Refining disclosure controls for the Census of Fatal Occupational Injuries (CFOI)

Danny Friel

FCSM Research & Policy Conference
October 26, 2023

Co-authors

- Alyssa Gillen
- Julie Krautter
- Yvan Saastamoinen

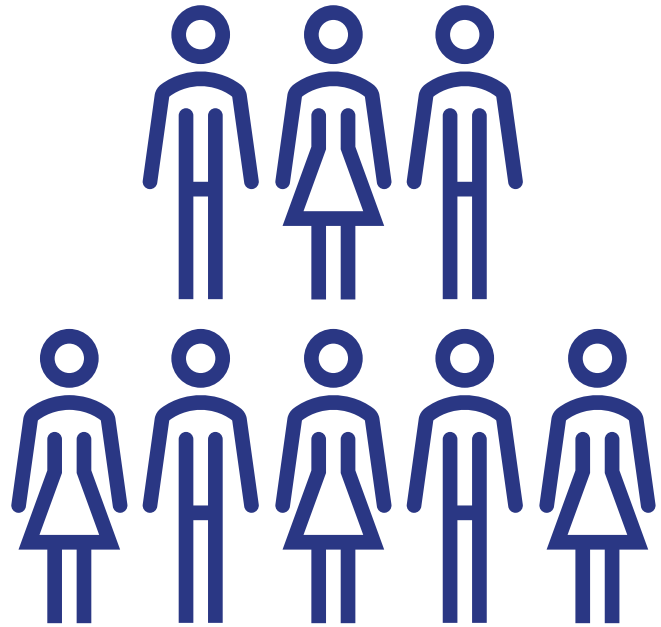
The views expressed here are those of the authors and do not necessarily reflect the views or policies of the Bureau of Labor Statistics or any other agency of the U.S. Department of Labor.

Road map

- What is disclosure control?
- Disclosure control for CFOI
  - Primary and secondary suppressions
  - Hypercube method
- Proposed refinements to the hypercube
- Discussion

Disclosure control

- The direct or indirect release of sensitive or private information about a survey or census unit
- Data users have access to other information
- Goal: minimize practical risk



Disclosure control

- Commonly used methods include:
  - Threshold rules
  - Dominance rules
  - Noise
- When a cell fails one or more rules, it is *suppressed*

Poured concrete foundation and structure contractors	23811	28	-	11	-	7	5	-	
Residential poured concrete foundation and structure contractors	238111	4	-	-	-	-	-	-	
Nonresidential poured concrete foundation and structure contractors	238112	6	-	-	-	-	-	-	

Census of Fatal Occupational Injuries (CFOI)

- Publishes a complete count of fatal injuries each year
- Protecting CFOI data is challenging
  - No sampling
  - Fatal injuries are rare events
  - Exact counts are important
  - Sixteen grouping variables for cells (industry, occupation, gender, …)

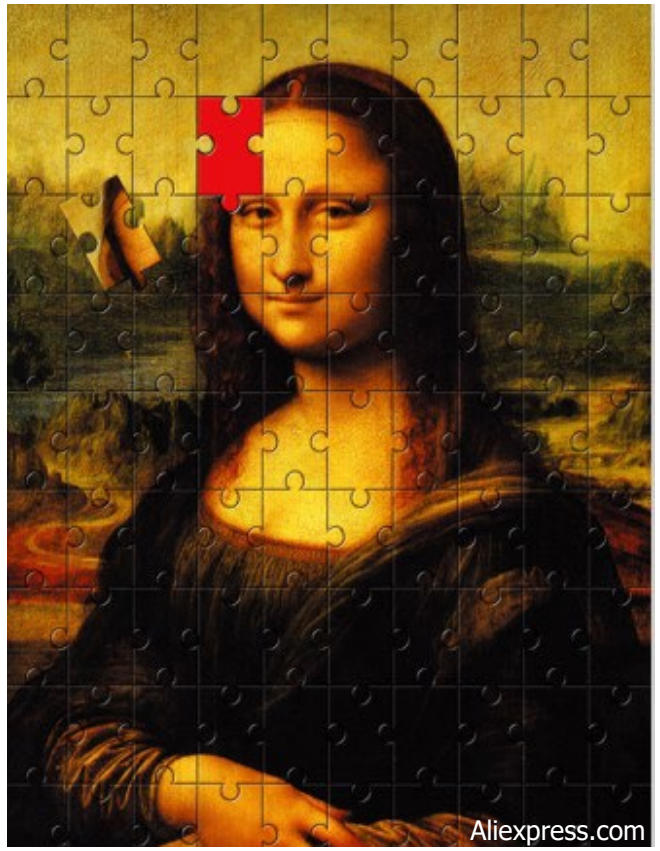
Primary suppression

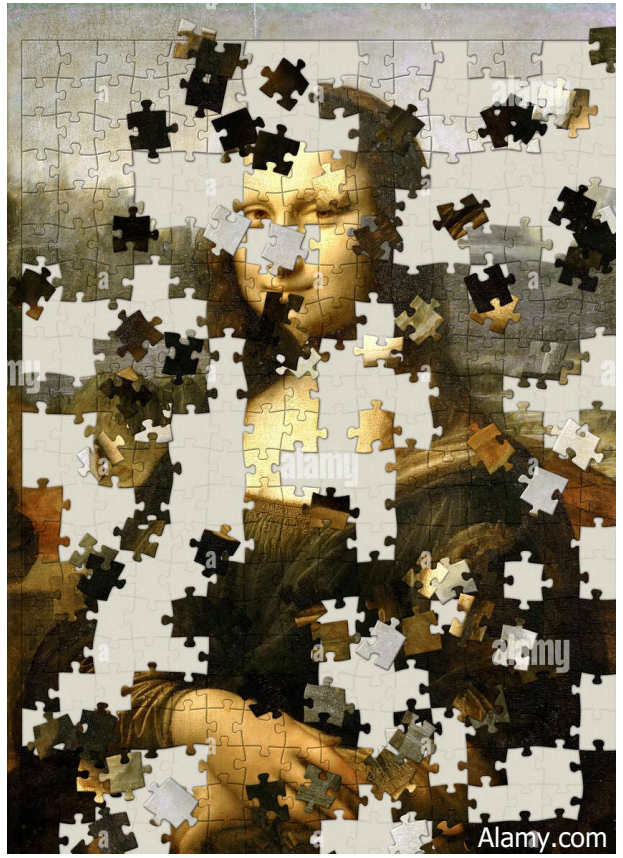
- When disclosure rules are violated, we suppress the cell

No suppressions	
Occupation	**Number of fatal injuries**
All occupations	**100**
Occupation 1	80
Occupation 2	18
Occupation 3	2

Sensitive cell is suppressed	
Occupation	**Number of fatal injuries**
All occupations	**100**
Occupation 1	80
Occupation 2	18
Occupation 3	--

The limits of primary suppression





Primary vs. secondary suppression

Primary suppression only	
The count for occupation 3 doesn't meet publishability criteria	
Occupation	**Number of fatal injuries**
All occupations	**100**
Occupation 1	80
Occupation 2	18
Occupation 3	--

Even though this cell is suppressed, we have enough information to compute its value: 100 – 80 – 18 = 2

Primary and secondary suppressions	
The count for occupation 2 is suppressed as well	
Occupation	**Number of fatal injuries**
All occupations	**100**
Occupation 1	80
Occupation 2	--
Occupation 3	--

With two cells suppressed, we don't have enough information to compute either value. Possible values include 20 and 0, 19 and 1, 10 and 10, 15 and 5...

Table differencing

Occupation 2	Number of fatal injuries
Full-time	12
Part-time	6

Table differencing

Occupation	Number of fatal injuries
1	80
2	--
3	--
Total	100

Occupation 2	Number of fatal injuries
Full-time	12
Part-time	6

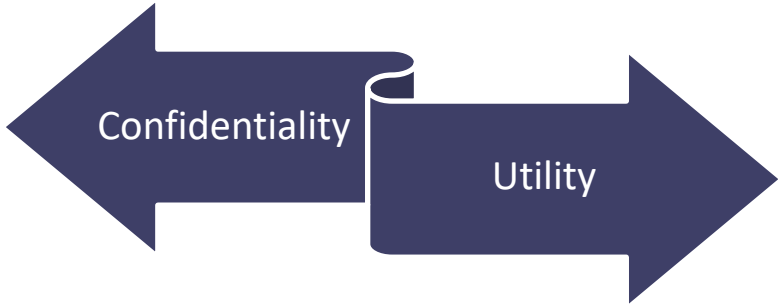
The CFOI Hypercube

- Scans for:
  - Primary suppressions
  - Secondary suppressions (within-tables)
  - Secondary suppressions (across tables) for up to four CFOI variables

CFOI variables	Number of variables	Number of cells screened by hypercube
Fatal injuries by industry	1	**706**
Fatal injuries by industry, event/exposure	2	706 x 8 = **5,648**
Fatal injuries by industry, event/exposure, state	3	706 x 8 x 56 = **316,288**
Fatal injuries by industry, event/exposure, state, age	4	706 x 8 x 56 x 9 = **2,846,592**

Improving the hypercube

- Tradeoff:



  - Table A-1: Fatal occupational injuries by industry and event/exposure
    - 5% of cells are published in 2021
- Ad-hoc improvements
- Systematic improvements

Refining the hypercube: zeroes

- Zero counts are often captured by confidentiality rules
- Zero counts are very common (80%+) in CFOI; useful for policy
- Two questions:
  - How do zero cells impact the effectiveness of secondary suppressions?
  - Can the hypercube be configured to only suppress zeroes when they pose a substantial confidentiality risk?

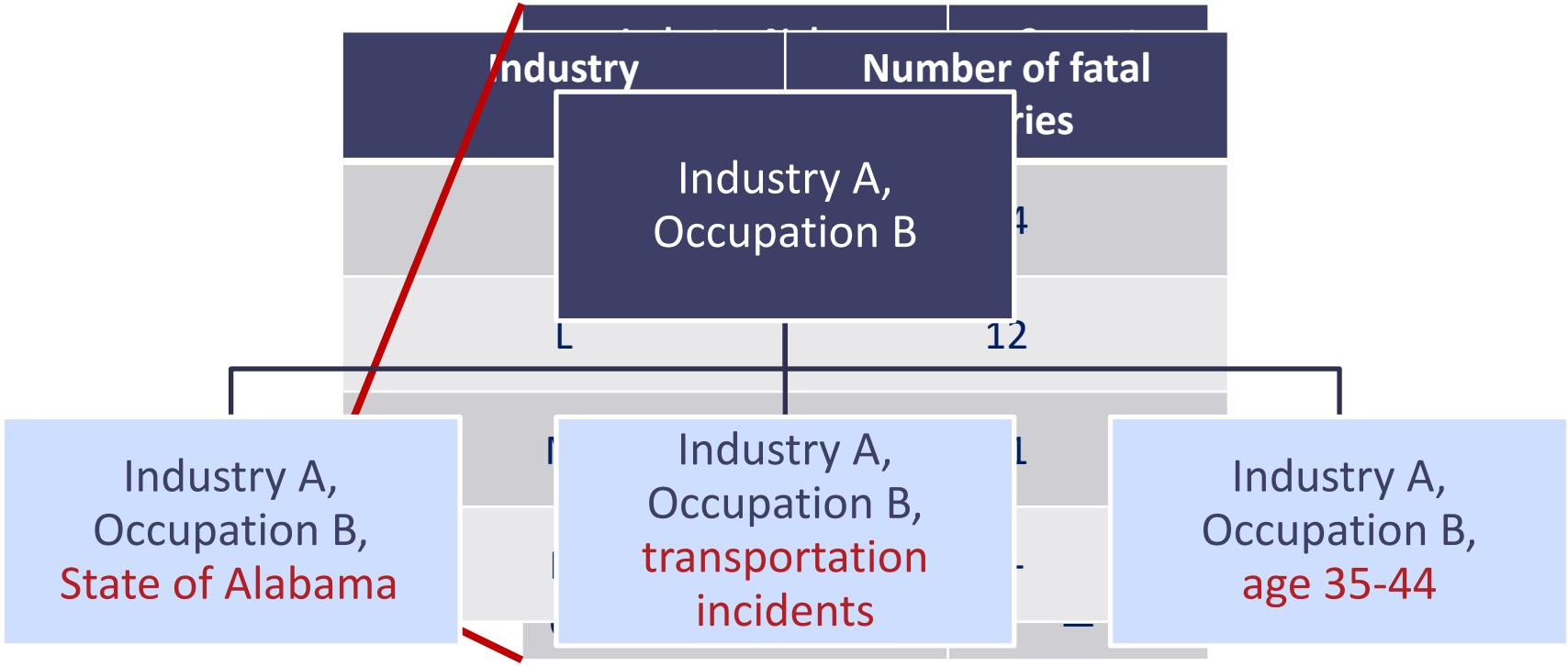
Adding zeroes to a data table (simulated data)

	All Events	Event 1	Event 2	Event 3	Event 4	Event 5	Event 6
Industry A	**21**	**4**	**--**	**8**	**0**	**4**	**--**
Industry A-1	**7**	--	--	--	0	1	--
Industry A-2	**12**	--	--	3	0	3	--
Industry A-3	**2**	~~0~~	0	--	0	0	~~0~~

Preliminary results

- Over 80% of zeroes may be publishable.
  - Zeroes must not be published if:
    - They can be used to derive the value of a suppressed nonzero cell
  - Zeroes may be published if:
    - They are the child of a non-zero cell or a suppressed zero cell
  - Zeroes can also be suppressed at random, to infuse more uncertainty
    - E.g., if there are > 3 zeroes in a table row, up to 2 may be randomly suppressed

Proposal 2: Limiting child suppressions



Summary

- Zero-count cells and child cells must be displayed selectively
  - Zeroes and child cells can be used to derive sensitive information
  - Cells that are non-sensitive in one table may be sensitive in another
  - Additional uncertainty can be infused as needed
- Post-processing time and resources may limit options

Contact Information

Danny Friel
Office of Compensation and Working Conditions
Friel.Daniel@bls.gov