

Randomized Safety Inspections and Risk Exposure on the Job: Quasi-experimental Estimates of the Value of a Statistical Life

By JONATHAN M. LEE AND LAURA O. TAYLOR

ONLINE APPENDIX

A. Sample Construction

To comply with the Occupational Safety and Health Act of 1970, states may either adopt their own inspection program or operate under the purview of the Federal OSHA inspection program. The research presented herein focuses on the 28 states and the District of Columbia that operate under the Federal OSHA inspection program since a common and transparent scheduling system for conducting inspections is available for these states. A map of the Federal OSHA program states is provided in Figure A1. Collectively, the Federal OSHA program states contain almost 60% of the manufacturing plants in the U.S.

Figure A2 describes how the sample of plants in Federal OSHA program states are further refined for estimation. The number of plants at each step of the sample selection process are hypothetical due to U.S. Census Bureau confidentiality requirements, but do reflect general statistics for the data. First, plants in OSHA Federal Plan states with fewer than 11 employees are dropped from the sample since OSHA does not include these plants in their programmed inspections. This restriction eliminates approximately 50% of manufacturing plants. For plants with 11 or more employees, we also drop any plant who received their first visit from OSHA inspectors prior to 1987, the start of our study period.

Prior to 1987, the federal OSHA program employed a two-step procedure for determining which plants would receive a comprehensive inspection. First, OSHA would randomly select plants to which it sent inspectors. Upon arrival at

the plant, the inspectors review the plant's injury logs. If the plant's reported injury rates were below the national average for all manufacturing plants, no inspection was conducted and, importantly, plants knew this rule. Thus, while plants were still randomly selected to be visited, the decision to conduct an actual inspection was not random. We thus delete any plant whose injury logs were reviewed prior to 1987, thus ensuring randomization of the plants that are dropped from our sample based on their prior-inspection history. This restriction reduces the sample of plants with 11 or more employees by approximately 40 percent.

B. Overview of the OSHA to COM Matching Procedure

OSHA inspection data identifies each plant by name and address. The U.S. Census Bureau assigns each plant a unique longitudinal plant identifier (PPN) that can be easily linked to other Census datasets such as the Census of Manufacturers and Annual Survey of Manufacturers. The universe of plants and their PPNs are available in the U.S. Census Bureau's *Business Register*. OSHA inspection data are matched to the plants in the *Business Register* using an iterative probability matching process that is based on an algorithm developed by Fellegi and Sunter (1969). The matching is conducted within two-digit SIC industry classification, but is imperfect because it is based on company name and plant address.¹ Although the *Business Register* contains a unique longitudinal plant identifier, we repeat the matching process between OSHA data and each *Business Register* for each year between 1987 to 1997 in order to improve match rates in situations where plants report multiple addresses across years.

¹ There are numerous reasons for an imperfect name and address match between OSHA and the Business Register, including ownership changes, misreported or miscoded addresses, and large plants with multiple addresses.

The first step of the matching algorithm involves standardizing the name and address records by converting these to a common case and street suffix abbreviation. The records are then parsed along several dimensions in order to identify matches based on the full name and address and also based on subsets of these fields (e.g. city, state, zip code, and street number). Each matching field is assigned a weight that reflects the fact that certain criteria such as a full address match are more likely to uniquely identify “true” matches in comparison to matching along a sub-string of the address.² The algorithm then assigns a matching score to each possible plant-record combination in the *Business Register* and OSHA inspection data that is equal to the sum of the weighted number of field sequences for which the two records are a match.

In addition to plant name and address, the comparison space is augmented to include whether or not the two records are a match using the DQMatch function in SAS.³ A threshold for the matching score is then chosen so that plants above

² Specifically, the weight assigned to each matching criteria is equal to the ratio of the probability that two records are matched along a field given that they are “true” matches divided by the probability that two records are matched given that they are not true matches. Determining true and untrue matched probabilities for each field requires calibrating the matching algorithm using a reference dataset that is a carefully reviewed subset of the data consisting of plants that are manually checked to identify either certain matches or certain non-matches. We thank Wayne Gray for sharing code upon which our matching is based. See Scholz and Gray (1993, 1990) and Gray and Mendeloff (2004) for examples of the code implementation using Census and BLS data.

³ SAS 9.1.2 Data Quality Server: Reference, Cary, NC: SAS Institute Inc., 2004. DQMatch also parses and standardizes the data as described above. It then uses a phonetic based matching algorithm for organization names and addresses that is tailored to the U.S. English language. Details are provided in the online documentation for SAS DataFlux Data Management Studio available online at:

the threshold are considered matches, and plants below the threshold are not. The threshold is calibrated using the carefully reviewed subset of data such that less than 1% of untrue matches are mistakenly treated as a match. A lower-bound threshold is also established to eliminate records that are almost certainly not matches. The lower-bound threshold is set at the lowest matching score for all true matches in the carefully reviewed reference data. Records with matching scores above the lower bound (certain non-matches) and below the upper bound (certain matches) are then manually reviewed by the researchers to identify additional matches that were missed by the automated algorithm. The initial algorithm match rates were approximately 60%.

Table A1 compares available data for matched and unmatched plants using the OSHA IMIS database. In most all cases, the means are quantitatively very close for matched and unmatched plants, and 95% confidence intervals overlap for all variables. However, we do note that the matched plants have a higher rate of unionized plants (although not statistically significant): 30% for matched versus 19% for unmatched. In addition, matched plants have a lower rate of “failed inspections.” In some cases, inspectors would arrive at a plant and could not conduct an inspection because the plant had an incorrect address, was closed, was recorded with an incorrect industry classification, or some other reason. For matched plants, only 9% had a failed inspection, while 24% of unmatched plants were recorded with a failed inspection. It is not surprising that unmatched plants have a higher rate of failed inspections since an incorrect address in the OSHA database would fail the inspection and also prevent us from finding a match for the plant in the Census data (assuming Census had the correct address). Lastly,

<http://support.sas.com/documentation/onlinedoc/dfdmstudio/2.6/dmpdmsug/dfUnity.html>

(last accessed July, 2015).

we note that the number of violations and the number of serious violations is somewhat higher for matched plants, but again, 95% confidence intervals overlap.

C. Balance Tests

Table A2 summarizes tests for covariate balance among inspected and uninspected plants. Specifically, for all inspected (I) and uninspected (UI) plants in year t , the mean difference among differences for covariate X is computed as $\Delta = E[X_t^I | ind_i, state_s] - E[X_t^{UI} | ind_i, state_s]$, conditioning on the plant's industry (ind_i) and state ($state_s$). To be consistent with OSHA's randomization policy, we exclude plants with fewer than eleven employees or which received a comprehensive inspection within a time period specified by OSHA guidelines.

All available plant-level employment and characteristic variables reported in Table 1 of the main text are tested for balance in each year. A variable is considered balanced if we fail to reject $\Delta = 0$ at the 10% level. Note that while inspections are tracked on an annual basis, the COM occurs once every five years. As a result, balance is tested in the COM variables among inspected and non-inspected plants in the year the COM takes place, and in each of the four years after the COM wave in question. For instance, using the 1992 COM wave, balance tests are conducted on the 1992 COM variables assigning plants to inspected or uninspected groups based on whether or not they received a programmed inspection in 1992. Balance is also tested in the 1992 COM data assigning plants to inspected or uninspected groups based on whether or not they received a programmed inspection in 1993, 1994, 1995, or 1996. Thus, each wave of the COM data is tested for balance five times using inspection status in each of five years.

In Table A2, covariates for which the null hypothesis is rejected at the 10% level are listed directly in the table. If a variable name does not appear, we could not reject the null of equality of means for that covariate for that year's inspected

and uninspected plants. Table A3 provides an example of the underlying tests using the year 1997, and reports the means of each covariates tested in 1997 and the difference between the means (and standard error of the difference). Table A2 highlights that prior to 1987, in years when OSHA's stated policy was to complete an inspection only after reviewing a plant's safety logs, many covariates fail to balance. In contrast, during the eleven years of the study sample, only the covariate *Single Unit Plant* fails the balance test and in only two years.

D. Additional Results

Table A4 corresponds to Table II of the main text and reports coefficient estimates for time-varying plant characteristics as described in equations (3) and (4). The models presented in Table A4 correspond to the model results reported in the first column of Table II. In general, aside from receiving a programmed OSHA inspection, time-varying plant characteristics are not significant predictors of plant-level fatality risks (see column 1, Table A4). The sign and significance for the time-varying plant characteristics are generally aligned with a priori expectations in the wage equation (see column 2, Table A4). Specifically, larger plants (as measured by total employment) pay higher wages, although plants with more production workers pay lower wages, holding total employment constant. We also find that increases in the cost of materials, production worker productivity, and turnover are associated with increases in wages. In addition, single-unit plants are also estimated to pay higher average wages in comparison to multi-unit plants.

REFERENCES

- Fellegi, Ivan P and Alan B Sunter.** 1969. "A Theory for Record Linkage." *Journal of the American Statistical Association*, 64(328), 1183-210.
- Gray, Wayne B and John M Mendeloff.** 2004. "The Declining Effects of Osha Inspections on Manufacturing Injuries, 1979-1998." *Industrial and Labor Relations Review*, 57, 571-87.
- Scholz, John T and Wayne B Gray.** 1993. "Does Regulatory Enforcement Work? A Panel Analysis of Osha Enforcement." *Law & Society Review*, 27(1), 177-213.
- _____. 1990. "Osha Enforcement and Workplace Injuries: A Behavioral Approach to Risk Assessment." *Journal of Risk and Uncertainty*, 3(3), 283-305.

Table A1. Summary Statistics for Matched and Unmatched Plants.^a

Variable	Matched (std. dev.)	Unmatched (std. dev.)
<i>Plant Characteristics</i>		
<i>Number of employees</i>	237.2 (1,433)	222.1 (1,212)
<i>Inspection year</i>	1991 (3.808)	1991 (3.213)
<i>Unionized plant=1</i>	0.305 (0.506)	0.191 (0.428)
<i>OSHA Inspection Data</i>		
<i>Inspectors could not conduct an inspection=1^b</i>	0.092 (0.430)	0.238 (0.366)
<i>Received a programmed inspection=1</i>	0.438 (0.587)	0.460 (0.498)
<i>Received a complaint inspection=1</i>	0.328 (0.555)	0.329 (0.469)
<i>Received an accident inspection=1</i>	0.021 (0.183)	0.028 (0.154)
<i>Received an inspection classified as “other” = 1</i>	0.210 (0.471)	0.183 (0.398)
<i>Number of violations found during an inspection</i>	4.640 (6.977)	3.688 (5.889)
<i>Number of violations found that were classified as “serious”</i>	2.595 (4.738)	2.021 (4.026)

^a Summary statistics are for plants listed in the OSHA IMIS database (available at http://ogesdw.dol.gov/raw_data_summary.php (last accessed April, 2018))

^b Inspection could not be conducted because, after arriving at a plant’s address, the inspector found the address was incorrect, the plant was closed, the industry classification was incorrect, or another reason that disqualified the plant from inspection on that visit.

Table A2. Randomization Tests for Federal Programmed Inspections.

COM Wave	Year of Inspection	Variables Failing to Balance^a
1982	1982	<i>Total # Employees, Cost of Materials, Single Unit, Productivity, Average # Production Workers</i>
	1983	<i>Single Unit</i>
	1984	<i>Total # Employees, Average # Production Workers, Single Unit</i>
	1985	<i>Total # Employees, Average # Production Workers, Single Unit</i>
	1986	<i>0</i>
1987	1987	<i>Single Unit</i>
	1988	<i>0</i>
	1989	<i>0</i>
	1990	<i>0</i>
	1991	<i>0</i>
1992	1992	<i>0</i>
	1993	<i>0</i>
	1994	<i>0</i>
	1995	<i>Single Unit</i>
	1996	<i>0</i>
1997	1997	<i>0</i>

^a If a variable name is listed, there was a statistically significant difference between the means of inspected and non-inspected plants at the 10% level.

^b A “0” indicates that all variables were balanced between the two samples in this year.

Table A3: Comparison of Covariates for 1997 Randomization Test.

Variable Name	Inspected Plants	Non-inspected Plants	Difference (std. err.)^a
<i>Hourly Wage</i>	\$13.09	\$13.21	-0.116 (0.14)
<i>Other Workers' Wages</i>	\$22.99	\$22.96	0.0341 (0.55)
<i>Number of Employees</i>	102	97.4	5.02 (6.44)
<i>No. Production Workers</i>	77.8	72.7	5.08 (5.09)
<i>Cost of Materials</i>	11.16	10.16	1.00 (1.68)
<i>PW Productivity</i>	\$133	\$132	0.908 (6.97)
<i>Single Unit Plant</i>	62.2%	61.7%	0.502 (1.50)
<i>Turnover</i>	9.47%	9.25%	0.224 (0.74)

^a In parentheses are the standard errors for the difference in the means of each variable for inspected and non-inspected plants. No differences are statistically significant at the 10% level.

Table A4: IV estimates of plant-level changes in risks and wages in response to receiving an OSHA inspection.^a

Variable Name	Estimated Coefficients	
	(standard errors)	
	Fatality Rate	Wages
	(Equation 3)	(Equation 4)
<i>Number of Employees</i>	-0.00006 (0.0001)	0.00044* (0.00024)
<i>No. Production Workers</i>	0.00014 (0.00019)	-0.00252*** (0.00041)
<i>Cost of Materials</i>	0.00089 (0.00347)	0.0135** (0.0069)
<i>PW Productivity</i>	0.00006 (0.00005)	0.00038*** (0.00009)
<i>Single Unit Plant</i>	0.0859 (0.149)	0.174** (0.0779)
<i>Turnover</i>	-0.0383 (0.0956)	0.296*** (0.0541)
<i>Programmed Inspections</i>	-1.338*** (0.338)	----
<i>Fatality Rate</i>	----	0.204*** (0.0715)
R-squared	0.296	0.484

^aStatistical significance at the 1%, 5%, and 10% level are represented by ***, **, and *, respectively. The models presented correspond with the first column of results presented in Table II and are based on 252,800 observations (65,300 plants). Industry-by-state-by-year fixed effects and plant-level fixed effects as specified in equations (2) and (3) are also included. Standard errors are clustered at the plant level. The number of observations (number of plants) are reported rounded to the nearest 100 due to U.S. Census Bureau confidentiality requirements.

Figure A1: Distribution of Federal and State Plan OSHA States.*



*Note: Alaska and Hawaii are both State Plan OSHA States

Figure A2. Example of the data construction process.

Sample Description <i>(Sample restriction in italicized text within each grey box)</i>	Approximate # Plants
Total U.S. Manufacturing Plants (1997)	362,000
<i>Delete plants in states with plans administered by the state OSHA programs (approximately 42% of plants)</i> ↓	
Plants located in states with Federal OSHA programs	210,000
<i>Delete plants that have fewer than 11 employees because they are exempt from OSHA programmed inspections (approximately 39% of plants)</i> ↓	
Plants located in states with Federal OSHA programs and with more than 10 employees	128,100
<i>Delete plants that have been randomly visited by OSHA (and possibly inspected) prior to 1987, the start of the study period (approximately 45% of plants).</i> ↓	
Plants located in states with Federal OSHA programs, with more than 10 employees, and who have not been inspected prior to 1987.	70,500
<i>Delete plants that received a complaint inspection during the 1987 to 1997 study period (approximately 8% of plants).</i> ↓	
Plants located in states with Federal OSHA programs, with more than 10 employees, who have not been inspected prior to 1987, and have not received a complaint inspection.	64,900
<i>Delete plants with an accident inspection prior to being scheduled for a federal programmed inspection during the 1987 to 1997 study period (approximately 0.1% of plants).</i> ↓	
Plants located in states with Federal OSHA programs, with more than 10 employees, who have not been inspected prior to 1987, have not received a complaint inspection or had a death prior to being scheduled for a programmed inspection. These data would comprise the final estimation sample.	64,800

^a Due to US Census Bureau confidentiality requirements, the numbers in this table are not based on actual data used in this research. They are hypothetical data, but based roughly on the publicly available statistics for manufacturing industries in 1997. As reported in the main text, the final estimation sample contains approximately 65,300 plants.