

- there were no obligatory follow-up inspections and as a result the RDD study could not assess the impacts on regulatory compliance,
- the RDD study requires a much larger sample size than the RCT to detect the same impact, and
- RDD estimates are local around the cutoff, while the RCT is a global estimate.[40]

Research questions

The research questions of the RCT and RDD studies are listed in Section 6. While an RCT study assesses impacts on regulatory compliance, an RDD study does not. In addition, the RDD study addresses Treatment-On-the-Treated (TOT) questions, which the RCT does not. The intention-to-treat (ITT) questions asked by the RDD are arguably more policy-relevant because they directly address the impact of the policy, with defined thresholds and assigned treatments based on those thresholds. However, the RDD results only have local validity near these thresholds. On the other hand, the RCT results can be extrapolated more broadly, because the study participants were selected randomly from a larger population.

Study years

Because the RDD study worked with ODI data, which was discontinued in 2011, it cannot examine treatments that happened after 2006, five years earlier than the RCT study's treatment year of 2011. The more recent the research date is, the more applicable its findings are to present conditions because there is likely to be less of a difference between unobserved characteristics of two worksites that are closer to each other in time.

Number and characteristics of worksites

Using all ODI data in the RDD study, we could work with a much larger number of worksites than in the RCT, but, as shown above in the power calculations for both studies, RDD requires a much larger sample.

The distribution of the RDD study sites is somewhat different from the distribution of RCT sites. The DART rates are somewhat lower in the RDD data because the RCT study dataset only contains sites above the Primary/Secondary thresholds, while the RDD covers a wider range. The distribution of the number of employees is different across the two studies because the 2008 SST Directive only targeted sites with more than 40 employees, while the RCT dataset only excluded worksites under 20 employees. (See Table 5).

Comparison of findings

The subtle differences in the research questions and the implementation of the two studies left the potential for the findings from the studies contradicting each other. For example, differences could have occurred if the effectiveness of the enforcement policies changed over time.

[40] Since the RDD analysis restricts the sample to sites with injury/illness rates near the cutoff, the estimates are asymptotically true as we approach the cutoff. There is a possibility for an impact for sites further from the cutoff but only an RCT, not an RDD, would be able to estimate such an impact.

While some point estimates were in the opposite direction between the RCT and RDD, we did not find contradictions in the overall conclusions that can be drawn from the results. Neither of the studies detected any statistically significant impacts of the enforcement actions (high-rate letters and inspections) on health and safety outcomes. The lack of statistically significant impacts is due to the different limitations discussed above, which impacted the data, design, and implementation of the study.

Table 32 below makes basic comparisons between the findings of the two causal studies. The values shown are the differences in DART case rate units, where one unit represents one incident per 100 full time employees that resulted in lost or restricted days or job transfer due to an injury or illness related to work.

Table 32. Comparison of impact directions and sizes on injury and illness rates between RCT and RDD

In DART case rate units	**RCT (% of control mean)**	**RDD (% of comparison mean)**
Letter treatment	Primary: 0.30 (4.4%) Secondary: -0.18 (-2.9%) Combined: -0.05 (-0.8%)	0.13 (3.7%)
Inspections treatment - Primary	-0.99 (-14.1%)	TOT: -0.075 (-1.62%) ITT: -0.01 (-0.2%)
Inspections treatment Secondary	-0.48 (-7.6%)	TOT: -46.073 (>1,000%) ITT: 0.31 (8.3%)

Note: none of these impacts was statistically significant at the 5% level.

2. Implications

The lack of significant impacts found by the RCT and the RDD designs does not mean that OSHA's enforcement actions had no beneficial impact on workplace health and safety or regulatory compliance. The inconclusiveness of this study does not provide evidence that the program is ineffective. The results of this impact evaluation allow for ruling out very large impacts, but do not allow for coming to conclusions about smaller impacts, if they exist. Therefore we focus our discussion on the study's implications for further research, particularly the avenues for improving future evaluations of this program, as opposed to discussing any policy implications of the study findings.

Implications for future research

Avenues for improvement in future analyses of the impact of inspections, letters, and other enforcement actions include: increasing the rigor of RCT implementation, working with larger sample sizes, and improving data collection.

There are several potential ways careful implementation of the randomized controlled trial can increase the likelihood of detecting an impact. Future studies could ensure that control group sites are isolated from the full scope of the potential treatment. For example, assuring that worksites that receive high-rate letters are not published on the list of high-rate letter sites. Careful implementation of future evaluations will require process controls that are planned based on the theory of change and that are supervised and enforced by the Federal agency. These responsibilities should include appropriate communication about evaluation implementation requirements to all of the stakeholders, stewardship to ensure reproducible data maintenance and documentation, and enforcement of implementation regulations within the agency and on contractors.

One of the biggest limitations of the present evaluation is the relatively small sample size. A larger sample size in future studies would improve the ability to detect impacts and to determine precisely how big those impacts are. A larger sample size could be attained economically by choosing the same overall sample size but fewer study groups. It could also be aided by aligning the follow-up inspection schedule so that study groups would not differ so dramatically in the time from treatment to follow-up, allowing consolidation of study groups in the analysis.

Finally, an improved process for data collection would help alleviate the sample attrition problem. This process could involve a separate data collection tool for the study sites. This would ensure that missing data elements are identified and corrected quickly rather than being lost in the data collection tools, which contain information for all sites with which OSHA interacts. A survey, such as the ODI, sent specifically to the study sites could also be used rather than Form 300A data since Form 300A data are often not kept or collected.

Improving these three items in future studies would provide DOL with better results that would make it more likely that an impact could be detected, if it exists.

BIBLIOGRAPHY

The United States House of Representatives, House Committee on Education and the Workforce, Subcommittee on Workforce Protections. 2010. The United States House of Representatives, House Committee on Education and the Workforce, Subcommittee on Workforce Protections. *Testimony of David Michaels on "Protecting America's Workers Act (PAWA): Modernizing OSHA Penalties"*. March 16,.

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

Becker, Gary S, and George J Stigler. 1974. "Law Enforcement, Malfeasance, and Compensation of Enforcers." *The Journal of Legal Studies* 3 (1):1-18.

Calonico, S., M. D. Cattaneo, and R. Titiunik. 2014. "Robust data-driven inference in the regression-discontinuity design." *Stata Journal* 14 (4):909-946.

Camerer, Colin F, George Loewenstein, and Matthew Rabin. 2011. *Advances in Behavioral Economics*. Princeton, NJ: Princeton University Press.

Cave, George. 1987. "Sample Sizes for Social Experiments." The American Statistical Association.

Diamond, Peter A., Hannu Vartiainen, and Yrjö Jahnssonin säätiö. 2007. *Behavioral Economics and Its Applications*. Princeton, NJ: Princeton University Press.

Eastern Research Group. 2004. Evaluation of OSHA's Impact on Workplace Injuries and Illnesses in Manufacturing Using Establishment-Specific Targeting of Interventions: Programmed Inspections and High Hazard Notification Letters. Final Report. Lexington, MA.

Gray, Wayne B, and John T Scholz. 1991. Do OSHA Inspections Reduce Injuries? A Panel Analysis. In *NBER Working Paper No. 3774*.

Gray, Wayne B., and John Mendeloff. 2002. The Declining Effects of OSHA Inspections on Manufacturing Injuries: 1979 to 1998. National Bureau of Economic Research.

Harrington, Winston. 1988. "Enforcement Leverage When Penalties are Restricted." *Journal of Public Economics* 37 (1):29-53.

Heckman, James, Neil Hohmann, Jeffrey Smith, and Michael Khoo. 2000. "Substitution and dropout bias in social experiments: A study of an influential social experiment." *Quarterly Journal of Economics*:651-694.

Imbens, Guido, and Karthik Kalyanaraman. 2011. "Optimal bandwidth choice for the regression discontinuity estimator." *The Review of Economic Studies*.

Juras, Randall, Peter Honnef, Jacob Klerman, Amy Minzner, and Meghan Shea. 2015. Evaluation of the OSHA On-site Consultation Program (OSC): The Effect of High-Rate Letters on OSC Requests. Abt Associates.

Levine, D. I., M. W. Toffel, and M. S. Johnson. 2012. "Randomized Government Safety Inspections Reduce Worker Injuries with No Detectable Job Loss." *Science* 336 (6083):907-911. doi: 10.1126/science.1215191.

Little, Roderick J. A. 1986. "Survey Nonresponse Adjustments for Estimates of Means." *International Statistical Review/Revue Internationale de Statistique* 54 (2):139-157.

Mendeloff, John, and Wayne B. Gray. 2005. "Inside the Black Box: How do OSHA Inspections Lead to Reductions in Workplace Injuries?" *Law & Policy* 27 (2):219-237.

Mischke, Christina, Jos H. Verbeek, Jenny Job, Thais C. Morata, Anne Alvesalo-Kuusi, Kaisa Neuvonen, Simon Clarke, and Robert I. Pedlow. 2013. "Occupational Safety and Health Enforcement Tools for Preventing Occupational Diseases and Injuries." *Cochrane Database of Systematic Reviews* (8). doi: 10.1002/14651858.CD010183.pub2.

Myerson, Roger B. 1991. *Game Theory: Analysis of Conflict*. Cambridge, MA: Harvard University Press.

National Opinion Research Center. 1996. Final Report: Sample Design for a Statistically Valid Evaluation of Accuracy and Completeness of an Establishment's OHSA Mandated Employee Records Chicago, IL: National Opinion Research Center.

Reardon, Sean F, and Joseph P Robinson. 2012. "Regression discontinuity designs with multiple rating-score variables." *Journal of Research on Educational Effectiveness* 5 (1):83-104.

Ruser, John W., and Robert S. Smith. 1991. "Reestimating Osha's Effects: Have the Data Changed?" *The Journal of Human Resources* 26 (2):212-235.

Stock, James H., and Mark W. Watson. 2015. *Introduction to econometrics*. Third edition update. ed, *The Pearson series in economics*. Boston: Pearson.

Thompson, Steven K. 2002. Sampling. 2nd ed, Wiley series in probability and statistics. New York: Wiley.

U.S. Department of Labor. "Occupational Safety and Health Administration. Freedom of Information Act - Archive of Frequently Requested OSHA Documents." Accessed August 26. https://www.osha.gov/as/opa/foia/archive-foia.html.

U.S. Department of Labor. Research Synthesis: Evidence on the Effects of OSHA Activities. Washington, DC: U.S. Department of Labor.

Viscusi, W. Kip. 1992. *Fatal Tradeoffs: Public and Private Responsibilities for Risk*. New York, NY: Oxford University Press.

Viscusi, W. Kip. 1993. "The Value of Risks to Life and Health." *Journal of Economic Literature* 31:1912-1946.

Viscusi, W. Kip, Joseph Emmett Harrington, and John M. Vernon. 2005. *Economics of Regulation and Antitrust*. Fourth ed. Cambridge, MA: MIT Press.

Weil, David. 1996. "If OSHA is So Bad, Why is Compliance so Good?" *The RAND Journal of Economics*:618-640.

Weil, David. 2001. "Assessing OSHA Performance: New Evidence from the Construction Industry." *Journal of Policy Analysis and Management* 20 (4):651-674.

Wong, Vivian C, Peter M Steiner, and Thomas D Cook. 2013. "Analyzing regression-discontinuity designs with multiple assignment variables a comparative study of four estimation methods." *Journal of Educational and Behavioral Statistics* 38 (2):107-141.

APPENDICES

Appendix A Enforcement and Data Instruments

High-Rate Letter (2011)

Dear_______:

Last year, the Occupational Safety and Health Administration (OSHA) surveyed employers to collect workplace injury and illness data. The Agency used these data to identify the approximately 14,600 workplaces with the highest Days Away from work, Restricted, or Transferred (DART) rates; your workplace was one of those identified. This means workers in your establishment are being injured at a higher rate than in most other businesses in the country.

I am writing you to indicate my concern about the high DART rate at your establishment and to identify ways that you can obtain assistance in addressing hazards in your workplace. OSHA recognizes that your elevated DART rate does not necessarily indicate a lack of interest in safety and health. Whatever the cause, a high rate is costly to your company in both personal and financial terms. In addition, you should be aware that OSHA may target up to 4500 general industry workplaces identified in the survey for inspection in the next year.

Over the years OSHA has found that many employers lack expertise in the field of workplace safety and health and welcome assistance by experts in this field. You may wish to consider hiring an outside safety and health consultant, talking with your insurance carrier, or contacting your state's workers' compensation agency for advice. Your workers can help identify hazards and find solutions. In addition, if you have a union at your site, please discuss with them how to reduce hazards in your workplace.
An excellent way for employers with 250 or fewer workers to address safety and health in their workplaces is to ask for assistance from OSHA's consultation program. This program is administered by a state agency and operated separately from OSHA's enforcement program. The service is free and confidential, and there are no fines even if problems are found. Designed for small employers, the consultation program can help you identify hazards in your workplace and find effective and economical solutions for eliminating or controlling them. In addition, the OSHA state consultant can assist you in developing and implementing a safety and health management system for your workplace.

In your state, the OSHA consultation program may be contacted at:
Name, Project Manager
Name of Organization
Address - first line
Address - second line
Telephone number

I encourage you to consider these suggestions as well as visit OSHA's home page at www.osha.gov for information to help you ensure safe and healthful working conditions in your establishment. As it was last year, a list of all the employers receiving this letter will be available from the OSHA home page.

Sincerely,
David Michaels, PhD, MPH
Assistant Secretary

ODI Survey Form

OSHA Work-Related Injury and Illness Data Collection Form, 2011

U.S. Department of Labor
Occupational Safety and Health Administration

OMB No. 1218-0209
Approval Expires 4/30/2013
OSHA Form 196B
(1/2012)

Place Label Here

Please make any necessary corrections to your establishment site address, SIC, and NAICS.

To submit your data electronically, access our electronic survey at http://www.osha.gov/form196/collection.htm

Public Law 91-596 requires you to participate in the data initiative collection.

OSHA estimates that it will take you, on average, 10 minutes to complete the forms in this data collection, including the time you'll spend reviewing the instructions, searching and gathering the data needed, and completing and reviewing the collection of information. Persons are not required to respond to the collection of information unless it displays a currently valid OMB control number. If you have any comments regarding these estimates or any other aspects of this data collection, send them to:

U.S. Department of Labor
Occupational Safety and Health Administration
Directorate of Evaluation and Analysis
Office of Statistical Analysis
Room N-3644
200 Constitution Ave. N.W.
Washington, D.C. 20210

DO NOT SEND THE COMPLETED FORM TO THIS ADDRESS. Send the completed form to the Return Address indicated on the mailing label.

Dear Employer:

The U.S. Department of Labor, Occupational Safety and Health Administration (OSHA) is working with State agencies to compile work-related injury and illness data from employers within specific industry and employment size specifications. The information will be used to focus OSHA activities (inspections, outreach, consultations, technical assistance, and leveraging programs) and to measure the performance of the Agency in meeting its goal of reducing workplace injuries and illnesses.

We are asking for the totals from your 2011 ***Summary of Work-Related Injuries and Illnesses*** (OSHA No. 300A), as well as information about hours worked and employment at your establishment. The Occupational Safety and Health Act, 29 U.S.C. §§ 657 & 673, and reporting regulations at 29 C.F.R. Part 1904 authorize OSHA to collect the requested information. Please note that establishments that fail to submit a completed 2011 survey may be subject to OSHA enforcement actions, including the issuance of a citation and assessment of penalties.

> At this time the Bureau of Labor Statistics (BLS) and its State partners are conducting the 2011 Survey of Occupational Injuries and Illnesses, Part 1 of which solicits information very much like what OSHA is collecting. Be aware that employers who receive the BLS survey as well as the OSHA data collection form are required by law to respond to both of them, since these are separate data collection efforts. **However, if you have already received the BLS survey, OSHA affords you an option intended to streamline the effort involved in responding to both collections: That is, you may either (1) complete the OSHA form in its entirety, in addition to the BLS form, or (2) simply send OSHA a copy of your responses to the BLS survey (Parts 1A and 1B), which OSHA will accept as your response to the Agency's collection.**

We recognize that responding to our questions may be time consuming for some employers and have made every effort to reduce the completion time while still obtaining the necessary information. In this spirit, we now provide two means of submitting your establishment information: (1) by mail or fax, using this hard-copy form, or (2) via the Internet, using a secure electronic version of this form available on our Web site. Instructions for use of the electronic form are displayed at the Web site. You can access an electronic survey form by pointing your browser to **http://www.osha.[illegible]form1[illegible]/collection.htm** and then when prompted, inputting your establishment-specific ID number and password (provided in the label on the cover of this form). If you choose this option, use your browser's print function to print a copy for your records. If you need help completing the enclosed survey form or if you have questions, please call the phone number printed on the cover.

OSHA has initiated a comprehensive approach to monitoring and improving data quality. As part of this approach, OSHA will audit the injury and illness records of a randomly chosen sample of establishments included in this data collection. We will continue to evaluate this initiative and will build on the lessons learned to improve OSHA's ability to protect the health and safety of America's workers sensibly and appropriately. We invite your comments as we proceed with this effort. Thank you for helping us collect accurate information and for participating in the effort to make America's workplaces safer and healthier.

Occupational Safety and Health Administration
U.S. Department of Labor

Who must complete this form?

All establishments that receive this form should complete and return it or respond via the Internet **within 30 days**, even if they had no work-related injuries and illnesses recorded on their 2011 OSHA No. 300.

What else do you need?

- Information from your 2011 ***Summary of Work-Related Injuries and Illnesses*** (OSHA No. 300A).

What do you need to do?

- Check the address information printed on the cover. Make any corrections necessary on the hardcopy or Web site.
- Complete this form **only** for the establishment noted on the cover.
- Complete pages 3 and 4. You can either photocopy your OSHA Form 300A or you can transcribe the entries from your OSHA Form 300A to this survey form.
- On the last page, fill in the name of the person we should call with questions and sign the form.
- Return this form in the enclosed envelope, fax, or respond via Internet within **30 days** of the date your establishment received it.

2

Establishment Information

Using your completed calendar year 2011 *Summary of Work-Related Injuries and Illnesses (OSHA Form 300A)*, copy the establishment information into the boxes below. If these numbers are not available on your OSHA Form 300A, or if your establishment does not keep records needed to answer (1) and (2) below, you can estimate using the steps that follow.

1. For the reporting site identified on the cover: Enter the annual average employment for 2011. (You can copy this from your OSHA Form 300A.)

[ ]

Annual average number of employees for 2011

If needed: Steps to estimate employment

STEP 1: Add the number of employees your establishment paid in every pay period during 2011. **Include all employees:** full-time, part-time, temporary, seasonal, salaried, and hourly.

Acme Construction pays its employees 26 times each year. During 2011,

In this pay period	Acme paid this many employees
1	30
2	0
3	35
↓	↓
25	36
26	32
	830 (sum)

STEP 2: Divide the sum by the number of pay periods your establishment paid in 2011. **Include** any pay periods when you had **no** employees.

Because Acme has 26 pay periods, Acme would divide its sum by 26.
830 divided by 26 = 31.92

STEP 3: Round the answer to the next highest whole number. Write the rounded number in the box marked *Annual average number of employees.*

Acme would round 31.92 to 32 and write that number in the box marked *Annual average number of employees.*

2. For the reporting site identified on the cover: Enter the total hours worked for 2011. (You can copy this from your OSHA Form 300A.)

[ ]

Total hours worked by all employees in 2011

Note: *Total Hours Worked* should exclude vacation, sick leave, holidays, and other non-work time.

If needed: Steps to estimate total hours worked

STEP 1: Find the number of full-time employees in your establishment for 2011.

ABC Company had 15 full-time employees during 2011.

STEP 2: Multiply this number by the number of hours worked for a full-time employee in a year. This is equal to the number of full-time hours worked:

ABC Company's 15 full-time employees worked an average of about 1,760 hours each, after excluding vacation, sick leave, holidays, and other non-work time. *(The hours worked for a full-time employee in a year may be different at your reporting site)*

15 (full-time employees) times 1,760 (hours worked by a full-time employee in a year) equals 26,400 full-time hours.

STEP 3: Add the number of any overtime hours and the number of hours worked by other employees (part-time, temporary, seasonal) to the amount in Step 2:

ABC Company's full time employees worked a total of 1,500 hours of overtime. In addition, 3 part time employees worked a total of 2,715 hours during 2011. Adding these hours to those from Step 2:

Full-time hours from Step 2		26,400
Overtime hours	+	1,500
Part-time hours	+	2,715
Total hours worked by all employees in 2011	=	30,615

3. Check any conditions that might have affected your annual average number of employees or total hours worked during 2011:

- ❑ Strike or lockout
- ❑ Shutdown or layoff
- ❑ Seasonal work
- ❑ Natural disaster or adverse weather conditions
- ❑ Shorter work schedules or fewer pay periods than usual
- ❑ Longer work schedules or more pay periods than usual
- ❑ Other reason: ____________________
- ❑ Nothing unusual happened to affect our employment or hours figures

3

Did you have ANY occupational injuries or illnesses during 2011?

- ❑ Yes. Go to the next section, *Summary of Work-Related Injuries and Illnesses, 2011.*
- ❑ No. Go to *Sign and return this form* below.

Summary of Work-Related Injuries and Illnesses, 2011

Using your completed calendar year 2011 *Summary of Work-Related Injuries and Illnesses (OSHA Form 300A)*:

1. Copy the establishment summary information into the spaces below.
2. If you prefer, you may enclose a photocopy of your *Summary of Work-Related Injuries and Illnesses (OSHA Form 300A).*
3. If any total is zero on your OSHA Form 300A, write "0" in that total's space below.

Number of Cases

Copy these totals from columns (G), (H), (I), and (J):	Total number of deaths (column G)	Total number of cases with days away from work (column H)	Total number of cases with job transfer or restriction (column I)	Total number of other recordable cases (column J)
	________	________	________	________

Number of Days

Copy these totals from columns (K) and (L):	Total number of days away from work (column K)	Total number of days of job transfer or restriction (column L)
	________	________

Injury and Illness Types

Total number of . . . from column (M)

(1) Injuries	________	(4) Poisonings	________
(2) Skin disorders	________	(5) Hearing loss	________
(3) Respiratory conditions	________	(6) All other illnesses	________

Sign and return this form

Fill in the name, title, phone number and fax number of the person we should call with questions about this form. Then sign and date the form.

Printed name	( ) Telephone number	Ext.	( ) Fax number	E-mail address (optional)
________	________		________	________

Signature	Title	Today's date
________	________	________

Use the envelope included with this packet to mail the original form to us. If the return envelope is missing, send the package to the Return Address on the front cover. Remember to keep a photocopy for your records.

4

OSHA Form 300A

OSHA's Form 300A (Rev. 01/2004)

Summary of Work-Related Injuries and Illnesses

Year 20__ __

U.S. Department of Labor
Occupational Safety and Health Administration

Form approved OMB no. 1218-0176

All establishments covered by Part 1904 must complete this Summary page, even if no work-related injuries or illnesses occurred during the year. Remember to review the Log to verify that the entries are complete and accurate before completing this summary.

Using the Log, count the individual entries you made for each category. Then write the totals below, making sure you've added the entries from every page of the Log. If you had no cases, write "0."

Employees, former employees, and their representatives have the right to review the OSHA Form 300 in its entirety. They also have limited access to the OSHA Form 301 or its equivalent. See 29 CFR Part 1904.35, in OSHA's recordkeeping rule, for further details on the access provisions for these forms.

Number of Cases

Total number of deaths	Total number of cases with days away from work	Total number of cases with job transfer or restriction	Total number of other recordable cases
______	______	______	______
(G)	(H)	(I)	(J)

Number of Days

Total number of days away from work	Total number of days of job transfer or restriction
______	______
(K)	(L)

Injury and Illness Types

Total number of . . .
(M)

(1) Injuries	______	(4) Poisonings	______
(2) Skin disorders	______	(5) Hearing loss	______
(3) Respiratory conditions	______	(6) All other illnesses	______

Post this Summary page from February 1 to April 30 of the year following the year covered by the form.

Public reporting burden for this collection of information is estimated to average 58 minutes per response, including time to review the instructions, search and gather the data needed, and complete and review the collection of information. Persons are not required to respond to the collection of information unless it displays a currently valid OMB control number. If you have any comments about these estimates or any other aspects of this data collection, contact: US Department of Labor, OSHA Office of Statistical Analysis, Room N-3644, 200 Constitution Avenue, NW, Washington, DC 20210. Do not send the completed forms to this office.

Establishment information

Your establishment name ______________________

Street ______________________

City ______________ State _____ ZIP _______

Industry description (*e.g., Manufacture of motor truck trailers*)
____

Standard Industrial Classification (SIC), if known (*e.g., 3715*)
___ ___ ___ ___

OR

North American Industrial Classification (NAICS), if known (e.g., 336212)
___ ___ ___ ___ ___ ___

Employment information

(*If you don't have these figures, see the Worksheet on the back of this page to estimate.*)

Annual average number of employees __________

Total hours worked by all employees last year __________

Sign here

Knowingly falsifying this document may result in a fine.

I certify that I have examined this document and that to the best of my knowledge the entries are true, accurate, and complete.

______________________ Company executive ______ Title

( ) - ______________ Phone ______ / / Date

Appendix B Theories of Compliance

Noncompliance as well as compliance entails added costs for many employers. Upon inspection, OSHA sometimes imposes penalties on violators of its standards, in addition to sometimes issuing orders to abate specific workplace hazards. An immediate impact of unsafe workplaces and workplace injuries/illnesses is that safety and health-related incidents render some workers unable to work, or unable to work as much, reducing productivity. Replacement workers for the injured may be less experienced and thus less productive. High rates of injury/illness also may lead to higher rates of turnover, and the average new worker will also be less experienced and less productive. Workers might have higher wage demands and receive compensating wage differentials in exchange for working in unsafe conditions.[41] The employer takes all of these factors into consideration in assessing its expected cost of current hazard and associated injury/illness rates. The expected cost of compliance, on the other hand, depends on the cost of abating hazards.

Changes in an employer's expected cost of noncompliance can be brought about by changes in its perceived probability of being inspected, changes in its perceived average number of violations, and changes in its knowledge of average penalty per violation. Alternatively, changes in an employer's expected cost of compliance can be brought about by recommendations the employer receives about its safety practices and the information it receives from OSHA's free consultation program. Both of these types of information have a downward effect on the cost of achieving compliance with OSHA standards. If an employer's expected costs of noncompliance increase while its expected costs of compliance stay the same or decrease, the employer should become more compliant.

Several researchers have reported that actual levels of worksite compliance with OSHA regulations are considerably higher than one would expect, given the levels of OSHA inspection probabilities and monetary penalties that explain employer behavior according to a strict short-term Becker-Stigler enforcement model.[42] For example, Weil contrasts two popular images of OSHA: the "toothless tiger" whose budget is sufficient to inspect the average workplace only once a century, and the "ominous ogre" forcing expensive compliance costs on often struggling businesses.[43] Weil concludes that each image is partly accurate. Data indicate that the probability of inspection is very low, that very few of the limited number of inspections each year are re-inspections of violators, and that monetary penalties are quite modest – the average penalty in 2010 was $1,000.[44] (2010) Nonetheless, compliance with specific OSHA standards in the woodworking industry Weil studied was quite high. Weil concluded that average hazard abatement costs in his sample were far above expected costs of noncompliance, and that the Becker-Stigler regulatory model does not sufficiently explain OSHA compliance. He speculated that substantial risk aversion by employers was a possible explanation for the failure of the risk-neutral Becker-Stigler model. Viscusi came to similar conclusions using a broader sample of industries.[45]

Findings from these kinds of empirical studies are crucial for OSHA enforcement strategy. If factors other than short-term risk-neutral monetary costs and monetary outcomes largely explain employer compliance behavior, then it is possible that carefully timed and worded messages from OSHA, combined with inspections and threats of inspections, could improve employer compliance and thus health and safety outcomes for workers.

---

[41] (Viscusi 1993)
[42] (Mendeloff and Gray 2005)
[43] (Weil 1996)
[44] (2010)
[45] (Viscusi 1992)

To reconcile empirical findings of over-compliance compared to the level of compliance predicted by the short-term Becker-Stigler model, we consider a game theoretic model of compliance proposed by Harrington.[46] Harrington recognizes that OSHA enforcement and worksite compliance is a repeated game, not just a short-term one-time occurrence. Harrington argues that, if the regulator's strategy tends to concentrate long-run inspection resources on non-cooperators, then profit maximization could lead site managers to over-comply in the short run, to brand themselves as cooperators, and to avoid high-priority subsequent enforcement. In this framework, high-rate letters might help spur compliance in the short run.

The impact of the SST program on incidence of workplace safety/health is posited to work both directly and indirectly: directly by creating incentives for improved workplace safety/health, and indirectly through follow-on effects of improved compliance. First, because SST targets worksites with high injury/illness rates, SST in effect increases the costs of such high injury/illness rates, and, as a result, worksites also have an incentive to become safer, at least according to the DART and DAFWII case rate metrics that SST uses. OSHA also makes public worksites' high injury/illness rates, which could create costs from increased wage demands and similar activity. Second, workplace safety/health may improve with improvements in regulatory compliance. Past non-experimental controlled before/after and quasi-experimental studies have found that inspections can reduce injuries.[47,48] However, other studies have shown ambiguous or mixed results about such a connection.[49]

Letters

OSHA sends high-rate letters to selected worksites to inform them that they have particularly high injury/illness rates, to warn them that OSHA inspections may be forthcoming, to note potential benefits of improving workplace safety and health conditions and costs of not doing so, and to give several sources that employers can contact to help them assess and abate their hazards.

High-rate letters are in part a threat of inspection. As such, their effectiveness is predicated in part on levels of actual inspections, and may reinforce the effect of actual inspections by promoting awareness of the credible threat of inspections to come. OSHA sends about 15,000 letters each year, and conducts between 2,500 to 3,000 SST inspections each year.[50] (U.S. Department of Labor) Therefore, the probability that a worksite that receives a letter will also receive an inspection is between 17% and 20%. These figures are public information and included in the letters themselves, so employers can easily calculate this themselves. This information may increase their perceived probability of inspection, with consequent effects on their levels of compliance. This would lead to the "awareness of possibility of inspection" output under "information-related outputs", and potentially also "voluntary cleanup efforts".

In addition to credible threats of inspection, high-rate letters offer information to employers who might not have realized their worksites are relatively unsafe, and tell employers where they can get help in assessing and correcting hazardous conditions. Thus high-rate letters combine the stick of credible threats of inspection with the carrot of information and offers of assistance, to help nudge worksites toward compliance. This may lead to "improved knowledge" under 'information-related outputs'.

In particular, the letters include information about OSHA's free and confidential On-site Consultation Program and a list of other organizations that the employer can contact (e.g. state workers' compensation agencies, insurance companies, and outside safety consultants). To the extent that the employers were not

[46] (Harrington 1988)
[47] (Gray and Scholz 1991)
[48] (Levine, Toffel, and Johnson 2012)
[49] (Mischke et al. 2013)
[50] (U.S. Department of Labor)

aware of the available resources for improving safety that are noted in the letter, the information given in the letters may lower the expected cost of compliance for these worksites. This is represented in Figure 1 by the “request outside consultation” output.

Although OSHA usually sends out only one letter, high-rate worksites subsequently may receive additional safety-related messages from other entities. The online list of identities of OSHA high-rate-letter recipients provides safety consultants and equipment vendors with pre-screened lists of possibly highly-motivated sales prospects. High-rate letters could trigger possibly thousands of marketing messages and sales calls from the safety industry, multiplying OSHA's messaging efforts many times. In addition, industry associations, unions, and other groups might target workplaces on the list for action, such as union organizing efforts, which might help pressure worksite managers to invest in hazard abatement.

Inspections

Inspections impose direct costs on employers in the form of penalties for violations of OSHA’s workplace safety and health standards. OSHA also orders some employers to abate hazards, and imposes fines for failing to abate the hazards identified. Penalties and hazard abatement orders increase the costs of noncompliance, giving employers incentives to comply with OSHA standards. These are represented in Figure 1 by the outputs “pay monetary penalties” and “abate hazards.” To the extent that the posited positive relationship between violations and the incidence of injury or illness holds, increased compliance also would lead to reductions in injuries and illnesses.

In addition, inspections inform worksites of conditions that violate OSHA workplace safety and health standards. Such information may be costly to gather; thus, the provision of such worksite-specific information to the employer can decrease the expected cost of compliance. Second, like letters, OSHA inspection results are public information. OSHA itself publicizes enforcement actions and hazards found at worksites through press releases on their website. Therefore, inspection results can lead to increased costs of noncompliance through possible impacts on workers’ compensation insurance coverage and reputations for dangerous workplace conditions.

Many studies have reported positive empirical relationships between inspections – especially inspections that impose monetary penalties for violations – and subsequent compliance with OSHA regulations. These are straightforward implications of the conventional model of regulatory compliance: costs imposed by OSHA drive employers to increase compliance with OSHA standards.

However, the policy context for OSHA inspections is that they are rather expensive (the average cost in 2014 was about $5,500) and thus resources for them have been very limited for many years.

Appendix C RCT study limitations

RCT - lack of letter treatment implementation information

Uncertainty about treatment implementation in the letter treatment study stemmed from a lack of information on which worksites actually received the letters (high-rate letter lists indicate which worksites are supposed to get letters, but the actual return receipts of those letters are not available). It was therefore impossible to distinguish sites that were actually treated from those sites that were intended to be treated.

The treatment variable is the indicator of belonging to the treatment group, not of receiving a high-rate letter. Therefore, the RCT models estimate the differences between study groups, not between high-rate letter recipients and non-recipients. If study crossover occurred, meaning some treatment group sites did not receive a letter, and/or if control group sites did receive a letter, the study cannot account for this. The presence of excess crossover results increases the likelihood of an inconclusive comparison of treatment and control data. Unfortunately, the level of study crossover cannot be determined from the data.

Avenues of the high-rate letter impact

Re-examining the theory of change helps us understand how the publication of control group sites' names on the DOL website reduces the chance for capturing the impact of high-rate letters. The actual receipt of a high-rate letter, along with contact from third parties, encourages worksites to improve compliance. This contact results from DOL publishing the list of worksites that are high-rate letter recipients. The theory of change of the letter treatment (see 0) separates out a number of avenues by which high-rate letters could impact regulatory compliance of worksites. Below we examine if and how these could explain why we found no significant effects.

First, a threat of inspection could induce compliance. Employers may fear that inspectors will cite them for violations, which often come with an attached penalty, leading them to improve the safety of their workplace. Improved safety reduces the probability of a violation. As such, the effectiveness of letters is predicated in part on levels of actual inspections and previous experiences. The number of high-rate letters that OSHA sends and the number of SST inspections is public information. The high-rate letters also indicate the number of inspections conducted in total. Based on that information, all worksites can calculate the probability that they will be inspected. Threat of inspection will probably only reduce the probability of a violation if the lack of safety is new information to the worksite (see below).

Second, high-rate letters offer information to employers who might not have realized their worksites are relatively unsafe. Those to whom lack of safety is new information may change their behavior and reduce the probability of violation. However, a relatively small proportion of worksites had not received a high-rate letter before the study. As Table 33 presents, the number of worksites that received their first high-rate letter in 2011 ranges from 55 of 211 (26%) to 156 of 420 (37%). Thus, for most firms in the sample, this was not their first high-rate letter. Furthermore, it is safe to assume that not all of these worksites were entirely unaware of their own safety problems, so the effect of the impact of receiving a letter is probably small.

Table 33. Sites published on the high-rate letter list prior to 2011

Treatment	SST List	Group	Not listed on high-rate letter list prior to 2011	Listed on high-rate letter list prior to 2011
Letter	Primary	Treatment	148 (35.2%)	272 (64.8%)
		Control	58 (28.4%)	146 (71.6%)
	Secondary	Treatment	114 (27.1%)	306 (72.9%)
		Control	55 (26.1%)	156 (73.9%)
	Total	**Treatment**	**262 (31.2%)**	**578 (68.8%)**
		Control	**113 (27.2%)**	**302 (72.8%)**
Letter Plus Inspection	Primary	Treatment	156 (37.1%)	264 (62.9%)
		Control	70 (32.4%)	146 (67.6%)
	Secondary	Treatment	114 (27.1%)	306 (72.9%)
		Control	56 (26.8%)	153 (73.2%)
	Total	**Treatment**	**270 (32.1%)**	**570 (67.9%)**
		Control	**126 (29.6%)**	**299 (70.4%)**

High-rate letters also tell employers about OSHA's free consultation program, where they can get help in assessing and correcting hazardous conditions (Juras et al. 2015). It can be safely assumed that the participation in such a consultation program will reduce the probability of post-consultation violations. However, the number of sites participating in voluntary programs is negligible.

Finally, OSHA high-rate letter recipients are published online each year.[51] Several other sources post articles online when the letters are published; some include advice or suggest sources for advice to letter recipients. These sources include safety journals and newsletters, such as the *EHS Journal: Practical Solutions for Environmental, Health and Safety Professionals*[52] and J.J. Keller's *OSHA Safety Training Newsletter*.[53] Safety consultants and equipment vendors can access this list as a source for leads for sales. These consultants and vendors may contact the worksites with marketing messages and sales calls, reiterating OSHA's message to the worksite to increase compliance. Additionally, unions and other industry associations may contact the worksite to encourage union organizing efforts. Contact from these third-party entities could encourage employers to invest more in safety at their worksite. Unfortunately, both treatment and control sites appeared on this public list, and therefore both received the marketing messages and sales calls, and were equally likely to be targeted by industry associations and unions to take action. Therefore this impact evaluation was not able to isolate one of the key avenues of effect and therefore the probability of detecting a difference between treatment and control groups was radically reduced. The only difference between the two groups is reduced to physically receiving the letter.

[51] The list of letter recipients can be accessed on OSHA's website: https://www.osha.gov/as/opa/foia/archive-foia.html
[52] Stacey Lucas. "U.S. OSHA Targets 15,000 Facilities with High Incident Rates." *EHS Journal*. April 4 2010.
[53] Judie Smithers. "OSHA notifies employers with high injury rates." *J.J. Keller's OSHA Safety Training Newsletter*. Vol. 19/No. 6, June 2012.

Of the avenues listed above, publication of worksites' names and the resulting sales calls, union and industry association initiatives seem to be of key importance in increasing regulatory compliance and thereby health and safety outcomes.

As seen in Table 34 and Table 35 below, between 85% and 88% of the treatment group worksites were published online as high-rate letter recipients, in line with the study design. However, between 80% and 91% of the control group worksites in both the letter only and the letter plus inspection treatment groups were also published online as high-rate letter recipients.

Table 34: Study Sites Listed on the High-Rate Letter List Online, by Treatment and Group

Treatment	Group	Total Number of Sites	Sites that were on the high-rate letter list in 2011		Sites that were not on the high-rate letter list in 2011	
			N	%	N	%
Letter	Control	415	347	84%	68	16%
	Treatment	840	726	86%	114	14%
Letter + Insp	Control	425	371	87%	54	13%
	Treatment	840	722	86%	118	14%

Table 35: Study Sites Listed on the High-Rate Letter List Online, by Treatment, List and Group

Treatment	List	Group	Total Number of Sites	Sites that were on the high-rate letter list in 2011		Sites that were not on the high-rate letter list in 2011	
				N	%	N	%
Letter	Primary	Control	204	163	80%	41	20%
		Treatment	420	356	85%	64	15%
	Secondary	Control	211	184	87%	27	13%
		Treatment	420	370	88%	50	12%
Letter + Insp	Primary	Control	216	196	91%	20	9%
		Treatment	420	363	86%	57	14%
	Secondary	Control	209	175	84%	34	16%
		Treatment	420	359	85%	61	15%

Reasons for Sample attrition for regulatory compliance outcomes

Many of the worksites that were lost for the regulatory compliance outcome assessment of the study (lacking outcomes) had some data in the inspection records. These data identified the reasons why post-2011 and post-2012 inspections could not be found for the letter and letter plus inspection studies, respectively[54].

[54]The scope of each inspection is coded in the data. Inspection scopes can be: complete, partial, records only, and no inspection. "No inspection" scope inspections after 2011/2012 were considered non-respondents, because they would not provide measures of regulatory compliance outcomes.

Note that reasons for non-response, summarized in Table 36, are described in a cascading order: The number of sites with previous inspections was calculated from non-respondent sites that, for example, did not did not have "Out of business/Site not found" as an explanation:

1. **Out of business/Site not found:** These are sites where the inspector could not find the worksite or it was out of business when he/she arrived. This is considered natural attrition.
2. **Previous inspection:** Once inspected, worksites are exempt from being inspected for the following 24 months. Many of the worksites in the study were inspected in the two years prior (2010 or 2011 for the letter study, 2011 or 2012 for the letter plus inspection study), which made them exempt from inspections. This is understandable since some inspections are unavoidable such as the inspections conducted due to accidents. However, a large majority of these inspections were recorded as SST inspections which–with careful implementation planning by IMPAQ International–should not have happened. Follow-up inspections should have taken place one to two years after the treatment. The prevalence of previous inspections explains a large proportion of non-response.
3. **Voluntary program:** These are worksites that participated in voluntary consultation programs provided by OSHA and as such are also exempt from being inspected for that period and for the following 24 months. Their occurrence is negligible.
4. **Ten or fewer employees:** Worksites in certain industries that have 10 or fewer employees are exempt from inspections. Occurrence of such sites is very low.[55]
5. **Other:** This category in the table covers worksites that had a "no inspection" scope inspection, and had "other" as a reason in the "why no inspection" data field. No further explanation is provided in the publicly available inspections dataset for why a "no inspection" scope inspection was conducted. This category also includes worksites that had "SIC not on PG" and "Non-Exempt Consult" as the reason for why no inspection occurred. The prevalence of such worksites is relatively low.
6. **No inspection found:** These are worksites for which post-2011 inspections could not be found in the records at all. Although we used extensive and thorough matching procedures to identify inspections for all worksites, it is possible (although unlikely[56]) that these inspections exist, but were not found by the matching procedure. Most or all of these sites could in fact have not received inspections at all after 2011 or 2012.

[55] To be included in the study, the worksite had to start out with 20 or more employees. The worksites would have had to drop to 10 or fewer employees between inclusion in the study and the follow-up inspection.

[56] The matching procedure found ODI records for each of the worksites in the study which indicates a fairly high quality matching procedure. Although the procedure did not find post-2011 inspection records for all worksites, given the high quality of the procedure, we can be reasonably sure that most of these inspection records were not entered into OIS.

Table 36: Reasons for Non-response

SST Study	Group	Total	Out of Business/ Site not Found	Previous Inspections		Voluntary Program	Ten or Fewer Employees	Other	No Inspection Found
				SST	Non-SST				
Letter	Treatment	187	15.9%	34.4%	7.4%	1.6%	1.6%	6.9%	32.3%
	Control	70	27.9%	25.0%	5.9%	0.0%	5.9%	4.4%	30.9%
	TOTAL	257	19.1%	31.9%	7.0%	1.2%	2.7%	6.2%	31.9%
Letter Plus Inspection	Treatment	226	21.8%	59.4 %	5.7%	0.4%	0.9%	10.0%	1.7%
	Control	235	6.9%	6.9%	6.9%	0.0%	1.3%	1.3%	76.7%
	TOTAL	461	14.3%	33.0%	6.3%	0.2%	1.1%	5.6%	39.5%

Appendix D Exploratory Analyses - Impacts on Additional Outcomes

Table 37. Effect of treatment on additional outcomes, Primary worksites, RCT

Primary	Letter				Letter + Insp			
	Estimates and CI		S.E.	Control/Trt. Means	Estimates and CI		S.E.	Control/Trt. Means
Number of Violations After Treatment	0.21 [-0.87,1.29]		0.55	C: 5.14 T: 5.36	-1.71 [-3.03,-0.39]	*	0.67	C: 5.10 T: 3.39
Number of Other-than-SeriousViolations After Treatment	0.02 [-0.32,0.37]		0.17	C: 1.28 T: 1.30	-0.48 [-0.93,-0.03]	*	0.23	C: 1.36 T: 0.88
Other-than-Serious Violation After Treatment (0/1)	-0.03 [-0.12,0.07]		0.05	C: 0.55 T: 0.52	-0.05 [-0.16,0.06]		0.06	C: 0.47 T: 0.42
Number of Serious Violations After Treatment	0.19 [-0.71,1.09]		0.46	C: 3.87 T: 4.06	-1.23 [-2.30,-0.16]	*	0.54	C: 3.75 T: 2.52
Serious Violation After Treatment (0/1)	0.02 [-0.06,0.11]		0.04	C: 0.7 T: 0.72	-0.08 [-0.18,0.03]		0.05	C: 0.69 T: 0.61
Penalty Amount After Treatment	-$1,851.69 [-5189.65,1486.26]		$1,698.83	C: 9364.42 T: 7512.72	-$6,150.54 [-11730.67,-570.41]	*	$2,838.38	C: 12336.40 T: 6185.86
Penalty Received After Treatment (0/1)	0.03 [-0.05,0.12]		0.04	C: 0.72 T: 0.75	-0.07 [-0.17,0.03]		0.05	C: 0.73 T: 0.65
DAFWII Rate	0.76 [-0.12,1.65]		0.45	C: 3.16 T: 3.92	-0.19 [-1.38,1]		0.60	C: 3.35 T: 3.16
Rate of Days Away from Work	27.63 [-13.94,69.21]	*	21.13	C: 91.27 T: 118.9	-48.27 [-99.54,3.01]	*	26.04	C: 123.28 T: 75.01

* p<0.05

Table 38. Effect of treatment on additional outcomes, Secondary worksites, RCT

Secondary	Letter			Letter + Insp		
	Estimates and CI	S.E.	Control/Trt. Means	Estimates and CI	S.E.	Control/Trt. Means
Number of Violations After Treatment	0.45 [-0.52,1.43]	0.50	C: 3.46 T: 3.91	0.15 [-0.85,1.16]	0.51	C: 2.41 T: 2.56
Number of Other-than-Serious Violations After Treatment	0.23 [-0.04,0.51]	0.14	C: 0.96 T: 1.19	0.06 [-0.27,0.39]	0.17	C: 0.76 T: 0.82
Other-than-Serious Violation After Treatment (0/1)	0.08 [-0.02,0.17]	0.05	C: 0.46 T: 0.54	0.04 [-0.07,0.16]	0.06	C: 0.36 T: 0.41
Number of Serious Violations After Treatment	0.22 [-0.63,1.07]	0.43	C: 2.5 T: 2.72	0.09 [-0.84,1.03]	0.47	C: 1.65 T: 1.74
Serious Violation After Treatment (0/1)	0.03 [-0.05,0.12]	0.04	C: 0.54 T: 0.57	0.06 [-0.06,0.18]	0.06	C: 0.40 T: 0.46
Penalty Amount After Treatment	$1,779.87 [-1060.72,4620.45]	$1,445.72	C: 5089.92 T: 6869.79	-$242.17 [-3873.27,3388.93]	$1,846.80	C: 4923.09 T: 4680.92
Penalty Received After Treatment (0/1)	0.04 [-0.05,0.12]	0.04	C: 0.62 T: 0.66	0.05 [-0.08,0.17]	0.06	C: 0.47 T: 0.52
DAFWII Rate	0.04 [-0.65,0.72]	0.35	C: 2.80 T: 2.84	0.41 [-0.52,1.33]	0.47	C: 2.83 T: 3.23
Rate of Days Away from Work	26.70 * [-7.63,61.03]	17.45	C: 82.19 T: 108.89	14.01 * [-21.73,49.74]	18.16	C: 92.07 T: 106.08

* p<0.05

Appendix E Statistical Power and Minimum Detectable Impacts in the RDD study

ITT ANCOVA Sample Size Formula

To measure the effect of assignment to a program on one continuous outcome with a completely randomized two-group experiment, let

> *α=significance level (10%, two-tailed);*
> *β=1-statistical power (1-80%);*
> *c=the fraction of the sample which belongs to the control group;*
> *z(x)= the inverse of the cumulative standardized normal distribution function.*
> *Y_i= the continuous outcome variable;*
> *Z_i= a k-vector of covariates measured just before assignment to treatment or control, plus the running variable R;*
> *S_i= a dummy variable which is zero for those assigned to control status and unity for those assigned to treatment; and*
> *ρ= intraclass correlation for the outcome attributable to running variable R_i.*

In the fitted regression equation $\hat{y}_i = S_i\,\delta_0 + Z_i\,\delta_{1:k} + \delta_{k+1}$, which explains fraction $\overline{R^2}$ of sample outcome variance, the first coefficient is interpreted as the sample impact of assignment to the treatment. Its expected value is the population effect of treatment, δ_0.

An expression for its variance, derived in the Appendix of (Cave 1987) may be manipulated as outlined to yield the sample size formula:

$$n \geq 4\left(\frac{z(1-\alpha)+z(1-\beta)}{\delta_0}\right)^2 \frac{(1-\overline{R^2})Var(y)}{1-R_{sz}^2}\frac{1}{4c(1-c)}\frac{1}{(1-\rho)} \tag{7}$$

where *n* is a positive integer, Var(y) is the sample variance of the outcome, and R_{sz}^2 is the proportion of the variation in S_i explained by a regression of S_i on Z_i and a constant. R_{sz}^2 has expected value zero[57] if assignment to treatment truly is random; the multiplicative factor ($1/(1-R_{sz}^2)$) is a sample size increasing randomization design effect analogous to a survey design effect.

The multiplicative factor involving *c* is a sample split inflation factor which takes the value unity when *c*=0.5, and *n* is the total number of usable data points required for analysis.

The factor ($1 - \overline{R^2}$) in the numerator is a variance deflation factor reflecting the degree of success for baseline covariates in soaking up outcome variance.

An additional variance inflation factor comes into play if the units assigned to treatment or control status are aggregates of the units whose outcomes are to be explained. For the SST RDD analysis, such clustering may have come about if DART, DAFWII, or LWDII rates had been rounded to one decimal place rather than computed to several more places when prioritizing worksites for inspections.

[57] To ensure the internal validity of inferences about effects, it is important to test this hypothesis for every sample and subsample of complete data used in the analysis of a social experiment.

The factor (1/(1- ρ)) is an assignment variable intraclass correlation factor that reflects the degree to which worksites had exactly the same measured priority for inspection. If there were such "ties" in priorities, then prioritizing inspections involved groups of worksites rather than individual worksites. Such clustering raises required sample sizes to measure a given effect size. For example, if OSHA rounded DART, DAFWII and LWDII rates to one decimal place before assigning worksites to enforcement groups, that would results in "ties", clusters of worksites at each decimal place hence largely decreasing the power of RDD and increasing minimum detectable impacts (MDIs). However, we know that OSHA applies the DART rates calculated by the ODI survey instrument to high precision[58], and hence there will be no ties in the assignment variable and the (1/(1- ρ)) becomes one. In the tables below, we present both the ITT MDIs assuming rounding to one decimal place and the ITT MDI assuming using the high precision versions of the assignment variables.

Multiplying inequality (7) by $\frac{{\delta_0}^2}{n}$ and taking square roots of both sides, turns it into MDI formula. Whatever increased required sample size for the previous formula now increases MDI:

$$\delta_0 \geq \sqrt{\frac{4}{n}\left(z(1-\alpha)+z(1-\beta)\right)^2 \frac{(1-\overline{R^2})Var(y)}{1-R_{sz}^2} \frac{1}{4c(1-c)} \frac{1}{(1-\rho)}} \tag{8}$$

TOT ANCOVA Minimum Detectable Impact Calculations

The TOT MDI formula is exactly the same as the ITT MDI formula, with one substitution:

Instead of the assignment dummy S_i, the predicted treatment variable $\hat{T}_i$ appears on the right hand side of the estimating equation.

$$\delta_0 \geq \sqrt{\frac{4}{n}\left(z(1-\alpha)+z(1-\beta)\right)^2 \frac{(1-\overline{R^2})Var(y)}{1-R_{\hat{T}z}^2} \frac{1}{4c(1-c)} \frac{1}{(1-\rho)}} \tag{9}$$

The sample split inflation factor remains the same as for the ITT MDI calculations, but the other factors

must be recalculated. The sample size increasing randomization design effect: $\frac{1}{1-R_{\hat{T}z}^2}$, and the variance inflation factor: (1 – $\overline{R^2}$) refer to a different regression equation.

[58] OSHA informed Summit about its non-rounding practice in an email on Monday, July 27.

Appendix F Exploring Baseline Differences in Covariates

The tables below present results of baseline comparisons between treatment and control groups. All tables show comparisons between the worksites outcomes were available for (the ones that remained in the analysis dataset), weighted after sample attrition adjustments.

Table 39. Baseline differences in covariates, RCT, letter treatment, all worksites, regulatory compliance outcomes

Letter-All-Regulatory Compliance	Coefficient	Standard Error
Industry		
Other	0.04	(0.03)
Mining, manufacturing	0.00	(.)
Region		
Boston, New York, Philadelphia, Atlanta	0.00	(.)
Chicago	0.03	(0.04)
Dallas, Kansas City, Denver, Seattle, San Francisco	-0.04	(0.04)
Employees		
0-99	0.00	(.)
100 or more	-0.03	(0.03)
Penalty Inspection Combinations		
No inspection, no penalty	0.00	(.)
Inspection, but no penalty	-0.05	(0.05)
Inspection and penalty	0.04	(0.04)
DART Rate		
0-9.99	0.00	(.)
10-19.99	0.091*	(0.03)
20 or more	-0.13	(0.08)
High Rate Letters		
None	0.05	(0.04)
1 to 5	0.00	(.)
6 or more	-0.02	(0.04)
Constant	0.62*	(0.04)
Observations	998	

Standard errors in parentheses, * $p < 0.05$

Table 40. Baseline differences in covariates, RCT, letter plus inspection treatment, all worksites, regulatory compliance outcomes

Letter + Insp- All - Regulatory Compliance	Coefficient	Standard Error
Industry		
Other	-0.03	(0.04)
Mining, manufacturing	0.00	(.)
Region		
Boston, New York, Philadelphia, Atlanta	0.00	(.)
Chicago	-0.02	(0.05)
Dallas, Kansas City, Denver, Seattle, San Francisco	-0.03	(0.05)
Employees		
0-99	0.00	(.)
100 or more	0.03	(0.05)
Penalty Inspection Combinations		
No inspection, no penalty	0.00	(.)
Inspection, but no penalty	-0.04	(0.06)
Inspection and penalty	-0.06	(0.05)
DART Rate		
0-9.99	0.00	(.)
10-19.99	0.02	(0.04)
20 or more	-0.03	(0.09)
High Rate Letters		
None	0.01	(0.05)
1 to 5	0.00	(.)
6 or more	0.01	(0.06)
Constant	0.70*	(0.05)
Observations	804	

Standard errors in parentheses, * $p < 0.05$

Table 41. Baseline differences in covariates, RCT, letter treatment, all worksites, health and safety outcomes

Letter - All - Health and Safety	Coefficient	Standard Error
Industry		
Other	0.00	(.)
Mining, manufacturing	-0.05	(0.04)
Region		
Boston, New York, Philadelphia, Atlanta	0.00	(.)
Chicago	0.03	(0.04)
Dallas, Kansas City, Denver, Seattle, San Francisco	-0.05	(0.05)
Employees		
0-99	0.00	(.)
100 or more	-0.02	(0.04)
Penalty Inspection Combinations		
No inspection, no penalty	0.00	(.)
Inspection, but no penalty	-0.02	(0.06)
Inspection and penalty	0.05	(0.04)
DART Rate		
0-9.99	0.00	(.)
10-19.99	0.05	(0.04)
20 or more	-0.14	(0.09)
High Rate Letters		
None	0.04	(0.04)
1 to 5	0.00	(.)
6 or more	-0.02	(0.05)
Constant	0.67*	(0.05)
Observations	703	

Standard errors in parentheses, * $p < 0.05$

Table 42. Baseline differences in covariates, RCT, letter plus inspection treatment, all worksites, health and safety outcomes

Letter + Insp- All - Health and Safety	Coefficient	Standard Error
Industry		
Other	0.01	(0.05)
Mining, manufacturing	0.00	(.)
Region		
Boston, New York, Philadelphia, Atlanta	0.00	(.)
Chicago	-0.05	(0.05)
Dallas, Kansas City, Denver, Seattle, San Francisco	-0.04	(0.06)
Employees		
0-99	0.00	(.)
100 or more	-0.01	(0.06)
Penalty Inspection Combinations		
No inspection, no penalty	0.00	(.)
Inspection, but no penalty	-0.07	(0.08)
Inspection and penalty	-0.09	(0.05)
DART Rate		
0-9.99	0.00	(.)
10-19.99	0.03	(0.05)
20 or more	-0.06	(0.12)
High Rate Letters		
None	-0.04	(0.06)
1 to 5	0.00	(.)
6 or more	-0.05	(0.07)
Constant	0.74*	(0.06)
Observations	572	

Standard errors in parentheses, * $p < 0.05$

Table 43. Baseline differences in covariates, RCT, letter treatment, primary worksites, health and safety outcomes

Letter-Primary-Health and Safety	Coefficient	Standard Error
Industry		
Other	-0.0807	(0.08)
Mining, manufacturing	0	(.)
Region		
Boston, New York, Philadelphia, Atlanta	0	(.)
Chicago	0.0256	(0.06)
Dallas, Kansas City, Denver, Seattle, San Francisco	0.0262	(0.07)
Employees		
0-99	0	(.)
100 or more	-0.109	(0.06)
Penalty Inspection Combinations		
No inspection, no penalty	-0.103	(0.06)
Inspection, but no penalty	-0.079	(0.09)
Inspection and penalty	0	(.)
DART Rate		
0-9.99	0	(.)
10-19.99	0.0448	(0.06)
20 or more	-0.0899	(0.10)
High-rate letters		
None	0.0771	(0.06)
1 to 5	0	(.)
6 or more	0.0783	(0.07)
Constant	0.709*	(0.06)
Observations	345	

Standard errors in parentheses, * $p < 0.05$

Table 44. Baseline differences in covariates, RCT, letter plus inspection treatment, primary worksites, health and safety outcomes

Letter + Insp- Primary - Health and Safety	Coefficient	Standard Error
Industry		
Other	-0.113	(0.10)
Mining, manufacturing	0	(.)
Region		
Boston, New York, Philadelphia, Atlanta	0	(.)
Chicago	-0.0409	(0.08)
Dallas, Kansas City, Denver, Seattle, San Francisco	0.188*	(0.08)
Employees		
0-99	0	(.)
100 or more	-0.157	(0.08)
Penalty Inspection Combinations		
No inspection, no penalty	0.0638	(0.07)
Inspection, but no penalty	-0.0681	(0.11)
Inspection and penalty	0	(.)
DART Rate		
0-9.99	0	(.)
10-19.99	0.023	(0.07)
20 or more	0.064	(0.13)
High-rate letters		
None	-0.0194	(0.07)
1 to 5	0	(.)
6 or more	0.0555	(0.11)
Constant	0.646*	(0.08)
Observations	275	

Standard errors in parentheses, * $p < 0.05$

Table 45. Baseline differences in covariates, RCT, letter treatment, primary worksites, regulatory compliance outcomes

Letter-Primary- Regulatory Compliance	Coefficient	Standard Error
Industry		
Other	-0.0357	(0.07)
Mining, manufacturing	0	(.)
Region		
Boston, New York, Philadelphia, Atlanta	0	(.)
Chicago	0.0199	(0.05)
Dallas, Kansas City, Denver, Seattle, San Francisco	-0.0313	(0.05)
Employees		
0-99	0	(.)
100 or more	-0.0753	(0.05)
Penalty Inspection Combinations		
No inspection, no penalty	-0.0845	(0.05)
Inspection, but no penalty	-0.0878	(0.08)
Inspection and penalty	0	(.)
DART Rate		
0-9.99	0	(.)
10-19.99	0.0476	(0.05)
20 or more	-0.113	(0.09)
High-rate letters		
None	0.0889	(0.05)
1 to 5	0	(.)
6 or more	0.0265	(0.07)
Constant	0.700*	(0.05)
Observations	498	

Standard errors in parentheses, * $p < 0.05$

Table 46. Baseline differences in covariates, RCT, letter plus inspection treatment, primary worksites, regulatory compliance outcomes

Letter + Insp - Primary - Regulatory Compliance	Coefficient	Standard Error
Industry		
Other	-0.0444	(0.09)
Mining, manufacturing	0	(.)
Region		
Boston, New York, Philadelphia, Atlanta	0	(.)
Chicago	-0.00782	(0.07)
Dallas, Kansas City, Denver, Seattle, San Francisco	0.0732	(0.06)
Employees		
0-99	0	(.)
100 or more	-0.0599	(0.07)
Penalty Inspection Combinations		
No inspection, no penalty	0.0543	(0.06)
Inspection, but no penalty	-0.00718	(0.09)
Inspection and penalty	0	(.)
DART Rate		
0-9.99	0	(.)
10-19.99	0.0687	(0.06)
20 or more	0.0567	(0.11)
High-rate letters		
None	-0.0382	(0.06)
1 to 5	0	(.)
6 or more	-0.0372	(0.09)
Constant	0.627*	(0.06)
Observations	409	

Standard errors in parentheses, * $p < 0.05$

Table 47. Baseline differences in covariates, RCT, letter treatment, secondary worksites, health and safety outcomes

Letter-Secondary - Health and Safety	Coefficient	Standard Error
Industry		
Other	0	(.)
Mining, manufacturing	0.0334	(0.04)
Region		
Boston, New York, Philadelphia, Atlanta	0	(.)
Chicago	0.0325	(0.05)
Dallas, Kansas City, Denver, Seattle, San Francisco	-0.0524	(0.05)
Employees		
0-99	0	(.)
100 or more	0.0106	(0.04)
Penalty Inspection Combinations		
No inspection, no penalty	0	(.)
Inspection, but no penalty	-0.0584	(0.06)
Inspection and penalty	-0.0294	(0.05)
DART Rate		
0-9.99	0	(.)
10-19.99	0.089	(0.05)
20 or more	0.0457	(0.05)
High-rate letters		
None	0	(.)
1 to 5	0.0106	(0.06)
6 or more	0.656*	(0.04)
Constant	631	0.00
Observations	345	

Standard errors in parentheses, * $p < 0.05$

Table 48. Baseline differences in covariates, RCT, letter plus inspection treatment, secondary worksites, health and safety outcomes

Letter + Inspection - Secondary - Health and Safety	Coefficient	Standard Error
Industry		
Other	0	(.)
Mining, manufacturing	0.015	(0.08)
Region		
Boston, New York, Philadelphia, Atlanta	0	(.)
Chicago	-0.0473	(0.08)
Dallas, Kansas City, Denver, Seattle, San Francisco	-0.0397	(0.09)
Employees		
0-99	0	(.)
100 or more	0.0676	(0.07)
Penalty Inspection Combinations		
No inspection, no penalty	0	(.)
Inspection, but no penalty	-0.103	(0.10)
Inspection and penalty	-0.132	(0.08)
DART Rate		
0-9.99	0	(.)
10-19.99	-0.00992	(0.09)
20 or more	n/a	n/a
High-rate letters		
None	-0.0133	(.)
1 to 5	0	-0.0982
6 or more	0.0543	-0.0766
Constant	0.723*	0
Observations	297	

Standard errors in parentheses, * $p < 0.05$

Table 49. Baseline differences in covariates, RCT, letter treatment, secondary worksites, regulatory compliance outcomes

Letter-Secondary-Regulatory Compliance	Coefficient	Standard Error
Industry		
Other	0	(.)
Mining, manufacturing	-0.0177	(0.05)
Region		
Boston, New York, Philadelphia, Atlanta	0	(.)
Chicago	0.037	(0.05)
Dallas, Kansas City, Denver, Seattle, San Francisco	-0.0526	(0.05)
Employees		
0-99	0	(.)
100 or more	0.0139	(0.04)
Penalty Inspection Combinations		
No inspection, no penalty	0	(.)
Inspection, but no penalty	-0.0496	(0.06)
Inspection and penalty	-0.0329	(0.05)
DART Rate		
0-9.99	0	(.)
10-19.99	0.0954*	(0.05)
20 or more	n/a	n/a
High-rate letters		
None	0.00798	(.)
1 to 5	0	-0.0547
6 or more	0.0231	-0.0436
Constant	0.676*	0
Observations	631	

Standard errors in parentheses, * $p < 0.05$

Table 50. Baseline differences in covariates, RCT, letter plus inspection treatment, secondary worksites, regulatory compliance outcomes

Letter + Insp - Secondary - Regulatory Compliance	Coefficient	Standard Error
Industry		
Other	0	(.)
Mining, manufacturing	0.03	(0.06)
Region		
Boston, New York, Philadelphia, Atlanta	0	(.)
Chicago	-0.0404	(0.07)
Dallas, Kansas City, Denver, Seattle, San Francisco	-0.12	(0.07)
Employees		
0-99	0	(.)
100 or more	0.024	(0.06)
Penalty Inspection Combinations		
No inspection, no penalty	0	(.)
Inspection, but no penalty	0.0295	(0.09)
Inspection and penalty	0.0132	(0.07)
DART Rate		
0-9.99	0	(.)
10-19.99	-0.0194	(0.08)
20 or more	n/a	n/a
High-rate letters		
None	-0.015	(.)
1 to 5	0	-0.0833
6 or more	-0.0468	-0.0686
Constant	0.692*	0
Observations	395	

Standard errors in parentheses, * $p < 0.05$

Table 51: Baseline Differences in Covariates, RDD, Letter List

RDD-Letter	Coefficient	Standard Error
Number of Employees	0.000	(0.00)
Region		
Region 1	0	(.)
Region 2	-0.0509	(0.03)
Region 3	-0.00482	(0.03)
Region 4	-0.0208	(0.03)
Region 5	-0.0295	(0.03)
Region 6	-0.00216	(0.03)
Region 7	-0.0757*	(0.03)
Region 8	0.0159	(0.04)
Region 10	-0.0412	(0.08)
Hours Per Employee	-0.0000540	(0.00)
Industry		
Manufacturing	0.000	(.)
Nursing	0.0527	(0.02)
Other	0.0449*	(0.02)
Penalty Before 2006	0.000	(0.00)
High-rate letter Before 2006	0.0285*	(0.00)
Constant	0.485*	(0.06)
Observations	4,925	0.00

Standard errors in parentheses * $p < 0.05$

Table 52: Exploring Baseline Differences in Covariates, RDD, Primary List

RDD-Primary	Coefficient	Standard Error
Number of Employees	0.000	(0.00)
Region		
Region 1	0	(.)
Region 2	0.0517	(0.05)
Region 3	-0.0603	(0.05)
Region 4	-0.0282	(0.05)
Region 5	-0.053	(0.05)
Region 6	-0.0892	(0.05)
Region 7	-0.0531	(0.06)
Region 8	0.0273	(0.07)
Region 10	-0.039	(0.13)
Hours Per Employee	-0.000049	(0.00)
Industry		
Manufacturing	0.000	(.)
Nursing	-0.0213	(0.05)
Other	0.00887	(0.03)
Penalty Before 2006	0.000	(0.00)
High-rate letter Before 2006	0.00565	(0.01)
Constant	0.502*	(0.09)
Observations	1,425	0.00

Standard errors in parentheses * $p < 0.05$

Table 53: Exploring Baseline Differences in Covariates, RDD, Secondary List

RDD-Secondary	Coefficient	Standard Error
Number of Employees	0.000	(0.00)
Region		
Region 1	0	(.)
Region 2	-0.0549	(0.03)
Region 3	-0.0031	(0.03)
Region 4	-0.0705*	(0.03)
Region 5	-0.0173	(0.03)
Region 6	-0.00161	(0.03)
Region 7	-0.0945*	(0.04)
Region 8	-0.052	(0.05)
Region 10	-0.019	(0.08)
Hours Per Employee	-0.0000285	(0.00)
Industry		
Manufacturing	0.000	(.)
Nursing	0.0193	(0.02)
Other	-0.0721*	(0.02)
Penalty Before 2006	0.000	(0.00)
High-rate letter Before 2006	0.0276*	(0.00)
Constant	0.440*	(0.06)
Observations	3,983	0.00

Standard errors in parentheses [*] $p < 0.05$

Appendix G Data

The data used in the SST impact evaluation came from a number of different data sources, each of which had different information relevant to the impact evaluation. Data from these different sources were combined to create an analysis dataset upon which the impact evaluation analysis is based. The data sources used in the impact evaluation are:

1. Study group assignments
2. ODI survey responses
3. Inspection results
4. High-rate letter lists
5. Form 300A data

Data from these data sources provided the following three types of information:

1. Outcomes, used to measure impacts
2. Baseline characteristics of worksites
3. Diagnostic information on the success of administration of the evaluation

Study Group Assignments (RCT)

The SST impact evaluation study group assignments for the 2,520 study sites form the core of the analysis dataset. The dataset contains each site's study group assignment, along with DUNS, site name, address, city, state, and ZIP, which we used to link with sites' corresponding records in other data sources. This dataset is based on the 2010 ODI survey data and thus these data contain similar identifying information.

ODI Survey Responses

The ODI survey is a yearly survey of about 80,000 worksites in high-hazard industries on their workplace injuries and illnesses. Data collected include number of deaths, number of cases of injuries and illnesses causing lost work days, number of cases with restricted work activity or job transfers, and days lost due to the same. Injuries and illnesses are broken down into injuries, skin disorders, respiratory conditions, poisonings, hearing loss, and all other illnesses. Worksite information reported includes DUNS number, name, address, location, industry, total hours worked by employees, and number of employees.

The ODI survey was gathered from 1996 to 2012 for calendar years 1995 to 2011. In 2012, the survey was cancelled.

RCT

The 2009 ODI survey results were used to determine the sites in SST program year 2011 from which the study sites were selected. Thus, all RCT study sites will have ODI survey results for calendar year 2009. These survey results will provide many of the baseline characteristics of the study sites: injury/illness rate, number of employees, hours worked, and so on.

Any ODI survey results available for the study sites for years prior to 2009 is used to add historical injury/illness rates as baseline characteristics. Since the ODI survey's last year coincided with the beginning of this evaluation, it could not be used as a source of outcomes.

RDD

Data from 1998-2011 are used from the ODI responses to create the RDD analysis dataset. The ODI survey responses provide both the assignment variables and the outcomes.

Inspection Results

The third data source to be used in the impact evaluation analysis is OSHA's data concerning inspections. OSHA's inspections data are stored in the OSHA Information System (OIS) database, which was implemented in 2011. Prior to that, its inspection data were recorded in the Integrated Management Information System (IMIS) database. Inspections data from both of these databases were combined and used together in the impact evaluation analysis. Inspections data, taking these data sources together, are available from the present going back to October 2004.

The inspections data provided information on all inspections, including SST inspections. The data include information on violations, penalties, and the type of inspection.

Inspections data provided information on regulatory compliance outcomes, baseline characteristics, and information allowing the assessment of the success of evaluation administration. Results of follow-up inspections provided outcomes. Data from inspections that occurred prior to SST impact evaluation treatment provided information on prior OSHA enforcement and will serve as part of the baseline characteristics. We used the completion of assigned treatment inspections and absence of unplanned inspections to assess the success of administration; we used the number of sites not assigned inspections but still inspected to determine the extent of study crossover and control services.

The RDD analysis only used the inspection data to identify which worksites that were on the inspection lists were actually inspected. We did not analyze inspection results such as violations and penalties through the RDD because the worksites in the control group are very rarely inspected.

High-Rate Letter Lists

OSHA makes public the lists of sites that are sent high-rate letters. We used the high-rate letter lists from each year of SST for the analysis. Each year, OSHA sets a letter threshold which is based on the DART rate. Worksites that reported a DART rate above this threshold are assigned to receive a high-rate letter and placed upon the high-rate letter list. This list not only helps OSHA with their mailing process, but it is also posted online. The online posting allows anyone who is interested—including the worksites themselves—to determine which worksites had higher-than-average DART rates.

For the RCT, these lists were used to assess the administration of the SST impact evaluation. To avoid interfering with the impact evaluation, worksites in the impact evaluation are not supposed to receive any other SST actions during the study period; receiving a letter or inspection after treatment would be a failure of administration.

For the RDD, the lists were used to determine which worksites were placed on the list and which were not so that these two groups can be compared.

Analysis Dataset Creation

We combined information from study group assignments, the ODI survey, inspections, and high-rate letter list data to create an analysis dataset. The analysis dataset includes data on outcomes, baseline characteristics, and information which we used to assess the administration of the evaluation. We linked corresponding records from each data source by DUNS number, where it was possible. Since DUNS was not available in every one of these datasets, and in some it is not always recorded, linking by DUNS was not always possible. In these cases, we linked corresponding records by matching names and addresses as closely as possible.

We matched names and addresses using edit distance, which measures the “distance” between two text strings, to find records with close names and addresses. We did this permissively to avoid failing to match records that do in fact correspond. We then manually reviewed all links from the study group assignments to records in the ODI, inspections, and high-rate letters datasets to verify that they are appropriate.

The research team also manually reviewed all cases where corresponding records were expected but not found. All study sites should have had ODI survey results for calendar year 2009. Lack of records of previous inspections (or letters) is possible and acceptable as some sites may not have been inspected before.

To create the analysis dataset for the RDD, we:

- Consolidated 1998-2011 ODI data into one observation per worksite by putting each year’s survey responses in separate columns.
- Linked inspections data and high-rate letter lists using the same record linkage methodology as the used for creating the RCT analysis dataset with one difference. There was no manual review of the record linkage process; it was all done automatically.

For the RCT, the record linkage process divided potential linked records into three groups: automatic matches, potential matches requiring manual review, and non matches. We did this by scoring each potential linked record and setting up score thresholds. In this case, we defined the matching score thresholds for two groups: matches and non matches. Due to much larger sample sizes than in the RCT, it was not efficient to manually review the matches. Instead, the threshold was set to ensure that the matches are sufficiently scrutinized using the matching program. There was still a manual quality check of the matching procedure only to see if the automatic process worked as intended.