Using Linked Data Sources To Predict Occupations With Respirator Or Mask Use Based On Establishment Characteristics

Danny Friel

Office of Compensation and Working Conditions

U.S. Bureau of Labor Statistics

April 6, 2022

About this project

- **Collaborators:**
  - ▶ Michelle Myers, Bureau of Labor Statistics
  - ▶ Dee Zamora, Bureau of Labor Statistics
  - ▶ Xingyou Zhang, Bureau of Labor Statistics
  - ▶ Katherine N. Yoon, National Institute for Occupational Safety & Health
  - ▶ Megan Casey, National Institute for Occupational Safety & Health
  - ▶ Emily J. Haas, National Institute for Occupational Safety & Health

- *CDC COVID-19 funds were provided to initiate survey development and pilot testing. Resources to execute the survey have not yet been allocated. This effort was conducted under CDC-BLS IAA 21FED21000007OHL which ended September 30, 2021 due to lack of funds.*

Agenda

- Introduce the Survey of Respirator Use and Practices (SRUP)

- Respondent burden associated with occupational data collection

- Proposed solution: predicting occupations via an algorithm

  ▶ Key factors:

  – Occupations likely to have respirator or mask use

  – Occupations likely to be present at each establishment

  ▶ Algorithm specifications

Survey of Respirator Use and Practices (SRUP)

- Objective: To obtain accurate national data about the management of respiratory hazards and respirator use
  - ▶ Last conducted in 2001
  - ▶ Part 1: Establishment-level questions
 - ▶ Part 2: Occupation-level questions

Puzzle: occupation-level data collection

- For each respondent, we need to identify <u>one occupation</u> for which information will be sought

- In an ideal setting, we could simply choose an occupation from those present at each establishment

  ▶ Since we don't have this information, we would have to:

  – ask respondents for a complete list of occupations

  – accurately enter them into the data collection interface, and

  – choose one for data collection

  – *very burdensome*
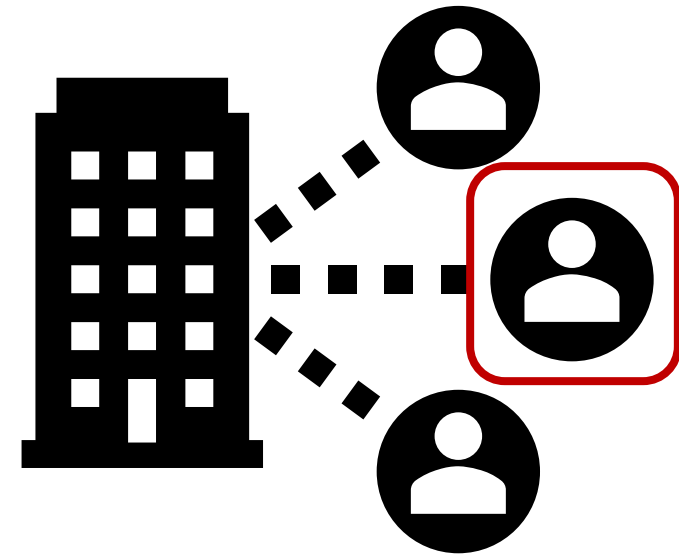
Predicting occupations at SRUP establishments

- To reduce respondent burden, we proposed using a menu of occupations
  - ▶ The menus would have held 14 algorithm-based predictions
  - ▶ The respondent would indicate which (if any):
    - Are **present** at the establishment
    - Use **respirators or masks**
- The algorithm would be successful if each respondent indicates that <u>at least one occupation</u> on their menu meets the two criteria
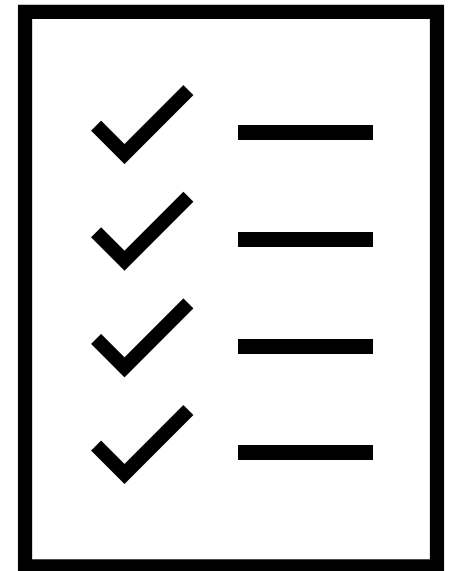
Key factors for occupation prediction

- Factor 1: Select occupations that are likely to use respirators or masks

- Factor 2: Select occupations that are likely to be present at the SRUP establishment at hand

- The algorithm generates a selection pool of predictions and prioritizes them according to these factors

Identifying priority occupations

■ Factor 1: Menus should contain occupations *likely to wear respirators and/or masks*

■ The Occupational Information Network (O*NET) database identifies occupations with documented use of personal protective equipment (PPE)

  ▶ We classified occupations into three priority groups for the SRUP

Priority group	Type of PPE use (O*NET)	Number of occupations
1	Respirators or masks	164
2	Other PPE	128
3	None	556

Identifying "similar" establishments

- Factor 2: Occupations on each menu should be *likely to be present* at the establishment at hand

- We identified establishments **similar to** those in the SRUP sampling frame

  - ▶ i.e., those with the same **NAICS** (industry) **code, size class, or state**

  - ▶ Sources: Occupational Employment and Wage Statistics (OEWS) program and the Survey of Occupational Injuries and Illnesses (SOII)
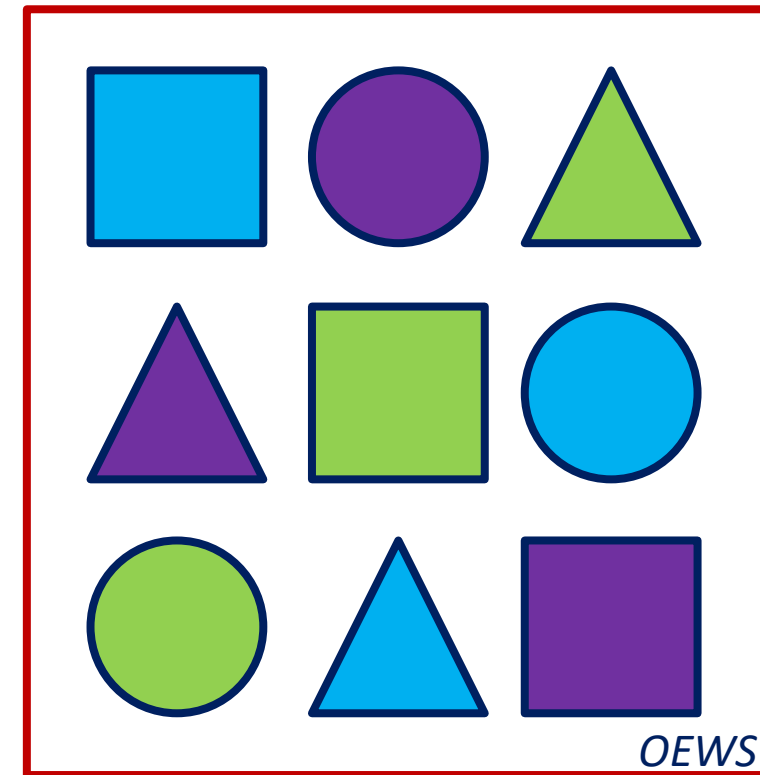
Establishments with uncommon characteristics

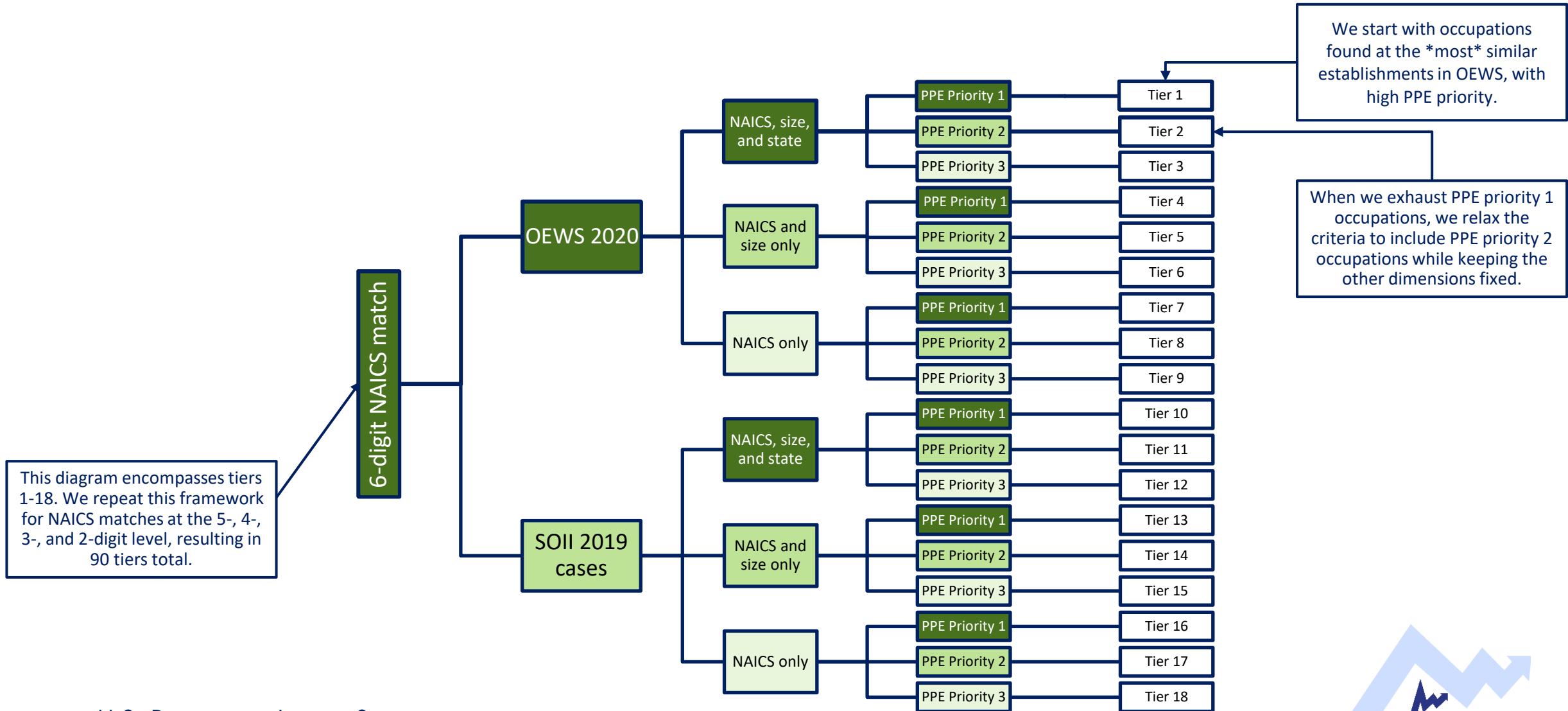- Some NAICS/size/state combinations have only a few observations, resulting in < 14 occupations

- We relax the criteria systematically to generate more options:

  1. PPE priority (priority 1, 2, or 3)

  2. Establishment similarity (NAICS/size/state, NAICS/size, or NAICS only)

  3. Data source (OEWS or SOII)

  4. NAICS level (6-, 5-, 4-, 3-, or 2-digit match)

Anatomy of the occupation selection pool

We start with occupations found at the *most* similar establishments in OEWS, with high PPE priority.

When we exhaust PPE priority 1 occupations, we relax the criteria to include PPE priority 2 occupations while keeping the other dimensions fixed.

This diagram encompasses tiers 1-18. We repeat this framework for NAICS matches at the 5-, 4-, 3-, and 2-digit level, resulting in 90 tiers total.

Constructing the menu of 14

Start

Are there **more than 14** occupations in the selection pool for tier 1?

— No →

Are there **exactly 14** occupations in the selection pool for tier 1?

— No →

When there are **fewer than 14** occupations in the selection pool for tier 1, select all of them.

Yes ↓

14 occupations are randomly selected from the selection pool for tier 1

↓

The menu of 14 is complete and the algorithm terminates.

Yes ↓

All 14 occupations are selected from the selection pool for tier 1

↓

The menu of 14 is complete and the algorithm terminates.

↓

The algorithm continues, moving to the next tier.

↓

Are there enough empty slots in the menu of 14 to add all occupations at the current tier?

— Yes →

All occupations are selected from the selection pool at the current tier.

No ↓

Randomly select enough occupations from the current tier to fill the empty slots in the menu of 14.

→

The menu of 14 is complete and the algorithm terminates.

BLS

Contact Information

Danny Friel
Office of Compensation and Working Conditions
U.S. Bureau of Labor Statistics
friel.daniel@bls.gov

Disclaimer: The findings and conclusions in this presentation are those of the author(s) and do not necessarily represent the official position of the National Institute for Occupational Safety and Health, Centers for Disease Control and Prevention.

BLS