Review

A Systematic Literature Review of the Effectiveness of Occupational Health and Safety Regulatory Enforcement

Emile Tompa, PhD,1,2,3* Christina Kalcevich, MA,1 Michael Foley, MA,4 Chris McLeod, PhD,1,5 Sheilah Hogg-Johnson, PhD,1,3 Kim Cullen, PhD,1 Ellen MacEachen, PhD,3,6 Quenby Mahood, MA,1 and Emma Irvin, BA1

Background  We aimed to determine the strength of evidence on the effectiveness of legislative and regulatory policy levers in creating incentives for organizations to improve occupational health and safety processes and outcomes.

Methods  A systematic review was undertaken to assess the strength of evidence on the effectiveness of specific policy levers using a “best-evidence” synthesis approach.

Results  A structured literature search identified 11,947 citations from 13 peer-reviewed literature databases. Forty-three studies were retained for synthesis. Strong evidence was identified for three out of nine clusters.

Conclusions  There is strong evidence that several OHS policy levers are effective in terms of reducing injuries and/or increasing compliance with legislation. This study adds to the evidence on OHS regulatory effectiveness from an earlier review. In addition to new evidence supporting previous study findings, it included new categories of evidence—compliance as an outcome, nature of enforcement, awareness campaigns, and smoke-free workplace legislation. Am. J. Ind. Med. 59:919–933, 2016. © 2016 Wiley Periodicals, Inc.

KEY WORDS: regulation; legislation; occupational health and safety; OHS; regulatory effectiveness

INTRODUCTION

Occupational health and safety (OHS) regulation takes different forms across jurisdictions and over time. Responding to involved stakeholders, government authorities strive to devise ways to protect the health and safety of workers that will be effective in the context of competitive business environments, contemporary labor market structures, available resources, labor-management power issues, perspectives of health and safety professionals, and political will. OHS policy levers (i.e., the specific means used to promote compliance)1 used by regulatory

---

1 We use the term “policy lever” to refer to the specific means by which a regulator attempts to encourage action in a particular direction, as opposed the broader notion of “policy” as the operationalization of legislation.
authorities include a variety of approaches such as administrative monetary penalties, prosecutions, orders to comply, injunctions, inspections and audits, and consultations. For regulations to be effective, it is critical that they address key health and safety risks that are amendable by workplace parties, are clearly communicated to organizations, are enforceable by regulators, and create incentives for compliance [Mendeloff, 1979; Scholz and Gray, 1997; Shapiro and Rabinowitz, 1997; Baggis et al., 2003; Wright and Genna, 2005; Bluff, 2011; Safe Work Australia, 2013].

The literature on OHS regulatory policy levers is large and diverse, uses different statistical methods and levels of data aggregation, and considers experiences from various time periods and jurisdictions. As a result, it is difficult to compare and contrast findings from different studies, discern the quality of evidence, and identify the overarching strength of evidence on particular levers. Nonetheless, collating and synthesizing the evidence on the effectiveness of OHS regulatory policy levers is critical, given the substantial resources invested by the public sectors in regulatory enforcement and by organizations seeking to be compliant.

While earlier studies reviewed parts of this diverse literature [e.g., Kralj, 2001; Mendeloff, 2001], Tompa et al. [2007] was the first to use a structured approach to identifying, evaluating, and synthesizing the evidence base. More recent reviews have complemented these evidence syntheses by considering the literature on the socio-psychological factors of stakeholders that bear on their ability to address OHS issues [Bluff, 2011], and others have focused on why some regulatory efforts are successful while others are not [Safe Work Australia, 2013].

The Tompa et al. [2007] review found strong evidence that the first hand experience of citations and penalties (known as specific deterrence) reduces injuries. In contrast, the first hand experience of inspections had only limited to mixed evidence. Similarly, there was limited to mixed evidence that the probability of inspections, citations, and penalties (known as general deterrence) reduces injuries. The review considered evidence up until 2004, but the literature has grown substantially since then, with investigations of new legislation and more detailed inquiries into existing practices. The current systematic review follows from the earlier review, expanding the scope to include studies that consider intermediate outcomes, specifically compliance, as well as other OHS policy levers such as consultations and awareness campaigns. Also included is the introduction of smoke-free workplace legislation.

The overarching question guiding this review is “what is the strength of evidence on the effectiveness of OHS policy levers in creating incentives for organizations to improve occupational health and safety processes and outcomes.” We follow with a detailed description of the systematic search strategy, study inclusion criteria, quality assessment, and evidence synthesis approach. The results provide evidence synthesis profiles across nine clusters of studies that were retained in the final review. Our discussion section elaborates on the policy implications of our findings, expounds on the importance of context, details the strength and limitations of our review, and provides suggestions for the way forward for research. We end with a brief summary of findings and implications in our conclusion.

In addition to the research question noted above, the review had the following objectives:

- To identify peer-reviewed literature on the effectiveness of OHS policy levers;
- To evaluate the quality of identified studies and synthesize the evidence;
- To engage a stakeholder committee throughout the process to assist with identifying the research questions, scope of the study as well as interpretation of findings; and
- To disseminate the review findings locally, nationally, and internationally.

**METHODS**

**Stakeholder Engagement**

Consistent with stakeholder engagement recommendations [Keown et al., 2008], an advisory committee was formed to provide feedback on the key question guiding the review, the scope of the literature search, the interpretation of the findings, and the formulation of policy implications. The committee consisted of three senior policy makers with extensive regulatory enforcement experience from the Ontario Ministry of Labour (Canada), three senior academics with a specialty in OHS issues (from the United States, United Kingdom, and Australia), two senior industry OHS service providers and an injured workers’ advocate (all three from Canada). The first advisory committee meeting was held in the early stages of the study after some preliminary groundwork on searches had been completed. A second one was held near the end of the study after preliminary synthesis profiles had been formulated.

**Systematic Literature Search**

This review, which focused on quantitative studies, was undertaken in conjunction with a qualitative review on the same topic. The quantitative review considered evidence on effectiveness, while the qualitative review considered planning and implementation of regulations. The two reviews had a common search process, but once full studies were identified as meeting preliminary inclusion criteria, the two reviews were completed independently.
Selection of Studies for Inclusion

The review was based on a systematic search of several peer-reviewed journal databases, specifically MEDLINE, EMBASE, PsychINFO, ABI Inform, Health and Safety Science Abstracts, ASSIA, EconLit, Sociological Abstracts, Wilson Social Science Abstracts, and Index to Legal Periodicals. Databases were chosen to capture all disciplinary areas that might have studies meeting the inclusion criteria. The search methodology followed a modified PICO (population, intervention, comparison, and outcome) format, in which studies had to contain at least one term in each of four categories (regulatory focus, setting, policy lever, and context) in order to be retrieved. The search strategy combined sets of keywords using an “AND” term between categories, and an “OR” term within categories so that citations would have to include at least one term from each of the categories in order to be pulled from the database. Wildcard characters were used extensively to ensure different spelling and forms of words were captured. Table SI in the supplemental materials provides details on the final keywords used.

A hand search was also done on the journal “Policy and Practice in Health and Safety” and on the website SafeWork Australia/RegNet because of their focus on research related to OHS regulatory mechanisms. As well, the research team compiled a list of 19 content experts from seven countries and solicited their suggestions of studies for consideration that were already published in peer-reviewed journals, in press, or accepted for review.

Searching for additional studies was an iterative process. For example, all references of studies selected for inclusion in the review were scanned to identify incremental relevant studies. Other literature review studies on the topic of OHS regulation were also scanned. This latter category included a review by Bluff [2011] and a Cochrane collaboration review by Mischke et al. [2013].

Selection of Studies for Inclusion

To be eligible for inclusion in the review, studies had to be in English, published between January 1990 and June 2013 in a peer-reviewed journal, and be longer than two pages. The latter was to filter out short discussion pieces.

At the title and abstract screening stage there were two criteria for study inclusion. First, studies had to consider directives related to OHS legislation and/or regulation made by a government authority. Second, studies had to evaluate OHS legislation and/or regulations using quantitative and/or qualitative methods. Studies that met the two criteria were then classified as quantitative, qualitative, or mixed methods.

The title and abstract screening stage began with a pilot test in which the entire research team reviewed the same set of 120 studies based on the above noted criteria. After all reviewers tested the criteria, the team met to discuss discrepancies and fine tune the technical guidelines document accordingly. A second meeting was held with the research team after alternating pairs of two reviewers assessed another 1,500 titles and abstracts to further discuss any outstanding issues. Following this piloting, each reviewer was assigned a separate batch of titles and abstracts. A second reviewer audited 20% of citations across the entire frame of studies pulled by the electronic searches to ensure accuracy of the selection process. Agreement for passing citations onto the next stage was 99.5%. In cases where only a title was available for review, the entire paper was retrieved and reviewed before deciding on its eligibility for inclusion.

The full study screening stage used the same criteria for inclusion as the title and abstract stage. The difference here was that the full article was reviewed to ensure it met the criteria. At this stage, studies were assigned to a single reviewer. Again, 20% of the full article screenings were subjected to a second review. Agreement was 96% at this stage for inclusion. All studies selected for inclusion were again screened by both the project coordinator and the principle investigator as part of a quality check.

Studies included in the quantitative and mixed methods categories were reviewed by multiple quantitative team members to identify those that met the following incremental inclusion criteria: (i) the study had a temporal element (i.e., study must either use data from multiple points in time, or ask respondents about past experiences); (ii) the study design was rigorous (i.e., study could not be just descriptive, it had to use multiple regression modeling methods or had to have a quasi-experimental design, including before/after or a concurrent control group); and (iii) the study considered final outcomes, such as injuries and illnesses, or intermediate outcomes, such as compliance or reduced exposures (studies examining only monetary outcomes were not included).

Compliance and reduced exposures are seen as intermediate outcomes since the ultimate goal of OHS regulatory enforcement is to reduce work injuries and illnesses. Essentially, compliance and reduced exposures are simply a means to that goal.

Relevant regulatory policy levers included regulation enacted and enforced at any level of government: country, state/province, sector, workplace, and/or individual level; and regulatory levers that focused on enforcement (e.g., stop work orders, injunctions, prosecutions, monetary penalties, warnings, orders, tickets, inspections), voluntary activities (e.g., voluntary guidelines, consultations, certifications, health, and safety group membership) and/or mandated

2 Monetary outcomes are used in studies focusing on insurance costs. We excluded these, as we were interested in studies focusing on occupational health and safety performance rather than insurance costs. We note that there were only a few such studies.
TABLE I. Summary Table of Policy Levers and Outcomes

<table>
<thead>
<tr>
<th>Policy levers</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction of OHS legislation</td>
<td>Compliance</td>
</tr>
<tr>
<td>Introduction of smoke-free workplace</td>
<td>Exposures</td>
</tr>
<tr>
<td>legislation</td>
<td></td>
</tr>
<tr>
<td>Inspection sequence</td>
<td>Awareness (of campaign)</td>
</tr>
<tr>
<td>Inspection activity: general deterrence of inspections and penalties</td>
<td>Health (respiratory and sensory symptoms)</td>
</tr>
<tr>
<td>Inspection activity: specific deterrence of inspections and penalties</td>
<td>Health behaviours (cigarette smoking)</td>
</tr>
<tr>
<td>Nature of enforcement: consultative activities</td>
<td>Injuries and fatalities</td>
</tr>
<tr>
<td>Nature of enforcement: autonomy supportive inspector style</td>
<td>Truck crashes</td>
</tr>
<tr>
<td>Nature of enforcement: state-versus federal-level enforcement</td>
<td></td>
</tr>
<tr>
<td>Awareness campaigns</td>
<td></td>
</tr>
</tbody>
</table>

activities (e.g. right to refuse unsafe work, requirement for a joint health, and safety committee). Table I provides an overview of the policy levers and outcomes considered in the review.

Quality Assessment and Data Extraction

Quality assessment and data extraction were performed concurrently by two reviewers. The two reviewers independently scored each assigned study and then met to discuss it. A consensus rating was not required. Rather, the discussion was meant to ensure that both reviewers considered the full range of issues relevant to each study. Studies were evaluated based on a quality appraisal protocol developed by Tompa et al. [2007] that consisted of 10 items in two parts (study quality and policy lever relevance). Each item was ranked on a Likert scale from one to five. See Table II for details of the 10 items. Ratings from the two reviewers were averaged for part one and part two separately. The lower of the two scores was used for the overall quality rating score. The final quality rating for a study was grouped into one of three categories: high (70% or greater), medium (50–70%), or low (50% or less). Only high and medium quality studies were retained for evidence synthesis.

Evidence Synthesis

The evidence-ranking algorithm used to synthesize evidence across studies was based on a qualitative methods approach known as “best evidence synthesis” developed by Slavin [1986, 1995] and used in other published reviews [e.g., Tompa et al., 2007; Rivilis et al., 2008; Tullar et al., 2010]. Best-evidence synthesis identifies the strength of a relationship based on the quantity, quality, and consistency of the evidence available to support a relationship between variables. Part and parcel to the approach is the notion of precedence. If a certain quality standard was required in the past, then current requirements should be the same or higher. Thus, as the literature grows and advances, the bar for each quality level may be set higher. The best evidence synthesis approach is well suited for the subject matter and literature of this review because of the broad range of study designs and analytic approaches. Quantitative methods such as meta-analysis cannot be employed when statistical methods used by different studies are too varied. Best evidence synthesis aims to provide the same methodological rigor to evidence synthesis as meta-analysis by clearly and concisely articulating the synthesis criteria.

For each policy lever and outcome category, we ranked the evidence supporting the hypothesized relationship on a five-level scale consisting of strong evidence, moderate evidence, limited evidence, no evidence, and mixed evidence.\(^3\) Evidence on a policy lever was tested against the criteria for the highest level (strong evidence), and, if it was not met, the criteria for the next highest level (moderate evidence) was considered. If it was not met, the subsequent level (limited evidence) was considered. If the evidence did not meet the criteria for any of these three levels, it defaulted to one of the two categories, no evidence or mixed evidence. The former arose if there were no studies or only low-quality studies. The latter arose if there was more than one high- or medium-quality study and the studies provided conflicting evidence. Table III below provides details on the evidence synthesis algorithm.

RESULTS

Literature Search Results

The literature search, which included electronic databases, hand searches, and references from content experts, identified 11,947 unique titles and abstracts across 13 sources. Table IV below provides details by source. The cell counts are prior to the removal of duplicates from across the different sources.

Of the 11,947 titles and abstracts, 2,360 passed to full study review. Of these studies, 282 were retained in the quantitative or mixed method categories. At this stage the incremental quantitative rigor criteria were invoked to screen the 282 studies, 61 of which passed the screening and moved onto the quality assessment stage. Three additional articles

---

\(^3\) A significance level of 5% was used as the cutoff for evidence of a relationship between a policy lever and an outcome. Though some studies included considered 10% as significant, we treated these as not significant.
were identified in the reference lists of these studies, increasing the count to 64 that were assessed for quality. Of these studies, 43 were rated high or medium quality and were retained for evidence synthesis. Figure 1 below provides a flow chart of the number of studies retained at each stage.

The Tompa et al. [2007] review had identified 24 studies on the effectiveness of regulatory policy levers published from 1970 to 2004. Sixteen of those studies were published from 1990 onward, though four were not from peer reviewed journal publications and one study was not in English. The remaining 11 studies from that review were identified in our search and had passed through the same quality appraisal protocol as the newly identified studies. Of the 11 studies, seven dropped in quality ranking from high to medium or from medium to low quality, due to a higher standard imposed by the team in response to a noted higher level of quality identified in the overall literature. This is consistent with the best-evidence synthesis approach, which takes into consideration precedence and the state of the literature. Ultimately, seven of the original studies received a ranking of medium or high quality and were included in this review.

Included studies were grouped into nine thematic clusters, defined by the policy lever being evaluated. The clusters were: (i) introduction of OHS legislation; (ii) introduction of smoke-free workplace legislation; (iii) inspection sequence (defined below); (iv) inspection activity: general deterrence of inspections and penalties; (v) inspection activity: specific deterrence of inspections with/without penalties; (vi) nature of enforcement: consultative activity; (vii) nature of enforcement: autonomy supportive inspection style; (viii) nature of enforcement: state- versus federal-level enforcement; and (ix) awareness campaigns. See Table SII in the supplemental material for details. Table V provides a high level summary of the studies by cluster and the synthesis statements related to each.

### Evidence Synthesis Results

#### Introduction of OHS legislation

There are nine studies in this cluster (one of high quality and eight of medium quality). Studies considered the introduction of a mix of different legislation, some enabling legislation to promote good practices and empower workplace parties, and others regulations as mechanisms for creating compliance obligations. Specifically, the legislation includes a hearing conservation program, chemical exposure mitigation, universal precautions for blood borne pathogens, ergonomics regulation, lockout/tagout requirements, internal responsibility systems, and training requirements. Some studies in this cluster considered final outcomes of injury, illness, and fatality rates, while others considered intermediate outcomes of exposure and compliance rates. For the former, there is “moderate evidence” that the introduction of OHS legislation has an effect on final outcomes (based on one high quality study and five medium quality studies). For the latter, there is “limited evidence” that the introduction of legislation improves intermediate outcomes, that is, reduces exposure rates and/or increases compliance rates (based on three medium quality studies). Four of the studies in this cluster were in manufacturing, two in health care, one in forestry, and one in multiple sectors. Six of the studies were undertaken in the United States, two in Canada, and one in Spain.

#### Introduction of smoke-free workplace legislation

There are six studies in this cluster (five of high quality and one of medium quality). Studies all considered the introduction of different forms of smoke-free workplace legislation in North America and Europe. Here too some studies considered final outcomes, specifically respiratory
and sensory symptoms. Others considered intermediate outcomes, specifically reductions in smoke exposure, and reductions in cigarette consumption. For final outcomes, there is “moderate evidence” that smoke-free workplace legislation reduces respiratory and/or sensory symptoms (based on two high quality and two medium quality studies). Four studies considered respiratory symptoms as an outcome and three sensory symptoms. For intermediate outcomes, there is “strong evidence” that smoke-free workplace legislation reduces smoke exposure and/or cigarette consumption (based on four high quality and one medium quality studies).

Four of the included studies on smoke-free workplace legislation examined the impact of national laws in Europe. Ayres et al. [2009] examined the health changes before and after smoke-free workplace legislation was passed in Scotland in 2006. Allwright et al. [2005] examined the health changes in bar workers resulting from smoke-free workplace legislation in the Republic of Ireland in 2004 and compared them to changes in workers in Northern Ireland (which at the time did not have smoke-free workplace legislation). Larsson et al. [2008] examined the health effects of a smoke-free workplace legislation, enacted in 2005, on bar, restaurant and gaming workers in Sweden. The fourth European study examined the impact of a Spanish federal smoke-free workplace legislation that exempted hospitality venues from following the ban. Firms were given the choice to choose a smoke-free policy, a partial smoking restriction, or have no restrictions at all. The authors compared exposure and health effects following the ban with those in hospitality workers in Portugal and Andorra, which had no smoke-free workplace legislation.

### TABLE III. Evidence Synthesis Algorithm

| **Strong evidence** |  
| Minimum study quality: high. |
| Minimum number of studies: three. |
| Consistency criteria: if there are only three high-quality studies, all of them must report consistent findings. If there are four or more high-quality findings, all of them must report consistent results unless there is a specific methodological reason that could explain a divergent result. The majority (>50%) of medium-quality studies must concur with the findings from the high-quality studies. |
| If the above criteria are not met, then the criteria for establishing moderate evidence are applied. |

| **Moderate evidence** |  
| Minimum study quality: medium or less than three high-quality studies. |
| Minimum number of studies: three; they can be a mixture of medium- or high-quality studies. |
| Consistency criteria: at least three studies must report consistent findings, and the majority (>2/3) of all the studies must report consistent findings. |
| If the above criteria are not met, then the criteria for establishing limited evidence are applied. |

| **Limited evidence** |  
| Minimum study quality: medium. |
| Minimum number of studies: one. |
| Consistency criteria: fewer than three studies report consistent findings, with the majority (>50%) of the studies reporting consistent findings. |
| If the above criteria are not met, then there is no evidence or mixed evidence. |

| **No evidence** |  
| No high- or medium-quality studies are available from which to draw conclusions. |

| **Mixed evidence** |  
| The findings from medium- and high-quality studies are contradictory. |

### TABLE IV. Titles and Abstracts Identified by Source

<table>
<thead>
<tr>
<th>Source</th>
<th>N</th>
<th>Source</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medline</td>
<td>3,450</td>
<td>Sociological Abstracts</td>
<td>195</td>
</tr>
<tr>
<td>EMBASE</td>
<td>4,190</td>
<td>Wilson Social Science Abstracts</td>
<td>184</td>
</tr>
<tr>
<td>PsycINFO</td>
<td>733</td>
<td>Index to Legal Periodicals</td>
<td>149</td>
</tr>
<tr>
<td>ABI Inform</td>
<td>4,000</td>
<td>Hand-search: Policy and Practice in Health and Safety</td>
<td>19</td>
</tr>
<tr>
<td>ASSIA</td>
<td>85</td>
<td>Content experts</td>
<td>19</td>
</tr>
<tr>
<td>EconLit</td>
<td>279</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Three of the studies in this cluster were undertaken in North America. One study, that of Bondy et al. [2009], examined the impact on bar workers of a municipal smoke-free workplace bylaw in Toronto, Canada in 2004. The authors compared secondhand smoke exposure of bar workers in Toronto to the exposure of workers in Windsor, Canada, a nearby city without smoke-free workplace legislation. The final smoke-free workplace legislation study, that of Moskowitz et al. [2000], considered legislative changes in California. It differs from the other studies in that it looked at the impact of local ordinances on smoking cessation rates and workplace smoking policies.

**Inspection sequence**

There were four studies in this cluster, all of which were undertaken in US jurisdictions. The studies considered how inspection sequence influences compliance rates based on changes in compliance violations cited. For this cluster, there is “moderate evidence” that the first inspection has the largest impact on compliance rates for all violations, with subsequent inspections having a declining impact (based on two high quality and two medium quality studies). All four studies had similar findings for all and for serious compliance violations. Ko et al. [2010] also considered four different time periods within 1972–2006 range and found similar results across the periods. As all studies used micro level data on the plant/site, this relationship would be considered specific deterrence.

**Inspection activity: General deterrence of inspections and penalties**

This cluster considers the impact of the probability of inspections through aggregate/industry levels inspection activity. There are three studies in this cluster, all of medium quality. Studies considered the final outcomes of lost-time injury rate, lost workdays, and the fatality rate. Based on this cluster, there is “limited evidence” of no general deterrence effect on lost-time injuries at the aggregate level (based on three studies—two studies found no effect and one study found an effect). There is “limited evidence” of a general deterrence on fatalities and lost workdays (based on one study for each outcome). The three studies were from North American jurisdictions; two from the US [Scholz and Gray, 1990; Ruser and Smith, 1991], and one from Alberta, Canada [Auld et al., 2001]. One study used aggregated data [Auld et al., 2001], and two used micro data at the firm level [Scholz and Gray, 1990; Ruser and Smith, 1991].

**Inspection activity: Specific deterrence of inspections with/without penalties**

This cluster has the largest number of studies, with 13 studies. Most considered specific deterrence in the form of inspections with/without penalties, while some also considered other types of specific enforcement activities such as consultations or details of the inspection activity such as programmed versus complaint inspections and the value of fines imposed. One study focused on early versus late inspections [Ruser and Smith, 1991], and another focused on compliance reviews in the transportation sector [Chen, 2008]. All considered the effects of the policy lever on final outcomes.

This cluster provides “strong evidence” that specific deterrence from inspections with penalties reduces final outcomes (based on nine studies, six of which were high quality, and three of medium quality). Outcomes considered in the studies include all injuries, health care only injuries, lost-time injuries, workdays lost, musculoskeletal disorders (MSD) and non-MSD injuries, and restricted activity days. All studies found a deterrence effect on most final outcomes.

The cluster provides “moderate to limited evidence” of no effect from specific deterrence of inspections without penalties on final outcomes (based on nine studies, six of which were high quality, and three of medium quality). A range of outcomes was considered in these studies, including
<table>
<thead>
<tr>
<th>Cluster</th>
<th>No. of studies by outcome</th>
<th>Study, jurisdiction, sector</th>
<th>Evidence synthesis</th>
</tr>
</thead>
</table>
| **Introduction of OHS legislation**                                    | Final outcomes (six studies); injuries and fatality | • Davies et al. [2008]: British Columbia, Canada; forestry  
• Arocena and Nunez [2009]: Spain; manufacturing  
• Bulzacchelli et al. [2007]: United States; manufacturing  
• Finger and Gamper-Rabindran [2013]: United States; chemical manufacturing  
• Lewchuk et al. [1996]: Ontario, Canada, manufacturing  
• Monforton and Windsor [2010]: United States; mining                   | Moderate evidence that the introduction of OHS legislation has an effect on final outcomes |
|                                                                       | Intermediate outcomes (three studies); exposure and compliance | • Foley et al. [2009]: Washington State, United States; multiple sectors  
• LaMontagne et al. [2004]: United States; health care  
• Ramsey and Glenn [1996]: Tennessee; United States, health care  | Limited evidence that the introduction of legislation improves intermediate outcomes |
| **Introduction of smoke-free workplace legislation**                    | Final outcomes (one study); respiratory and sensory symptoms | • Ayres et al. [2009]: Scotland; bar/restaurant  
• Allwright et al. [2005]: Republic of Ireland; bar/restaurant  
• Fernandez et al. [2009]: Spain; hospitality  
• Larsson et al. [2008]: Sweden; bar/restaurant  
• Allwright et al. [2005]: Republic of Ireland; bar/restaurant  
• Bondy et al. [2009]: Ontario, Canada; bar/restaurant  
• Fernandez et al. [2009]: Spain; hospitality  
• Larsson et al. [2008]: Sweden; bar/restaurant  
• Moskowitz et al. [2000]: California, United States; multiple       | Moderate evidence that smoke-free workplace legislation improves final outcomes |
|                                                                       | Intermediate outcomes (two studies); smoke exposure, cigarette consumption | • Gray and Jones [1991b]: United States; manufacturing  
• Weil [2001]: United States; manufacturing  
• Gray and Jones [1991b]: United States; manufacturing  
• Ko et al. [2010]: United States, manufacturing                       | Strong evidence that smoke-free workplace legislation improves intermediate outcomes |
| **Inspection sequence**                                                 | Intermediate outcomes (four studies); compliance | • Auld et al. [2001]: Alberta, Canada, construction  
• Ruser and Smith [1991]: United States, manufacturing  
• Scholz and Gray [1990]: United States, manufacturing  | Moderate evidence that the first inspection results in the largest improvement in compliance |
| **Inspection activity: general deterrence of inspections and penalties** | Final outcomes (three studies); lost-time injuries | • Auld et al. [2001]: Alberta, Canada, construction  
• Ruser and Smith [1991]: United States, manufacturing  
• Scholz and Gray [1990]: United States, manufacturing  | Limited evidence (three studies) of no general deterrence effect on lost-time injuries  
Limited evidence of a general deterrence effect on fatalities (one study) and lost workdays (one study) |
<table>
<thead>
<tr>
<th>Cluster</th>
<th>No. of studies by outcome</th>
<th>Study, jurisdiction, sector</th>
<th>Evidence synthesis</th>
</tr>
</thead>
</table>
| Inspection activity: specific deterrence of inspections with/without penalties | Final outcomes (13 studies): injuries and truck crashes | - Foley et al. [2012]: Washington State, United States; multiple sectors  
- Gray and Scholz [1991]: United States; manufacturing  
- Gray and Scholz [1993]: United States; manufacturing  
- Gray and Mendeloff [2005]: United States; manufacturing  
- Haviland et al. [2010]: Pennsylvania, United States; manufacturing  
- Haviland et al. [2012]: Pennsylvania, United States; manufacturing  
- Levine et al. [2012]: California, United States; high risk industries  
- Mendeloff and Wayne [2005]: United States, manufacturing  
- Nelson et al. [1997]: Washington State, United States; construction  
- Ruser and Smith [1991]: United States; manufacturing  
- Scholz and Gray [1990]: United States; manufacturing  
- Scholz and Gray [1997]: United States; manufacturing  
- Chen [2008]: United States; trucking | Strong evidence (nine studies) that specific deterrence from inspections with penalties reduces final outcomes  
Moderate to limited evidence (nine studies) of no effect from specific deterrence of inspections without penalties on final outcomes  
Limited evidence (one study) that specific deterrence in a compliance review of motor safety performance reduces truck crashes |
| Nature of enforcement: Consultative activities | Final outcomes (three studies): injuries | - Baggs et al. [2003]: Washington State, United States; multiple sectors  
- Foley et al. [2012]: Washington State, United States; multiple sectors  
- Hogg-Johnson et al. [2012]: Ontario, Canada; manufacturing | Strong evidence/limited evidence that consultative activity has no effect on final outcomes |
| Nature of enforcement: Autonomy supportive inspection style | Intermediate outcomes (one study): compliance | - Burston et al. [2010]: Alberta, Canada; multiple sectors | Limited evidence that an autonomy-supportive inspector style reduces visits to achieve compliance |
- Morantz [2009]: United States, construction | Limited evidence that state (versus federal) enforcement results in lower fatalities and higher injuries |
| Awareness campaigns | Final outcomes (one study): injuries | - Mancini et al. [2005]: Italy, metal workers  
- Gadomski et al. [2006]: New York State, United States; agriculture (children)  
- Gadomski et al. [2006]: New York State, United States; agriculture (children)  
- Bjorkdahl et al. [2008]: Sweden; multiple sectors  
- Stokols et al. [2001]: California, United States; multiple sectors | Limited evidence that awareness campaigns improve final outcomes |
| Awareness campaigns | Intermediate outcomes (two studies): awareness and compliance | | Moderate evidence that awareness campaigns improve intermediate outcomes |
all injuries, health care only injuries, lost-time injuries, workdays lost, MSD and non-MSD injuries, and restricted activity days. Some studies found deterrence effects [Foley et al., 2012; Levine et al., 2012], while others found effects only under specific conditions such as in fixed-site industries [Scholz and Gray, 1990], inspections initiated by a worker [Scholz and Gray, 1997], and inspections that were not superficial record checks [Gray and Scholz, 1991], or found a smaller effect than with citations [Nelson et al., 1997]. This inconsistency is the reason for the dual level of evidence. Some studies in this cluster considered inspections without identifying whether there were citations or penalties [Ruser and Smith, 1991; Nelson et al., 1997; Levine et al., 2012].

One study is treated separately in that it focused on truck crashes as an outcome in the transportation sector. This study alone provides “limited evidence” of specific deterrence from a compliance review of motor safety performance with regard to reduced truck crashes.

**Nature of enforcement: Consultative activity**

Three studies are in this cluster, all of which are rated high quality. The three considered final outcomes. For this cluster there is either “strong evidence or limited evidence” that consultative activity has no effect on injury rates (based on three high quality studies). The reason for the two-sided synthesis profile is that one study in this group, that of Foley et al. [2012], found mixed results in which one outcome was significant and negative for fixed site locations (lost workday for non-MSD claims) and several other outcomes were significant and negative for non-fixed site locations. The three high quality studies in this cluster included two based on data from Washington State [Baggs et al., 2003; Foley et al., 2012], and one Canadian study from Ontario [Hogg-Johnson et al., 2012].

**Nature of enforcement: Autonomy supportive inspection style**

The one study in this cluster considered the intermediate outcome of compliance. For this cluster there is “limited evidence” that an autonomy-supportive style (e.g., one in which an inspector provides a rationale and choices versus deadlines and pressure) reduces the number of visits to achieve compliance (based on one high quality study). The one study in this cluster is based on data from Alberta [Bursten et al., 2010].

**Nature of enforcement: State- versus federal-level enforcement**

Two US studies examined whether state enforcement was more or less effective than federal level enforcement. The premise of these studies was that different styles of enforcement may be provided by different levels of government. The two studies in this cluster are both of medium quality and both considered final outcomes. For this cluster there is “limited evidence” that state enforcement results in lower fatality rates compared to federal enforcement (based on two medium quality studies). One of the studies also considered injury rates and found a positive and significant result, suggesting that state enforcement is associated with higher injury rates compared to federal enforcement.

**Awareness campaigns**

The studies included in this cluster are in different sectors and focused on different OHS risks. There are four studies in this cluster, two of which considered final outcome and three which considered intermediate outcomes. For the former group, there is “limited evidence” that awareness campaigns reduce injuries (based on two medium quality studies). For the latter group, there is “moderate evidence” that awareness campaigns improve compliance (based on three medium quality studies). Among the four studies, one evaluated an eye injury campaign for metal workers in Italy [Mancini et al., 2005], a second evaluated a noise awareness campaign in Sweden [Bjorkdahl et al., 2008], a third evaluated a child labor in agriculture campaign in New York [Gadomski et al., 2006], and a fourth evaluated a train-the-trainer program in California [Stokols et al., 2001].

**DISCUSSION**

Our findings have important implications for both policy and research. On the policy side, the finding that several legislative and regulatory policy levers are effective in reducing injuries and/or increasing compliance provides evidence for supporting such activities. Specifically, among nine clusters (some with sub-categories) we found strong evidence in three clusters and moderate evidence in five. In terms of generalizability, the introduction of OHS legislation cluster spanned several sectors in the United States, Canada and Spain, and so may be broadly applicable to other developed countries in North America, Europe, and Australia. The introduction of smoke-free workplace legislation cluster also spanned several countries, and therefore may be broadly applicable to developed countries, particularly in the hospitality sector. Two clusters have studies exclusively from the United States, and largely in manufacturing—that of inspection sequence and specific deterrence of inspections—and thus may be less generalizable to other sectors and countries. The consultative activity cluster and the awareness campaign one have two countries in each and with studies undertaken in multiple sectors.
suggested they may be generalizable to other developed countries.

The strong evidence of an effect from actual inspections with penalties and moderate to limited evidence of no effect from inspections without penalties reinforces the importance of regulators being out in the field identifying and citing/penalizing non-compliance. The limited evidence for a general deterrence effect found in our review is consistent with this interpretation. The literature review by Tompa et al. [2007] had similar findings with regards to general and specific deterrence and their impact on injury outcomes. Our update with more recent studies [e.g., Gray and Mendeloff, 2005; Mendeloff and Wayne, 2005; Haviland et al., 2010; Foley et al., 2012] adds to this evidence, and the addition of studies with compliance as an outcome [Gray and Jones 1991a,b; Weil, 2001; Ko et al., 2010] provides incremental support for this finding. Essentially, firms may not have the capacity to digest information about inspection activities in the field; they may only react when the adverse experience of an inspection with citations/penalties is first hand.

Bluff [2011] refers to competing theory of firm behavior to explain why general deterrence may not be effective. For it to be effective, firms would need to be rational, long-run optimizers, and knowledgeable about the probability and the financial implications of being inspected, whereas in reality, firms may have bounded rationality and have limited capacity to process information. If this is the case, regulators need to heighten awareness in the field by actively communicating the consequences of non-compliance, and possibly make information about non-compliers easily available to the general public. Focused awareness campaigns and inspection blitzes might also be a way to provide acute awareness on a particular hazard.

Studies that considered compliance as an outcome offer important insights. The moderate evidence that the first inspection has the largest impact on compliance is corroborated across all four studies included in the review [Gray and Jones, 1991a,b; Weil, 2001; Ko et al., 2010] and has important implications for the efficient use of inspectorate resources. If subsequent inspections to a site have substantially lesser impacts, as the literature suggests, then an intensive regime of multiple inspections to a site may not be the best use of resources. Some jurisdictions have attempted such intensive enforcement strategies.

In another cluster in this review, moderate evidence was found that awareness campaigns increase compliance, reinforcing the importance of communicating regulatory obligations to stakeholders. What is not clear is the relationship between compliance and final outcomes. The limited evidence that awareness campaigns reduce injuries would suggest that the relationship is not definitive. That finding is based on only two medium quality studies. Clearly, more research is needed on the relationship between intermediate and final outcomes. In fact, Bluff [2011] emphasizes the importance of better understanding of motivations, attitude, perceptions, and skills in order to determine how particular strategies, mechanisms, and approaches can be best used to achieve compliance and ultimately better final outcomes. A recent review by Safe Work Australia [2013] also attempts to address issue of how and why interventions work.

The findings for consultative activity (strong or limited evidence depending on context) provide some preliminary insights relevant to the move towards voluntary guidelines in some jurisdictions. Of three studies considering final outcomes, all of which were high quality, only one study by Foley et al. [2012] found significant effects in some contexts. The study by Hogg-Johnson et al. [2012] did not find an effect, but noted that not all firms received consultative services as was originally planned in the program, and many only received a “light touch.” The findings in this cluster are quite consistent with the findings that specific deterrence is much more effective than general deterrence, and suggest that consultation in the absence of specific deterrence might be interpreted by organizations as that there are no consequence for non-compliance. Clearly, more research is needed in this area to better understand whether consultations, if implemented in a comprehensive and extensive fashion, have an impact on outcomes. Even more pressing is the need for studies on the effectiveness of voluntary guidelines, since no studies were identified on this topic. Another avenue of research could investigate the mix of policy levers that are most effective when used together.

The moderate evidence of an effect on final outcome from the introduction of OHS legislation suggests, at face value, that legislation may not always be the best approach to addressing new and emerging health and safety issues. The limited evidence for an impact on intermediate outcomes further reinforces this interpretation. But the studies in this cluster were quite heterogeneous in the type of legislation being introduced, and this may be the reason for the limited impact. They were also incremental to an existing broad legislative framework and related regulatory enforcement, and were designed to increase protection related to a specific hazard. This is different from the introduction of broad OHS legislative frameworks in the 1970s and 1980s in many developed countries that were the subject of effectiveness studies thereafter. Another issue is that such studies need to consider a longer measurement time period following the introduction of legislation in order to capture the long-run impact. Essentially awareness, compliance and ultimately injury outcomes may take more time to improve than might have been expected by researchers and regulators.

Clearly, given the right context, the introduction of legislation can be effective, as was the case with smoke-free workplace legislation. Looking at this particular example,
some lessons might be learned that can be generalized to other areas. Possibly timing, public sentiment, and a broad awareness of the serious health implications may be some of the important ingredients. Also noteworthy were the concerns on the part of the restaurant and entertainment industries with regards to the implication of such legislation on business, concerns which were not vindicated. Also important was the “across the board” nature of the legislation in most jurisdictions. In fact, in the one study in Spain where a choice was available between total, partial, or no restriction [Fernandez et al., 2009], exposure reductions to second hand smoke varied from substantial for total ban establishments, modest for partial ban establishments, to inconsequential for no ban ones.

The number of studies on the effectiveness of OHS regulation has increased notably since the review by Tompa et al. [2007] as has the quality. More studies are using micro-level data and robust statistical methods to address industry- and organizational-level behavioral responses to regulation and its enforcement. Earlier, less structured reviews that drew on an older evidence base noted quality concerns and concluded that, overall, the evidence suggests OSHA has resulted in only a modest improvement in workplace health and safety in the United States [Kralj, 2001; Mendeloff, 2001; Thomason, 2001].

In our review, the criteria of publication since 1990 onward eliminated some of the weaker (and older) studies. Our expansion of the inclusion criteria to intermediate outcomes has allowed us to explore a broader and richer literature, such as the nature of enforcement and policy levers such as awareness campaigns. The inclusion of studies evaluating smoke-free workplace legislation provided an example of successful introduction that might provide lessons for regulation in other OHS areas.

Regarding the way forward for research, we would encourage policymakers and researchers to work together to build in policy evaluation, particularly with the introduction of new legislation, changes in enforcement strategies, and the roll out of awareness campaigns. This would lend itself to better planned study designs, in some cases the possibility of randomization or staggered introduction. Also, longer measurement time periods may be needed with new legislation in order to ensure there is time for stakeholders to become aware of changes and respond accordingly. The effectiveness of voluntary guidelines is a relatively uncharted area that urgently needs exploration.

More exploration is also needed of the context and conditions for successful legislation and policy. Related to this latter issue, the construct of how OHS policy levers create incentives for organizations to improve OHS processes and outcomes has direct implications for how programs to address OHS within organizations are arranged and implemented. It is noteworthy that such programs themselves have been the subject of systematic reviews [e.g., Robson et al., 2007]. Consideration of the effectiveness of such programs and how they are affected by policy levers and other environmental factors is an area warranting investigation.

**CONCLUSIONS**

There is a substantial body of evidence on the effectiveness of legislative and regulatory policy levers at improving intermediate and final outcomes. We identified strong evidence of the following: (i) specific deterrence from inspections with penalties results in a decrease in injuries; (ii) consultative activity has no effect on injury outcomes with some exceptions; and (iii) the introduction of smoke-free workplace legislation reduces exposure to second hand smoke. We identified moderate evidence of the following: (i) a first inspection has the largest impact on compliance rates; (ii) specific deterrence from inspections without penalties has no effect on injuries except in particular contexts; (iii) awareness campaigns improve compliance; (iv) the introduction of OHS legislation as no effect on injury outcomes; and (v) the introduction of smoke-free workplace legislation reduces respiratory and/or sensory symptoms.

This study adds substantially to the evidence base identified in an earlier review. In addition to new evidence supporting previous study findings, it included new categories of evidence–compliance as an outcome, nature of enforcement, awareness campaigns, and workplace smoking legislation. The evidence is of value for informing policy decision making in the OHS field, and provides insights into areas warranting further exploration in future research.

**AUTHORS’ CONTRIBUTIONS**

Emile Tompa, Principal Investigator: Provided oversight of the review, participated in the design of the study and all phases of the systematic review process, and was the lead pen of the manuscript. Christina Kalcevich, Project Co-ordinator: Coordinated all aspects of the review, participated in the design of the study and all phases of the systematic review process, prepared materials for meetings and presentations, drafted the methods section of the manuscript, and reviewed drafts of the manuscript. Michael Foley, Co-investigator: Participated in the design of the study and all phases of the systematic review process, and reviewed drafts of the manuscript. Chris McLeod, Co-investigator: Participated in the design of the study and all phases of the systematic review process, and reviewed drafts of the manuscript. Sheilah Hogg-Johnson, Co-investigator: Participated in the design of the study and all phases of the systematic review process, and reviewed drafts of the manuscript. Kim Cullen, Co-investigator: Participated in the design of the study and
all phases of the systematic review process, and reviewed drafts of the manuscript. Ellen MacEachen, Co-investigator: Participated in the design of the study, the title and abstract and article inclusion selection process, and reviewed drafts of the manuscript. Quenby Mahood, Co-investigator: Participated in the design of the study, provided library search and systematic review guidance, and reviewed drafts of the manuscript. Emma Irvin, Co-investigator: Participated in the design of the study and all phases of the systematic review process, provided library search and systematic review guidance, and reviewed drafts of the manuscript.

ACKNOWLEDGMENTS

We thank our advisory committee members—Wayne De L’Orme, Anne Duffy, Anne-Marie Feyer, Steve Mantis, Carol Sackville-Duyvelshoff, Barbara Silverstein, Carmine Tiano, David Walters, and Michael Zachs—for their valuable guidance at key junctures during the process of completing this review.

FUNDING

Funding for this review was provided by the Institute for Work and Health, which is an independent, not-for-profit research institute. The Institute receives funding from the Prevention Office of the Ministry of Labour in Ontario, Canada.

ETHICS APPROVAL AND INFORMED CONSENT

Ethics review and approval was not sought for this systematic literature review study, as it does not involve human subjects or animals.

DISCLOSURE (AUTHORS)

The authors declare that there are no conflicts of interest.

DISCLOSURE BY AJIM EDITOR OF RECORD

Rodney Ehrlich declares that he has no competing or conflicts of interest in the review and publication decision regarding this article.

REFERENCES


Gadomski A, Ackerman S, Burdick P, Jenkins P. 2006. Efficacy of the North American guidelines for children’s agricultural tasks in


SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher’s web-site.