3-1993

Analysis of Persistence in Employer Injury Rates: Final Report

H. Allan Hunt
W.E. Upjohn Institute for Employment Research, hunt@upjohn.org

Upjohn Institute Technical Report No. 93-002

ANALYSIS OF PERSISTENCE IN EMPLOYER INJURY RATES

Upjohn Institute Technical Report No. 93-002

Final Report

Submitted to

Office of Statistics
Occupational Safety and Health Administration
U.S. Department of Labor
under
Purchase Order B9F24231

by

H. Allan Hunt
Assistant Executive Director
W.E. Upjohn Institute for Employment Research
300 S. Westnedge Avenue
Kalamazoo MI 49007-4686
(616) 343-5541

March 1993

Table of Contents

Executive Summary iv

I. Background of the Issue 1
Introduction 1
Previous Studies 1

II. Data and Methods 7
Matched MESC/BWDC 1986-88 Database 7
SET Sample Database 7
Methods 8

III. Correlation Analysis of Injury and Disability Measures 10
MESC/BWDC Database - WC Claims Correlation Analysis Across Time 10
SET Database - Correlation Analysis Across Measures 15
SET Database - Correlation Analysis of OSHA Log Data 16
Conclusion 19

IV. Multivariate Analysis of Injury and Disability Incidence 20
MIOSHA Recordable Injuries 21
Lost Workday Cases 26
Total Lost Workdays 29
Marginal Impact of Lagged Performance Variables 32

V. Conclusions and Research Agenda 34
Summary and Conclusions 34
Future Research Needs 35

References 38

Technical Appendix: Data Sources 40
MESC Data Summary 40
BWDC Data Summary 41
Success of the Match 42
The Matched MESC/BWDC 1986-88 Database 44
SET Sample Database 46
Survey Administrative Procedures 49
Response Bias Analysis 53

List of Figures

4.1 Actual and Predicted Recordable Injuries 26
4.2 Actual and Predicted Lost Work Day Cases 29
4.3 Actual and Predicted Lost Work Days 31

List of Tables

3.1 WC Claim Rate Correlation Analysis, 1986-88 10
3.2 WC Claim Rate Correlation Analysis, 1986-88 11
3.3 WC Claim Rate Correlation Analysis, 1986-88 12
3.4 WC Claim Rate Correlation Analysis, 1986-88 13
3.5 WC Claim Rate Correlation Analysis, 1986-88 13
3.6 WC Claim Rate Correlation Analysis, 1986-88 14
3.7 Correlation Analysis Across Performance Measures 16
3.8 Lost Workday Case Rate Correlation Analysis, 1986-89 17
3.9 MIOSHA Recordable Rate Correlation Analysis, 1986-89 18
3.10 Lost Workday Rate Correlation Analysis, 1986-89 19
4.1 Negative Binomial Regression of 1989 Recordables 23
4.2 Negative Binomial Regression of 1989 Lost Work Day Cases 28
4.3 Negative Binomial Regression of Lost Work Days 30
4.4 Marginal Impact of Lagged Performance Variables 33
A.1 BWDC Incurred Claims - Claim Means 41
A.2 BWDC Incurred Claims - Employer Means 42
A.3 Match Rates, Incurred 1986 Data, Adding Zero Claim Firms 43
A.4 Summary of MESC/BWDC Matched Sample Size 44
A.5 Summary Statistics from MESC/BWDC Matched Database, 1986-88 45
A.6 SET Sampling Strategy 48
A.7 SET Survey Response Rates 50
A.8 Summary Statistics from SET Survey Sample 51
A.9 Establishment Characteristics from SET Survey Sample 52
A.10 Comparison of Respondents with Non-Respondents 54

Executive Summary

This report seeks an empirical answer to a very specific question, "Do injury and disability incidence rates at the establishment level persist through time?" The answer should help to answer a more operational question, "Would OSHA inspection strategy benefit from knowledge of past injury rates at the establishment level?" This report will bring new evidence to bear through the analysis of two Michigan databases. A review of earlier research literature is offered as background for the analysis. Generally speaking, previous studies have shown that there is a good deal of persistence in the frequency of lost workday cases over time. For example, an unpublished study by Ruser (1992) shows that establishments that had worse than average injury records in 1979 were still worse than average five years later.

The current study uses administrative data on Michigan workers' compensation claims from 1986 through 1988 to investigate consistency of firm performance over time. Simple correlation analysis of the incidence of workers' compensation claims for approximately 25,000 Michigan firms shows that the degree of correlation varies with the employment level of the firm. Firms with less than 100 employees generally do not have sufficient exposure to develop consistent claim rates and therefore actually show negative correlation between claim rates in adjacent years. However, for firms with more than 100 employees, the analysis shows that average correlations of .70 for claim rates in adjacent years characterize Michigan firms. This means that about 49 percent of the variation in workers' compensation claim rates in a given year is predictable based on past experience.

A special random sample of 220 Michigan establishments developed for the Safety, Education and Training Division (SET) of the Michigan Department of Labor was also analyzed for correlations, both across performance measures and over time. The SET sample demonstrated a significant level of correlation among different performance measures for the same year, ranging from .35 to .66 for various combinations of MIOSHA recordables, MIOSHA lost workday cases, total MIOSHA lost workdays, and workers' compensation claims. Comparison of adjacent year rates of recordable injuries and lost workday cases revealed correlations of approximately .80 across the 1986-1989 period. The conclusion is that there is substantial consistency of performance through time at the establishment level, with nearly 65 percent of the variation being predictable based on past performance.

A more sophisticated multivariate analysis of the SET database offers even greater insight into the value of knowing past performance levels. Negative binomial regression models were estimated for the number of recordables, the number of lost workday cases, and the total number of lost workdays. A set of characteristics of the establishments, including industry, employment level, multiplant firm status, percent low tenure workers, union status, hourly wage, and self-insured status, are used to explain the level of performance on the three performance measures in 1989. The previous year's level of performance is included

as an explanatory variable in the regression function to determine the contribution it could make. Results show that knowing last year's performance level substantially improves prediction results; increasing the percent of variance explained (predictability) from 40 to 60 percent for total lost workdays and from 39 to 70 percent for lost workday cases.

Even more impressive, an analysis of the marginal contribution of past performance levels in explaining current performance was performed. Adding last year's performance level to a model that included only industry and employment level triples the explanatory power of the model for lost workday cases and total lost workdays. Adding additional establishment specific variables to the model increases the explanatory power by another 6 to 16 percent. Thus, the past performance level is the single most powerful predictor available of current injury and disability performance.

The conclusion is that there is a substantial degree of persistence in injury and disability performance across time at the establishment level. Further, based on these analyses, it is obvious that OSHA would benefit from access to establishment level data in targeting inspections, or other interventions designed to improve occupational safety and health performance.

I. Background of the Issue

Introduction

Currently, industries (at the 4-digit SIC level) are targeted for OSHA inspection based on the combined lost workday injury and illness rate at the national or the state level. The rates are produced from the Bureau of Labor Statistics' (BLS) Annual Survey of Occupational Injuries and Illnesses. Industries are rank-ordered, then all establishments within the top industry are selected for inspection before moving to the next ranked industry. Establishments receiving a complete inspection are exempted for the next two years.

OSHA has undertaken a series of studies designed to answer questions about potential refinements to the current safety inspection targeting system. Different approaches and various data sources are being explored for ranking industries and selecting establishments. Among the alternatives is the use of establishment-specific injury rates to direct inspection resources toward more hazardous firms. Both the National Academy of Sciences and the General Accounting Office have urged OSHA to consider the use of establishment-specific data.

This report seeks the empirical answer to a very specific question, "Do injury and disability incidence rates at the establishment level persist through time, or should they be regarded as essentially random?" The answer should help to inform a more operational question, "Would OSHA inspection strategy benefit from knowledge of past injury rates at the establishment level?" This report brings new evidence to bear on the question through the empirical analysis of two original databases from Michigan.

The report is organized in five major sections. This first section addresses the background of the issue and the other empirical studies that have contributed to its definition. The second section briefly describes the data and methods that this report brings to bear on the problem. The third section presents the findings of the study, consisting of simple correction analyses to illustrate the persistence of injury and disability performance at the establishment level across time, as well as the relationship between the different measures. The fourth section contains a more complex multivariate analysis, which models the performance of establishments on injury and disability measures and evaluates the contribution of past performance to such a model. The fifth section discusses the conclusions from the study and lays out some future research needs. There is also a technical appendix which describes the data used in the study in more detail.

Previous Studies

There have been previous empirical studies of this issue. Gray and Scholz studied this issue in the context of trying to determine whether OSHA inspections reduce subsequent

injuries.[1] Their data come from the Bureau of Labor Statistics (BLS) Annual Survey of Occupational Injuries and Illnesses matched to OSHA's Management Information System for 6,842 manufacturing plants with annual data from 1979 to 1985.[2] This enables them to study the connection between inspections and injuries for a database of large establishments in OSHA states over a 6-year period. There are a few problems with their database that should be recounted before results are reviewed. First, it is noteworthy that the average employment level of establishments in the database is about ten times the level of all manufacturing establishments (523 workers compared to 54). Further, probably because of their size, the plants in their database are six times more likely to have been inspected (.51 inspections per plant versus .08 for all establishments).[3]

Their outcome measures are the percentage change in the number of lost workday injuries and the percentage change in the total lost workdays per 100 workers. These percentage change measures are used to avoid the high correlation between the number of injuries (and lost workdays) in different years in the sample plants which makes conventional estimation techniques inefficient or inaccurate. They reported a correlation coefficient of .68 between the number of injuries in 1979 and injuries in 1985, but only .06 between the percentage change in injuries.[4] However, for the purposes of this project, it is the correlation of levels of injury that is most interesting.

Unfortunately, Gray and Scholz only have three independent variables to include in their multivariate models of the change in injuries and lost workdays; namely, the change in number of employees, the change in hours of work, and the number of inspections resulting in penalty assessments.[5] Presumably, the change in employment would represent the addition of new workers who were not yet familiar with their jobs and, hence, more susceptible to injury. The change in hours variable could be claimed to represent the influence of overtime schedules which might lead to workers less cautious due to fatigue. Gray and Scholz implicitly control for industry and size of firm, because this influence is "built-in" to their fixed effect model.

Gray and Scholz find that the addition of fixed effects variables (essentially a dichotomous variable for each site) to represent plant level characteristics provides a significant improvement in estimating the impact of penalty inspections on injury performance. As they put it:

---

[1]Gray and Scholz (1991).

[2]See Gray and Scholz (1991), pp. 5-6.

[3]Gray and Scholz (1991), p. 6.

[4]Gray and Scholz (1991), p. 7.

[5]They also demonstrated that inspections without penalty had no effect on injury performance.

> We find a significant negative relationship between OSHA enforcement activity at a particular plant and the change in injuries at that plant over subsequent years. We find no evidence that this relationship is the result of endogeneity [injuries causing inspections] of inspections, the negative autocorrelation of injury changes, or plant-specific fixed effects. Based on...[these results], a plant that is inspected (and penalized) in a given year experiences a 22 percent decline in injuries over the following three years, and a 20 percent decline in lost workdays.[6]

They also found that the effects of an OSHA penalty inspection do not last beyond three years; in other words, whatever change can be induced by an OSHA penalty inspection will have been accomplished within a 3-year period.

Another set of studies focused specifically on the policy issue of targeting OSHA inspections. John Ruser of the BLS, working with a database similar to Gray and Scholz, has raised the question of whether injury rates persist through time at the same establishments. If they do, the efficiency of OSHA inspections might be substantially improved by selecting establishments for inspection based on their past performance. On the other hand, if injury rates are dominated by stochastic processes (i.e. they are essentially random events), then there would be no point in targeting OSHA inspections on particular establishments in this way, since past performance would not be a reliable guide to future performance.

Ruser (1991) in a path-breaking methodological and empirical study of the connection between workers' compensation benefits and the incidence of occupational injuries and illnesses concluded that "...it is evident that it is not reasonable to assume independence (of injury rates) over time within an establishment."[7] Ruser reported, for his BLS/OSHA panel data on large establishments, correlations of lost workday case rates ranging from .70 to .93, depending on the specific model estimated, across a 6-year period.[8] These are very high correlations and they argue that future performance is predictable at normal social science levels of certainty.

In a subsequent unpublished paper, Ruser (1992) has applied the methodology developed in the 1991 paper to the analysis of the persistence of establishment injury rates, specifically relating the results to the potential use of establishment level data for OSHA inspection targeting.[9] Ruser estimated a negative binomial model including variables

---

[6]Gray and Scholz (1991), p. 20.

[7]Ruser (1991), p. 343.

[8]Ruser (1991), p. 343.

[9]But note that OSHA inspections have not been included as an explanatory variable in the analysis.

representing the average real weekly wage for production workers, the average real weekly wage-loss benefit for workers' compensation claimants (estimated from weekly wages at the plant), a time trend variable, the annual change in employment, the average weekly overtime hours, the percentage of production workers, and the percentage female workers among the plant workforce.

These covariates were used as an alternative to the plant level fixed effect variables used in the earlier article. Thus, they can be thought of as an attempt to open up the "black box" of the fixed effect at the establishment level by estimating the influence of particular characteristics of the establishment or its workforce. These equations were estimated for each of four establishment size classes (1-99 employees, 100-249 employees, 250-499 employees, over 500 employees) over the period 1979 to 1984 (annual observations).

Ruser estimated this model using maximum-likelihood estimators and then calculated the residuals, i.e. the difference between the predicted number of lost workday cases from the estimated equation and the actual number of lost workday cases for each establishment. Then he discarded all observations with negative residuals in 1979 or 1980, i.e. where the actual level was better than predicted by the model, and analyzed only the establishments that had performed worse than expected (positive residuals) for these two years.[10] When Ruser charted the mean annual residuals for those establishments selected in this way, he reported that the residuals tended to decline over time, but not all the way to zero.

> Importantly, the profiles are not consistent solely with the explanation that injury rates are high due to a bad stochastic draw. If this were the case, the mean residuals would be zero in the years following the year used to select high risk establishments.[11]

Further, Ruser observes that the greatest decline occurs in the first year, which probably reflects the contribution of those establishments who did just happen to have an unusually bad year.

In addition, it was found that results differed by size of establishment, both in terms of the initial displacement and the rate of recovery. For those plants with more than 500 employees, and a mean positive residual of just over 4 lost workday cases in 1979, the residual declined to nearly 2 lost workday cases by 1984 (or about 50 percent). So, after five years time, these establishments are still worse than average by about two cases per year. For plants with 100 to 249 employees, the decline in the residual was from nearly 7

---

[10]There is no obvious reason for this self-imposed selection, unless Ruser believed that the persistence of poor performance should differ significantly from the persistence of good performance for some reason.

[11]Ruser (1992), p. 7.

lost workday cases in 1979 to about 2.5 by 1984 (or about 64 percent).[12] The negative binomial model estimates yielded significant coefficients for almost all the covariates, except weekly overtime hours, and the signs were generally as predicted. While there was some variability, parameter estimates were generally consistent across the size categories as well.[13]

Next, Ruser used the lagged dependent variable (lost workday cases from the previous year) in estimating the one year persistence of the positive residuals from each years' operating results for all establishments with poor results (positive residuals) in 1979. For establishments with more than 500 employees, these estimates show that the overall one year "regression to the mean" was about 25 percent. For establishments with less than 100 employees, it was about 50 percent. Ruser calculated that these estimates meant that it would take from three to eight years for the injury rate to return to normal, with the larger establishments taking longer.

Ruser also lengthened the lag in the dependent variable to see whether the results depended on a specific lag structure. With his panel data, he was able to test lags of up to five years, but found no significant relationship beyond three years.[14] These estimates showed smaller, but still generally significant effects when compared to the one year lag. The time to reconvergence on initial values generally was longer when estimated with longer lags; from 8 to 19 years depending on size.[15] Ruser selected the model with three years of lagged data (third order autoregressive process) as the preferred model. Ruser's conclusion from this analysis is:

> The general conclusion that one can draw from the third order autoregressions is that injury data can aid in targeting inspections. Though for many size classes I cannot reject the hypothesis that the residuals ultimately return to zero, the time it takes for them to approach the steady state is sufficiently long that injury data can identify establishments with high residuals, even when the data is (sic) produced with a lag.[16]

The results of this work by Ruser clearly support targeting OSHA inspections. As Ruser put it:

---

[12]Ruser (1992), figure 1.

[13]Ruser (1992), table 2.

[14]The same finding as Gray and Scholz (1991).

[15]Ruser (1992), table 4.

[16]Ruser (1992), p. 12.

> Given the currently available covariates and the length of time that the residuals remain above zero, there is certainly value in collecting establishment injury rates to use to target inspections. At the same time, an effort should be made to identify and collect additional covariates, so that we may better understand why some establishments are riskier than others. With better covariate information, the role of injury rates themselves as a targeting tool will be clarified.[17]

The remainder of this report is aimed at furthering this inquiry into the persistence of OSHA injury rates and the role of establishment level characteristics (covariates) in explaining that persistence.

[17]Ruser (1992), p. 16.

II. Data and Methods

Two new Michigan data sources are brought to bear on these questions in the current study. One draws on administrative data for the population of firms from the unemployment insurance and workers' compensation systems in Michigan and the other was specifically developed to study the determinants of establishment level performance in Michigan on injury and disability measures.

Matched MESC/BWDC 1986-88 Database

Beginning with the population of Michigan establishments according to the Michigan Employment Security Commission (MESC) (approximately 158,000 establishments in 1988), and matching them to the list of firms with at least one workers' compensation wage-loss claim in each year according to the Bureau of Workers' Disability Compensation (BWDC), yields a matched sample of about 15,000 firms for each year.[18] When these are aggregated over the 3-year period from 1986 to 1988, a total of 28,462 firms are found to have incurred at least one workers' compensation wage-loss claim in the three year period. From these, a total of about 26,000 firms are available for analysis in each year, after imputation of zero claims, elimination of missing data, etc.[19]

This matched MESC/BWDC database contains firm identification information, industry (4-digit SIC), location (by County or Labor Market Area), employment levels (both continuous and categorical), aggregate wage payments (for the second quarter), type of ownership (public or private), number of workers' compensation wage-loss claims, and total workers' compensation indemnity payments for each year, 1986 through 1988. In section III, this database will be used to investigate the persistence of workers' compensation claims rates per 100 employees over the period from 1986 to 1988.

SET Sample Database

The SET sample database was developed during a 3-year study of the determinants of employer injury and disability rates sponsored by the Safety Education and Training Division of the Michigan Department of Labor.[20] This database resulted from a mail survey of a random sample of 220 Michigan establishments with more than 100 employees in seven

[18]Note that Michigan has a 7-day waiting period for workers' compensation wage-loss benefits.

[19]See Technical Appendix for full details of the matching process and discussion of the imputed values that were assigned.

[20]See Hunt, Habeck, VanTol and Scully (1993) for the full study.

industries (2-digit SIC level) conducted in the first half of 1991.[21] The survey achieved a response rate of 46 percent, and is generally representative of the Michigan manufacturing population of establishments with over 100 employees, although respondents are known to be significantly larger and experience about 20 percent fewer workers' compensation claims than non-respondents. The survey protocol was specifically designed to collect data thought to be useful in discriminating between employers with good and poor disability performance records, based on the findings of an earlier study by the same research team.[22]

This database of 220 Michigan establishments contains very substantial self-reported detail about the firm and its disability prevention and management policies and practices.[23] The variables used in the analysis presented here include: industry employment level, whether the establishment was part of a multiplant firm, average hourly wage for production workers, organized union presence at the plant, percent of employees with less than one year tenure, and workers' compensation insurance status (self-insured or carrier). These variables will be employed in a multivariate analysis of the level and persistence of injury and disability rates at the establishment level in the fourth section of this report.

Methods

Two different levels of analysis will be presented in this report. The first, presented in section III, will be a simple correlation analysis of the injury and workers' compensation claim rates per 100 workers for both databases identified above. We will examine the correlation between workers' compensation claim rates in available years, 1986 through 1988, for the 25,000 firms in the MESC/BWDC database. We use workers' compensation claim rates for this purpose because it is the only measure of disability performance available for the entire population of firms. This analysis will focus largely on differences by employment level of the firms, which will be shown to be critical to the basic policy questions raised here.

Then, we will turn to the SET sample database to examine the correlation across different injury and disability measures for the same year, and across the years 1986 through 1989 for the same measure. In the first analysis, we will use the MIOSHA recordable rate, MIOSHA lost workday case rate (excluding restricted day cases), the total MIOSHA lost workday rate, and the workers' compensation wage-loss claim rate to examine the consistency of performance across the different measures. The subsequent analyses will focus on the MIOSHA recordable rate, MIOSHA lost workday case rate, and the total

[21]See Technical Appendix for more detail on the SET sample survey.

[22]See Habeck, Leahy, Hunt, Chan and Welch (1991) for an account of these findings.

[23]See Hunt, Habeck, VanTol and Scully for more detail on the design of the instrument and survey procedures that resulted in the SET sample database.

MIOSHA lost workday rate for the years 1986 through 1989 to specifically examine the persistence of injury and disability performance through time.

Last, a multivariate analysis of these MIOSHA variables will be presented for the SET sample database of 220 establishments. Using a "count" model (that uses the number of incidents rather than the rate per 100 workers) similar to the one developed by Ruser (1992), we will examine the persistence of injury and disability rates across time, controlling for establishment characteristics. We will use the negative binomial regression model to estimate these multivariate relationships, then correlate the predicted level of injuries from the model with the actual level reported by the establishment. Correlation analysis of these two levels will enable us to determine how valuable it would be to know previous performance levels in predicting current performance.

III. Correlation Analysis of Injury and Disability Measures

Using the two databases described briefly in the previous section, an analysis of the correlation of injury and workers' compensation claims performance will be reported.[24] This analysis is considerably less sophisticated than the multivariate empirical work that will be presented later, but it will focus directly on the issue of the degree of persistence of establishment performance across time. Therefore, these results will be easily interpretable, but they do not control for other influences that may be important. However, this simple analysis does offer significant insight to the question of whether knowing last year's performance is valuable in predicting current performance on these critical safety and disability dimensions. When confirmed by the multivariate analysis to be presented in section IV, they can be regarded as authoritative.

MESC/BWDC Database - WC Claims Correlation Analysis Across Time

Table 3.1 reports the correlation coefficients across the 1986 to 1988 period for the MESC/BWDC matched database of some 25,000 Michigan firms. This database represents the full population of firms that experienced at least one workers' compensation claim during the three years.[25] The table shows that for this total database, the workers' compensation (WC) claim rate per 100 workers is very slightly negatively correlated across time, i.e., the higher the claim rate in 1987, the lower the claim rate in 1988.

Table 3.1 WC Claim Rate Correlation Analysis, 1986-88†
MESC/BWDC Matched Database

Mean	Variable	Correlation with 1987 Claim Rate	Correlation with 1988 Claim Rate
6.291	1986 Claim Rate	-0.011	-0.021
	(n)	(23738)	(22737)
6.001	1987 Claim Rate		-0.025
	(n)		(24286)

† Claim Rate is the number of workers' compensation wage-loss claims per 100 employees. The number of firms (n) is shown beneath the correlation coefficient in parentheses.

[24]A full description of the data sources is presented in the Technical Appendix.

[25]Subject to the match difficulties described in the Technical Appendix. The fact that we are dealing with the population of firms means that it is not appropriate to use inferential statistics because we are reporting on the population.

However, these coefficients are very low at around ρ = -.02, so it would appear that table 3.1 shows that there is not strong persistence through time of WC claim rates when the firm is the unit of analysis. But this analysis ignores the fact that most firms had zero claims in every year, since only 25,000 of 150,000 firms are even in table 1. When these imputed zero claim firms are included, the correlation coefficients rise to about .13 to .14. Since this requires strong assumptions about the firms that are missing from table 3.1, these results are not being presented in tabular form. However, table 3.1 is still misleading, because it is dominated by the experience of small firms. This will be clearly demonstrated in the tables that follow, which control for the employment size of the firm.

Table 3.2 reports the same coefficients for the population of about 22,000 firms with at least one workers' compensation claim during the three year observation period (1986-88) and less than 100 employees in 1988. Two observations are in order. First, well over 80 percent of the firms in the total WC claims population are firms with less than 100 employees. Second, there is a clear negative correlation between WC claim rates in different years for this group of firms (ρ = -.04); but it seems likely that this result is a consequence of the relatively low incidence of workers' compensation claims in general and the relatively small size of these firms in particular.

Table 3.2 WC Claim Rate Correlation Analysis, 1986-88†
MESC/BWDC Matched Database
Firm Size < 100 Employees

Mean	Variable	Correlation with 1987 Claim Rate	Correlation with 1988 Claim Rate
6.904	1986 Claim Rate	-0.032	-0.043
	(n)	(20151)	(19237)
6.560	1987 Claim Rate		-0.048
	(n)		(20636)

† Claim Rate is the number of workers' compensation wage-loss claims per 100 employees. The number of firms (n) is shown beneath the correlation coefficient in parentheses.

To illustrate this point, consider the following example. If the average annual workers' compensation claim rate is 3.0 claims per 100 workers, a typical firm with an employment level of 100 employees could be expected to have about 3 claims per year, on the average. But a firm with an employment level of 25 would only be expected to have .75 claims per year, and a firm of 10 employees would only be expected to have .3 claims per year. Of course, workers' compensation claims occur only in integers (i.e. no fractions), so

this last firm could expect about 1 claim every three years. In fact, the modal number of claims for firms of less than 100 employees is zero.[26] So, for the great mass of small firms, having a workers' compensation claim is a fairly unusual experience. In fact, if a small firm does experience a claim in one year, it is less likely that they will experience a claim in the following year, because their actual expected claim incidence is less than one per year. Thus the negative correlation between claims incidence in adjacent years reflects this indivisibility problem for small firms, compounded with a relatively low expected incidence.

To illustrate the impact of this problem table 3.3 repeats the analysis of table 3.2, but it only includes firms with less than 100 employees who had at least one workers' compensation claim in each of the three years. This sub-group of small firms (about 13 percent of the firms in table 3.2) shows positive correlation of the workers' compensation claim rate in adjacent years (1986 with 1987 and 1987 with 1988) of about $\rho = .53$, and across a two year time period (1986 to 1988) of $\rho = .41$. So when the firms with zero claims in some years are omitted, the small negative correlation coefficients are replaced with sizeable positive coefficients. This demonstrates how important it is to control firm size when analyzing the persistence of injuries.

Table 3.3 WC Claim Rate Correlation Analysis, 1986-88†
MESC/BWDC Matched Database
Firm Size < 100 Employees
Firms with Claims in Each Year

Mean	Variable	Correlation with 1987 Claim Rate	Correlation with 1988 Claim Rate
7.701	1986 Claim Rate	0.541	0.408
	(n)	(2774)	(2774)
7.492	1987 Claim Rate		0.526
	(n)		(2774)

† Claim Rate is the number of workers' compensation wage-loss claims per 100 employees. The number of firms (n) is shown beneath the correlation coefficient in parentheses.

Table 3.4 reports the correlation coefficients for approximately 2,350 firms with 100 to 249 employees in 1988. Here the correlation coefficients range from .54 to .68 across the years of observation for firms with at least one claim in the three year observation period.

[26]The mode is the single value which contains the highest number of observations.

Table 3.5 covers about 800 firms with 250 to 499 employees, except the correlation coefficients are even higher. Table 3.5 shows that adjacent year correlation coefficients for these mid-size firms are approximately $\rho = .83$. These are large correlation coefficients and it is also revealing that they increase with size of firm, as the law of large numbers would suggest. Thus, when the influence of small firms (who are unlikely to sustain workers' compensation wage-loss claims) is reduced, very substantial persistence from year to year is revealed in the incidence of claims.

Using the fact that the squared correlation coefficient gives an approximation to the percent of variance the two variables share in common, it is possible to say that the WC claim rate in 1987 accounts for over 60 percent of the variance in WC claim rate in 1988 for firms with 250 to 499 employees. This is substantial persistence by any reasonable standard.

Table 3.4 WC Claim Rate Correlation Analysis, 1986-88†
MESC/BWDC Matched Database
Firm Size 100-249 Employees

Mean	Variable	Correlation with 1987 Claim Rate	Correlation with 1988 Claim Rate
2.947	1986 Claim Rate	0.538	0.611
	(n)	(2224)	(2167)
2.846	1987 Claim Rate		0.678
	(n)		(2273)

† Claim Rate is the number of workers' compensation wage-loss claims per 100 employees. The number of firms (n) is shown beneath the correlation coefficient in parentheses.

Table 3.5 WC Claim Rate Correlation Analysis, 1986-88†
MESC/BWDC Matched Database
Firm Size 250-499 Employees

Mean	Variable	Correlation with 1987 Claim Rate	Correlation with 1988 Claim Rate
2.664	1986 Claim Rate	0.824	0.758
	(n)	(766)	(753)
2.624	1987 Claim Rate		0.831
	(n)		(774)

† Claim Rate is the number of workers' compensation wage-loss claims per 100 employees. The number of firms (n) is shown beneath the correlation coefficient in parentheses.

Finally, table 3.6 reports the same analysis for the 618 firms in the MESC/BWDC matched database with more than 500 employees. The correlation coefficients for this group of firms range from $\rho = .65$ to .78; quite a high level for observations at this level of disaggregation. It is puzzling, however, that the correlation coefficients are lower for firms with over 500 employees than for firms with 250 to 499 employees. It would seem logical that any bias introduced by the integer count of WC claims would contribute less to the variance of the WC claim rate the larger the firm. Perhaps the greater presence of multi-establishment firms in this size category has provided an offsetting influence.

Table 3.6 WC Claim Rate Correlation Analysis, 1986-88†
MESC/BWDC Matched Database
Firm Size 500+ Employees

Mean	Variable	Correlation with 1987 Claim Rate	Correlation with 1988 Claim Rate
2.299	1986 Claim Rate	0.668	0.777
	(n)	(593)	(576)
2.244	1987 Claim Rate		0.652
	(n)		(598)

† Claim Rate is the number of workers' compensation wage-loss claims per 100 employees. The number of firms (n) is shown beneath the correlation coefficient in parentheses.

In search of a single number to express the degree of persistence of WC claim rates across this large sample of Michigan firms, the following exercise is offered. Suppose we average all the adjacent year correlation coefficients (1986 with 1987 and 1987 with 1988) from firms with more than 100 employees (namely from tables 3.4, 3.5 and 3.6.). The result is an unweighted average correlation coefficient of $\rho = .699$. Repeating the exercise for observations that are two years apart (i.e. 1986 with 1988) yields an unweighted average coefficient of $\rho = .715$. The conclusion is that the average correlation coefficient for WC claims in different years in the population of Michigan firms of over 100 employees is approximately .70. Thus, about 50 percent of the variance in WC claim rates is accounted for by the claim rate in the previous year.

Thus, this simple correlation analysis shows very high correlation coefficients for firm level workers' compensation claim rates across time. Furthermore, these numbers are consistent with those reported by Ruser and Gray and Scholz for an entirely different database. Gray and Scholz found a correlation coefficient of $\rho = .68$ between the number of

lost workday injuries in 1979 and the number in 1985 (a 6-year internal).[27] Ruser reported a range of ρ = .70 to .93 for different models and different years.[28] This confirmation of those earlier results with a different performance measure and a different dataset is very reassuring. The question is, "How relevant are these findings on workers' compensation claim rates to the OSHA operational problem of selecting the targets for enforcement inspections?" The answer depends on how close the link is between OSHA injury measures and workers' compensation claims. The next section will address this issue utilizing the SET database described earlier.

SET Database - Correlation Analysis Across Measures

The SET sample survey included questions on the number of workers' compensation claims incurred in 1989, but it also requested MIOSHA log data for 1986, 1987, 1988, and 1989. The instrument carefully directed the respondent's attention to the specific items from the MIOSHA Form 200 (Log and Summary) that were requested. These data enable a comparison across different measures of injury incidence, injury severity, and workers' compensation claims incidence. The SET database contains measures of the number of recordable injuries, the number of lost workday cases, the total number of lost workdays, and the number of workers' compensation claims incurred. The data were gathered in this way to permit analysis of the flows from one level to the next (i.e. how many injuries lead to lost workday cases?) and to study the impact of policy and practice interventions on these flows.[29]

Table 3.7 displays these correlations across performance measures for the 220 surveyed firms for calendar year 1989. The general conclusion is that there is a substantial degree of correlation between these four different performance measures in the same establishment. In particular, the correlation of the MIOSHA recordable rate per 100 employees with: (1) the MIOSHA lost workday case rate per 100 employees is ρ = .629; (2) with the workers' compensation claim rate per 100 employees is ρ = .405; and (3) with the total MIOSHA lost workday rate is ρ = .346. These are substantial correlations, although less impressive than were shown when comparing the same measure across time in the earlier analyses. Addressing the issue of whether the workers' compensation claim rate is a relevant measure for discussing injury incidence, the answer is clearly yes. Table 3.7 shows that the correlation between the WC claim rate and the lost workday case rate in 1989 is ρ = .663. This is nearly comparable to the correlations demonstrated between adjacent years in the WC case rate earlier (average ρ = .699). So it seems clear that for this random sample of

---

[27]Gray and Scholz (1991), p. 7.

[28]Ruser (1991), p. 343.

[29]See Hunt, Habeck, VanTol and Scully (1993) for a full explanation of the strategy.

Michigan establishments, the workers' compensation wage-loss claim rate measure performs similarly to the lost workday case rate measure.[30]

Table 3.7 Correlation Analysis Across Performance Measures†
SET Sample Database

Mean	Variable	Correlation with LWDCAS89	Correlation with LWDRTE89	Correlation with CLMRTE89
22.56	MIORTE89	0.629**	0.346**	0.405**
	(n)	(205)	(204)	(193)
6.08	LWDCAS89		0.617**	0.663**
	(n)		(205)	(194)
133.5	LWDRTE89			0.522**
	(n)			(193)

† The number of firms (n) is shown beneath the correlation coefficient in parenthesis.

MIORTE89 = MIOSHA Recordables per 100 Workers in 1989
LWDCAS89 = MIOSHA Lost Workday Cases per 100 Workers in 1989
LWDRTE89 = MIOSHA Lost Workdays per 100 Workers in 1989
CLMRTE89 = Workers' Compensation Claims Incurred per 100 Workers in 1989

**Correlation coefficient significant at the 99 percent confidence level.

SET Database - Correlation Analysis of OSHA Log Data

Table 3.8 displays the correlation analysis across the 1986 to 1989 period for the lost workday case rate from the SET sample. There is more variability in these tables than in the earlier ones for the MESC/BWDC matched database, because of the stratified sampling design and because the number of observations is so much smaller. However, the advantage of having OSHA log performance measures outweighs the disadvantage of stochastic noise. Table 3.8 shows that the correlation between adjacent years (1986 with 1987, 1987 with 1988, and 1988 with 1989) in the lost workday case rate is very high; greater than $\rho = .80$ for two of the three pairs. The substantially lower correlation of 1987 with 1988 presumably reflects random variation.

[30]In fact, MIOSHA uses the workers' compensation claim data to target inspections in Michigan.

The table also seems to show a more dramatic fall-off in correlation than we saw earlier after the first year, particularly from 1986 to 1988 and 1989. This was not reported by either Ruser or Gray and Scholz for their 1979 to 1985 panel data, which would have had substantially more 2-year and 3-year period comparisons available.[31] However, too much should not be made of this finding unless it can be confirmed elsewhere because it is likely due to sampling error or response bias in this relatively small sample. The 1986 comparison point in the SET sample restricts the sample to the smallest number of observations (n = 157) due to non-response on the earlier year MIOSHA log data.

Table 3.8 Lost Workday Case Rate Correlation Analysis, 1986-89†
SET Sample Database

Mean	Variable	Correlation with LWDCAS87	Correlation with LWDCAS88	Correlation with LWDCAS89
7.26	LWDCAS86	0.847**	0.314**	0.237**
	(n)	(157)	(156)	(157)
6.96	LWDCAS87		0.530**	0.496**
	(n)		(176)	(178)
6.47	LWDCAS88			0.835**
	(n)			(188)

† The number of firms (n) is shown beneath the correlation coefficient in parentheses.

LWDCAS86 = MIOSHA Lost Workday Case Rate per 100 Workers in 1986
LWDCAS87 = MIOSHA Lost Workday Case Rate per 100 Workers in 1987
LWDCAS88 = MIOSHA Lost Workday Case Rate per 100 Workers in 1988
LWDCAS89 = MIOSHA Lost Workday Case Rate per 100 Workers in 1989

**Correlation coefficient significant at the 99 percent confidence level

Table 3.9 continues the examination of OSHA log data from the SET sample with the number of MIOSHA recordables per 100 workers across the 1986 to 1989 period. As we have seen before, the table shows a very high level of correlation between adjacent year measures of the outcome variable (ranging from ρ = .705 to .895, average ρ = .808) and showing a slight fall-off as the time period between observations increases. Thus, this

[31]Recall however, that both Ruser (1991) and Gray and Scholz (1991) determined that three years was the optimal lag structure.

evidence argues that recordables are nearly as consistent as lost workday cases across time.[32]

Table 3.9 MIOSHA Recordable Rate Correlation Analysis, 1986-89†
SET Sample Database

Mean	Variable	Correlation with MIORTE87	Correlation with MIORTE88	Correlation with MIORTE89
22.72	MIORTE86	0.705**	0.599**	0.552**
	(n)	(156)	(155)	(156)
23.39	MIORTE87		0.824**	0.761**
	(n)		(177)	(179)
22.97	MIORTE88			0.895**
	(n)			(189)

† The number of firms (n) is shown beneath the correlation coefficient in parentheses.

MIORTE86 = MIOSHA Recordables per 100 Workers in 1986
MIORTE87 = MIOSHA Recordables per 100 Workers in 1987
MIORTE88 = MIOSHA Recordables per 100 Workers in 1988
MIORTE89 = MIOSHA Recordables per 100 Workers in 1989

**Correlation coefficient significant at the 99 percent confidence level

Table 3.10 displays the analysis for the MIOSHA severity measure, lost workdays per 100 workers. The overall level of correlation of adjacent year observations is somewhat lower for this measure, presumably reflecting the additional element of variation that has been introduced with the number of lost workdays per case. This variable is very difficult to predict, so it makes the correlations between years for this severity measure less reliable than the others.[33]

[32]Results from Hunt, Habeck, VanTol and Scully (1993) show that their determinants are different, however.

[33]The adjacent year correlations for lost workdays per lost workday case range from $\rho = .179$ to .342, indicating that there is less consistency in this measure across years. This was confirmed in our SET project, where we were unable to do better than chance in predicting the level of lost workdays per case.

Table 3.10 Lost Workday Rate Correlation Analysis, 1986-89†
SET Sample Database

Mean	Variable	Correlation with LWDRTE87	Correlation with LWDRTE88	Correlation with LWDRTE89
139.0	LWDRTE86	0.525**	0.550**	0.373**
	(n)	(156)	(155)	(156)
129.1	LWDRTE87		0.666**	0.534**
	(n)		(175)	(177)
129.2	LWDRTE88			0.568**
	(n)			(188)

† The number of firms (n) is shown beneath the correlation coefficient in parentheses.

LWDRTE86 = MIOSHA Lost Workdays per 100 Workers in 1986
LWDRTE87 = MIOSHA Lost Workdays per 100 Workers in 1987
LWDRTE88 = MIOSHA Lost Workdays per 100 Workers in 1988
LWDRTE89 = MIOSHA Lost Workdays per 100 Workers in 1989

**Correlation coefficient significant at the 99 percent confidence level

Conclusion

The conclusions of this simple correlation analysis are clear. There is substantial consistency between the same measure of employer performance in different years.There is also a significant degree of consistency in employer performance across the different measures of injury and disability incidence. Thus, while this analysis is not sophisticated, it does not really have to be. The relevant question is whether knowing this year's injury rate is useful in predicting next year's injury rate; the answer is obviously yes. The next section will address the determinants of injury and disability levels with a more rigorous multivariate methodology, but the basic policy question has already been answered here.

IV. Multivariate Analysis of Injury and Disability Incidence

A supplementary analysis that extends the results of the last section into the more complex multivariate realm will be presented here. Utilizing the SET survey database described briefly in section II, and informed by the literature presented in section I, we can provide a more thorough answer to the basic question of this report, "Do injury and disability incidence rates persist through time at the same establishment?" We can also provide a tentative answer to the direct policy question, "How valuable would it be to OSHA to know establishment level performance on injury and disability measures?"

Taking a lesson from Ruser, we will use the negative binomial regression model to estimate models of the number of MIOSHA recordable injuries, the number of MIOSHA lost workday cases, and the total number of MIOSHA lost workdays in 1989 using the SET sample database described above.[34] Also like Ruser, we will use the lagged dependent variable (i.e. the 1988 number of recordables, lost workday cases, or total lost workdays) as one of the arguments in the model.[35]

As reported earlier, the SET survey database was designed to collect information directly from employers that would be useful in determining the institutional correlates of establishment performance on injury and disability measures.[36] For this analysis, some exogenous variables thought to affect injury and disability likelihood were selected for inclusion in the model. They included: a dichotomous (yes-no) variable to indicate whether the establishment is part of a multiple plant firm; the percent of employees at the establishment with less than one year of tenure; a dichotomous variable indicating the presence of an organized union at the plant; a dichotomous variable for self-insured status for workers' compensation; the hourly wage level for production workers; and variables representing the industry (dichotomous variables for 2-digit SIC) and employment level of the establishment (both continuous and dichotomous).

---

[34]These count data models are statistically preferable in cases where a large number of observations take the legitimate value of zero, (i.e. no lost workday cases in a given year) and/or where an ordinary least squares or weighted least squares model would yield predictions outside the legitimate range of measurement for the dependent variable (i.e. a predicted lost workday case rate of less than zero for given establishment characteristics). The negative binomial model is useful where the dependent variable is a non-negative integer, but where the Poisson distribution assumptions of equal distribution mean and variance cannot be maintained.

[35]Unlike Ruser, we will not analyze only the residuals from such an estimated model. It is the level of injuries, not changes in the level, that are of most interest to OSHA. By focussing on the "unexpected" variance in lost workday case incidence, Ruser maximizes the influence of the stochastic process. We prefer to focus on predicting the current level of performance, and describe changes in the performance of the estimated model with and without last year's performance level included.

[36]See Hunt, Habeck, VanTol and Scully (1993) for full details on data collection and variable definition for the SET project.

The multiple plant variable is expected to have a negative sign, as an establishment that is part of a larger entity will bring greater policy and implementation resources to bear on its injury and disability problems. The percent of workers with tenure less than one year should have a positive sign, as the literature amply documents that lack of experience with the specific job is a major determinant of injury frequency. The presence of a union might have either a positive or negative effect on injuries and disabilities. On the one hand, the union likely will have bargained over safety performance, or established a joint union-management safety committee, which should reduce the number of injuries. On the other hand, the presence of the union, and the job security it brings to individual workers, may increase the reporting of injuries and disabilities over that found in an unorganized environment.

The self-insured variable is expected to have a negative sign, since the greater financial incentives for reducing injuries and workers' compensation claims should, other things equal, mean better performance on injury prevention activities. However, this effect could be offset by superior loss-control consultation services provided by workers' compensation insurance carriers. The wage level is expected to have a negative sign also, as a general proxy for the quality of the working environment.[37] Higher wage employers have been demonstrated to follow a more investment oriented strategy in dealing with their employees, and this is expected to carry over into safety and prevention efforts as well. However, the economic literature demonstrates that there may be compensating wage differentials for the probability of injury, which could produce a positive effect.

The effect of past performance is incorporated in these models by including last year's value for the dependent (outcome) variable. The variable is expressed as the product of 1988 performance level interacted with a set of dichotomous size variables (for small, medium, and large establishments) to allow the effect of past performance to vary with the employment level. This specification provides separate estimates of the influence of last year's performance for each size group. It is expected that the impact of past performance would differ by size, both because of scale effects and inherent stochastic differences.

MIOSHA Recordable Injuries

Table 4.1 reports the negative binomial regression estimate for the number of OSHA recordables in 1989, as reported by the 220 establishments in the SET sample survey.[38] Because of the complex mathematical form of the negative binomial equation, the coefficients are not always easily interpretable.[39] However, a general approximation is that for a one

---

[37]See Ruser (1991) for a similar finding.

[38]The Michigan OSHA definitions are identical to Federal requirements.

[39]Presentation of the specification of the negative binomial model is beyond the scope of this report. See Ruser (1991) for a discussion of the econometric issues.

unit change in the value of each of the independent variables, the coefficient reports the percentage change in the dependent variable; in the case of table 4.1, the number of recordables in 1989.

Thus, for example, the table reports that the coefficient for self-insurance is .393. Since self-insured is a dichotomous (yes or no) variable, this means that self-insured establishments reported, on the average, 39 percent more recordables than non self-insured establishments, controlling for the other characteristics in the regression. Further, the asterisks indicate that this result is statistically significant at the 99 percent level of confidence. This is an unexpected result, since traditional assumptions about greater incentives for safety in self-insured establishments have caused us to believe that such firms would show better performance.[40] However, there has been no other database that measured both employment level and self-insured status directly, so this finding is really unprecedented.

We can proceed similarly for the other dichotomous (yes or no) variables in the analysis. This includes the multiple plants variable, which shows a coefficient of -.178, or an average 17.8 percent reduction in recordables for plants that reported they were part of a larger, multi-plant firm. This result is significant at the 90 percent confidence level. The presence of a union at the establishment is associated with a reduction of 6 percent in the number of recordables, but this result cannot be distinguished from zero based on these results (i.e. not statistically significant).

[40]See Smith (1992) for a useful summary of this literature.

Table 4.1 Negative Binomial Regression of 1989 Recordables
SET Survey Sample

Mean	Standard Deviation	Variable	Coefficient	Std. Error	t-statistic & Significance
		Intercept	.1639	.3571	.459
13.310	27.949	1988 Recordables X Small	.0058	.0017	3.418**
32.011	62.614	1988 Recordables X Medium	.0056	.0006	9.040**
80.576	269.16	1988 Recordables X Large	.0013	.0002	8.148**
.668	.472	Multiple Plants	-.1783	.1034	-1.724ρ
10.557	12.831	Percent Tenure <1 Year	.0051	.0057	.884
.620	.487	Union	-.0610	.1062	-.575
.462	.500	Self-Insured	.3930	.1033	3.806**
10.378	2.787	Wage	-.0282	.0211	-1.335
.049	.216	SIC 25	-.3883	.2712	-1.431
.168	.375	SIC 30	-.0681	.1853	-.368
.239	.428	SIC 34	.0969	.1705	.569
.130	.338	SIC 35	-.1493	.2103	-.710
.163	.370	SIC 37	-.0738	.1863	-.396
.130	.338	SIC 80	-.8326	.1756	-4.742**
5.910	.934	Employment (ln)	.6952	.0661	10.519**
		Alpha	.2667	.0322	8.278**

n = 184
Mean log-likelihood = -4.972
r (Y, $\hat{Y}$) = .782

ρ = significantly different from zero at 90% confidence level
* = significantly different from zero at 95% confidence level
** = significantly different from zero at 99% confidence level

The set of dichotomous industry control variables indicate the comparison between the various 2 digit industry groups and the reference category, SIC 20. It is necessary to designate one of the industries as the reference category and to measure the effect of the others as divergence from the reference category to avoid over-determining the system. The selection of the reference category is arbitrary, but it will affect the significance of the coefficients, since they are measured relative to the influence of the reference category. Further, the reference category is not listed in the table, since its coefficient is effectively set at 0. The table indicates that SIC 80 and SIC 25 have substantially lower numbers of recordables than SIC 20, while SIC 34 shows a higher number of recordables. The statistical significance of these coefficients is not particularly important since they are all relative measures. However, it is important to remember that the effects of industry are controlled in this analysis.

The continuous variables are somewhat more difficult to interpret than the dichotomous variables. The hourly wage for production workers coefficient indicates that for a $1.00 increase in the hourly wage, a reduction of 2.8 percent in the number of recordables was observed across the SET sample establishments (this result is only significantly different from zero at an 80 percent confidence level). The percentage of low tenure workers at the establishment is not strongly associated with the number of recordables; the non-significant coefficient indicates that for each percentage point increase in this measure, there is about .5 percent increase in the number of recordables.

The employment level variable is highly significant and positively related to the number of recordables. Since it was measured in log form, the interpretation is somewhat more direct because the coefficient gives the change in recordables with respect to changes in employment. Thus, the coefficient indicates that a one percent increase in employment is associated with .7 percent increase in recordables. This is the familiar finding that larger firms perform better on injury and disability prevention than smaller firms.[41] Since the number of recordable injuries increases less rapidly than employment, the injury rate must be declining with size.

The coefficients that incorporate the previous year's number of recordables (lagged variable) are interacted with size categories, to allow for the possibility that the effects are different for large, medium, and small establishments. Large establishments are defined to be those with more than 500 employees, medium establishments range from 250 to 499 employees, and small establishments have from 100 to 249 employees. All three lagged dependent variables show highly significant positive coefficients. This confirms that last year's performance on recordables (lagged variable) has a very significant effect on this year's performance.

[41]Compare to Ruser (1991), table 2.

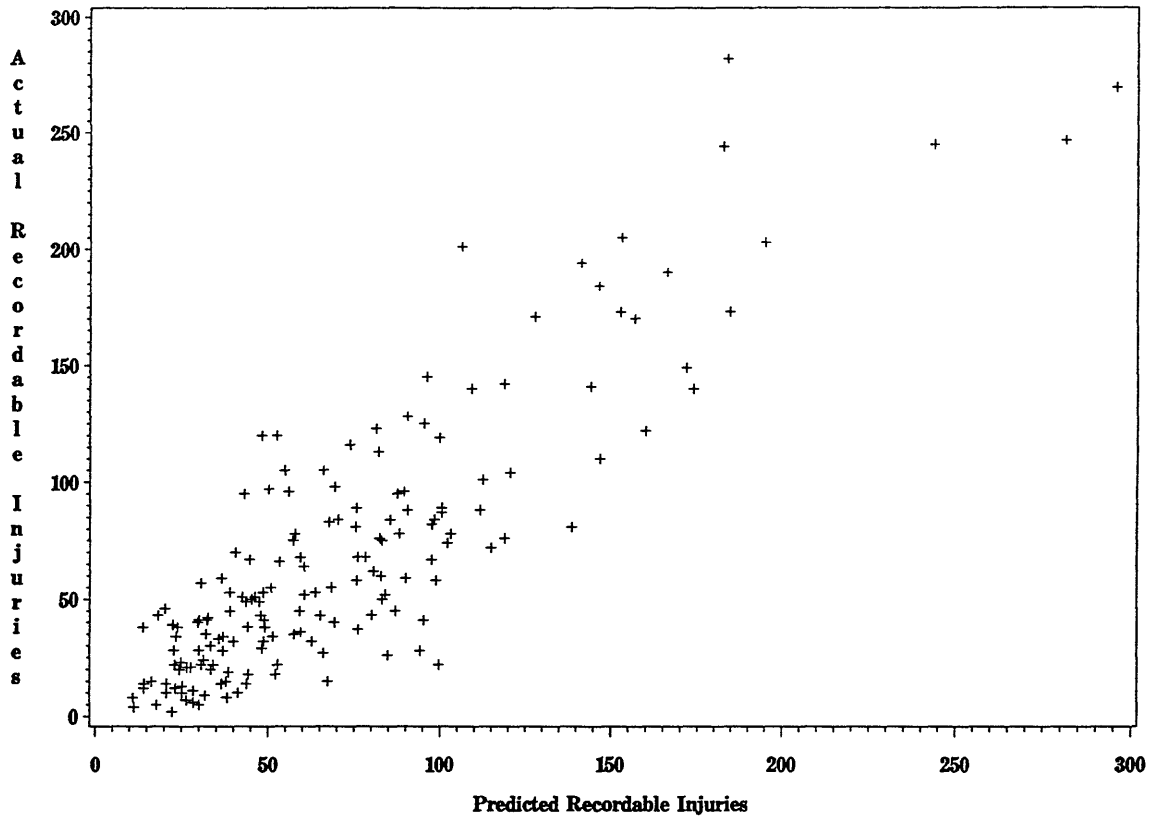
The small firm (100 to 249 employees) coefficient indicates that for each additional recordable in 1988, the number of recordables in 1989 increases by .58 percent for small firms. For medium size firms (250 to 499 employees), each additional recordable in 1988 is associated with an increase in 1989 recordables of .56 percent. For large firms (over 500 employees), each additional 1988 recordable is associated with .13 percent increase in 1989 recordables. When evaluated at the mean of the lagged dependent variable (recordables in 1988), these coefficients imply elasticities of .077 for small establishments, .181 for medium sized establishments, and .104 for large establishments. Thus a 10 percent higher level of recordables in 1988 is associated with 1 percent greater recordables (.10 x .104) for 1989 for large employers, almost 2 percent greater for medium sized (.10 x .181) and less than 1 percent greater (.10 x .077) for small employers, respectively.

The overall level of explanatory power of the negative binomial regression model will be represented in two ways, with a correlation coefficient and a plot of actual versus predicted values. First, the correlation (Pearson product-moment correlation coefficient) between the actual number of recordables in 1989 reported by the sample establishments and the predicted number of recordables from the negative binomial regression model is .782. The square of this correlation coefficient gives the percentage of the variance that is explained by the correlation relationship. The model does quite well in predicting the number of recordables, accounting for approximately 61 percent of the total variance in 1989 recordables present in the sample.

Figure 4.1 displays these results graphically for the bulk of the sample. The figure shows the scatter diagram for the actual and predicted number of recordables. The figure was truncated at 300 recordables, so the largest establishments are not represented here, but the figure shows a good cluster around the 45 degree line which would represent perfect prediction of 1989 recordables.

While this work should be regarded as somewhat exploratory, the model appears to provide a good fit and the coefficients generally have the expected signs. The biggest surprise is in the coefficient for self-insured employers, but our hypothesis on this coefficient was derived from other work that did not have a direct measure of self-insured status. Our results show that self-insured establishments have higher injury rates when controlling for other influences. Thus, the evidence presented here provides confirmation of the basic judgment reached in the correlation analysis that the number of recordables is substantially consistent from year to year.

Figure 4.1 Actual and Predicted Recordable Injuries
Values Less Than or Equal to 300



Lost Workday Cases

Table 4.2 reports the negative binomial regression estimate for the lost workday cases reported by the establishments in the SET survey sample.[42] The table shows that many of the relationships are similar to those shown for recordables. Self-insurance and employment size are significantly and positively related to the number of lost workday cases in 1989. Establishments that are part of multi-plant firms, with higher wages, and those with union representation have lower levels of lost workday cases, but none of these coefficients is statistically significant. The percent low tenure workers is not associated with the number of lost workday cases.

The coefficients for the lagged dependent variable (1988 lost workday cases) interacted with size class show similar patterns as well. All are highly significant and

[42]For the purposes of the SET survey, restricted days were not included in this measure.

positive, indicating a high level of continuity from 1988 to 1989 in the number of lost workday cases. For small firms, one additional lost workday case in 1988 is associated with an increase of 3.3 percent in 1989 lost workday cases. For medium size firms, the figure is 2.9 percent, and for large firms, one additional 1988 lost workday case is associated with .75 percent more lost workday cases in 1989. When evaluated at the mean of the lagged dependent variable (lost workday cases in 1988), these coefficients imply elasticities of .146 for small establishments, .231 for medium sized establishments, and .125 for large establishments. Thus a 10 percent higher level of lost workday cases in 1988 is associated with 1.2 percent greater lost workday cases (.10 x .125) in 1989 for large employers, about 2.3 percent greater for medium sized (.10 x .231) and between 1 and 2 percent greater (.10 x .146) for small employers, respectively.

The table indicates that the correlation between the actual and predicted levels of lost workday cases is .836, meaning that the estimated model accounts for about 70 percent of the variance in lost workday cases. This is nearly the same level of correlation between adjacent year measures of lost workday case rates reported earlier in section III. It indicates that the model is very successful in predicting the level of lost workday cases, doing even better with lost workday cases than for recordables.

Table 4.2 Negative Binomial Regression of 1989 Lost Work Day Cases
SET Survey Sample

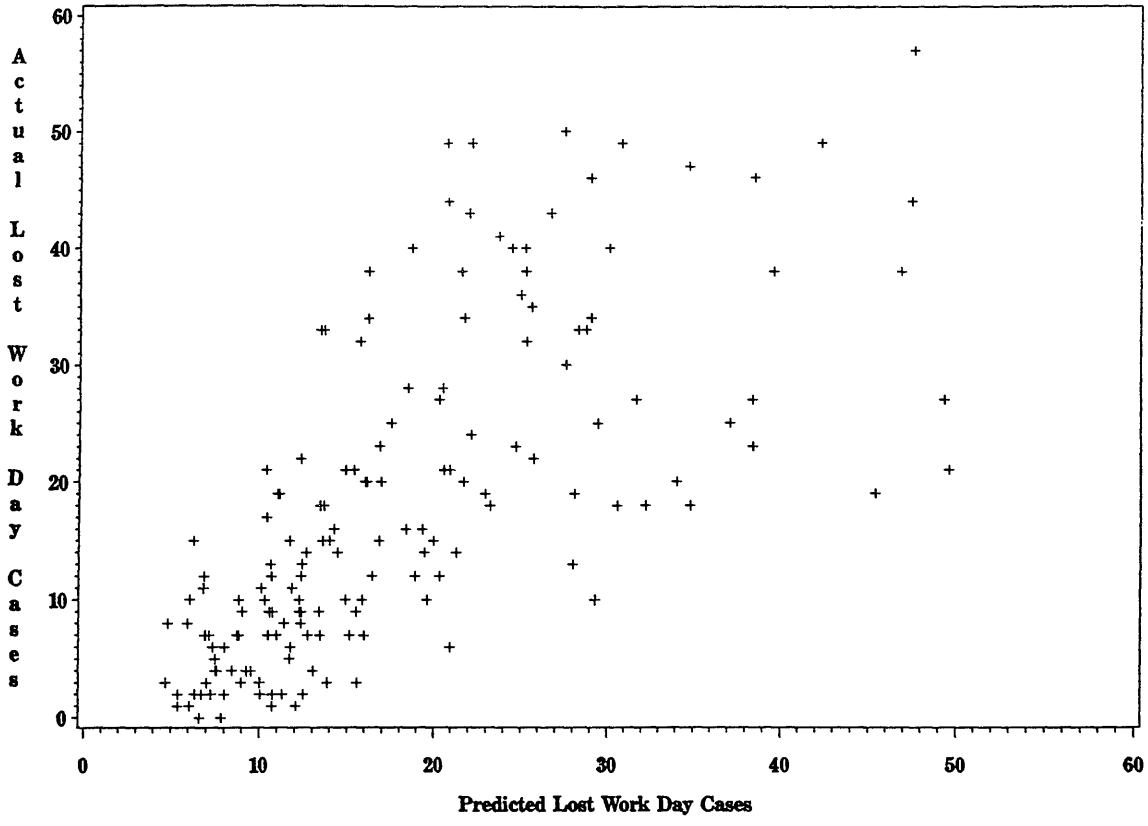
Mean	Standard Deviation	Variable	Coefficient	Std. Error	t-statistic & Significance
		Intercept	-.5027	.4077	-1.233
4.397	9.560	1988 LWD Cases X Small	.0333	.0047	7.022**
7.978	14.485	1988 LWD Cases X Medium	.0289	.0033	8.680**
16.609	45.170	1988 LWD Cases X Large	.0075	.0017	4.477**
.674	.470	Multiple Plants	-.1234	.1043	-1.183
10.611	12.850	Percent Tenure <1 Year	.0004	.0060	.074
.620	.487	Union	-.0181	.1145	-.158
.462	.500	Self-Insured	.2649	.0992	2.669**
10.384	2.785	Wage	-.0336	.0254	-1.321
.049	.216	SIC 25	-.4273	.2433	-1.756ρ
.168	.375	SIC 30	-.2857	.1695	-1.685ρ
.245	.431	SIC 34	-.2474	.1693	-1.461
.125	.332	SIC 35	-.2083	.2074	-1.004
.163	.370	SIC 37	-.3967	.1773	-2.238*
.130	.338	SIC 80	-.3538	.1966	-1.800ρ
5.904	.938	Employment (ln)	.5986	.0804	7.444**
		Alpha	.2778	.0428	6.486**

N = 184
Mean log-likelihood = -3.736
r (Y, $\hat{Y}$) = .837

ρ = significantly different from zero at 90% confidence level
* = significantly different from zero at 95% confidence level
** = significantly different from zero at 99% confidence level

Figure 4.2 presents the scatterplot for actual and predicted values of lost workday cases for those establishments with less than 60 lost workday cases in 1989. Again, the figure shows that the model is very successful in predicting the number of lost workday cases for establishments in the SET sample, based on last year's performance and a set of establishment characteristics.

Figure 4.2 Actual and Predicted Lost Work Day Cases
Values Less Than or Equal to 60



Total Lost Workdays

Table 4.3 reports the same negative binomial regression results for total lost workdays in 1989, as reported by the establishments in the SET sample survey. The results are again quite similar, although the relationships are generally not as striking as for lost workday cases. Presumably this represents the relatively random contribution of disability duration, since total lost workdays are simply the product of the number of lost workday cases and the number of lost workdays per case.[43] Self-insured employers and larger employers have significantly more lost workdays. There is no significant impact demonstrated from multiplant firm status, wage level, or union representation. The percent low tenure workers is positively associated with lost workdays, but only at an 85 percent statistical confidence level. The industry control variables show no particularly significant pattern of coefficients.

[43]In the SET project, it was discovered that disability duration could not be predicted with the variables available for analysis. See Hunt, Habeck, VanTol and Scully (1993) for details.

Table 4.3 Negative Binomial Regression of Lost Work Days
SET Survey Sample

Mean	Standard Deviation	Variable	Coefficient	Std. Error	t-statistic & Significance
		Intercept	2.7572	.6848	4.026
82.336	192.32	1988 Lost Work Days X Small	.0015	.0004	3.440**
169.04	300.44	1988 Lost Work Days X Medium	.0015	.0003	4.664**
352.11	865.49	1988 Lost Work Days X Large	.0005	.0002	3.167**
.678	.469	Multiple Plants	.0339	.1521	.223
10.560	12.867	Percent Tenure <1 Year	.0078	.0053	1.479
.623	.486	Union	.0297	.1892	.157
.464	.500	Self-Insured	.6145	.1564	3.928**
10.384	2.793	Wage	-.0112	.0350	-.319
.049	.217	SIC 25	.3180	.4139	.768
.169	.376	SIC 30	-.0726	.2668	-.272
.240	.429	SIC 34	.0221	.2402	.092
.126	.332	SIC 35	.0957	.2969	.322
.164	.371	SIC 37	.0852	.2730	.312
.131	.338	SIC 80	-.3392	.2874	-1.180
5.910	.937	Employment (ln)	.4189	.1330	3.150**
		Alpha	.6840	.0574	11.912**

n = 183
Mean log-likelihood = -7.030
$r(Y, \hat{Y}) = .773$

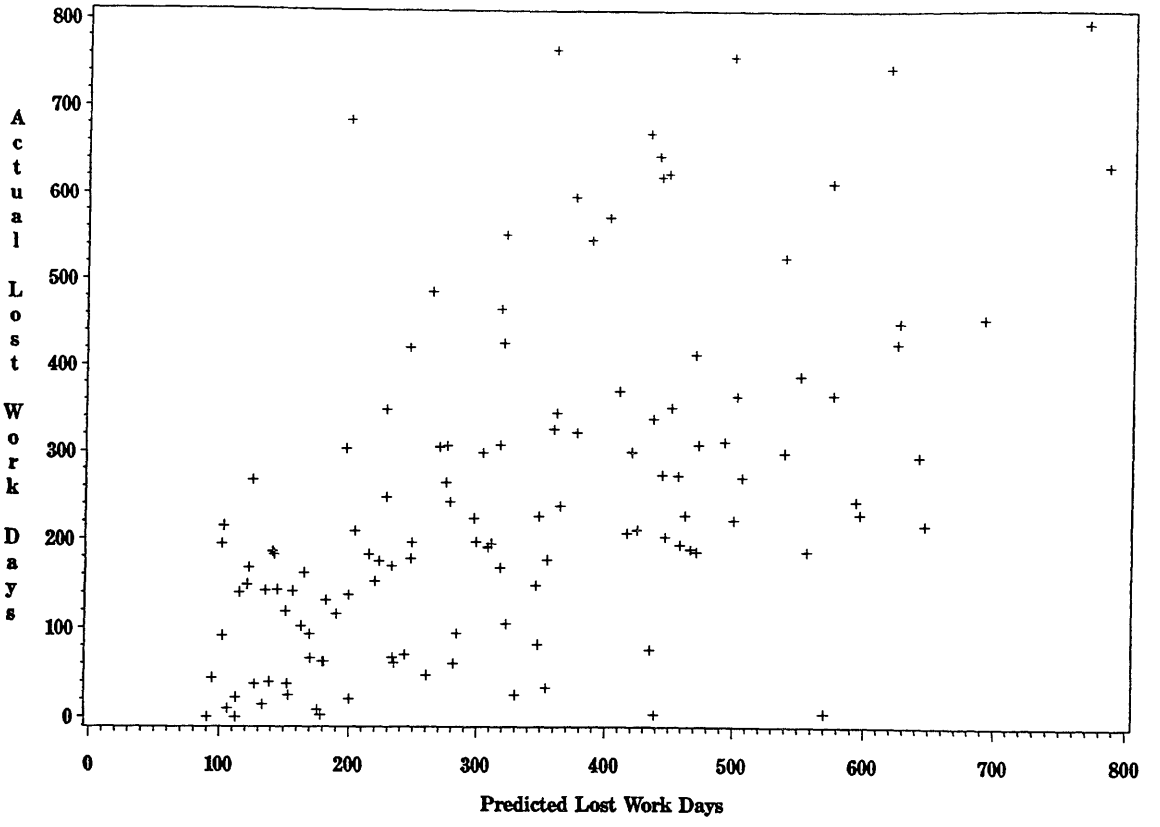
ρ = significantly different from zero at 90% confidence level
* = significantly different from zero at 95% confidence level
** = significantly different from zero at 99% confidence level

The lagged dependent variable (1988 lost workdays) interacted with the size classes shows highly significant and positive association with lost workdays across the sample. These coefficients appear smaller in an absolute sense than the earlier ones, but this largely results from the scaling of lost workdays relative to lost workday cases; since lost workdays has a sample mean of 20 times that of lost workday cases.

For small and medium size firms, table 4.3 indicates that one additional lost workday in 1988 is associated with .15 percent more recordables in 1989. For large firms, an additional lost workday in 1988 is associated with .05 percent increase in lost workdays in 1989. When evaluated at the mean of the lagged dependent variable (total lost workdays in 1988), these coefficients imply elasticities of .124 for small establishments, .258 for medium sized establishments, and .168 for large establishments. Thus a 10 percent higher level of lost workdays in 1988 is associated with 1.7 percent greater lost workdays (.10 x .168) in 1989 for large employers, about 2.6 percent greater for medium sized (.10 x .258) and just over 1 percent greater (.10 x .124) for small employers, respectively.

The correlation coefficient between actual and predicted lost workdays for the sample establishments is .773, which means that about 60 percent of the variance in 1989 lost workdays is explained by the model. Figure 4.3 presents the scatterplot for the actual and predicted lost workdays for those establishments with less than 800 lost workdays in 1989. The figure indicates that knowing the level of lost workdays for last year, together with the other establishment variables, is of very significant value in predicting the level of lost workdays for this year.

Figure 4.3 Actual and Predicted Lost Work Days
Values Less Than or Equal to 800



Marginal Impact of Lagged Performance Variables

One last way to assess the value of knowing last year's level of performance in predicting this year's injury and disability performance is to estimate the same model without the lagged dependent variable terms. Then we can compare the results from the model including the lagged dependent variables with the model that does not include these variables. This can be pushed one step farther by also dropping the covariates available in the SET database. Thus, the model can be estimated using only industry and employment level, essentially the variables that OSHA currently utilizes in selecting establishments for safety and health inspections.[44] We have performed these exercises and found that the models with the lagged level of performance variables perform remarkably better.

Table 4.4 demonstrates the marginal impact of the lagged performance variables in the negative binomial regression models. The first column in the table reports the correlation coefficient and the squared correlation coefficient (percentage of variance explained) between the actual and predicted performance values for the establishments in the SET sample when only industry and employment size are used as arguments. The models are able to predict only from 17 to 24 percent of the variance with this specification.

The second column in the table reports the same statistics (r and r^2) when the lagged performance variable is added to the model with industry and employment size. For recordables, the proportion of variance explained more than doubles. For lost workday cases and total lost workdays, the proportion of variance explained more than triples. In each case, more than half the variance in performance is explained with such a simple model, and in the case of lost workday cases, it approaches two-thirds.

Last, the covariates from the SET survey are added, to arrive at the full models reported earlier in tables 4.1 through 4.3. The addition of the covariates contributes another 6 to 16 percent to the variance explained by the simpler models. These results demonstrate that knowledge of past performance is extremely valuable information for predicting the number of recordables, lost workday cases, and total lost workdays at the establishment level. In fact, it is more useful than knowing the industry and employment level, at least as those are measured here.

[44]Although OSHA uses more detailed industry data than represented here which should improve predictive power.

Table 4.4 Marginal Impact of Lagged Performance Variables*

	SIC & Size Only	Plus Lagged Performance	Plus SET Covariates
Recordables	r .488	.746	.782
	r^2 .238	.556	.612
LWD Cases	r .445	.812	.836
	r^2 .198	.659	.699
Lost Workdays	r .410	.717	.773
	r^2 .168	.514	.597

*As estimated in negative binomial regression models. The table reports the correlation between actual and predicted performance levels for establishments in the SET sample.

These analyses provide powerful confirmation to the conclusions of the last section, namely that there is a substantial degree of consistency of establishment performance across time, and it would be extremely valuable to OSHA to know previous performance results when developing enforcement interventions.

V. Conclusions and Research Agenda

Summary and Conclusions

All the research results discussed in this report, both original and secondary, are consistent with the conclusion that injury and disability incidence levels are relatively predictable at the establishment level, at least for establishments with over 100 employees. This is true even though these rates are characterized by great variability overall. The literature review indicated previous studies had demonstrated that injury and disability performance at the establishment level was relatively stable across periods of up to six years. Gray and Scholz (1991) reported a correlation coefficient of .68 between the number of injuries in 1979 and the number of injuries in 1985. Ruser (1991) found correlations of lost workday case rates ranging from .70 to .93 across the same six year period, depending on the specific model estimated.

The simple correlation analysis presented here on the two Michigan databases (section III) shows that there is very considerable persistence in establishment level performance over 2- to 3-year periods on a number of different measures of injury and disability performance, including MIOSHA recordables, MIOSHA lost workday cases, total MIOSHA lost workdays, and workers' compensation wage-loss claims incidence. Analysis of the population of some 25,000 Michigan firms with workers' compensation wage-loss claims in the period 1986 through 1988, demonstrated that among firms with more than 100 employees, average correlation coefficients of about .70 between observations two years apart were typical. This means that about one-half of the variation in workers' compensation wage-loss claim rates is predictable on the basis of previous rates.

Additional correlation analyses of a special random sample of 220 Michigan establishments prepared for the Safety, Education and Training (SET) Division of the Michigan Department of Labor were also presented. These data included MIOSHA log statistics that were reported by the respondents in a voluntary mail survey conducted in the first half of 1991. This sample, generally representative of Michigan manufacturers with over 100 employees, also showed very substantial persistence of injury and disability rates from year to year. For example, adjacent year correlation of MIOSHA recordables and lost workday case rates were generally around .80, meaning that nearly 65 percent of the variation is predictable based on previous performance. MIOSHA lost workday rates showed adjacent year correlations of about .60, reflecting greater variation in the duration of injuries. A correlation analysis of the SET sample data also revealed substantial consistency across these different measures of injury and disability incidence. The implication from these analyses is that it would be very useful to know the past performance of an establishment in trying to predict its future performance. However, since no other factors were controlled in these simple correlation analyses, there is still room for doubt about the possible contribution of other variables that might be significant in predicting injury and disability performance.

In a more sophisticated test, using multivariate analysis techniques to include other potentially important variables, it was demonstrated that knowing the past performance of establishments on specific injury and disability measures can double or triple the predictability of performance, as measured by the proportion of explained variance, over knowing only the industry and employment level. Negative binomial regression models were estimated that contained only the industry and employment level of establishments. These models explained from 17 to 24 percent of the variation in performance levels across establishments, depending on the specific measure. Adding last year's actual level of performance on lost workday cases to the model more than triples the proportion of explained variance, from 20 percent to 66 percent. The same impact is seen for total lost workdays, where the proportion of explained variance increases from 17 percent to 51 percent. This is a direct and powerful demonstration of the value of information about past performance on injury and disability measures. Thus, the conclusion is that it would be extremely valuable to OSHA to know past performance in determining appropriate interventions at the establishment level.

These results, when combined with the earlier literature reviewed, make it clear that OSHA could benefit from using an establishment level database to help guide the deployment of inspection, or other intervention, resources. Without such data, it seems difficult or impossible to develop a systematic plan to attack the country's occupational safety and health problems. In addition, our general understanding of the determinants of good safety and health performance could be enormously enhanced by such data. The possibility of using such micro data to guide improvements in current OSHA practice and procedure seems very great. But the likelihood that the availability of micro-level data would spur additional research, and demonstration projects, on the efficacy of various OSHA interventions promises even greater potential performance.

Future Research Needs

All empirical work with random sample data is subject to sampling variability, so replication or other verification of the results presented here is important to ensure that there is not something specifically wrong with this random sample that biased our findings. This confirmation is particularly important when the sample is relatively small and has been subjected to intensive analysis, as in this case.

In addition, it would be reassuring to reproduce these results for another state (or states) to guarantee that the outcome is not specific to the Michigan environment, or to the seven industries selected for analysis in the SET project. While these industries represent the most hazardous parts of Michigan manufacturing relatively well, this may not be true for the nation as a whole, and it almost certainly is not true for other individual states with a different economic base.

Neither of the original databases analyzed here included establishments with less than 100 employees. The applicability of these findings to smaller firms, inherently less

predictable in their performance, should also be determined. The threshold of 100 employees was not developed as a research finding, but emanated from earlier sample feasibility decisions that were made in another project. It may be that more careful examination of this issue would lead to selection of a threshold other than 100 employees. OSHA should pursue this issue to determine the practical limits of predictability.

The cross-sectional regression analysis in this report relied upon retrospective reporting of data, which was collected from sampled establishments by means of a mailed questionnaire. On the one hand, this direct access to employers provided the opportunity to gather more detail than would be possible from any known administrative data source. But, this is also a disadvantage of these data in that they are self-reported and not subject to easy verification. In addition, the retrospective data collection design limited the number of years of data that could be collected. The willingness and ability of firms to respond falls off rapidly as the number of years requested rises.

It would be extremely productive to analyze multiple year data on a sample of firms, generally called "panel data," to gain a longer time perspective than can be provided with the cross-sectional data used here. The SET database gathered information about performance measures over a four year period, but establishment characteristics were gathered for one year only. Neither of the databases used here provided an opportunity for panel analysis, like that performed by Ruser and Gray and Scholz. But such data should be routinely available for interested researchers and policymakers. They could be made available in a public-use format, as is done in other instances, to maximize the analytical use that is made of such data.

It also would be interesting to replicate the results achieved by Gray and Scholz and by Ruser on their panel data, especially since they used such different methods of analysis. But it would be even more valuable to analyze a longer-term panel data set that would encompass the entire period from 1978 to 1991 and as many states as possible, hopefully including some state-plan venues. This would provide the opportunity to test for systematic differences across states in performance outcomes. If these differences can be linked to specific policy initiatives or structural characteristics, such analytical results might then lead to new policy initiatives informed by state-specific experience. Development of a new matched BLS-OSHA database should be a high priority.

There is also a general need for more empirical research on safety and disability performance. There has as yet been no satisfactory explanation for the lack of aggregate improvement in lost workday rates over the last 30 years, despite the passage and implementation of OSHA. There also has been very little investigation of interstate differences in performance on injury and disability measures. Such studies might provide very provocative policy hypotheses about the reasons for such differences, which then could be investigated for additional policy interest.

It is critically important to develop such "macro," or aggregate level, understanding of our safety and health performance as a complement to the "micro," or establishment level focus of this report and the other research reviewed here. As indicated earlier, the development of more adequate data for analysis, and making them more readily available will also have significant policy development implications that cannot be foreseen at this time. As OSHA moves forward into the 21st century, it is critically important that our occupational safety and health efforts be as effective as they can possibly be. Rigorous, policy-oriented, empirical research can be a strategically important element in helping to focus those efforts more productively.

References

Barkume, Anthony, and Ruser, John. (1992). "Do High Workplace Injuries Reveal Violations of Safety Standards? Some Evidence on the Persistence of High Injury Rates in the Workplace." U.S. Department of Labor, Bureau of Labor Statistics.

Gray, Wayne B., and Scholz, John T. (1991). "Do OSHA Inspections Reduce Injuries? A. Panel Analysis." NBER Working Paper Series, no. 3774. Cambridge, MA: National Bureau of Economic Research.

Habeck, Rochelle V., Leahy, Michael J., Hunt, H. Allan, Chan, Fong, and Welch, Edward M. (1991). "Employer Factors Related to Workers' Compensation Claims and Disability Management," Rehabilitation Counseling Bulletin, 34(3).

Hunt, H.A. (1988). The Incidence of Workers' Compensation Claims in Michigan. Report submitted to Bureau of Workers' Disability Compensation, Michigan Department of Labor. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

Hunt, H.A., Habeck, R.V., VanTol, B., and Scully, S.M. Disability Prevention Among Michigan Employers. Report submitted to Safety Education and Training Division, Michigan Department of Labor. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research, (forthcoming).

Moore, Michael J., and Viscusi, W. Kip. (1990). Compensating Mechanisms for Job Risk. Princeton: Princeton University Press.

Occupational Injuries and Illnesses Survey, Michigan 1988. (1990). MIOSHA Information Division, Michigan Department of Labor in cooperation with U.S. Department of Labor, Bureau of Labor Statistics.

Pollack, Earl S., and Keimig, Deborah Gellerman, eds. (1987). Counting Injuries and Illnesses in the Workplace: Proposals for a Better System. Washington, D.C.: National Academy Press.

Ruser, John W. (1992). "Self-Correction Versus Persistence of Establishment Injury Rate Shocks: Evidence and Implications for the Use of Microdata for Inspection Targeting." Unpublished paper. U.S. Department of Labor, Bureau of Labor Statistics.

Ruser, John W. (1991). "Workers' Compensation and Occupational Injuries and Illnesses." Journal of Labor Economics, 9(4), pp. 325-350.

Scholz, John T. and Gray, Wayne B. (1990). "OSHA Enforcement and Workplace Injuries: A Behavioral Approach to Risk Assessment." Journal of Risk and Uncertainty, 3, pp. 283-305.

Smith, Robert S. (1992). "Have OSHA and Workers' Compensation Made the Workplace Safer?" in Research Frontiers in Industrial Relations and Human Resources, David Lewin, Olivia S. Mitchell, Peter D. Sherer, eds. Madison, WI: Industrial Relations Research Association, pp. 557-586.

Smith, Robert S. (1979). "The Impact of OSHA Inspections on Manufacturing Injury Rates." Journal of Human Resources, 14, pp. 145-170.

Wokutch, Richard E. (1990). Cooperation and Conflict in Occupational Safety and Health: A Multination Study of the Automotive Industry. New York, NY: Praeger.

Technical Appendix: Data Sources[45]

This Technical Appendix describes the two Michigan data sources used in this report. It is designed to provide sufficient detail about the methods used to develop these data that the reader can judge the credibility of the data for the purposes of the report.

MESC Data Summary

The Michigan Employment Security Commission (MESC) administers the unemployment insurance system in the state of Michigan. Since coverage for this system is nearly universal (employers with one employee for 20 weeks, or $1,000 in wage payments), it is a good source of information on the population of firms or establishments. The MESC ES-202 quarterly data provide administrative support for the unemployment insurance system, and contain information about the employment level and wage payments for the reporting unit, usually an establishment, but sometimes a firm with multiple establishments.[46]

MESC provided establishment level data from the ES-202 reports for the second quarter of 1986, 1987 and 1988 including employment (head count), aggregate wage payments, industry code (SIC), location by city, county and labor market area, firm ownership (public or private), and discrete employment size categories (0-3, 4-9, 10-19, 20-49, 50-99, 100-249, 250-499, 400-999, 1000 and over. The tape also included the Federal Employer Identification Number (FEIN), which enables a match to the data from the Michigan Bureau of Workers' Disability Compensation on workers' compensation claims.[47]

The MESC data contains establishment level detail, which can be aggregated to firm level by means of a multiple establishment firm code. In the second quarter of 1988, for example, there were 158,507 establishments who reported wages to MESC; aggregating to the firm level, there were 153,211 firms reporting wages. These data provide a good foundation for analysis, but they must be matched to another administrative database with information on workers' compensation claims in order to address the relevant policy issues of this report.[48]

---

[45]Ken Kline and Leslie Lance assisted in developing this description of the database used in this report.

[46]Thanks to the Commission and staff member Abel Feinstein for facilitating access to these confidential data. They are used here with the permission of the Commission, but of course they have no responsibility for the analysis we have performed.

[47]Unfortunately, this FEIN identifier is not an edited field for MESC. It will be shown later that this leads to problems in trying to match these administrative data with those from the workers' compensation system.

[48]These data were initially requested from MESC to support the development of a sampling frame for the SET survey described below.

BWDC Data Summary

The source for the workers' compensation component of the database is the Bureau of Workers' Disability Compensation (BWDC) administrative database, called COMPMAST. This is an enormous database (over 2.7 million records) that contains all workers' compensation claims for approximately a 10-year period, with each administrative form treated as a unique record. The records aggregate into claims through the identifying claimant social security number and date of injury. At our request and direction, BWDC extracted from COMPMAST a database for analysis that contains all claims with injury dates in 1986, 1987 and 1988.

As shown in table A.1, there were approximately 70,000 workers' compensation wage-loss claims identified for each year. All these claims involved some wage-loss, or at least allegations of wage-loss, meaning that a minimum of seven (7) days were lost from work. This database does not include any information about payments to claims involving medical costs only. The number of claims declines from 1986 to 1988 because the latent, litigated claims have not all been filed as of December 1989, when these data were extracted. This can be seen in the proportion litigated statistic, which declines from 17 percent in 1986 to 6 percent in 1988. When the cohort is complete, the 1988 figure will be closer to 17 percent than 6 percent.

Table A.1 BWDC Incurred Claims - Claim Means

	1986	1987	1988
Number of claims	72,534	70,930	69,994
Compensation/claim ($)	5790.02	4758.99	3372.76
Duration (days)	129	116	92
Proportion litigated	0.174	0.135	0.063
Proportion redeemed	0.130	0.094	0.038
Weekly earnings ($)	401.46	405.25	416.62
Compensation rate ($)	229.28	233.62	237.86
Proportion male	0.702	0.695	0.697
Claimant age	36.7	36.6	36.5

Table A.1 shows the average compensation per claim for each year. The duration of claims is calculated from the beginning of compensation to the end of compensation, excluding periods that were not compensated for some reason. The database also contains the weekly earnings of the claimant, as reported by the employer, and the weekly compensation rate that was actually paid while the claimant was on workers' compensation (final rate paid if there was a change).

Table A.2 accumulates these same data on an employer unit basis and reports means for the employer database. Note that the means change somewhat because of the implicit equal weighting for each employer. This results in lower means for earnings and compensation rates as the large, high-wage employer influence is diluted by the enormous number of smaller, low-wage firms. As will be discussed later, these employer units were generally firm level rather than establishment level, although there were exceptions. BWDC does not express any preference as to how multi-plant employers report their claims, and it turns out that most such employers report from one centralized office.

Table A.2 BWDC Incurred Claims - Employer Means

	1986	1987	1988
Number of firms	18,911	18,600	18,477
Number of claims	3.85	3.81	3.78
Compensation/employer ($)	21926.34	17883.20	12629.05
Duration (days)	129	112	97
Weekly earnings per claim ($)	331.17	342.39	353.79
Comp rate per claim ($)	196.80	203.37	208.22

The process which we have discussed thus far has identified workers' compensation claims from an administrative data set, the wage-loss compensation that employers have paid on those claims as of December 31, 1989, the characteristics of the claimants involved, and the FEINs of those employers against whom the claims were filed. However, at this point, other than the FEIN, the database contains no firm specific information. Industry classification, firm employment and wage levels (other than for injured workers) are not available in the BWDC data. To derive firm specific information, it was necessary to match the firm level BWDC data to the Michigan Employment Security Commission (MESC) ES-202 data.[49] The following section discusses the match process and its outcome.

Success of the Match

The major difficulty in matching these two administrative databases was in conclusively identifying the same reporting unit. There is no necessary uniformity in the way that employers report data for unemployment insurance and workers' compensation purposes. In particular, firms are not consistent in reporting by establishment, or at an aggregate firm level across these two programs. Because BWDC data were more likely to be reported at the firm level, the matched database generally adopted this strategy as well.

[49]It is worth noting here that the BWDC data do contain some establishment level detail by means of a 3-digit identifier attached to the FEIN. MESC also has a 3-digit identifier for multi-plant firms but the MESC/BWDC establishment identifiers are not the same. Only the 9-digit FEIN (firm level) is the same across the two data sets. Therefore, a match at the establishment level is not possible.

However, there were some exceptions to this rule. In addition, MESC does not edit the FEIN employer ID field in their data entry process. So, an unknown proportion of firms could not be matched simply because the FEIN was entered incorrectly.

In 1986, there were 143,351 firms in the MESC database and 15,106 of these firms (10.5 percent) incurred at least one workers' compensation claim, according to the successfully matched BWDC data. The breakdown of match rates by firm size category shows the predictable result that match rate increases with firm size. Due to the infrequent nature of injuries involving seven or more lost workdays, especially in some white collar industries, smaller firms that did not match probably did not incur a workers' compensation claim in that year and therefore, legitimately have a zero claim rate for that year. However, with the disparity between the BWDC and MESC data sets, it would not be a safe assumption to assign a zero claim rate to all MESC firms that did not match to the BWDC data. Therefore, we have imputed a zero claim rate to firms in the MESC data set that did not match in a particular year only if the firm successfully matched in at least one of the other years. Since these firms had at least one match in the 1986-88 period, we can be more confident that the failure to match in other years is due to the firm incurring no claims, and not due to discrepancies between the MESC and BWDC data sets. Table A.3 shows the frequency distribution of firms in the 1986 data set after adding firms with zero imputed claims according to the rule described above. The match rate was 80 percent for firms of more than 49 employees and 87 percent for firms of 100 employees or more.

Table A.3 Match Rates, Incurred 1986 Data, Adding Zero Claim Firms

Number of Employees	Total Firms	Matches to BWDC	Percent of Total Firms
NA	42		37.1
0 - 3	75,601	3,342	4.4
4 - 9	28,558	3,801	13.3
10 - 19	18,377	5,226	28.4
20 - 49	12,155	6,238	51.3
n < 50	134,691	18,607	13.8
50 - 99	4,492	3,277	72.9
100 - 249	2,619	2,219	84.7
250 - 499	879	799	90.9
500 - 999	379	352	92.9
1,000 +	249	238	95.6
n > 49	8,618	6,885	79.9
n > 99	4,126	3,608	87.4
Total	143,351	25,495	17.8

Table A.4 summarizes the total number of firms for each year. It shows that a total of 28,462 firms incurred at least one workers' compensation wage-loss claim in the three years, 1986 through 1988. The data show that in any given year, about 10 percent of MESC population firms incurred at least one workers' compensation claim, and another 7 percent could be assumed to have no claims, meaning the firm matched in at least one of the other years and was present in the MESC database. Conversely, over 80 percent of firms in the MESC database did not incur a compensation claim in the 1986 through 1988 period or didn't match due to discrepancies between the MESC and BWDC data.

Table A.4 Summary of MESC/BWDC Matched Sample Size

	1986	1987	1988
Total MESC firms	143,351	149,496	153,211
Firms with:			
One or more claims	15,106	14,982	15,135
Zero claims	10,389	11,458	11,076
Missing	2,967	2,022	2,251
Total incurred sample	28,462	28,462	28,462
Total analysis sample	25,495	26,440	26,211
Percent of MESC firms	17.8	17.7	17.1

The Matched MESC/BWDC 1986-88 Database

The preceding process yields a database containing the workers' compensation claims experience of the vast majority of firms that had any workers' compensation activity for 1986, 1987 and 1988. Table A.5 summarizes this matched database. The addition of the employment data provides a critically important variable, the workers' compensation claim rate per 100 employees. We will use this variable to examine the persistence of firms' disability performance. The average claim rate for firms experiencing at least one claim in the years 1986 through 1988 was 2.62, 2.50 and 2.46 per 100 employees respectively.[50] The data set also contains firm level information on industry, employment size (as shown previously), location by county, labor market area, SMSA, and type of ownership.

Note that the number of observations depends upon the variable, reflecting whether it is defined over the 3-year period, or for each year, and missing data. Thus there are a total

[50]Given that the large number of firms that are not represented here probably did not have any claims in three years, these figures severely overstate the true average.

of 28,674 firms for the calculation of average claimant age for 1986-88, but only 14,100 for average duration of disability claim in 1986.

Table A.5 Summary Statistics from MESC/BWDC Matched Database, 1986-88

Variable Description	Year	Number of Firms	Mean Value Per Firm
Total compensation ($)	- 1986	15,106	24,277
	- 1987	14,982	19,543
	- 1988	15,135	13,634
Duration (days)	- 1986	14,100	127
	- 1987	14,193	112
	- 1988	14,792	96
Number of claims	- 1986	25,495	2.512
	- 1987	26,440	2.325
	- 1988	26,211	2.332
Employment	- 1986:2	25,198	96
	- 1987:2	25,906	93
	- 1988:2	25,630	95
Avg weekly earnings ($)	- 1986	14,129	336.95
	- 1987	14,200	349.96
	- 1988	14,792	361.18
Avg weekly comp rate ($)	- 1986	14,116	199.78
	- 1987	14,193	207.16
	- 1988	14,785	211.60
Avg claimant age 1986-88		28,674	35

It is not possible to determine whether this database precisely represents the population of firms with at least one workers' compensation claim in the 3-year period. It seems likely that 80 to 90 percent of the BWDC population of firms is present in the database, but it is impossible to prove this.[51] However, the sources of error that are known among larger firms seem randomly distributed, like FEIN errors and differences in coding of multiple establishment enterprises. The fact that a minority of MESC firms are present in

[51]See Hunt (1988) for a fuller discussion of the frustrations of the matching process in a previous study.

this database is due to the fact that the majority of firms rarely experience a workers' compensation claim. It is also clear that nearly 90 percent of firms with 100 employees or more and 80 percent of those with 50 employees or more are present in the database. It seems adequately representative of those firms that have significant workers' compensation activity, with the adequacy probably increasing with the frequency of such activity.

SET Sample Database

The other database analyzed here was developed during a study of the impact of injury and disability prevention activities among a random sample of Michigan employers.[52] This database was created as part of a 3-year study supported by the Safety, Education and Training (SET) Division of the Michigan Bureau of Safety and Regulation. Under the Michigan state plan for Occupational Safety and Health, the SET Division is a complement to the MIOSHA division, which performs the compliance inspections. SET provides voluntary consultation and education services to employers and employee groups throughout Michigan. A network of 34 SET Consultants around the state are available to work with employers one-on-one in tackling their occupational safety and health problems without any recriminations from the enforcement side of the agency.

The SET study involved the W.E. Upjohn Institute for Employment Research and Michigan State University, College of Education in a collaborative research partnership with SET to intensively study the policy and practice dimensions of accident prevention, injury management, and workers' compensation claim prevention activities on the part of Michigan employers. The study was an extension of an earlier pilot study that found some very intriguing correlations between workers' compensation claims incidence and indicators of organizational climate and behavior.[53] The follow-on study sought to quantify the payoff to various employer practices using cross-sectional comparisons. While drawn from a relatively small random sample of 220 establishments, these data are ideally suited to probing the specific employer characteristics associated with different levels of performance on injury and disability measures.

The industries selected for study included: SIC 20, Food Production; SIC 25, Furniture and Fixtures; SIC 30, Rubber and Miscellaneous Plastics; SIC 34, Fabricated Metals; SIC 35, Machinery, except Electrical; SIC 37, Transportation Equipment; and SIC 80, Health Services. The goal was to select industries with high exposure to injuries, substantial numbers of establishments, and significant employment levels in Michigan to maximize the representability of the sample and to increase the credibility of the study. In addition, the need for diversity among industries to maximize the generalizability of the findings was a concern. Thus SIC 20, Food Production, and SIC 80, Health Services,

[52]Hunt, Habeck, VanTol, and Scully (1993).

[53]Habeck, et. al. (1991).

were important to give the sample more diversity, even though they imposed some special sampling problems. The final selections give the study coverage of six of the eight most hazardous industries according to MIOSHA, plus the most hazardous of the service industries (SIC 80, which ranks 21st overall). In total, these seven industries contained 23,156 establishments employing 955,400 people in Michigan as of 1988.[54]

Study of administrative data revealed that injury and workers' compensation claims incidence rates for small firms were extremely variable from year to year. This reflects the difficulty of observing an infrequent event without sufficient exposure, and creates a problem in distinguishing between levels of performance on a variable that may be dominated by stochastic considerations. Thus, establishments with less than 100 employees were omitted from the sampling frame on the grounds that it would not be possible to effectively distinguish between good and poor company performance in a 3-year period, given the inherent variability in these data. In insurance parlance, these smaller establishments were deemed to have low credibility in a study that took the establishment as the unit of observation.

The MESC population of establishments from the second quarter of 1988 was stratified by SIC code (2-digit level) and employment size (from 100 to 249 employees, from 250 to 499 employees, or over 500 employees). It was determined that the most efficient sampling design would provide for sampling from each industry stratum proportional to the expected injury rate.[55] This will provide the greatest concentration of observations in those strata with the greatest expected variance, assuming the variance is roughly proportional to the mean. Within an industry, the sample was spread equally among establishment size classes, subject to the maximum number of establishments available. Table A.6 reports the size of the MESC universe and the target sample for each of the 21 sampling strata.

---

[54]MIOSHA (1990), table 4.

[55]We are indebted to Dr. Stephen Raudenbush, Michigan State University, College of Education for this insightful addition to the study design.

Table A.6 SET Sampling Strategy

MESC Universe from MESC Establishment List

		SIC 20 Food Prod.		SIC 25 Furniture		SIC 30 Rubber & Plastics		SIC 34 Fab. Metals		SIC 35 Machinery		SIC 37 Trans. Equip.		SIC 80 Health Servs.		Totals	
Firm Size	Class	# Firms	Total Emplmt	# Firms	Total Emplmt	# Firms	Total Emplmt	# Firms	Total Emplmt	# Firms	Total Emplmt	# Firms	Total Emplmt	# Firms	Total Emplmt	# Firms	Total Emplmt
(6)	100-249	47	6,103	29	4,458	84	11,715	158	21,944	137	17,174	88	9,598	265	33,154	808	104,146
(7)	250-499	29	8,322	16	4,709	37	11,561	42	13,545	35	9,068	58	13,996	46	16,530	263	77,731
(8)	500+	15	23,043	7	19,052	13	11,241	22	15,676	25	31,082	92	345,767	60	106,910	234	552,771
	Totals	91	37,468	52	28,219	134	34,517	222	51,165	197	57,324	238	369,361	371	156,594	1,305	734,648

Industry Risk Rates - MIOSHA

	SIC 20 Food Prod.	SIC 25 Furniture	SIC 30 Rubber & Plastics	SIC 34 Fab. Metals	SIC 35 Machinery	SIC 37 Trans. Equip.	SIC 80 Health Servs.	Average
Industry Risk Rate*	9.28	8.82	10.9	10.21	5.66	6.15	2.55	7.65 Unweighted

*Lost workday claims per 100 employees from MIOSHA data for 1988

Sampling Proportional to Risk - Industry Specific Samples**

		SIC 20 Food Prod.		SIC 25 Furniture		SIC 30 Rubber & Plastics		SIC 34 Fab. Metals		SIC 35 Machinery		SIC 37 Trans. Equip.		SIC 80 Health Servs.		
Firm Size	Class	Ratio	Sample	Ratio	Sample	Ratio	Sample	Ratio	Sample	Ratio	Sample	Ratio	Sample	Ratio	Sample	Total
(6)	100-249	0.404	19	0.379	11	0.464	39	0.380	60	0.168	23	0.352	31	0.075	20	203
(7)	250-499	0.655	19	0.688	11	1.000	37	1.000	42	0.657	23	0.534	31	0.435	20	183
(8)	500+	1.000	15	1.000	7	1.000	13	1.000	22	0.920	23	0.337	31	0.333	20	131
	Totals	0.582	53	0.552	29	0.683	89	0.640	124	0.355	69	0.385	93	0.160	60	517

**This method uses the proportional risk factor to allocate sample points among industries and then allocates equally across size classes, subject to universe size, within an industry.

As shown in the table, sampling proportions ranged from .075 (1 in 13 operationally) to 1.00 (i.e. every establishment in those strata where the entire population did not fill the quota). The largest samples were selected in Fabricated Metals (124 establishments), Transportation Equipment (93 establishments), and Rubber and Plastics (89 establishments). The smallest sample was that for Furniture Manufacturing (29 establishments), followed by Food Production (53 establishments) and Health Services (60 establishments). The theoretical sample, based on the exact calculated sampling proportions, was arbitrarily set at 500 establishments. The actual sample drawn, after sampling proportions were integerized for operational purposes, was 517 establishments. As shown in the table, 203 of these establishments had from 100 to 249 employees, while 183 had 250 to 499 employees, and 131 had more than 500 employees.

Survey Administrative Procedures

When determining the structural layout of the mail survey and the survey administration procedures, specific principles from Dillman (1978) Mail and Telephone Surveys: The Total Design Method, were adopted to the extent feasible. The mail survey was pilot tested with nine companies, not included in the survey sample, early in 1991. Individuals completing the survey were asked to comment on the ease of completion, the appearance of the instrument, and the time involved in completing the survey. Nine content experts were also asked to review the survey instrument for content and clarity of survey items. Feedback received from the piloting and the review process was used to make a final edit of the instrument.

In most claims, the mail survey was addressed to the CEO, but in larger establishments to the Director of Human Resources, by specific name. Unfortunately, the sampling frame was nearly three years old by the time of the survey mailing, and substantial verification research was required to get the appropriate name for each establishment. The result was that a total of 58 of these establishments were deleted before the mailing of the survey, because they had gone out of business, were incorrectly identified, or were otherwise unreachable. These "prior deletions" were replaced, where possible, with an alternate establishment that was drawn by the same random process (these are termed "prior additions"). This process resulted in a net mailable sample of 507 establishments.

The survey was mailed on March 5, 1991 to the random sample of 507 establishments in Michigan. Even after the careful screening process, a total of 30 of these establishments were still not reachable or not appropriate for the survey.[56] This means that there were 477 total potential respondents to the mail survey. A return rate of approximately 16 percent was achieved in the first three weeks (81 respondents). A second full mailing occurred on March 26, 1991 to the 426 firms that had not yet responded. During the next six weeks 94

[56]Reflecting the dynamic population of firms, the age of the directories used for verification, and other largely random influences.

additional responses were received. In May, individual telephone follow-ups began to every firm that had not responded (332 firms).

The survey was completed on July 31, 1991. A total of 220 employers responded for an aggregate response rate of 46 percent. Table A.7 indicates the number of survey responses and response rates by strata. The major disappointments were the very small rate of response among establishments of 100-249 employees in the Health Services industry and among large establishments in Plastics and Rubber Manufacturing. All completed cell sizes are small in Furniture Manufacturing, but the population was also quite small in those strata.

Table A.7 SET Survey Response Rates

Industry

Size of Plant	20 Food	25 Furniture	30 Rubber	34 Metals	35 Machinery	37 Transp.	80 Health	Total
100-249								
Sample	17	9	33	58	22	29	20	188
Respondents	9	7	12	24	9	11	5	77
Response rate (%)	52.9	77.8	36.4	41.4	40.9	37.9	25.0	41.0
250-499								
Sample	15	11	35	39	21	27	19	167
Respondents	9	5	17	17	12	11	11	82
Response rate (%)	60.0	45.5	48.6	43.6	57.1	40.7	57.9	49.1
500+								
Sample	14	7	12	19	20	30	20	122
Respondents	7	3	4	13	7	14	13	61
Response rate (%)	50.0	42.9	33.3	68.4	35.0	46.7	65.0	50.0
Industry Totals								
Sample	46	27	80	116	63	86	59	477
Respondents	25	15	33	54	28	36	29	220
Response rate (%)	54.3	55.6	41.3	46.6	44.4	41.9	49.2	46.1

Table A.8 reports unweighted means from the self-reported SET survey database comparable to those of table A.5 for the MESC/BWDC matched administrative database. The concentration is on OSHA variables, but workers' compensation claims and employment figures are reported as well. It appears from these data that the lost workday case rate has been declining over the observation period, although there do not appear to be strong trends in either MIOSHA recordables or lost workdays per 100 workers. The number of missing values increases for each variable as one goes back in time, presumably reflecting the lack of easy access to records from earlier years.

The average employment level for establishments in the SET survey sample was around 600 in 1988, but this includes the influence of some very large plants of nearly 10,000 workers. This figure is comparable to the databases used by Ruser and Gray and Scholz. It is also noteworthy that the SET sample had about 3.6 workers' compensation claims per 100 workers in 1989, compared to 2.3 per 100 for the MESC/BWDC sample (1988 observation). This represents the influence of the sampling design, especially the hazardous industry focus of the SET sample, as it will be shown later that establishments with better than average performance were actually more likely to respond to the SET survey.

Table A.8 Summary Statistics from SET Survey Sample

Measure	Year	Number of Firms	Mean Value	Standard Deviation
MIOSHA Recordables per 100 Workers	1986	156	22.72	22.25
	1987	179	23.39	23.29
	1988	189	22.97	20.85
	1989	206	22.56	20.50
Lost Workday Cases per 100 Workers	1986	157	7.26	11.36
	1987	178	6.96	9.61
	1988	188	6.47	6.84
	1989	206	6.08	6.01
Lost Workdays per 100 Workers	1986	156	139.0	155.9
	1987	177	129.1	127.8
	1988	188	129.2	119.4
	1989	205	133.5	148.8
WC Claims per 100 Workers	1989	196	3.58	3.63
Employment	1986	183	591	988.8
	1987	193	592	967.3
	1988	201	601	949.4
	1989	216	635	965.4

Table A.9 presents some of the establishment characteristics from the SET survey sample. This type of detail, as reported by the employer-respondents, is what makes the SET database interesting for this analysis. While it is possible to construct average wages, and maybe even the percentage production workers, from administrative data, there is no way to gather data on the percentage of employees with tenure of less than one year, and other establishment specific items, without going out and gathering it. The percent of employees with tenure of less than one year should provide a good index of the number of employees who are exposed to an unusual likelihood of injury by virtue of their lack of familiarity with operating procedures and hazards of the job. It is expected to be positively related to the injury rate. The average wage will serve as a proxy for the general attractiveness of employment at the establishment.

The variable for union representation in the plant provides an opportunity to test for the effect of union activity on injuries and disabilities. The sign of such a term is uncertain. On the one hand, unions can be expected to bargain over safety and health issues, encourage joint safety committees, and engage in other supportive policy initiatives. On the other hand, it has been observed that the individual job security conveyed by union representation can encourage reporting of injuries, and thereby increase the apparent incidence of disability in the plant.

Table A.9 Establishment Characteristics from SET Survey Sample

Establishment Characteristic	Number of Firms	Mean Value	Standard Deviation
Average Wage ($/hour)	208	10.33	2.74
Percent Production Workers	217	70.0	17.4
Percent Tenure <1 Year	210	11.2	13.8
Union Representation (Proportion)	218	.615	.488
Self-Insured (Proportion)	218	.578	.495
Multiple Plants (Proportion)	218	.656	.476

The self-insured variable will differentiate between those employers who purchase their workers' compensation coverage from commercial insurance carriers and those who are self-insured. Some of the literature mentioned in section I of this paper focussed on indirect

tests of the effect of this variable, using employment size as a proxy for the degree of experience rating, with the assumption that large, self-insured employers are perfectly experience rated. We are able to test the effect of self-insured status directly. Last, the multiple plants variable will serve to differentiate between two plants of similar employment size which have different policy sophistication owing to a larger corporate support network behind one establishment. These characteristics (or covariates) are used in the multivariate analysis presented in section IV of this report. They enable some unique hypotheses testing on the SET sample survey database.

Response Bias Analysis

The highest response rate to the SET survey was for large establishments in the Fabricated Metals industry (68 percent) and the lowest was for small establishments in the Health Services industry (25 percent). The Plastics & Rubber industry and the Transportation Equipment industry had the lowest aggregate response rate at 41-42 percent, while Furniture Manufacturing and Food Production showed the highest responses at 54-56 percent. Given the substantial differences in response rate among the strata, the issue of potential non-response bias needs formal investigation. Because the SET project used the MESC administrative data to prepare the sampling frame, and because workers' compensation (BWDC) administrative data were also gathered for all firms in the universe that could be matched, it is possible to compare the matched MESC/BWDC data for respondent and non-respondent firms (not establishments) to the SET sample survey. This should provide the most definitive possible test for response bias.

Table A.10 shows that there are some interesting differences between respondents and non-respondents to the SET survey. In the first place, the respondents are much larger when employment is measured at the firm level. While both respondents and non-respondents exclude firms with less than 100 employees, the mean employment level for survey respondents was 2,043 in the second quarter of 1988, while for non-respondents it was only 549. The huge standard error for respondents prevents this difference from being statistically significant and indicates that the mean is dominated by a small number of very large firms. Recall that the mean employment level for establishments (not firms) in the SET sample was reported as 601 in 1988 in table A.9.

Table A.10 Comparison of Respondents with Non-Respondents

SET Survey Sample - Firm Level Data

	Respondents			Non-Respondents			t-statistic
	n	Mean	Std. Error	n	Mean	Std. Error	
Employment, 1988:II	196	2,043	1,018.3	214	549	172.9	-1.45
WC Claim Rate, 1986-88	196	3.69	.194	215	4.68	.278	2.92**
Claim Duration, 1986-88 (days)	195	104.2	3.76	214	101.4	4.29	-0.49
Average Indemnity/Payroll, 1986-88 ($)	196	$.60	0.41	215	$.72	.045	1.81

**Difference between respondents and non-respondents is statistically significant at 99 percent confidence level.

The respondents have significantly lower incidence of workers' compensation claims, about 3.69 per 100 employees versus 4.68 per 100 for non-respondents, or about 21 percent less. This difference is statistically significant, and not unexpected. Given that the survey was specifically probing establishment's injury and disability prevention activities, it is no surprise to find higher response rates among firms that are doing a good job than among those doing a poor job.

The table shows that the average duration of workers' compensation wage-loss claims is about the same for both groups at just over 100 days. It also indicates that indemnity payments (wage-loss benefits), like claims, were about 20 percent higher for non-respondents at $.72 per $100 versus $.60 per $100 for respondents.[57] The comparisons reported in table A.10 are for the unweighted SET sample, which means that the stratified sampling design contributes to these reported differences as well. This is because the sampling ratios were designed to be proportional to the injury rates of the different strata, and because the response rates also differ by strata. However, weighted sample results return the same basic comparisons between respondents and non-respondents.

A probit regression analysis was also performed comparing respondents to non-respondents to the SET survey. It yielded the same basic results as table A.10, with response status being significantly negatively related to the workers' compensation claim rate (t = - 1.99) and positively, but not quite significantly, related to size of firm (t = 1.90). It

[57]Note that this represents the wage-loss payments as of the date of report. It does not constitute an estimate of final wage-loss payments when all claims have been closed at some point in the distant future.

is also noteworthy that response status was somewhat geographically specific, with establishments in Grand Rapids (Michigan's second largest city) more likely, and establishments in Detroit less likely, to respond to the survey. There were no significant differences by industry.

The conclusion is that the SET sample is not perfectly representative of the underlying population, but that it is sufficiently representative to justify additional analysis. In part this judgment depends on the fact that the biases appear to work in the expected direction when comparing respondents to non-respondents. However, it will also be important to control for size of establishment to prevent dominance by the larger employers in the SET sample. The acceptance of the SET sample also depends on the richness of establishment level detail in the SET survey; it is just too promising to pass up. There is no other comparable database where information about establishment performance on the full range of OSHA outcome measures (recordables, lost workday cases, and total lost workdays) has been combined with data on workers' compensation experience and specific employer characteristics.