OSHA ENFORCEMENT REVIEW PROTOCOL

Highlights

- The objective of this systematic review is to determine the quality of existing causal evidence on the effectiveness of OSHA enforcement activities.
- In particular, the review considers research on whether OSHA inspections reduce the occurrence of workplace illnesses, injuries, exposure to hazards, and/or deaths.
- Only research with causal designs is reviewed for this topic area.

Introduction

The topic area for this evidence review protocol is enforcement activities conducted by the Occupational Safety and Health Administration (OSHA). Established by the Occupational Safety and Health Act of 1970, OSHA strives "to assure safe and healthful working conditions for working men and women by setting and enforcing standards and by providing training, outreach, education and assistance." [1] OSHA standards are the rules that describe steps employers must take to ensure a safer work environment. Some examples include preventing exposure to infectious diseases, harmful chemicals, and excessively loud noise; providing respirators or other safety equipment to employees; putting guards on dangerous machines; and providing training for dangerous jobs.[2]

OSHA enforces these standards through programmed and unprogrammed inspections of worksites. Programmed inspections target industries, workplaces, or occupations that have been identified as high-risk. Industries and firms are selected for these inspections based on observable characteristics, including past injury rates and citations. Conditional on these factors, some firms are chosen at random for programmed inspections while others are inspected with certainty.[3] In FY 2010, Federal OSHA conducted 24,759 programmed inspections.[4] In contrast, unprogrammed inspections are triggered by employee complaints, referrals from another source, such as the media or other government agency, and fatalities and catastrophes (more than three hospitalizations). Federal OSHA conducted 16,234 unprogrammed inspections in FY 2010.[5] In addition, OSHA investigates informal complaints by mail or fax; these complaints do not rise to the level of a formal complaint because, for example, they do not allege a serious hazard or violation or they have not been signed.

During an inspection, an OSHA compliance inspector conducts a walkthrough of the facility with an employer and an employee representative and inspects for hazards that could lead to employee illness, injury, or death. The OSHA inspector issues citations, some of which have monetary penalties attached to them, for all violations identified. Violations are categorized as other-than-serious, serious, willful, repeated and failure to abate; penalties range up to $7,000 for each serious violation and up to $70,000 for each willful or repeated violation.[6]

---

[1] http://www.osha.gov/about.html

[2] http://www.osha.gov/OSHA_FAQs.html

[3] https://www.osha.gov/Publications/osha2098.pdf

[4] http://www.osha.gov/dep/2010_enforcement_summary.html

[5] http://www.osha.gov/dep/2010_enforcement_summary.html

[6] http://www.osha.gov/OshDoc/data_General_Facts/factsheet-inspections.pdf

The effectiveness of OSHA enforcement activities on workplace safety is of primary interest to this topic area review. Specifically, the evidence review focuses on enforcement-related research questions including:

- Do OSHA enforcement activities reduce the occurrence of workplace illnesses, injuries, exposure to hazards, and/or deaths?

- Does compliance with OSHA standards affect a company's productivity or profitability?

The rest of this evidence review protocol sets forth the criteria by which existing research is determined to be eligible for review, rules for using causal evidence guidelines to rate the quality of causal evidence, and review procedures. Appendix A provides further details on the methods used to identify studies for potential inclusion in the review.

Eligibility Criteria

CLEAR staff identifies research studies for potential inclusion in the OSHA enforcement topic area using a broad literature search (see Appendix A for details). Each study identified through the literature search is evaluated against a set of eligibility criteria:

1. **Does the research examine an OSHA enforcement activity?** To be eligible for review, the research must examine some type of OSHA enforcement activity, including (but not limited to) inspections of worksites conducted in-person, by phone or by fax.

2. **Is it a study of effectiveness?** To be eligible for review, the research must use quantitative methods to assess the effectiveness of the OSHA enforcement activity. Research that solely describes the characteristics or implementation of OSHA enforcement activities, or is a case study of firms' experiences with them, is not eligible for review under this protocol.[7]

3. **Does the research examine a population of interest?** All workplaces covered by OSHA standards are eligible for review. Reviews in this topic area will encompass analyses of subgroups determined by firm or establishment size, industry, and whether a location is covered by federal OSHA standards or an approved state plan. However, the overall causal evidence rating will be determined based on the full sample.

4. **Does the analysis include at least one outcome of interest?** The goal of this review process is to determine the extent of the causal research evidence supporting OSHA's mission of improving workplace safety. Therefore, the outcomes of primary interest include indicators of workplace safety, such as: the number or rate of workplace illnesses, injuries, or fatalities; exposures to safety hazards from lack of machine guarding, lockout/tag out procedures, and required fall prevention protections; health hazards from exposure to harmful particles or chemicals, blood borne pathogens, or excessive noise; use of personal protective equipment; and measures of worker safety training or the quality of health and safety programs. In addition, analyses could include mediating variables, such as indicators of compliance with standards, number of violations issued or

[7] Causal studies in this topic area were reviewed according to the CLEAR Causal Evidence Guidelines, Version 1.0. The full set of guidelines can be found at http://clear.dol.gov. CLEAR also has guidelines for reviews of descriptive and implementation research; however, this topic area is limited to causal studies.

dollar amount in penalties levied. Indicators of firm profitability and productivity are also of interest. These could include, for example, an indicator of firm survival over the follow-up period, sales, savings in workers compensation costs, profits, or total employment.

5. **Was the research conducted in a relevant place?** All research must have been conducted using data from the United States, including the 50 states, the District of Columbia, territories, and tribal entities.

Causal Evidence Guidelines

This topic area does not include reviews of experimental causal research because no such research was identified as meeting eligibility criteria. In general, CLEAR assesses the quality of evidence for randomized controlled trials (RCTs) using an adaptation of the Institute for Education Science's What Works Clearinghouse (WWC) standards.[8] RCTs are the only type of studies that may receive a High causal evidence rating. This occurs if there are no obvious confounds to the RCT design and if the level of attrition in the RCT is low. If CLEAR determines that an RCT cannot be rated as providing high causal evidence, the research is reviewed using the nonexperimental causal evidence guidelines developed by CLEAR.

Nonexperimental Causal Evidence Guidelines Specific to the Topic Area

In collaboration with a Technical Work Group of experts, Mathematica Policy Research, developed a set of evidence guidelines to be used in reviewing nonexperimental studies with causal designs. These causal designs include instrumental variables, difference-in-differences, fixed and random effects, and other types of regression analyses. Research designs that meet the causal evidence guidelines receive a Moderate causal evidence rating; this rating indicates that there is evidence that the study establishes a causal relationship between the intervention being examined and the outcomes of interest, but there may be other factors that were not included in the analysis that also could affect the outcomes of interest. Designs that do not meet the guidelines receive a Low causal evidence rating, which indicates that we cannot be confident that the estimated effects are attributable to the intervention being examined.

Causal evidence guidelines for nonexperimental studies are tailored to the topic area of interest. In particular, the topic area protocol sets forth the specific types of control variables that need to be included in nonexperimental regression analyses (other than those using fixed effects) for a study to receive a Moderate causal evidence rating. The topic area protocol also describes whether changes in group composition should be a concern for the review.

Control variables

The control variables for the OSHA enforcement protocol were developed in consultation with two topic area experts from the TWG. The control variables are:

- Size of the firm, plant, or establishment
- State, for analyses including more than one state
- Sector (e.g., manufacturing, construction, service, maritime), for analyses including more than one sector

[8] See http://ies.ed.gov/ncee/wwc/InsidetheWWC.aspx for details.

- Lagged values of an indicator of workplace health and safety, such as the number or rate of injuries, illnesses, fatalities, or exposure to hazards. These could include lagged values of the dependent variable or lagged values of another variable that is highly correlated with it.[9] These data could come from a variety of sources, including workers compensation reports, insurance records, and self-reports.

Lagged values of an indicator of workplace health and safety are particularly important to establishing the comparability of the groups. For this topic area, an analysis that includes lagged values to capture differences in pre-existing levels or trends in outcomes between the treatment and comparison groups must also demonstrate that the number of lags used will account for any such differences. For example, if an analysis includes a single lagged value of the dependent variable, it must demonstrate that the difference between the treatment and comparison groups was not systematically changing prior to the treatment. An analysis that includes the first and second lags of the dependent variable must demonstrate that the difference in outcomes was constant or changed linearly over time before the inspections of treatment group firms. Authors can use a variety of different techniques, including graphical analysis and placebo tests, to show this. See Appendix B for further details.

Changes in group design

Although research designs in the OSHA Enforcement topic area commonly use state- or industry-level aggregate data, we do not require that authors demonstrate that the composition of the groups being compared does not change. Any changes in the composition or characteristics of firms in the aggregate due to an OSHA enforcement activity may be seen as an impact of that activity, and thus should be part of the measured treatment effect. For example, if high-risk firms in an industry receiving more inspections went out of business because of the higher enforcement propensity, the resulting drop in injury rates for this industry can be thought of as part of the impact of the OSHA activity. Therefore, studies need not demonstrate that interventions left group composition unchanged.

Review Procedures

Each research paper or report that is identified as being eligible for review against causal evidence guidelines is assigned to a reviewer who has been certified by CLEAR to understand and be able to apply its standards with fidelity. The reviewer reads the study in detail; applies the causal evidence guidelines to determine the design's causal evidence rating; and documents all aspects of the review in a standardized review guide. In particular, the review guide contains supporting information for the rating, details of the study sample and intervention, and any other pertinent information.

If the reviewer assigns a rating of High or Moderate causal evidence, a second reviewer also reviews the research to confirm such a rating is warranted. Any discrepancies between the two reviewers' ratings are resolved by the Principal Investigator (PI) to determine a final causal evidence rating. If the first reviewer assigns a rating of Low, the PI examines the review guide and confirms that the rating is appropriate.

[9] If the analysis includes a pre-intervention measure that is not obviously a good indicator of pre-inspection outcomes, the study should contain a clearly articulated discussion of why the particular measures were selected and justify their use.

When a research paper or report does not contain sufficient information to determine a causal evidence rating, CLEAR may contact the authors to gather this information; whether this step is undertaken depends on the age of the study and the quantity of information that would need to be gathered (so as not to overly burden authors). Authors receive a minimum of four weeks to respond, and reasonable requests for extensions are granted. If the information is provided by the authors, it is incorporated into the review and factors into the causal evidence rating. If the authors do not provide the relevant information, the design is given the highest rating that can be determined with the information available in the report.

APPENDIX A: LITERATURE SEARCH

Studies in this topic area are identified by conducting a literature search in Scopus, which covers 19,500 peer-reviewed journals, 400 trade publications, 360 book series, and "Articles-in-Press" from over 3,850 journals.[10] Studies that have not yet been published are identified by searching the Social Science Research Network (SSRN), which contains abstracts on over 464,100 scholarly working papers and forthcoming papers.[11]

The search parameters for both searches are:

- The document contained the words "OSHA," "occupational safety and health," or "workplace safety inspection" in the title or abstract.
- The document contained the word strings "inspection*," "regulat*," or "compliance" in the title. (Note: an asterisk indicates that all results containing the word string are returned in the search results. For instance, "regulat*" captures both "regulation" and "regulated.")

In addition, studies are identified by searching the websites of DOL's OSHA; the Office of the Assistant Secretary for Policy, Chief Evaluation Office; the Washington SHARP program; the Institute for Work and Health in Toronto; and the RAND Center for Health and Safety in the Workplace.

[10] http://www.info.sciverse.com/scopus/scopus-in-detail/facts

[11] http://www.ssrn.com/

APPENDIX B: DETAILS ON CONTROL VARIABLES AND METHODS

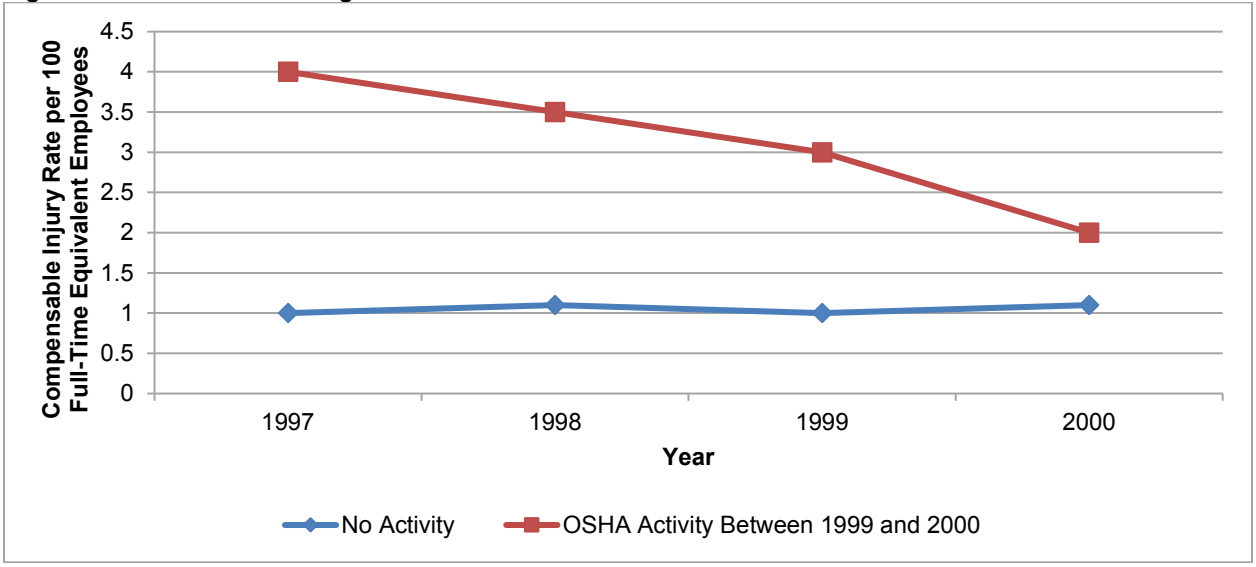
Regression methods that use a comparison design, in which statistical methods are used to create a comparison group that is as similar as possible to the group receiving the intervention, must match on the control variables listed in the protocol or, if they do not match on them, must include them as controls in the regression.

PIs can use their discretion to modify the requirements for sufficient controls. For instance, these controls might be waived in cases where the authors successfully argue that assignment into treatment and comparison groups was determined largely by chance. Additionally, if the intervention of interest is determined to be strongly influenced by prior values of the outcome variable, the controls listed in the protocol may not be considered sufficient.

Figure 1 demonstrates a situation where authors could successfully use two lags to control for pre-existing differences between firms in treatment and comparison groups. In the example shown in the figure, before the end of 1999, when the treatment group received some OSHA enforcement activity, injuries were declining in the treatment group but approximately constant in the comparison group. Because the differences evolve in a roughly linear way, controlling for two lags would be sufficient (but one lag would not be sufficient).

Regression methods that incorporate fixed effects must also demonstrate that the fixed effects are sufficient controls for differences between the treatment and comparison groups. This could be done by demonstrating that within-firm conditions were not systematically changing before the OSHA activities. Alternatively, the authors could demonstrate that before the OSHA activity, outcomes in treatment and comparison firms were not differentially changing.

Figure 1: Case Where Two Lags Are Sufficient



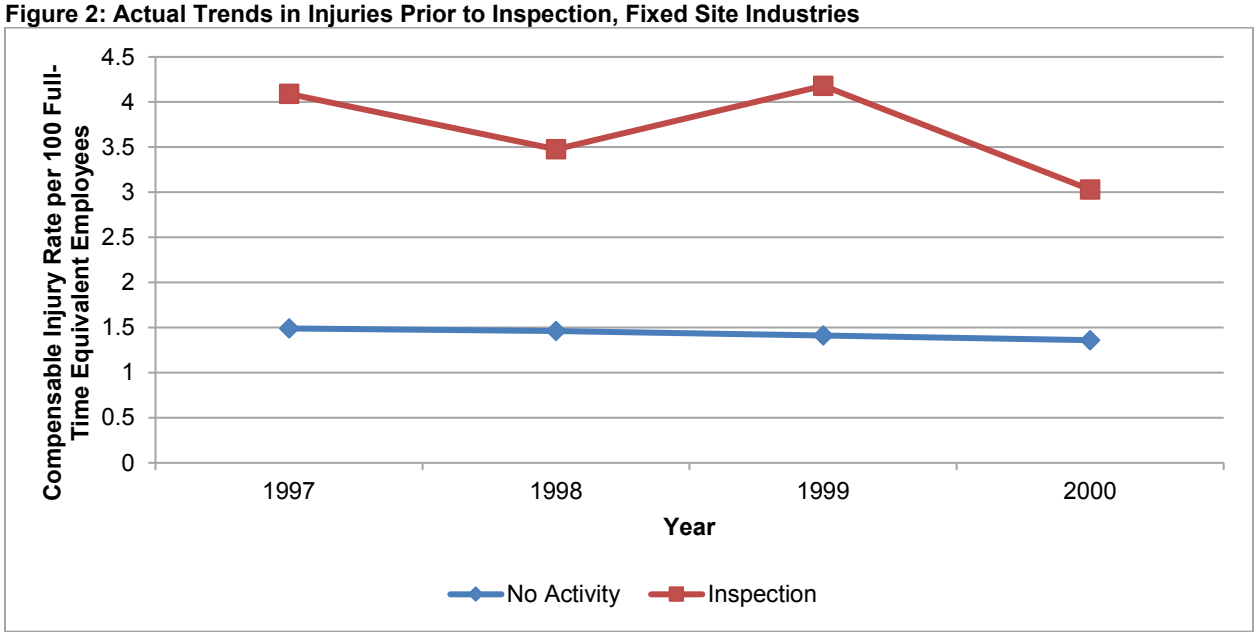
Classification of Common Study Designs

The literature on OSHA enforcement activities focuses on several non-experimental regression designs. To promote consistency across reviews, they are discussed here.

Comparisons of inspected and non-inspected firms. Studies in this classification compare outcomes for firms that received an inspection to outcomes for firms that did not receive an inspection. Many of these studies control for one or more lagged values of the outcome of interest or incorporate firm fixed-effects. However, firms that received complaint or other unprogrammed inspections likely had more underlying workplace hazards, on average, than firms that were not inspected. In the absence of inspections, the firms with more hazards and safety violations may have experienced different changes in outcomes as conditions deteriorated or because management would have made improvements to address unsafe working conditions. Therefore, studies that use this design to analyze unprogrammed inspections (alone or in combination with other inspections) are rated as low, regardless of the controls included in the regression.

Figure 2 further demonstrates why controls for lagged outcome values might not be sufficient in analyses of OSHA inspection activities. This figure contains the compensable injury rates over time for two groups of firms: those receiving an inspection in 2000 and those not receiving an inspection. The trend for comparison firms is roughly flat; however, the injury rate first increases, then decreases, in treatment firms. Unlike in Figure 1, these trends do not imply an obvious value for the injury rate in treated firms in the absence of an inspection. We might expect injury rates to rise, as they did in the previous period. But if the rise in injury rates represented a one-time change which prompted an inspection, injury rates might be expected to fall, even without intervention (as they did from 1997 to 1998). Using lags of the dependent variable in the regression will not capture these issues and thus cannot account for the pre-existing differences.

Figure 2: Actual Trends in Injuries Prior to Inspection, Fixed Site Industries



Note: Example based on Baggs, et al. (2003).

Analyses of only programmed inspections, which are not triggered by any specific adverse event in a workplace, are not subject to this critique. Industries and firms are selected for these inspections based on observable characteristics but there is also some random variation in which firms are

inspected under this program.[12] Because of the random component of the selection process, a careful examination of programmed inspections alone, which includes a rich set of control variables, might not suffer from the aforementioned issues.

Comparisons of firms receiving penalties or citations to other firms. Studies in this classification compare outcomes for firms that received an inspection which resulted in a penalty or citation to outcomes for firms that did not receive such sanctions. Many of these studies control for multiple lagged values of the outcome of interest or incorporate firm fixed-effects. However, similar to the studies of non-programmed inspections, firms that received penalties or citations likely had more underlying workplace hazards, on average, than unsanctioned firms. In the absence of penalties or citations, the firms with more hazards and safety violations may have experienced different changes in outcomes as conditions deteriorated or because management would have made improvements to address unsafe working conditions. Therefore, studies that use this design must examine only firms that received penalties and/or citations at random or use some underlying random variation in the receipt of penalties (which may not exist).

Comparisons of firms inspected at different times in year. Several studies compared outcomes for firms inspected earlier or later in the year. Because outcomes are measured over the course of a year, firms with earlier inspection dates were post-inspection for a larger proportion of the measured year. If the timing of inspections is random, the difference in injury rates between firms inspected early and late in the year may be attributed to the inspection. An examination of OSHA policies demonstrates that this is likely true before 1978. In 1978, OSHA developed a new system for determining which firms to inspect in response to the Supreme Court's ruling on *Marshall v. Barlow's Inc.* As a result of the ruling, OSHA created a system for prioritizing inspections so that firms more likely to be unsafe were inspected earlier in the year. This suggests that before 1978, firms inspected early and late in the year were likely not systematically different. Thus controlling for a single lag of the outcome variable is likely sufficient. However, in 1978 and later years, we cannot be confident that firms inspected early and late in the year are comparable. Thus, authors using later data cannot receive a moderate rating, even if they include the listed control variables in their regression analysis.

Sequence number studies. Studies using this identification strategy compare the number of citations a firm (or plant) received at their first, second, and higher-order inspections. As there is no reason to believe that conditions at a firm or plant are systematically different when the firm or plant receives an inspection for the first, second, or nth time, studies using this strategy will receive a rating of moderate causal evidence, as long as the study compares within-firm or within-plant differences (that is, a firm fixed-effect is included in a firm-level analysis; a plant fixed-effect is included in a plant-level analysis) and includes sufficient controls for the type of inspection received.

[12] https://www.osha.gov/Publications/osha2098.pdf