Metrics to Predict Contractor Performance

A Report to ORC and the Duke Energy Foundation

2008

Prepared by the RAND Center for Health and Safety in the Workplace

John Mendeloff
Rachel Burns
Hangsheng Liu
Matt Schonlau

This paper has not yet been peer-reviewed. Do not cite or distribute without permission.

Contact
John Mendeloff, RAND CHSW
412-683-2300 ext 4532
jmendel@rand.org

Metrics to Predict Contractor Performance

Executive Summary

This report provides findings of the analyses that the RAND Center for Health and Safety in the Workplace (CHSW) has carried out under its contract with ORC and the Duke Energy Foundation to study Metrics for Predicting Contractor Performance.

Because of the contract's focus on "serious injuries," we sought a measure that addressed only very serious events. We also needed to be able to link those events to particular workplaces. The only data that we knew of that met these criteria are the accident investigations (AIs) carried out by OSHA. AIs are carried out to investigate fatalities or "catastrophes" (events resulting in the hospitalization of 3 or more workers). Fatalities caused by highway accidents, assaults, and heart attacks were generally not investigated. The drawback of AIs for our purposes is that there are not that many of them. They are rare events. Rare events are generally difficult to predict.

At the suggestion of ORC, we also looked at the LWDII rate as an outcome. We were also encouraged to look at manufacturing as well as construction.

For both outcome measures (fatal accidents and LWDII rates) we looked at 3 predictors that appear to be frequently used by firms in assessing and trying to predict the safety records of prospective contractors: Have they had serious problems of non-compliance with OSHA standards? Have they had high LWDII rates? Have they had fatalities in recent years?

Highlights of the Findings

Table ES-I summarizes our findings.

To predict fatal or catastrophic events

- The prior record of non-compliance appears to have some value in construction, but not in manufacturing.
- The lost workday injury rate has no value for either sector.
- For both sectors, the occurrence of a fatal event one or two years before does have predictive value, but the impact is the opposite of what most users of this measure assume. Fatalities are lower, not higher, if such an event had occurred. It appears that serious accidents tend to spur renewed attention to safety. This finding suggests that firms should reconsider use of this measure as a negative indicator. However, that finding could still be consistent with the proposition that firms with a long history of deaths pose above average risks.

To predict lost workday injury rates.

- In manufacturing, non-compliance may be associated with higher injury rates; we were unable to study this issue in construction.

- As others have found, a manufacturing establishment's prior lost workday injury rate has considerable predictive value, although the value declines with smaller establishment size and with the length of time that has passed. We also find very similar effects for construction firms.
- In the construction industry, but not in manufacturing, prior deaths also appear to reduce injury rates in the following year or two.

We also review work on whether assessments of firms' management systems can predict future injury rates. Studies of existing management systems don't provide much guidance, but they do reveal some association between high ratings and lower injury rates. We also find a similar association in the 2004 data collected by ORC for over 60 large firms. Approximately 10% to 12% of the variation in the "days away from work"(DAW) injury case rate can be explained by the overall score used by ORC. The single most important of the 5 elements in that system was the risk score, measuring activities that the firm has undertaken to understand and address risks. In a cross-sectional analysis, a 10% increase in the risk score was associated with about a 15% reduction in the DAW injury rate.

We conclude that, if host firms are chiefly interested in predictions of fatal or catastrophic events, in the construction sector they should make use of a measure of serious violations like the one we examine here and, in both sectors, they should continue to ask about recent fatalities, except they should be viewed as a protective factor, not a risk factor. Lost workday injury rates had no value in predicting fatalities in either sector. For the present, it seems that scores of management systems have some predictive value, but further improvements are needed to ensure that the effort to use them is worthwhile. Of course, the insights obtained from the information collected under these management systems may have other values that can justify their use, apart from prediction.

•

Table ES-1
An Overview of the Findings

I) Fatal Accident as Outcome Explained by	**Construction**	**Manufacturing**
A) Prior Non-compliance	May Predict an Increase	No Predictive Value
B) Prior LWDII Rate	Unlikely Predictive Value	Unlikely Predictive Value
C) Prior Death	May Predict a Decrease	Strongly Predicts a Decrease
II) LWDII Rate as Outcome Explained by		
A) Prior Non-compliance	[Not feasible to study with available data]	May Predict Increase
B) Prior LWDII Rate	Strong Positive Prediction	Strong Positive Predictive
C) Prior Death	May Predict a Decrease	Uncertain Predictive Value

Acknowledgements

We acknowledge the financial support of the Duke Energy Foundation for this study. Steve Newell and Scott Madar of ORC provided useful guidance and comments on the draft. Elyce Biddle of the National Institute for Occupational Safety and Health provided information on the management audit data that ORC had collected. Professor David Weil graciously provided us with the data set he had created which linked OSHA construction inspections Stacy Fitzsimmons helped to prepare the manuscript.

Understanding of the Problem

Contracting is a large and growing phenomenon for many companies. Some "host" companies have, in recent years, increasingly accepted the principle that they have a responsibility to provide the same level of safety for contract workers as they do for their own employees. The National Summit on Contractor Safety in 2006 addressed the problems of improving contractor safety performance. The Summit found that, although many large firms had developed systems for prequalifying contractors, these systems had never been validated in terms of their ability to predict safety and health performance.[1] The absence of predictive evidence of effectiveness usually applies to all firms, not just to contractors. However, the attempt to develop criteria for assessing contractors does raise some unique issues. Perfect predictions will never be possible, of course, but we can clarify how well certain factors are related to different measures of future performance.

Improving the ability of prequalification procedures to predict safety outcomes is an important task. As Summit participants noted, the current focus in many firms is on screening contractors on the criteria of cost, quality, and timeliness. Safety is often secondary. It is plausible, we believe, that one reason why safety may be ranked low as a criterion is that firms lack confidence in their ability to predict safety performance. If you believe you can't identify who will be a good performer on safety, then there is little point in using it as a criterion for qualifying firms. Thus one result of more valid predictors could be a greater weight given to safety in the prequalification process.

Host firms face the problem of how to identify which contractors are likely to do better or worse in achieving these outcomes. Ideally, the predictive scheme would actually quantify the size of the differences among potential contractors. A less taxing requirement would be that it simply rank firms without providing any meaningful sense of the size of the differences. Finally, firms may be seeking only a single cut-off point to distinguish acceptable from unacceptable contractors. However, because safety is not the only criterion firms are considering when they select contractors, it would be preferable to know something about how big the differences are in order to be able to make intelligent decisions about possible trade-offs between safety and the other criteria. This discussion does highlight an important question: how good does predictive ability have to be in order to justify the efforts needed to achieve it.[2]

Possible Measures to Use

[1] National Summit on Contractor Safety, Summary Report, Duke Energy Foundation and Industrial Resource Counselors, Inc., February, 2007.

[2] For example, one study developed a complicated statistical procedure for predicting the safety performance of motor carriers. The study found that the 2.5% of firms with the worst scores had injury rates twice the average. Like the findings in our study, there may be some real value in this finding, but it is limited. See Leon N. Moses and Ian Savage, "Identifying Dangerous Trucking Firms," Risk Analysis (1996) 16:359-366.

What are the potential candidates to be considered for predicting performance? They come in several categories:

- Measures of earlier injury performance
- Measures of safety program activities or "management systems"
- Measures of top management and front-line supervisor commitment
- Measures of "safety climate" or "safety culture"
- Measures of worker participation in safety programs
- Measures of other firm characteristics (e.g., size)
- Measures of demographic characteristics of the workforce

Clearly, these categories overlap and there are debates about which of these subsume the others and which are causes and which are effects.

Commonly Used Pre-Qualification Measures

The more direct approach we follow here is to examine what appear to be, based on a review of the pre-qualification procedures for several large firms, some of the most commonly used measures that host firms (in the United States) ask potential contractors to provide[3]:

1) Experience Modification Rating (EMR)
2) Counts of Injuries in Each of the Broad Categories Mandated by OSHA: Injuries with Days Away from Work, Injuries with Restricted Work Activity Only; and Medical Only Cases. The first two comprise the Lost Workday Injury category (LWDI); and all three comprise the Total Recordable Rate.
3) Number of Fatalities
4) Violations Cited in OSHA Inspections During the Last 3 Years

Philips and Waitzman report that among the safety and health professionals they surveyed, 40% "indicated that past fatalities were the single most important predictor of current contractor risk of serious injuries, fatalities or other catastrophic accidents. Less than 20% felt that lost workday injury rates were the most important past indicator of present performance, and less than 20% felt that experience modification rates were the most important predictive indicator."[4]

We were not able to gain access to a data base with employer EMRs. Thus our study is limited to looking at 3 possible predictive measures:

- The firm's non-compliance with OSHA standards
- The firm's lost workday injury rate
- The firm's fatality experience in the preceding few years

We use each of these measures to predict two outcome measures:

---

[3] In addition, host firms generally ask questions about the elements of contractors' safety programs. Some hosts want to know if the required procedures are in written form and, if so, to obtain copies. Some want to know whether compliance with those procedures is documented.

[4] Peter Philips and Norman Waitzman, Report to Duke Energy Foundation and Industrial Resource Counselors, Inc.

- The occurrence of a fatality in a given year
- The lost workday injury rate in a given year

In addition, we were able to examine a data base collected by ORC which allows us to examine the relation between scores on the implementation of a management system and the corporate injury and illness rate.

Host firms may be interested not only in the relationship of this inspection history to future injury and illness measures, but also to the probability that inspections during the life of the contract will find serious non-compliance. Firms may also be interested simply in the probability that the contractor will be subject to a complaint inspection. Any inspection imposes costs on a contractor in terms of delays and lost productivity. In addition, a history of complaint inspections may indicate that a contractor has problems resolving internal disputes over safety conditions. If so, these problems might re-occur during the contract period. However, we did not include these outcomes in the set we were trying to predict in this study.

The host firm's contracting decision will generally be made after the decision about what services to contract for and where the work will be done. Thus, although injury rates can vary a great deal depending upon the type of project and the location, these factors will be fixed for any single decision. However, in our analysis, we do look separately at firms in construction and firms in manufacturing. For example, the methods that work best for construction projects may work less well for other goods and services.[5]

RAND has collected a number of data sets that are used in the analyses below. These data sets include:

- The OSHA IMIS data from 1972, which includes:
  --Characteristics of the establishments and the inspections
  --Particular standards cited and the severity of the violations
  --Exposure information from health inspections
  --Injury and illness rates collected from the logs of inspected establishments
  --Investigations of fatalities
- A subset of the IMIS that links inspections at 2030 construction firms from 1987 to 1993
- OSHA Data Initiative (ODI) from 1995-2005
- Pennsylvania Workers Compensation files from 1998 to 2006 (employers' first reports of injuries)

<u>The Predictive Measures Examined in this Research</u>

The Lost Workday Injury and Illness Rate

[5] To a lesser extent, the same argument may apply to location; for example, the importance of planning for providing emergency services may be more important for a project in an isolated rural area than for one in the middle of downtown.

The simplest kind of model uses prior outcomes to predict future values of the same type of outcomes. We often use injury rates in year 1 to predict injury rates in year 2. At firms where the expected number of injuries is large, this method can work fairly well, at least when there are no major changes in the nature of the work, the workforce, or environmental factors that affect reporting of injuries. In contrast, when a firm would normally report zero or only a few injuries each year, injury *rates* could fluctuate wildly from year to year.

At least two studies have been carried out that have examined the extent to which LWDI rates remain stable over time. A major motivation for these studies was to determine whether targeting workplaces on the basis of high injury rates in a previous year would take OSHA inspectors to "high" injury rate workplaces in the following year or two.[6] Not surprisingly, both studies found that persistence of high injury rates was greater when the employment base was larger and when the injury rate in the industry was higher.

In addition, we carry out new analyses of the persistence of injury rates from one year to another. The two studies cited above were carried out with data from before 1984. We examine the stability of lost workday rates in manufacturing establishments from 1995 to 1996 (using ODI data) and the stability of lost time injury rates in Pennsylvania from 2003 to 2006 for all industries.

Fatalities

Some firms (e.g., Alcoa) include a question in their prequalification survey that asks contractors whether they have had a fatality within the last few years. Does a prior fatality increase the likelihood of a future fatality? Are fatalities in prior years associated with higher or lower LWDI rates in future years? We can also reverse the question, asking whether higher LWDI rates in prior years are associated with a higher probability of fatalities in later years.

Non-compliance with OSHA Standards

We also look at the relationship between compliance with OSHA standards in previous years with a number of other outcomes. These tests are relevant because many firm use past regulatory compliance as a factor in their prequalification procedures. Does a record of non-compliance predict any of the outcomes that firms may care about? We focus on a measure of percentage of a OSHA inspections at a firm which cited serious violations (in construction) and the number of serious violations per 100 workers in programmed inspections (in manufacturing). The reason for using programmed inspections is that they are much more likely to cover the entire workplace than complaint inspections are and thus provide a better basis for comparing different facilities.

[6] John W. Ruser, "Self-Correction and Persistence of Establishment Injury Rates," (1995) Journal of Risk and Insurance 62:67-93; A. Allan Hunt, "Analysis of Persistence in Employer Injury Rates," (1993) Office of Statistics, OSHA, U.S. Department of Labor.

Safety and Health Management Systems

A recent systematic review of OSH management systems by the Institute of Work & Health noted that "OHSMs are generally distinguished from traditional occupational health and safety programs by being more proactive, better internally integrated and for incorporating stronger elements of evaluation and continuous improvement.[7] In a separate literature review on audit tools, the Institute authors noted that "an OHS audit is considered to be a systematic assessment of an OHS management system."[8] Below we review the findings of that study and also examine a data set collected by ORC from member firms that have reported using an assessment instrument developed by the firms.

<u>Other Factors Not Directly Addressed in this Study</u>

Firm Characteristics

Host firms may also want to consider the size or profitability of the contractor firm. Most analysts have assumed that larger firms are safer than smaller ones, partly because they have stronger safety incentives from experience rating. However, a recent RAND study[9] indicated that much of the better performance at larger firms (at least with respect to fatalities) arises because their workers tend to work at larger establishments. The relationship between establishment size and risk was strong, robust, and negative; but, holding establishment size constant, the benefits of larger firms were smaller and appeared only for firms with over 1,000 employees. (This analysis did not include the construction industry.) There are a number of reasons for suspecting that more profitable firms have better safety records, although there is little evidence. [10]

Worker Demographics

It is well-known that workers who are new to a job tend to have higher injury rates. There is also evidence that younger workers (in their teens and early twenties) have higher rates, independent of experience. Older workers, in contrast, have more severe injuries. There is also evidence that married workers are safer than unmarried workers, at least through their twenties or thirties. Language differences within a workforce may be linked to more injuries. Better education and higher wages are linked to fewer injuries.

Knowledge about the demographic composition of a contractor's work force may enable a host firm to gain some insight about its likely injury performance.

Safety Culture/Safety Climate

---

[7] Robson L. et al. "The Effectiveness of Occupational Health and Safety Management Systems: A Systematic Review," Toronto. Institute for Work & Health, 2005.

[8] Bigelow P, Robson L. "Occupational Health and Safety Management Audit Instruments: A Literature Review," Toronto: Institute of Work & Health, 2005.

[9] John Mendeloff, Christopher Nelson, Kilkon Ko, Amelia Haviland, Small Business and Workplace Fatality Risk: An Exploratory Analysis" RAND Corporation TR-371 ICJ (2006).

[10] Ongoing work by Gray and Mendeloff shows that manufacturing establishments with lower injury rates in their industry have higher productivity, on average.

Guldenmund has written that “The trend in literature is to treat culture as a set of core values and climate as an expression of these values…”[11] Although there is a debate about whether and how to measure “safety culture,” safety climate has become an increasingly important concept.[12] Not only have researchers made major attempts to measure it, but many agree with Petersen that “such a survey provides a better predictor of the future safety record than any other indicator tested and helps to clearly target what needs to be done to improve safety systems in organizations.”[13]

One concern is that, even if the safety climate measure is valid when used in a firm for its own employees, it remains less clear what would happen if a firm undertook a safety climate survey in order to comply with requirements for bidding for work at a host firm. Strong incentives might lead a prospective contractor to cheat in order to obtain findings that made the safety climate look better than it really was. The extent to which this process may reduce the validity of the measures is one issue that needs to be explored.

[11] F.W. Guldenmund, “The Nature of Safety Culture: A Review of Theory and Research,” Safety Science (2000) 34:215-257.

[12] The concept of safety culture refers principally to the norms regarding the role and importance of safety in the organization, norms which govern expected behavior. Comments are often made that culture does not change quickly. This is interesting in light of various claims that certain interventions achieve very large reductions in injuries within a short period of time. Either “culture” can not be the causal factor here or culture must be able to change much more quickly than some writers assume.

[13] Dan. Petersen, “Safety Improvement,” Professional Safety . 5 (2005) 45-48.

Review of Findings

I. Does Non-compliance Predict a Higher Fatal Accident Rate?

A. Construction

We used a data set that had linked all OSHA inspections for the years 1987-93 at the 2060 largest construction firms.[14] For each firm, we calculated the percentage of inspections in each year which had cited a serious violation. We investigated how this measure was related to the number of OSHA investigations of fatal and catastrophic accidents in the same year. Then we calculated the average percentage of inspections citing a serious violation for the one year, two years, three years, and four years prior to the year of the fatality; and examined whether these measures were related to the number of future fatal accidents. The analyses controlled for the 3-digit SIC construction industry category because some industries are clearly more likely to incur fatalities than others, regardless of compliance (e.g., structural steel work versus carpentry). They also controlled for any time trend in fatalities.

The findings, shown below, indicate that the measures of non-compliance are always associated with an increase in fatal accidents; i.e., the coefficients on these variables are all positive. However, the ‘p’ values are all larger than .05, which is often used as a measure of whether we should conclude that the coefficients are truly different from zero (i.e., whether they are “statistically significant”). For example, for the 3 year measure of non-compliance, the ‘p’ value is 0.08, which means that we would expect to find an effect this large (i.e., this different from zero) simply due to chance 8 times out of 100. The standard of proof we use to decide whether an effect should be treated as “real” is, in large part, a policy decision. If we use a high standard of proof (e.g., a ‘p’ value below 0.05), we run a greater risk of concluding that an effect does not exist, when it really does. In contrast, with a lower standard of proof (e.g., 0.10 or 0.20), we run a greater risk of concluding that an effect does exist when it really does not. The choice we make should reflect how costly we think each kind of error could be. We think that failing to use a potentially useful screening device would probably generate greater losses than those that would be incurred from excluding some firms who were, in truth, not at higher risk. Therefore, we view the findings as indicating that the average non-compliance over 2, 3, or 4 years is probably a predictor of the occurrence of a fatal accident in construction. Certainly, it is much more likely than not that this is true.

[14] This data set was provided to us by its creator, Prof. David Weil. For a further description of the data , see his article “Assessing OSHA Performance: New Evidence from the Construction Industry,” Journal of Policy Analysis and Management (2001) 20:651-674.

Table 1
The Effect of Non-Compliance on the Number of Fatal and Catastrophic Accidents in Construction

Percentage of Inspections Citing Serious Violations	Impact of a Change from 0% Non-Compliance to 100%	'p' value
In the same year	1.6%	0.92
In the prior year	3.4%	0.86
In the prior 2 years	32.1%	0.20
In the prior 3 years	50.6%	0.08
In the prior 4 years	39.4%	0.18

Table 1 shows the findings expressed in terms of the impact on the number of fatal/catastrophic events at a firm in the year of having no serious violations versus having every inspection cite serious violations.[15] The size of the impact is essentially zero for non-compliance in the same year as the accident or in the prior year. Beginning with the average non-compliance for the two prior years, the effects become substantial. Among inspected firms, a firm which had been cited for a serious violation in every inspection during the prior 3 years would have 50% more fatal and catastrophic events than one that had never been cited.

But, in addition to the statistical uncertainties of these findings, we need to take into account two other points. First, most firms that have been frequently inspected will not be at either extreme of the spectrum—neither 0% nor 100% non-compliance. Thus, in practice, predicted differences are likely to be smaller than 50%.

Second, and more important, because fatal and catastrophic accidents are rare events even a 50% increase in their probability would raise the average probability per firm per year in our sample from about 5% to about 7.5%. The large construction firms in this sample probably have below average fatality rates, so these figures would be higher in other construction firms. The average revenue for the construction firms in the sample was $130M per year (in 1994 dollars). Because almost all of these firms work at multiple sites, the probability of a fatal or catastrophic injury at any one work site would be lower.

As a result, although average non-compliance over several years may indeed substantially increase the risk of a fatality, the overwhelming majority of firms with major non-compliance will still not have fatalities on any particular job. It is usually very hard to achieve high predictive accuracy when we are trying to predict rare events.

IB. Manufacturing

---

[15] Our analysis used a Poisson regression, which is appropriate when the dependent variable is a count of the number of events.

For this analysis, we used OSHA inspection data for Pennsylvania, where we had aggregated inspections that occurred at the same establishment. The sample included 26,115 inspections from 1972 though 2003. At 734 establishments, there had been at least 1 accident investigation, so the question was whether the workplaces with those fatal or catastrophic events had experienced unusually high rates of serious violations during OSHA inspections in the prior years.

As with the analysis for construction, we calculated the percentage of inspections at the establishment in each of the 4 years prior to the accident investigations (AIs) which had cited serious violations. These provided our measures of non-compliance for the establishment. We then again used a Poisson regression method to estimate the association between the number of AIs in a year and the percentage of inspections with a serious violation in the year prior to the AI, the 2 years prior, the 3 years prior, and the 4 years prior. The regression also controlled for the 2-digit SIC, the year, and the employment at the establishment (the measure of exposure).

Table 2 shows the estimated effects of non-compliance. The effects implied by the coefficients are very small and they don't reach conventional levels of statistical significance.

Table 2
The Relation of Non-Compliance to the Number of Fatal and Catastrophic Accidents at Manufacturing Establishments

	IRR (odds ratio)	P>lzl	# OBS	# Groups
Prior year	1.002	0.60	1056	799
Prior 2 years	1.000	0.93	1583	1088
Prior 3 years	0.996	0.23	1884	1252
Prior 4 years	0.999	0.74	2026	1360

II. Do Higher LWDII Rates Predict More Fatal Accidents?

A. Construction

For this analysis we examined a sample of several thousand construction inspections (outside of California) from 1996 through 1999.[16] We compared the prior injury rates at firms which experienced a fatal or catastrophic accident (as indicated by an OSHA accident investigation or AI) with the injury rates for the same years for firms that did not experience a fatal accident. The injury rate data came from the OSHA "log file." Since the mid-1990s, OSHA compliance officers have collected the injury rate data for the prior 3 years for the inspected firm. Although available only for inspected firms, this file provides the easiest method to link prior injury rates with inspection outcomes.

[16] California conducts a vastly larger number of AIs for non-fatal events than other states do, so we omitted it here in order to keep the data reflective of the rest of the nation.

We used a model that included each of the prior injury rates for 3 years to explain whether a fatal accident occurred at the workplace. We controlled for SIC at the 3 digit level and for the firm's employment, as reported in the log file. The data set for this test included 2,995 inspections. Of these, OSHA had investigated a fatal or catastrophic accident at 165 (5.5%).

We characterized the dependent variable according to whether a firm had at least one fatality or catastrophe (as indicated by an OSHA AI outside of California) in a year. We use a logistic regression method, which is appropriate when the variable is construed to take only the values of 0 and 1. In logistic models, we typically calculate the "odds ratio" as a measure of the independent effect of each variable on the probability of the event (a fatal or catastrophic event in this case) occurring. In the regression, the odds ratios for the effect of the prior injury rates were:

Table 3
The Effect of LWDII Injury Rates on Fatal Accidents in Construction

	Odds Ratio Point Estimate	'p' value	N OBS
LWDII rate (1 year prior)	0.991	0.18	6,805
LWDII rate (1+2 years prior)	0.994	0.35	7,048
LWDII rate (2+3 years prior)	0.992	0.37	3,962
LWDII rate (1+2+3 years prior)	0.992	0.25	3,084

The findings are that higher LWDII rates in prior years are associated with a slightly lower probability (by less than 1%) of a fatal or catastrophic accident. These effects are small and the "p" values suggest that they are not very precisely estimated, suggesting that the LWDII rate is probably not useful for predicting these events in construction.

IIB. Manufacturing

To examine the relation of LWDII rates to fatal events in manufacturing, we examined 10,541 observations from OSHA log data to see if LWDII rates in the prior year were associated with the probability of a fatal accident in the next year. In this file there were 279 fatal or catastrophic accidents. In our regression, we controlled for the 2 digit SIC category and for the number of employees. The estimated odds ratio for the injury rate variable was 0.991, indicating that a 1.0 increase in the rate was associated with a 1% increase in the probability of a fatal accident. (Results not shown.) The 'p' value for this estimate was 0.22. For LWDII rates two or three years prior, the estimated effects were smaller and the 'p' values were even less significant. As for construction, we conclude that any relationship between injury rates on the probability of fatal accidents is too small and uncertain to have predictive value.

III. Does a Prior Fatal Accident Predict Future Fatal Accidents?
A. Construction

We used the same construction data set used in our first analysis to examine whether the number of fatal accidents in one year (i.e., as indicated by an OSHA AI) has a relationship to the occurrence of a fatal accident in the next year. The data covered the 2,060 largest construction firms and their experience from 1987-93. We controlled again for SIC at the 3 digit level, used firm revenue as our measure of exposure, and included a time trend to account for changes in fatality rates. Controlling for these factors, the Poisson regression analysis asked whether the number of these events in one year is associated with the number in following years.

The results indicated that a fatal accident in the prior year was associated a 12% reduction in the probability of a fatal accident in the following year. The 'p' value for this estimate was only 0.23. So it appears that this relation is likely, but by no means certain. If true, the magnitude of the effect (a 12% reduction) is substantial. We conclude that there may be predictive value.

III. B. Manufacturing

For this analysis we used a file with all OSHA inspections from 1993 through 2003 at manufacturing establishments in Pennsylvania. We linked all inspections that occurred at the same establishment. The file had 26,115 plant-year observations. There were 673 establishments which had 1 AI in a year, 50 which had 2, 8 which had 3, and 2 which had 4. We used a Poisson regression to look at the impact of prior fatal accidents on the number of fatal or catastrophic events at an establishment. The number of establishments ranged from 3,258 for the analysis of the impact of deaths in the first prior year, to 5,072 for the analysis of all 3 prior years. We controlled for 2 digit SIC categories and also for any overall trend in the number of accidents over time

We did two analyses. First, we ran separate regressions for the impact of a fatal accident one year earlier, 1 or 2 years earlier, and or 1, 2, or 3 years earlier. Table 4 shows that we found statistically significant effects for the predictive value of fatal accidents 1 year, 2 years, and 3 years in to the future. In all cases, the occurrence of the prior fatal accident reduced the number of fatal accidents in the future year.

Table 4
Impact on the Number of Fatal Manufacturing Accidents of the Following:

Variable	IRR	'p' value	N OBS
Fatal accident 1 year prior	0.80	0.002	26,115
Fatal accident 1 or 2 years prior	0.83	0.008	26,115
Fatal accident 1, 2, or 3 years prior	0.88	0.054	26,115

Second, we ran a single regression with variables for fatalities in each of the 3 years. Thus the effect of deaths in each of the other years is held constant. The results are shown in Table 5.

Table 5
Impact on the Number of Fatal Manufacturing Accidents of the Following:

Fatal accident 1 year prior	0.82	0.016	26,115
Fatal accident 2 years prior	0.84	0.127	
Fatal accident 3 years prior	0.95	0.675	

In these analyses, the IRR represents an odds ratio (like relative risk) that shows how much the variable increases (if IRR>1) or decreases (if IRR<1) the number of fatal accidents. The last results indicate that fatal accidents in the last 1 or 2 years are strongly associated with a 16-20% decrease in the probability of a fatal accident at a manufacturing establishment. The explanation may be that the occurrence of a fatality leads firms to undertake corrective actions that reduce the risk below that of similar firms which had not had a fatality. After 2 years the effects decays heavily.

That analysis has the important implication that the occurrence of a fatality in the last two years should probably not be viewed as increasing the future risk, compared to other firms in the same industry and size category. In fact, just the opposite is true. However, it still could be true that a pattern of fatal accidents over many years is a predictor of more, not fewer, fatal accidents.

IV What is the Impact of Non-Compliance on LWDII Rates?
A. Construction?

As noted above, we were not able to carry out this analysis for construction. We had a file linking all inspections at the same construction firm only for the years 1987-93. However, we do not have firm-level injury rate data for those years.

IV.B. Manufacturing

Using Pennsylvania inspection data from 1993-2003, we estimated whether the percentage of inspections citing serious violations was related to the LWDII in the same and subsequent years. (We removed accident investigations from the inspections that we used to estimate non-compliance.) We controlled for 2-digit SIC industries and for a possible time trend that would affect the rates.

The results are summarized in table 6.

Table 6
Does Non-Compliance Predict LWDII Rates in Manufacturing?

Percentage of Inspections Citing Serious Violations	Coefficient	'p' value	N OBS
in the same year	0.0004	0.93	1283
in the prior year	0.0113	0.14	695
in the prior 2 years	0.0020	0.75	1040
in the prior 3 years	-0.0042	0.48	1118
in the prior 4 years	0.0005	0.94	1086

Only the effect in the prior year comes close to being statistically significant. The prior year coefficient of 0.0113 means that an increase in the percentage of non-compliance from, e.g., 30% to 40% would predict a 11.3% increase in the LWDII. Because those rates average about 3.0 per 100 FTE, the change would be about 0.3, raising the rate to about 3.3 per 100 FTE.

We also looked at another measure of non-compliance—the average number of serious violations per inspection—rather than the percentage of inspections citing serious violations. With this measure, we obtained the results in Table 7:

Table 7
Does Non-Compliance Predict LWDII Rates in Manufacturing?

# of Serious Violations per Inspection			
Non-compliance in the same year	0.173	0.11	1283
Non-compliance in the prior year	0.204	0.15	695
Non-compliance in the prior 2 years	0.175	0.14	1040
Non-compliance in the prior 3 years	0.013	0.89	1118
Non-compliance in the prior 4 years	0.048	0.62	1086

With this measure, we see somewhat stronger evidence that non-compliance may be associated with increased injury rates, both in the present and up to 2 years in the future. We view these results as indicating that non-compliance may predict future LWDII rates in manufacturing.

V. Are companies with higher injury rates in one year more likely to have higher injury rates in the following years?

A. Construction

Studies in the manufacturing sector (cited in footnote 3) have shown that there is a substantial positive correlation of injury rates across years and, not surprisingly, that it is greater for establishments with more employment.[17] With a small number of employees, injury rates can change dramatically as a result of only 1 or 2 injuries. However, less attention has been focused on the stability of injury rates in construction.

The study by Philips and Waitzman examined this issue in construction with a data set it obtained from PICS, a safety prequalification services company. They describe the results of regressing the average of the LWDII rate for the previous two years on the current LWDII rate. They do not provide the sample size for this analysis, but it produces a slope of 0.87, indicating that, on average, a firm with a rate 100% higher than another firm in the two year period will have a rate 87% higher in the third year. This analysis did not include any adjustment for industry. Lumping together firms in industries with quite different industry rates may tend to exaggerate the stability of rates. For example, roofing firms will generally have higher rates than carpentry firms, year after year. Yet from the point of view of a firm engaged in contracting, the issue is the stability of rates among roofing firms or among carpentry firms. They are not substitutes, so there is no reason to analyze them together. However, in practice, the difference may not be great. we compare the persistence of injury rates in construction separately for firms in low rate industries and in high rate industries.

Philips and Waitzman report (equation 1 on p. 110) the results of another regression that shows similar results. That regression on the log of the LWDII rate included variables for the log of the past LWDII, the log of the past number of days away from work, and the log of the number of requirements for prequalification. In this regression, the sample size is only 114.

We carried out an analysis of the stability of LWDII rates for construction firms using the OSHA log file for construction inspections carried out from 1996 to 1999. Thus this sample is limited to construction firms that were inspected by OSHA. We know that OSHA inspections focus more on larger construction sites and less on residential construction. Thus our sample is not necessarily representative of all construction firms; however, it probably resembles the types of firms that would be involved in bidding on projects of corporate clients.

A regression of the 1998 LWDII rate on the 1999 rate for 815 construction firms found a coefficient of 0.73 with a 'p' value of <0.0001. The adjusted R^2 for this regression was 0.26. So the coefficients are broadly consistent with those of Philips and Waitzman, although the percentage of the variation explained is lower.

In addition to the regression analysis, we also created a number of tables that display how firm LWDII rates change from year to year, using quartiles. Presented in this way, the results show more clearly what you can expect if you choose a firm that was one of the best in the prior year. If you are good in one year, what is the probability that you will be good in the next year? Table 8 presents some results from analysis. For example, our sample included 508 contractors with 50-99 employees. Of the 131 who

[17] See John W. Ruser, "Self-Correction versus Persistence of Establishment Injury Rates," Journal of Risk and Insurance 62(1995)67-93.

were in the best (lowest rate) quartile in 1998, 51% were in that same quartile in 1999 and 8% were in the worst quartile. With larger firms, the stability increases somewhat.

Table 8
Percentage of Construction Firms in the Best LWDII Quartile in 1998 Who Were in the Best Quartile in 1999, by Firm Size

Firm Size	N in Best Quartile in 1998	% in Quartile in 1999 Best	Worst
50-99	131	51%	8%
100-249	99	54	5
250-499	30	67	3

If we try to predict 1999 rank using 1997 data, a 2 year gap, we find the following:

50-99	63	40%	10%
100-249	61	54	13

Using earlier years weakens the correlation, although it remains better than chance even with a 4 year gap between years.

We also tried to take account of differences among construction industries in their LWDII rates. We put firms in low LWDII industries (SICs 15, 16, 172, and 173) in one group and those in high LWDII industries (SIC 171, 174-179) in another. Table 9 shows the results only for firms with 100-499 employees for 1998 to 1999.

Table 9
Percentage of Construction Firms with 100-499 Employees in the Best LWDII Quartile in 1998 Who Were in the Best Quartile in 1999, by Industry Injury Rate

Firm Size	N in Best Quartile in 1998	% in Quartile in 1999 Best	Worst
All Industries	129	57%	5%
Low Rate Industries	75	55	5%
High Rate Industries	37	49	5%

The results are consistent with our argument about there being less stability in rates when you focus within more homogenous industries, but the difference here is very small and doesn't alter the basic findings.

V. .B Manufacturing

We used the OSHA Data Initiative (ODI) to obtain LWDII rates for manufacturing establishment in Pennsylvania for 1996—1998 and used the average rate of the first two years to predict the rate in the last year. We repeated the process with rates from 2000 and 2001 to predict the rates in 2002. Only establishments with 60 or more full-time equivalent employees would have been in that data set all 3 years from 1996-1998. For the later years, the sample included all establishments with 50 or more. The 1997 and 1998 data excluded establishments which had rates below
We ran several regression models:

Table 10
Predicting Future Injury Rates Using Prior Injury Rates in Manufacturing

Independent variables:	Injury Rate Coefficient	'p' value	N	Adj. R^2
For 1998 Prediction:				
Only the av. injury rate in 96-97	0.65	<0.0001	746	0.30
Add control for employment	0.77	<0.0001	419	0.50
Add 2 digit SIC controls	0.64	<0.0001	746	0.30
For 2002 Prediction:				
Only the av. injury rate in 00-01	0.50	<0.0001	419	0.24
Add control for employment	0.58	<0.0001	439	0.31
Add 2 digit SIC controls	0.48	<0.0001	419	0.26

We also carried out the same type of quartile analysis that we did for construction firms. Some of those findings are shown in Table 11.

Table 11
Percentage of Manufacturing Establishments in the Best LWDII Quartile in One Year Who Were in the Best and Worst Quartile in Following Years, by Establishment Size

Establishment Size	N in Best Quartile in 1998	% in Quartile in 1999	
	(cut-off<2.7 per 100 FTE)	Best	Worst
50-99	375	50%	5%
100-249	367	60	2
250-499	165	69	2
2 Year Gap			
	N in Best Quartile in 1997	% in Quartile in 1999	
	(<2.8 per 100 FTE)	Best	Worst
50-99	190	50%	7%
100-249	209	58	6
250-499	107	58	3
3 Year Gap			
	(<2.8 per 100 FTE)	% in Quartile in 2000	
		Best	Worst
50-499	428	42%	11%
4 Year Gap			
	(<2.2 per 100 FTE)	% in Quartile in 2001	
		Best	Worst
50-499	333	40%	12%

VI. Do Fatal Accidents Predict LWDII Rates?
A. Construction?

Again the Philips and Waitzman study makes a contribution here. Using the data they obtained from PICS, they added a variable on the "number of fatalities in the last two years" to the regression explaining the LWDII. The estimated coefficient (Equation 2) was negative with a 'p' value of 0.115. There was a 7% decline in the current LWDII rate for every fatality in the past 2 years.

One limitation of this finding is that the sample was still only 114 firms. There were 30 fatalities in the sample, but one firm accounted for 11 of them. Omitting that firm, there were 19 deaths in 226 firm-years, just over 9%, a figure somewhat higher than the 4-5% we found in our sample.

As they and others have argued, firms may react to bad experiences and try to improve. Others have argued that a similar process works with injury rates; a spike in the rates may generate a reaction. In practice, it may be difficult to distinguish this reaction pattern from the purely statistical workings of regression to the mean—given the somewhat random nature of accidents, unusually high rates or unusually low rates tend to be followed by a return to normal.

It is also plausible that although, on average, fatal accidents spur reforms and thus are not associated with future deaths, firms which do incur repeated instances do often have poor programs and a poor safety culture.

We have no independent analysis of this issue for the same reason we could not address the effect of prior non-compliance on LWDII. We have data linking inspections (and thus AI inspections) only for 1987-93 and we don't have injury rate data for that period.

VI. B. Manufacturing

Using Pennsylvania data, we examined whether accident investigations for fatalities or catastrophes were correlated with subsequent LWDII rates. Our regression model again included 2 digit SIC industry control variables and a time trend variable.

Table 12
Do Prior Fatal Accidents Predict LWDII Rates in Manufacturing?

	Coefficient	'p' value	N OBS
Fatal accident in the prior year	-0.851	0.14	7,609
Fatal accident in the 2 years prior	-0.782	0.26	7,609
Fatal accident in the 3 years prior	-0.646	0.32	7,609

The number of fatal accidents ranged from 29 in the regression with 1 prior year to 57 in the regression for 3 prior years. These small numbers and 'p' values create at least moderate uncertainty about these effects. If the effect is real, the coefficients suggest a very large impact on the LWDII of a recent fatal accident in manufacturing. In the late 1990s, the mean LWDII in manufacturing was less than 4.0, which implies that prior accidents were associated with 15% to 20% reductions in LWDII rates. Of course, it is plausible that the industries with numerous fatal accidents are also industries with higher LWDII rates, in which case the impact would not be quite so great. Nevertheless, the impact would still be large, considerably larger than the effect estimated for construction.

Management Systems—Attempts at Measurement

Since the focus of this project is on metrics, this discussion of audit tools is very relevant. The Bigelow and Robson (2005) study cited above examined the relevant literature on the reliability and validity of 9 different instruments developed to assess OHS management. In general, inter-rater reliability for these measures was not satisfactory. Several were also incomplete in their coverage of the full range of management issues. Yet it did find some evidence that audit scores improved after interventions. Nevertheless the number of studies of even moderate validity was too low to allow any confident judgments about the value of the tools. One study compared 2 facilities; one looked at a hospital and another studied one municipality. One study examined 16 small or medium firms, but it looked only at implementation, not at the relation to injuries.

Although the Bigelow and Robson study listed the elements in each of the instruments, it did not review how they were measured or weighted. Its focus was on only the overall score and its reliability and validity. If a measure is not reliable, it can't be valid. In other words, if you look at a program and say it's good and I look at it and say it's bad, then there is no hope that our rating will correctly capture how good it is. It is true that reliability can generally be improved by training the raters about what to look at and how to score the different elements. The important point is that, unless the tool is reliable, you can't really conclude that a finding that it predicts injuries well in one context will mean that it will do so in another.

ORC itself has helped many of its member firms to implement a management system which has been in place for up to 3 years. Working with member firms, it developed a health and safety management system and a measurement tool whose elements were: management leadership, accountability, employee involvement, risk analysis and continuous improvement. Scores for each of these were based on from 3 to 9 separate questions. In 2004 69 firms provided scores for the tool as well as their rates for total recordable injuries and days away from work injuries. In 2005, only 33 firms provided all of that information. Twenty-nine firms provided all of the data in both years.[18]

Focusing on the 2004 data, we calculated the average score for each of the 5 elements. These were highly correlated with each other.

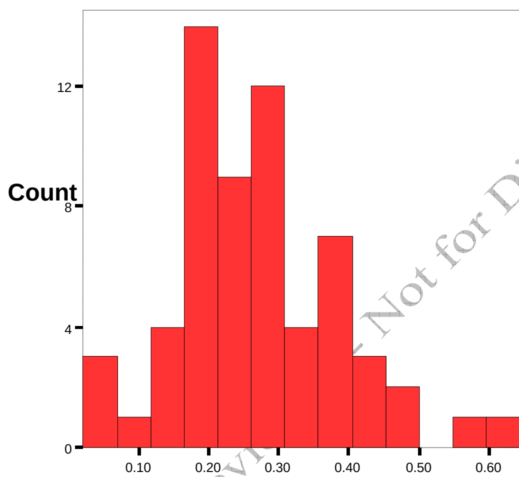
One interesting finding is that the firms varied greatly in the percentage of the total recordable cases that involved days away from work. Figure 1 shows this variation. Clearly, firms vary in their recordkeeping policies. As a result, the same firm may have a relatively low or high rate depending upon which injury measure is being used. Did firms which reported a low percentage of DAW cases have different total recordable rates than those which did not? No, there was no apparent relationship between the share of injury types and the total rates.

To what extent do the overall management system scores explain variation in injury rates? One question is whether to look at the relation between the absolute levels of these measures or at the relation between the percentage changes. In other words, do we want to estimate the effect of, e.g., a 1 point increase in the average management system score on the absolute change in the DAW or TR injury rate; or do we want to estimate the effect of, e.g., a 10% change in the score on the percentage change in the

[18] This analysis of the data collected by ORC builds upon an earlier analysis conducted by Elyce Biddle of NIOSH. Steve Newell of ORC also helped to provide a review of the data.

injury rates. The former assumes that the impact of a given change in the score will have the same absolute effect on the injury rate regardless of whether, e.g., the DAW rate is 0.2 or 2.0. This seems unlikely. The latter model, which implies that, e.g., a change in the score from 2 .0 to 2.4 (or 20%) will have the same percentage effect on the injury rate as a change from 3.0 to 3.6 (also 20%), seems more plausible. No model, of course, is going to be totally correct.

Figure 1
The Percentage of Total Recordable Cases that Involve Days Away from Work

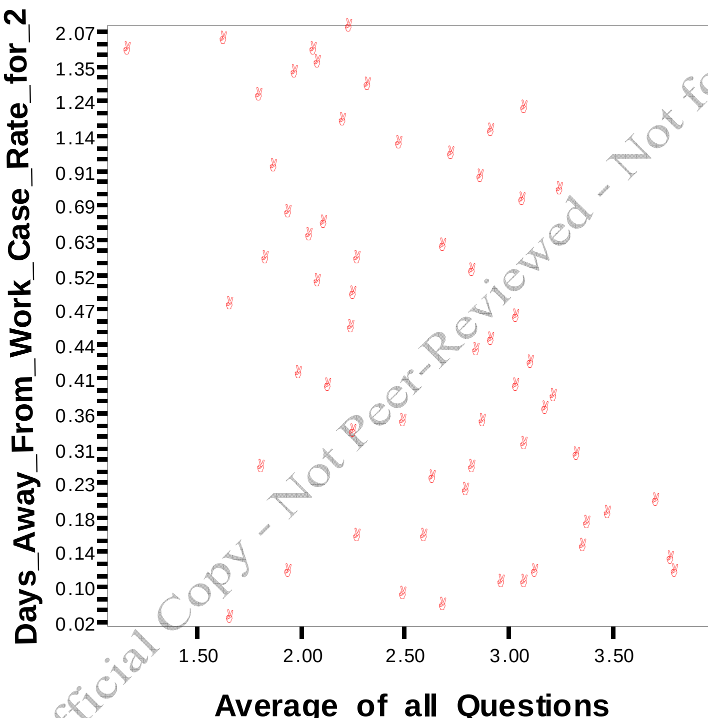


Before we show the results of that model, it may be helpful to refer to Figure 2, which plots the overall scores against the DAW rates. There is a pattern in which better scores tend to be associated with lower injury rates; however, there is a great deal of variation. Firms with the same management systems scores often have widely different injury rates.

To what extent do the overall management system scores explain variation in injury rates? One question is whether to look at the relation between the absolute levels of these measures or at the relation between the percentage changes. In other words, do we want to estimate the effect of, e.g., a 1 point increase in the average management system score on the absolute change in the DAW or TR injury rate; or do we want to estimate the effect of, e.g., a 10% change in the score on the percentage change in the injury rates. The former assumes that the impact of a given change in the score will have the same absolute effect on the injury rate regardless of whether, e.g., the DAW rate is 0.2 or 2.0. This seems unlikely. The latter model, which implies that, e.g., a change in the score from 2 .0 to 2.4 (or 20%) will have the same percentage effect on the injury rate as a change from 3.0 to 3.6 (also 20%), seems more plausible. No model, of course, is going to be totally correct.

Using the percentage change model, we find that a 10% increase in the overall score is associated with a 13% decrease in the DAW rate and a 10% decrease in the TR rate. This model explains just over 10% of the variation in the DAW rate and just under 9% of the variation in the TR rate (adjusted R-squared). If we could somehow attain greater accuracy and consistency in measuring both scores and injuries, these figures would all probably be higher. Inconsistent measurement of the management system will tend to bias the effect of the score toward zero. Inconsistent measurement of injuries will increase the uncertainty of the measurements and reduce the percentage of variation that will be explained.

Figure 2
The Relation Between "Days Away from Work" Injury Rate and Overall Average Score on Management Tool, ORC Firms, 2004 (N=61)



Because the 5 basic elements of the management systems score are highly correlated, we find that none are statistically significant when all are put in the model at the same time. It should be more feasible to identify the independent effects of each when looking at changes over time. Not all of the elements are likely to change in the same way at the same time, which would provide a better opportunity for that assessment. The predictive value of this measurement tool may seem low, but it is important to realize that other management audit instruments have generally performed with only limited success.[19]

One concern about the ORC and other management system scores is whether those who do the scoring are influenced by the injury rates. If the scorer gives lower scores because the firm has high injury rates or increases the score because the injury rate has decreased, the value of the tool is undermined. Instead of the score predicting the injury rate, in these cases the injury rate would also be determining the score. With such two-way causation, the relationship between scores and injuries becomes very hard to interpret. Ideally, the scorer would not know the injury rates, but this will probably be very difficult to realize.

[19] Philip L. Bigelow and Lynda S. Robson, "Occupational Health and Safety Management Audit Instruments: A Literature Review. Toronto: Institute for Work and Health, 2005