

Systematic review of the prevention incentives of insurance and regulatory mechanisms for occupational health and safety

by Emile Tompa, PhD,^{1,2} Scott Trevithick, MA,^{1,3} Chris McLeod, MA⁴

Tompa E, Trevithick S, McLeod C. Systematic review of the prevention incentives of insurance and regulatory mechanisms for occupational health and safety. Scand J Work Environ Health 2007;33(2):85–95.

Objectives The objective of this study was to determine the strength of evidence on the effectiveness of two policy levers—the experience rating of workers' compensation insurance and the enforcement of occupational health and safety regulation—in creating incentives for firms to focus on health and safety issues.

Methods An extensive systematic literature review was undertaken in an effort to capture both published and grey literature studies on the topic. Studies that met specific subject-matter and methods criteria underwent a quality assessment. A qualitative approach to evidence synthesis, known as “best-evidence” synthesis, was used. This method ranks the strength of evidence on a particular topic on the basis of the number, quality, and consistency of studies on the topic.

Results There was moderate evidence that the degree of experience rating reduces injuries, limited to mixed evidence that inspections offer general and specific deterrence and that citations and penalties aid general deterrence, and strong evidence that actual citations and penalties reduce injuries.

Conclusions Although experience rating is a key policy lever of those providing workers' compensation insurance, there is much to be learned about its merits. Few studies have concerned the topic, and most have used crude proxy measures or exploited natural experiments. There have been many more studies on the merits of regulation enforcement, even though here too measures were often crude. Nonetheless, this synthesis indicates that general deterrence is less effective in reducing injury incidence and severity, whereas specific deterrence with regard to citations and penalties does indeed have an impact.

Key terms citation; experience rating; inspection; occupational health and safety regulation; penalty; workers' compensation.

There are two broad avenues by which public policy attempts to influence the prevention of injuries⁵ in the workplace: experience rating of workers' compensation insurance premiums and enforcement of occupational health and safety (OHS) regulations. The empirical literature on these policy levers is large, and diverse statistical methods, different levels of data aggregation, and samples from various time periods and jurisdictions are used. As a result, it is difficult to compare and contrast the evidence and quality of evidence provided by different studies. Consequently, synthesizing the evidence for the purposes of informing policy is a formidable task.

However, the effort is warranted given that it is a highly significant area of research and policy making. To date, Kralj (1), Hyatt & Thomason (2), Mendeloff (3), and Thomason (4), among others, have reviewed parts of this diverse literature, but none have employed systematic review methodology. To our knowledge, this is the first attempt to perform a systematic review of this literature and one of very few attempts at systematic review in the economics literature in general.

Our review focuses on employer behavior, as this feature constitutes an important target for policies that aim at reducing injuries in the workplace. We consider

1 Institute for Work & Health, Toronto, Canada,

2 Department of Economics, McMaster University, Hamilton, Canada,

3 Department of History, University of Toronto, Toronto, Canada,

4 Centre for Health Services and Policy Research, University of British Columbia, Vancouver, Canada.

5 In this article we use the term “injuries” to refer to both injuries and illness unless expressly referring to both in order to make a distinction between the two.

studies investigating workplace injury experiences or proxies for these experiences, such as claims activity, and firm reporting of injuries. In addition to providing behavioral incentives for employers, some insurance and regulatory design features also provide behavioral incentives for workers. For example, increasing the wage replacement rate of workers' compensation benefits may encourage prevention on the part of employers while also providing an incentive for workers to file claims. A large amount of empirical literature examines the relationship between work injury claims and benefits or wage replacement rates. However, this literature is also beyond the scope of this review.⁶ The two key workers' compensation features that we review in this study are the introduction of experience rating of insurance premiums and the varying of the degree of experience rating. The two key aspects of OHS regulation reviewed are the introduction of regulation and the enforcement of regulation through inspections, citations, and penalties.

This paper proceeds as follows. In the next section we discuss the theoretical underpinnings of, and modeling and measurement issues in, the empirical literature on insurance and regulatory mechanisms. This section is followed by a detailed account of the methodology we used to search for, select, evaluate, and synthesize the evidence on the features of interest. In the subsequent section, we present our results, summarizing the number and quality of studies identified for each feature, as well as presenting the level of evidence found in the synthesis of these studies. We conclude with a discussion of the policy implications of our findings and recommendations for future research.

Behavioral theories, modeling and measurement

Behavioral theories

Workers' compensation insurance provides ex-post compensation for certain types of costs and losses incurred by workers who have sustained a work-related injury. With the introduction of compulsory insurance coverage in many jurisdictions, workers gave up the right to sue employers for financial compensation for work-related injury costs and losses. In turn, they received entitlement to no-fault insurance paid for by employers. Different

jurisdictions allow for coverage through various providers, usually some mix of private insurance, public insurance, and self-insurance.

Compulsory workers' compensation insurance through a third party can give rise to firm-level incentive problems. Specifically, if premiums are based solely on the average risk in the industry (sometimes described as a manual rating or pooled risk systems), then there is little incentive for firm-level preventative efforts. Insurance providers (whether public or private) typically attempt to encourage prevention efforts by tying a firm's insurance premiums to its claims activity. This approach is typically described as experience rating. Most providers allow for varying degrees of experience rating, often based on firm size, with a certain amount of risk pooling.

Another avenue governments often use to encourage firms to focus on injury prevention is OHS regulation.⁷ Prevention incentives are created by the enforcement of compliance with regulation through inspections, citations, and penalties. In some programs, firms may be randomly inspected, whereas, in others, particular firms are targeted in particular industries (eg, poor performing firms in high risk industries). Targeting is generally undertaken to increase the efficiency of inspectorate field time.

Both experience rating of insurance and enforcement of regulation can give rise to perverse incentives. For example, a firm may attempt to reduce claim costs by discouraging workers from making claims, appealing an excessive number of claims, and pressuring workers to return to work before it is safe to do so (6, 7). OHS regulation may give rise to the falsification of injury logs in jurisdictions in which targeting is based on industry performance or a firm's injury track record (8), disputing legitimate citations and penalties, and lobbying politicians and the public sector to reduce the power of regulatory authorities. This latter behavior is sometimes described as regulatory capture.

The ideal circumstances for effective regulation may not always be met and, as a result, outcomes may be less than optimal. One possible reason for less than optimal outcomes is that regulation may not address the root causes of injuries or it may address only some of the causes (9, 10). A second reason is that a regulator's ability to detect or punish noncompliance may not be sufficient. For example, inspectors may be limited in their

⁶ See Fortin & Lanoie (5) for a comprehensive review of this literature, which includes studies of cost shifting between social programs (eg, between unemployment insurance and workers' compensation) due to differential wage-replacement rates and shifting due to incentives such as health care coverage provided by workers' compensation in jurisdictions that do not have universal health care coverage.

⁷ The promotion of internal responsibility systems, often focused on the obligation of firms to have joint health and safety committees, might be thought of as a third category of public sector involvement in prevention (4). A similar arrangement, known as internal compliance programs, has been gaining popularity in the United States.

ability to detect noncompliance due to unfamiliarity with a firm's operation, time constraints imposed on an inspection, or a lack of experience. Even if inspections are effective, a regulatory authority may be restricted in its ability to enforce compliance due to a limited budget or the legal bureaucracy involved in punishment.

A third possible reason is that the threat of punishment may not be an effective deterrent for noncompliance. It may be that the penalties are not severe enough to encourage compliance (ie, the expected penalties are less than the cost of compliance),⁸ or firms may not have full information on the probability of inspection. More fundamentally, it may be that firms are not always risk neutral, do not always act rationally, or do not always seek to maximize profit (a key element for insurance and regulation being financial motivators) at the expense of other goals. Consequently, an actual inspection (specific deterrence) may provide an incremental incentive over and above the threat of inspection (general deterrence).

Some researchers have suggested that punishment through citations and penalties might actually create antagonism, rather than provide an incentive for compliance, and have suggested that fostering trust and cooperation may work better (12, 13). Consistent with this perspective, there has been a shift towards voluntary approaches in the United States, including an increased emphasis on consultations, although there are few studies on the topic and hence little evidence exists to support their effectiveness (14–16).⁹ Still other researchers have suggested that citations and penalties may be ineffective because they divert attention away from important health and safety concerns to the intermediate issue of compliance (10). Shapiro & Rabinowitz (14) have proposed a selective mix of cooperative and punitive enforcement strategies in an effort to distinguish between good and bad firms and between admirable efforts for compliance and flagrant disregard for regulations.

Modeling and measurement

Several modeling and measurement issues arise in analyses of the relationship between behavioral incentives

and outcomes. A key concern pertains to the limitations of a data set. The most common type of data used in empirical studies is administrative (eg, workers' compensation claims). This type of data can be prone to reporting bias. Underreporting is particularly a concern in jurisdictions in which claims can affect a firm's premiums or the probability of inspection. Furthermore, underreporting may be more heavily concentrated among certain types of injuries (eg, long-latency and multifactorial injuries). To address this issue, some researchers focus on injury claims and exclude illness claims. Others focus on particular types of injuries that are less prone to reporting bias (eg, death and acute trauma injuries).

In addition to the various insurance and regulatory design features, other contextual factors should be taken into consideration in the modeling (eg, sociodemographic characteristics of the labor force, unionization, capital–labor ratios, industry, stage of the business cycle, and time-period-specific effects). In some cases, explanatory variables may be endogenous (eg, wages may be, in part, determined by the riskiness of jobs, and this riskiness is reflected in the injury experiences of firms; consequently wages are explained by injuries rather than the reverse).¹⁰ The key features of interest can also be endogenous, as is the case with investigations in some jurisdictions that target high-injury firms. Related to this practice is the fact that some firms may experience abnormally high injury rates due to unusual events that are eventually addressed regardless of intervention. This is known as regression to the mean and is sometimes controlled by including the previous year's injury experience as an explanatory variable or by examining the percentage change in injury rates between years.

Study design and feature measurement also have an important bearing on the quality of evidence provided by studies. Before–after studies based on natural experiments may present an excellent opportunity to test the effectiveness of the introduction of a program. However, employing a dummy variable to identify the impact of a program's introduction provides no information about varying the degree of experience rating or the level of regulation enforcement. A count of the number of inspections or number of fines provides more insight,

⁸ Bartel & Thomas (10) provide data to support the proposition that expected penalties are modest—in 1975 the expected fine per violation was USD 0.52. In fact, expected penalties for violating many regulations of the Occupational Safety and Health Administration (OSHA) are far below the cost of compliance (11). In contrast, penalties imposed by the Environmental Protection Agency (EPA) for noncompliance are, on the average, substantially higher due to a larger number of inspectors per regulated establishment and higher fines. Hence, the EPA has historically been taken more seriously by the regulated community.

⁹ Baggs et al (15) and Smitha et al (16) have carried out the only two studies identified in this systematic literature review that included consultations in their empirical analysis. Both studies investigated other features of the regulatory environment, and Baggs et al (15) also investigated the mix of activities.

¹⁰ Endogeneity is most often addressed through instrumentation.

but still falls short since it says nothing about how the intensity of inspections, the types of citations, or the amount of fines affect incentives.¹¹

Another important issue is the temporal sequencing of explanatory and outcome variables. While many studies employ concurrent measures for both, a more intuitively appealing formulation would be to lag some of the explanatory variables. On a practical level, finding a relationship between variables when they are time lagged allows for a stronger inference about the causal relationship.

Methodology

Searches, review and extraction

The study sources included journal articles, books, conference proceedings, working papers, dissertations, and unpublished manuscripts. Specific databases were chosen to capture studies from each of these sources and to cover all disciplinary areas that might have studies meeting the inclusion criteria. The criteria were (i) the study should have investigated a relevant insurance or regulatory feature, (ii) the study should have been quantitative and should have used multiple regression techniques, and (iii) there should be a temporal element to the analysis (ie, the data were not strictly cross-sectional).

Searches were conducted in EconLit, MEDLINE, Sociological Abstracts, Wilson Social Science Abstracts, Ideas, Dissertation Abstracts International, and the University of Toronto Library Catalogue. Although searches began with the generic terms “workers’ compensation” and “occupational health and safety”, in each case, database-specific terminology and wild-card characters were employed to adapt the searches.

Journal and working paper search results were reviewed on the basis of the titles and abstracts of the studies. If a title and abstract were available and deemed a study for potential inclusion, the complete article was retrieved and reviewed for inclusion. If only a title was available and deemed a study for potential inclusion, an abstract, the introduction, or the beginning of the study was retrieved and reviewed before the entire article was reviewed for inclusion. Two reviewers participated in the stages of the process. The title and abstract review

was conducted by one reviewer, with periodic quality checks performed by the second reviewer. Article inclusion reviews were conducted independently by the two reviewers. In cases when a disagreement about inclusion of an article arose, the reviewers met to arrive at a consensus.

Book titles were first reviewed, and, if a book was deemed a potential source of studies for inclusion, it was followed by a review of the introduction and table of contents. Subsequently, the introductions to chapters with titles of interest were reviewed. This initial process was undertaken by one reviewer before the chapter was considered in its entirety for inclusion by both reviewers.

Studies that successfully passed the article review stage were blinded before they entered the data extraction phase. The following items were blinded: authors and their affiliation, publication venue, publication date, and any acknowledgments to individuals or institutions. The data extraction process consisted of two reviewers independently extracting data from each study, followed by periodic meetings to discuss study details and extraction data. The objective of the meetings was to arrive at a consensus on the data to be extracted and to verify the details of the study. The principal issue of discussion was the appropriate regression models from which to extract data. As a “rule of thumb”, the most fully specified models or the models deemed by the study authors to be the preferred specification were considered first candidates, although this practice was subject to the reviewers’ discretion. In some cases, article inclusion was also at issue, and, in a few cases, articles were excluded at this stage. Once a consensus was reached on the information to be extracted, the data were logged for final inclusion into the best evidence synthesis.

Quality assessment and evidence synthesis

Several issues related to measurement and statistical modeling were considered in the quality assessment of the studies. Since some of these issues are unique to the economics literature, quality assessment instruments developed for systematic reviews in the clinical literature could not be adopted in their entirety. Instead, we developed our own instrument based on several of the instruments used in the health sciences, on suggestions provided by Levine et al (17), on advice from our clinical colleagues supporting the Cochrane Back Review Group, and on our own knowledge of the nature

¹¹ Data limitations can also oblige a researcher to rely on less-than-perfect proxies. For example, in many studies in the United States information on experience rating at the firm level is not available to researchers due to the proprietary nature of insurance data; therefore, some researchers have proxied for the degree of experience rating using firm size or the interaction between firm size and wage replacement rates.

of studies in this literature. Our instrument had two components: part I (questions 1 through 6) contained questions focusing on the overall quality of the study and part II (questions 7 and 8) contained questions on the measurement validity and generalizability of the features of interest. Both sets of questions employed 5-point Likert scales. [See appendix 1 for a complete version of the quality assessment instrument.]

Both reviewers assessed each study using the quality assessment instrument. The evaluations were carried out individually, and the reviewers met to discuss their ratings of each study. The goal of this process was not to produce a consensus, but rather to ensure that the evaluations were based on a sound consideration of all relevant aspects of the studies. The reviewers had the opportunity to change their ratings at, or after, the meetings. After the individual reviewer's scores were finalized, the overall scores for the articles were obtained by averaging the two reviewers' scores. The final quality rating for the evidence of a relationship between a particular feature and outcome was based on the lower of the two scores (ie, the overall study score and the score for the feature). Scores were clustered into the following three quality rating categories: high (score of $\geq 70\%$), medium (score between 50% and 69%), and low (score below 50%). Only medium- and high-quality studies were retained for the evidence synthesis, which is a common practice in systematic reviews in the clinical literature. We followed this practice, on the premise that the inclusion of low-quality studies would dilute the value of the evidence synthesis (18).

We used Slavin's (19, 20) best-evidence synthesis approach, which is qualitative. Best-evidence synthesis assumes that the strength of a relationship varies depending on the quantity and quality of the evidence available to support a relationship between variables. It aims to provide the same methodological rigor to evidence synthesis as is used in meta-analyses by clearly and concisely articulating the synthesis criteria.

We ranked the evidence supporting the hypothesized relationship on a 5-level scale consisting of strong evidence, moderate evidence, limited evidence, no evidence, and mixed evidence. Evidence on a feature was tested against the criteria for the highest level (strong evidence), and, if they were not met, the criteria for the next highest level (moderate evidence) were considered. If they were 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. [The evidence-ranking algorithm can be found in appendix 2.]

Results

The main workers' compensation features identified in the eligible studies were (i) the introduction of experience rating and (ii) the degree of experience rating, whereas the main OHS features identified were (i) the introduction of regulations, (ii) the enforcement of regulations through inspections, and (iii) the enforcement of regulations through citations and penalties.

We synthesized evidence on the significance of the relationship between the preceding list of features and the outcomes of frequency and severity of injuries. Table 1 contains a list of studies by feature.

Studies providing evidence on the introduction of experience rating were based on natural experiments with before and after comparisons. A limited number of studies investigated this phenomenon. Of the six identified, five were of high or medium quality (2, 21–24) and were therefore included in a synthesis. [See table 2, to be found on the homepage of the *Scandinavian Journal of Work Environment & Health*, for details on these studies.] On the basis of these five studies (two of high quality and three of medium quality), we found moderate evidence that the introduction of experience rating was associated with a reduction in the frequency of injuries. One high-quality study and two medium-quality studies supported this hypothesis. One high-quality study and one medium-quality study provided qualified support in that the frequency of injuries decreased, although the severity increased for some or all types of injuries.

The evidence on the effectiveness of the degree of experience rating is based primarily on studies that employed various proxies for the degree of experience rating (eg, firm size interaction with benefit replacement rates), although one study employed a more direct measure—surcharges and rebates. Once again, few studies investigated this phenomenon. Of the five studies (25–29) included in the synthesis, all five were of medium quality. [See table 3, to be found on the homepage of the *Scandinavian Journal of Work Environment & Health*, for details on these studies.] On the basis of these five studies, we found moderate evidence that the degree of experience rating was associated with a reduction in the frequency or severity of injuries. Three studies provided consistent support for this relationship, while two studies provided support for only some measures.

As in the case with the introduction of experience rating, the evidence on the effectiveness of the introduction of OHS regulation was based on studies employing data from natural experiments. There were very few studies on this phenomenon, and only two were included in the synthesis (30, 31) [See table 4, to be found on the homepage of the *Scandinavian Journal of Work Environment & Health*, for details on these studies.] On the basis of two medium-quality studies, we found

Table 1. Overview of the studies included in the review.

Feature and study	Type of injury	Quality
Features related to workers' compensation regulations		
Introduction of experience rating		
Bruce & Atkins, 1993 (21)	Frequency	High
Hyatt & Thomason, 1998 (2)	Severity and severity	High
Cheadle et al, 1994 (22)	Severity	Medium
Deland, 1995 (23)	Frequency	Medium
Kralj, 1995 (24)	Severity	Medium
Degree of experience rating		
Kotz & Schafer, 1993 (25)	Frequency and severity	Medium
Lanoie & Strélski, 1996 (26)	Frequency and severity	Medium
Ruser, 1985 (27)	Frequency	Medium
Ruser, 1991 (28)	Frequency	Medium
Ruser, 1993 (29)	Frequency	Medium
Features related to occupational health and safety regulation		
Introduction of regulation		
Curington, 1988 (30)	Frequency	Medium
Lanoie, 1992 (31)	Frequency	Medium
Enforcement of regulation—inspections		
Chung, 1990 (12)	Frequency and severity	High
Gray & Mendeloff, 2005 (32)	Frequency	High
Gray & Scholz, 1993 (11)	Frequency	High
Ruser & Smith, 1991 (33)	Frequency	High
Scholz & Gray, 1990 (34)	Frequency and severity	High
Scholz & Gray, 1997 (13)	Frequency	High
Sha, 1995 (35)	Frequency	High
Smith, 1979 (36)	Frequency	High
Baggs et al, 2001 (37)	Frequency	Medium
Guo, 1999 (38)	Frequency and severity	Medium
Kim, 1991 (39)	Frequency	Medium
Lanoie, 1992 (40)	Frequency and severity	Medium
Lanoie & Strélski, 1996 (26)	Frequency and severity	Medium
McCaffrey, 1983 (41)	Frequency	Medium
Ruser, 1995 (42)	Frequency	Medium
Smitha et al, 2001 (16)	Frequency	Medium
Viscusi, 1979 (43)	Frequency	Medium
Viscusi, 1986 (44)	Frequency and severity	Medium
Enforcement of regulation—citations and penalties		
Gray & Mendeloff, 2005 (32)	Frequency	High
Gray & Scholz, 1991 (45)	Frequency	High
Gray & Scholz, 1993 (11)	Frequency and severity	High
Scholz & Gray, 1990 (34)	Frequency and severity	High
Scholz & Gray, 1997 (13)	Frequency	High
Bartel & Thomas, 1985 (10)	Severity	Medium
Bartel & Thomas, 1985 (46)	Severity	Medium
Guo, 1999 (38)	Frequency and severity	Medium
Kim, 1991 (39)	Frequency	Medium
Lanoie, 1992 (40)	Frequency and severity	Medium
Lanoie & Strélski, 1996 (26)	Frequency and severity	Medium
Robertson & Keeve, 1983 (47)	Frequency	Medium
Sha, 1995 (35)	Frequency	Medium
Smitha et al, 2001 (16)	Frequency	Medium
Viscusi, 1979 (43)	Frequency	Medium
Viscusi, 1986 (44)	Frequency and severity	Medium

mixed evidence that the introduction of OHS regulation was associated with a reduction in the frequency of injuries. Both studies found mostly nonsignificant relationships between the introduction of regulation and injury frequency, some increases in frequency and some decreases. Due to the small number of high- or medium-quality studies examining this association, the results to

date are inconclusive rather than in favor of a consistent lack of a relationship between the introduction of OHS regulation and a decrease in injuries.

Many studies investigated the relationship between inspections and injury frequency or severity. Altogether 18 studies (11–13, 16, 32–44) were included in the synthesis. [See table 5, to be found on the homepage of the *Scandinavian Journal of Work Environment & Health*, for details on these studies.] We have combined the synthesis of articles that investigated the general and specific deterrence of inspections. Some studies also separated inspections into those with and without citations and penalties, whereas others did not make this distinction. Studies in which the distinction was made were included with the evidence on citations and penalties, although in practice citations and penalties are incremental to inspections. The studies examining inspections provided a mixed picture of the association between inspections and injury outcomes. On the basis of 8 high-quality studies and 10 medium-quality studies, we found limited evidence that inspections were associated with a reduction in the frequency or severity of injuries.¹² While five of the high-quality studies found evidence that inspections in previous years reduced current year injuries, three high-quality studies did not find a relationship. Of these three, two looked at inspections without citations and penalties, and one looked at the effect of inspections early in the year versus inspections late in the year. Three medium-quality studies confirmed the relationship, but four medium-quality studies found no relationship, and three medium-quality studies found some evidence that inspections were associated with an increase in injury frequency or severity.

We identified 16 studies (10, 11, 13, 16, 26, 32, 34, 35, 38–40, 43–47) that considered citations or inspections with citations or citations with penalties. [See table 6, to be found on the homepage of the *Scandinavian Journal of Work Environment & Health*, for details on these studies.] These studies investigated the general or specific deterrence of citations or citations with penalties. With these studies, we examined the evidence on general and specific deterrence separately. On the basis of two high-quality studies and nine medium-quality studies, we found mixed evidence that an increase in the probability of being cited or penalized (ie, general deterrence) was associated with a reduction in the frequency or severity of injuries. While the two high-quality studies found a relationship, all nine medium-quality studies did not. In contrast, we found strong evidence that the experience of actually being cited or penalized (ie, specific deterrence) was associated with a reduction in injuries. This finding was based on five high-quality and two medium-quality studies. Three high-quality studies

¹² We include both general and specific deterrence in this synthesis.

provided consistent support, and two high-quality studies and one medium-quality study provided qualified support (eg, the significance of the association changed over time, or it was significant for only some types of citations and penalties). One medium-quality study did not find support for this relationship.

The empirical literature on the impact of workers' compensation and OHS regulation features provided different levels of evidence on the most commonly examined features directed at providing prevention incentives for employers. The following list is a synopsis of the evidence on these features:

- moderate evidence that the introduction of experience rating reduces the frequency of injuries—only some types of injuries may be reduced and severity may increase;
- moderate evidence that the degree of experience rating reduces the frequency or severity of injuries;
- mixed evidence that the introduction of OHS regulations is associated with a reduction in the frequency of injuries;
- limited-to-mixed evidence of the general and specific deterrence of inspections and the general deterrence of citations and penalties on the frequency or severity of injuries;
- strong evidence that actual citations and penalties reduce the frequency or severity of injuries.

Discussion

Our findings have important implications for both policy decision making and research. With regard to experience rating, we found only moderate support for a relationship between the degree of experience rating and injury outcomes; therefore there is still much to be learned about the merits of this feature. The results of a review by Kralj (1) were somewhat consistent with our conclusions. Other reviews by Hyatt & Thomason (2) and Thomason (4) reached stronger conclusions, suggesting that the literature provides relatively consistent evidence supporting a relationship between experience rating and injury frequency.¹³ All three reviews concluded that the evidence on the relationship between experience rating and injury severity is more ambiguous than in the case of injury frequency.

We propose two possible explanations for the latter finding. First, firms may engage in desirable and undesirable claims management practices in response to

the incentives created by experience rating (eg, accommodating injured workers, suppressing some types of claims). Second, firms may find it easier to prevent less severe injuries than to prevent more severe ones. Both responses can result in a lower frequency but a higher average of severity.

Experience rating is a key policy lever of workers' compensation insurance providers; yet the few studies that have investigated this feature have either exploited natural experiments, which have little to say about varying degrees of experience rating, or have used crude proxies for the degree of experience rating. Only one study employed a more direct measure (the amount of rebates and surcharges) in an aggregate level study using data from Germany. With so little evidence, and such imprecise measures, it is difficult to draw robust conclusions about the effectiveness of experience rating. What are needed are micro-level longitudinal studies with reasonably direct measures such as the participation (rating) factor, the proportion of rebates or surcharges, or the relative magnitude of divergence between individualized premiums and pooled risk or manual rates.

Specific aspects of experience-rating program design also merit investigation since they can vary dramatically. For example, do prospective and retrospective programs provide different incentives, and do the incentives differ for firms of different sizes? It has been recognized, for example, that, while providing considerably greater accuracy, complex retrospective programs are confusing for small and medium-size firms and thus may provide modest incentives for such firms to invest in safety. What is the impact of the length of the review period in a program (ie, the length of time that claims are monitored and affect a firm's premiums before being dropped from the firm's records)? Longer review periods provide a closer association between injury costs and a firm's premiums, but they are also less responsive and therefore provide less immediate incentive to reduce injury costs. Does experience rating affect different types of injuries differently? How effective an incentive is it with respect to musculoskeletal injuries and long-latency illnesses?

A greater number of studies has been undertaken on the merits of OHS inspections, citations, and penalties. However, even though the inspections feature was examined in a large number of studies, the inconsistent evidence suggests that additional investigation of this feature is warranted. Other reviewers of this literature also note the limitations of many empirical studies, particularly the earlier ones (3, 4). In spite of these limitations, Thomason (4) states that, overall, the evidence suggests that OSHA has, at best, resulted in a modest improvement in workplace health and safety in

¹³ Hyatt & Thomason (2) qualify their conclusions, stating that it may be due to claims management rather than to genuine reductions in frequency.

the United States. The review by Kralj (1) supports this conclusion. Mendeloff (3) elaborates on the importance of more recent firm-level studies that have found a positive relationship.

Similar to our suggestions on future workers' compensation research, we emphasize the need to use more precise measures of features. Wherever possible, a meaningful specification of the nature of inspections, citations, penalties and other OHS features need to be included. Measures relating to inspection type (eg, targeted, random, complaint-driven) and inspection quality or intensity (eg, the amount of time spent conducting the inspection), citation type, and the amount of penalty imposed would be better than simply the number of inspections, citations, and penalties. In addition, more research is needed on alternatives for inspections, citations, and penalties, such as voluntary activities (guidelines, consultations) and internal responsibility systems, as well as research on the optimal mix of different OHS policy levers.

Nonetheless, our synthesis indicates that general deterrence is less effective in reducing injuries, whereas specific deterrence with regard to citations and penalties does have an impact. This finding suggests that regulators need to "be in the field" undertaking investigations and actively seeking out cases of noncompliance for regulation to be effective. Clearly, the costs and benefits of increasing inspection and enforcement intensity should be assessed before such an initiative is undertaken.

There was some indication in the literature that, over the longer term, the effectiveness of enforcement has declined. Indeed, one of the articles (32) analyzed the effectiveness of enforcement over three time periods and found that enforcement had an important impact in the early years of OSHA, a less important impact in later years, and no impact in most recent times. Factors that have been proposed as explanations include changing technology, which has rendered standards originating from the 1960s less relevant, financial incentives created by the increasing costs of workers' compensation, which has diminished the incremental incentives of OSHA regulation, and deindustrialization of the economy and a concomitant shift towards services for which the nature of injuries is different (ie, a lower prevalence of acute trauma injuries and a higher prevalence of musculoskeletal injuries, particularly cumulative trauma disorders such as carpal tunnel syndrome and low-back pain). Indeed, musculoskeletal injuries comprise a large fraction of workers' compensation claims in many jurisdictions;

yet efforts to introduce ergonomic standards to address them have not been very successful.¹⁴ Another possible explanation of the declining effectiveness of regulation is that firms may have "captured" their regulators. If this is the case, then one possible way to circumvent it is by empowering both workers and employers through the promotion of internal responsibility systems. Unfortunately, there were only a few articles that investigated this feature, too few to warrant synthesis.¹⁵

Some study design issues warrant mention. First, whenever possible, future studies need to consider a fuller range of factors that bear on injury outcomes, including the various aspects of programs directed at firm behavior. In the studies we reviewed, few considered both workers' compensation and features of OHS regulation simultaneously. This is a critical flaw when one considers the overlapping incentives of these two programs. Second, the temporal element of behavioral responses to incentives requires better treatment. Many studies had contemporaneous explanatory and outcome variables. Those that lacked explanatory variables generally included information from only a short time period prior to the outcome. More broadly, factors such as technology, industrial mix of the economy, and human resource practices have changed over time, and, therefore, need to be accounted for in an analysis, particularly if the measurement time frame spans a long period of time.

Last, there is an obvious need for a standardized set of reporting conventions for the publication of studies. Many did not provide complete information on the basic characteristics that we sought to extract, including data type, sample characteristics, sample inclusion-exclusion criteria, the formulation of explanatory and outcome variables, and regression statistics. Clear and comprehensive reporting of these characteristics is a fundamental element of credibility. Clarity in presentation is also important to ensure reproducibility by other researchers. For policy decision makers, it is essential when the applicability of the findings in other jurisdictions is being assessed.

Acknowledgments

We would like to thank two anonymous referees for their insightful comments on an earlier draft of this manuscript.

14 The introduction of ergonomics standards by OSHA in 2000 was repealed by Congress in 2001.

15 Only a few studies investigated specific regulations such as legislation surrounding internal responsibility systems (specifically joint health and safety committees), the right to refuse unsafe work, the right to protective reassignment, and the requirement of a safety program.

References

- Kralj B. Occupational health and safety: effectiveness of economic and regulatory mechanisms. In: Gunderson M, Hyatt DE, editors. *Workers' compensation: foundations for reform*. Toronto: University of Toronto Press; 2001. p 187–218.
- Hyatt DE, Thomason T. Evidence on the efficacy of experience rating in British Columbia: a report to the Royal Commission on Workers' Compensation in BC. Victoria (Canada): Royal Commission on Workers' Compensation in BC; 1998.
- Mendeloff J. Regulatory approaches to preventing workplace injury. In: Sullivan TJ, Frank JW, editors. *Preventing and managing injury and disability at work*. London (UK): Taylor & Francis; 2001. p 207–29.
- Thomason T. Economic incentives and workplace safety. In: Sullivan TJ, Frank JW, editors. *Preventing and managing injury and disability at work*. London (UK): Taylor & Francis; 2001. p 185–205.
- Fortin B, Lanoie P. *Effects of workers' compensation: a survey*. Montreal (Canada): CIRANO; 1998. Report no 98s-04.
- Hyatt DE, Kralj B. The impact of workers' compensation experience rating on employer appeals activity. *Ind Relat*. 1995;34(1):95–106.
- Kralj B. Employer responses to workers' compensation insurance experience rating. *Ind Relat*. 1994;49(1):41–59.
- Ruser JW, Smith RS. The effect of OSHA records-check inspections on reported occupational injuries in manufacturing establishments. *J Risk Uncertain*. 1988;1(4):415–35.
- Mendeloff J. *Regulating safety: an economic and political analysis of occupational safety and health policy*. Cambridge (MA): MIT Press; 1979.
- Bartel AP, Thomas LG. Direct and indirect effects of regulation: a new look at OSHA's impact. *J Law Econ*. 1985;28(1):1–25.
- Gray WB, Scholz JT. Does regulatory enforcement work?: a panel analysis of OSHA enforcement. *Law Soc Rev*. 1993;27(1):177–213.
- Chung J-T. *The effectiveness of enforcement activities of the occupational safety program of Korea*. Washington (DC): The American University; 1990.
- Scholz JT, Gray WB. Can government facilitate cooperation?: an informational model of OSHA. *Am J Pol Sci*. 1997;41(3):693–717.
- Shapiro S, Rabinowitz R. Punishment versus cooperation in regulatory enforcement: a case study of OSHA. *Admin Law Rev*. 1997;49:713–62.
- Baggs J, Silverstein BA, Foley MP. Workplace health and safety regulations: impact of enforcement and consultation on workers' compensation claims rates in Washington State. *Am J Ind Med*. 2003;43:483–94.
- Smitha MW, Kirk KA, Oestestad KR, Brown KC, Lee S-D. Effect of state workplace safety laws on occupational injury rates. *J Occup Environ Med*. 2001;43(12):1001–10.
- Levine M, Walter S, Lee H, Haines T, Bolbrook A, Moyer V. Users' guides to the medical literature. *JAMA*. 1994;271(20):1615–9.
- Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJM, Gavaghan DJ, et al. Assessing the quality of reports of randomized clinical trials: Is blinding necessary? *Cont Clin Trials*. 1996;17:1–12.
- Slavin RE. Best-evidence synthesis: an alternative to meta-analytic and traditional reviews. *Educ Res*. 1986;November:5–11.
- Slavin RE. Best evidence synthesis: an intelligent alternative to meta-analysis. *J Clin Epidemiol*. 1995;48(1):9–18.
- Bruce CJ, Atkins FJ. Efficiency effects of premium-setting regimes under workers' compensation: Canada and the United States. *J Labor Econ*. 1993;11(1):S38–S69.
- Cheadle A, Franklin G, Wolfhagen C, Savarion J, Liu PY, Salley C, et al. Factors influencing the duration of work-related disability: a population based study of Washington State Workers Compensation. *Am J Public Health*. 1994;84(2):190–6.
- Deland CS. *Experience rating of workers' compensation in Alberta*. Calgary (Canada): University of Calgary, Department of Economics; 1995.
- Kralj B. Experience rating of workers' compensation insurance premiums and the duration of workplace injuries. In: Thomason T, Chaykowski RP, editors. *Research in Canadian workers' compensation*. Kingston (Ontario): Industrial Relations Press; 1995. p 106–22.
- Kotz H, Schafer HB. Economic incentives to accident prevention: an empirical study of the German sugar industry. *Int Rev Law Econ*. 1993;13(1):19–33.
- Lanoie P, Strélski D. L'impact de la réglementation en matière de santé et sécurité du travail sur le risque d'accident au Québec: de nouveaux résultats [The impact of workplace health and safety policies on accident risk in Québec: new results]. *Ind Relat*. 1996;51(4):778–801.
- Ruser JW. Workers' compensation insurance, experience-rating, and occupational injuries. *J Econ*. 1985;16(4):487–503.
- Ruser JW. Workers' compensation and occupational injuries and illnesses. *J Labor Econ*. 1991;9(4):325–50.
- Ruser JW. Workers' compensation and the distribution of occupational injuries. *J Hum Res*. 1993;28(3):593–617.
- Curington WP. Federal versus state regulation: the early years of OSHA. *Soc Sci Q*. 1988;69:341–60.
- Lanoie P. Safety regulation and the risk of workplace accidents in Quebec. *South Econ J*. 1992;58(4):950–65.
- Gray WB, Mendeloff J. The declining effects of OSHA inspections on manufacturing injuries, 1979–1998. *Ind Lab Relat Rev*. 2005;58(4):571–87.
- Ruser JW, Smith RS. Reestimating OSHA's effects: have the data changed? *J Hum Res*. 1991;26(2):212–35.
- Scholz JT, Gray WB. OSHA enforcement and workplace injuries: a behavioral approach to risk assessment. *J Risk Uncertain*. 1990;3(3):283–305.
- Sha S. *Accident rates, workers' compensation, and safety regulations*. Stony Brook (NY): State University of New York; 1995.
- Smith RS. The impact of OSHA inspections on manufacturing injury rates. *J Hum Res*. 1979;14(2):145–70.
- Baggs J, Silverstein BA, Foley MP. *Observed associations between WISHA activities and compensable claims rates. Olympia (WA): Safety and health assessment and research for prevention (SHARP) program; 2001. Report no 70-1-2001.*
- Guo S. *The influence of OSHA inspectors' detection capabilities on OSHA's effectiveness: evidence from a panel data, 1979–1985 [dissertation]*. Worcester (MA): Clark University; 1999.
- Kim KJ. *The political economy of OSHA regulation: a pooled time series analysis [dissertation]*. Athens (Georgia): University of Georgia; 1991.
- Lanoie P. *The impact of occupational safety and health regulation on the risk of workplace accidents: Quebec, 1983–87*. *J Hum Res*. 1992;27(4):643–60.
- McCaffrey DP. An assessment of OSHA's recent effects on injury rates. *J Hum Res*. 1983;18(1):131–46.

42. Ruser JW. Self-correction versus persistence of establishment injury rates. *J Risk Insur.* 1995;62(1):67–93.
43. Viscusi WK. The impact of occupational safety and health regulation. *Bell J Econ.* 1979;10(1):117–40.
44. Viscusi WK. The impact of occupational safety and health regulation, 1973–1983. *Rand J Econ.* 1986;17(4):567–80.
45. Gray WB, Scholz JT. Analyzing the equity and efficiency of OSHA enforcement. *Law Policy.* 1991;13(3):185–214.
46. Bartel AP, Thomas LG. The costs and benefits of OSHA-induced investments in employee safety and health. In: *Work* (eds. J. D. Appel, D. Appel). Workers' compensation benefits: adequacy, equity and efficiency. Ithaca (NY): ILR Press, New York State School of Industrial and Labor Relations, Cornell University; 1985. p 14–56.
47. Robertson LS, Keeve JP. Worker injuries: the effects of workers' compensation and OSHA inspections. *J Health Polit Policy Law.* 1983;8(3):581–97.

Received for publication: 11 May 2006

Appendix 1

Quality assessment instrument

Overall study quality

- Question 1: Does the study specify a theoretically correct relationship between the policy feature (explanatory variable) and the outcome variable (dependent variable)?
- Question 2: Are the characteristics of the study population properly defined, measured, and described?
- Question 3: Was the statistical methodology appropriate for the research question and study design?
- Question 4: Does the study establish an empirically correct relationship between the outcome and independent variables?
- Question 5: Was adjustment made for important covariates?
- Question 6: Are the results interpreted correctly?

Feature quality

- Question 7: How strongly would you rate the measurement validity of this feature?
- Question 8: How strongly would you rate the external validity of this feature?

Appendix 2

Best evidence synthesis guidelines

Strong evidence

- Minimum study quality: High
- Minimum number of studies: 3

- 1) If there are only three high-quality studies, all of them must report consistent findings.
- 2) 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.
- 3) 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 <3 high-quality studies

Minimum number of studies: 3; they can be a mixture of medium- and high-quality studies.

1) At least 3 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: 1

1) Fewer than three studies reporting consistent findings.

2) 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 were available from which to draw conclusions.

Mixed evidence

The findings from medium- and high-quality studies were contradictory.