

Incorporating OSHA Administrative Records in the Survey of Occupational Injuries and Illnesses

Brooks Pierce
March 9, 2018



A Nice Problem to Have

- Administrative records on workplace injuries have become available
- Possible windfall to our survey of same



Roadmap

- Compare the survey and administrative data
- Describe a simple data combination approach
- Describe a prototype record linkage



OSHA Records and SOII Data



OSHA Recordkeeping and Reporting

- Private sector employers must maintain injury logs
 - ▶ Separately for each establishment
 - ▶ Exemptions based on number of employees and industry
- Annual Summary form, posted at workplace
 - ▶ Annual totals for injuries of various types
 - ▶ Measures of exposure
- New: Electronic Reporting Requirement

Survey of Occupational Injuries and Illnesses (SOII)

- Collects information recorded on OSHA forms
- Federal-State cooperative program
- Large survey, oversamples higher risk parts of the economy
- Broader scope than OSHA reporters
 - ▶ Substantial overlap



Potential Concerns - External Data

- Completeness
- Representativeness
- Different process; different reports
 - ▶ Establishment versus broader aggregate
 - ▶ Particular fields
- Changes through time



Comparing the External and Survey Data



Injury Tracking Application (ITA) data

- Approximately 215K records for 2016 calendar year injuries
- Restrict to:
 - ▶ Private sector
 - ▶ Required to submit electronically
 - ▶ Fed-OSHA states that participate in SOII (21 states)
 - ▶ Miscellaneous edits and exclusions
 - ▶ 80K records, self-weighting prior to adjustments

SOI Records

- Impose same restrictions
- Approximately 35K records
- Weights reflect initial selection probabilities as well as the typical array of post-selection adjustments



Reporting Propensities

- ITA records (80K) are about 35% of the sum of SOII weights for the same population
 - ▶ Reasonable measure for an ITA “response rate”
- Coverage varies with observables, including response outcomes
- Patterns consistent with a selection story: more sophisticated agents report

Larger Establishments Report in ITA

# Employees	Ratio of ITA to SOII weights totals
20-49	23.3
50-249	45.2
250-999	68.1
1000+	86.5
All	35.0

ITA Records Differ in Outcomes

- Higher establishment case rates (4.7% versus 3.9%)
- Lower percentage reporting zero cases (23.7% versus 39.3%)
- Higher establishment percentage of job restriction cases, conditional on positive cases (27.7% versus 21.9%)
- Reweight ITA using SOII employment totals within state-industry-size cells – analog to SOII final benchmarking

Injury Rates by Source, Select States

State	All Cases			DAFW	DJTR
	SOII	ITA	ITA-SOII	ITA-SOII	ITA-SOII
TX	3.33	3.57	0.24	-0.00	0.24
IL	4.04	4.45	0.41	0.16	0.17
ME	6.88	7.35	0.48	0.09	0.27

Notes: ITA statistics utilize adjusted weights. ‘DAFW’ refers to cases resulting in one or more days away from work; ‘DJTR’ refers to cases with one or more days of job transfer or restriction, but zero days away from work.



How to Combine the Data?



Dual Frame / Small Area Models

$$\hat{Y}_d = \theta_d \hat{Y}_{SOII,d} + (1 - \theta_d) \hat{Y}_{ITA,d}$$

for domain d common to both source frames

Regression Adjustments to ITA

$$\hat{Y}_d = \theta_d \hat{Y}_{SOII,d} + (1 - \theta_d) X_{ITA,d} \hat{\beta}$$

- ▶ β via regressions relating SOII (lhs) and ITA estimates across d
- ▶ compositing weights θ_d depend on relative sizes of variances
 - SOII estimate variance
 - model error variance

Regression Adjustments

- N=1809 state x industry domains
- Hours worked (exposure measure): $\beta=.975$ (.004), $R^2=.97$
- Total cases: $\beta=.925$ (.006), $R^2=.92$
- Suggests an ITA total cases rate reduction of about 5%



Composite Estimates

- Resulting weights θ_d tend to be high
 - ▶ About 75% are in the [0.85,1.0] range
 - ▶ Typical composite estimate not far from SOII value
 - ▶ Variance improvements for some domains below state level
- Possible implication: revisit SOII sample allocations

Record Linkage



Example Record Linkage

- Link ITA records to SOII survey responses
 - ▶ Identify reporting differences by same units
 - ▶ Fellegi-Sunter type probabilistic linkage methods
 - ▶ Raw materials: (multiple) names and addresses
 - Do not use industry, employment, reported outcomes
 - Do not have EIN or other identifying information
 - ▶ Block on zip code
- Linking to SOII frame would give fuller set of linked records

Linkage Results

- Approximately 9000 linked records
- Substantial fractions report identical employment, hours worked, and total cases in the two sources
 - ▶ Among linked units reporting identical employment, about 12% report industry differently enough to move units across strata
- Linked units reporting different employment levels are less likely to report consistent case totals or industry
 - ▶ Incorrect links or different reporting?

Reporting Consistency

ITA-SOII Employment Differential	Percent with same:		
	Case total	3-digit NAICS	Stratum
< -5%	34.2	74.2	45.9
[-5%, 5%]	85.6	91.5	88.5
> 5%	51.1	82.2	82.2



Conclusions



Ongoing work

- Understand selection into ITA reporting
 - ▶ Further weights adjustments
- Implications for SOII sample allocation
- Linkage improvements
 - ▶ Gauge industry and employment misclassification (stratum assignment)
 - ▶ Are reports for different aggregates?

Contact Information

Brooks Pierce
pierce.brooks@bls.gov

