

# Non-healthcare occupational exposure to SARS-CoV-2 across industries in the United States before March 2020: an exploratory data analysis protocol

Michael Walsh\*

## Keywords

applied epidemiology, COVID-19, emerging infectious disease, exploratory data analysis, global health, occupational health

## Abstract

This protocol provides a way to summarize the baseline measures for non-healthcare occupational exposure to SARS-CoV-2 across Census coded industries in the United States before March 2020. The SOEM Industry Extension (SOEM-IE) will provide the baseline measures for non-healthcare occupational exposure to SARS-CoV-2 across Census coded industries in the United States before March 2020. Note, the SOEM-IE derives these baseline measures from two additional data sources. These include the SARS-CoV-2 Occupational Exposure Risk Matrix (SOEM) and the Employed Labor Force Query System (ELFQS). By way of statistical methods, this dataset provides a way to explore the baseline measures for non-healthcare occupational exposure to SARS-CoV-2 across Census coded industries in the United States before March 2020. The summaries generated by this protocol should be of immediate use to policymakers, practitioners, and researchers who are trying to understand various facets of the SOEM-IE data at the levels of workers and occupations. They should also prove useful to individuals and institutions seeking to understand the impact of interventions designed to mitigate occupational exposure to SARS-CoV-2 across industries in the United States.

## Protocol

This protocol involves the following sequence of activities. The first step involves the creation of a master spreadsheet file.<sup>2</sup> The second step involves the extraction of source data from the SOEM-IE. The third step involves summarizing the baseline measures of occupational exposure to SARS-CoV-2 across industries in the United States before March 2020 at the level of workers. The fourth step involves the summarizing the baseline measures of occupational exposure to SARS-CoV-2 across industries in the United States before March 2020 at the level of occupations. In the third and fourth step, a specific set of commands and calculations are used to generate the summaries. These commands include SUMIFS. Note, this sequence of activities is informed by an earlier one used to explore baseline measures for occupational exposure to SARS-CoV-2 across occupational groups in the United States before March 2020 (Walsh 2021a).

---

\* Correspondence: A00021985@sorbonne.ae

<sup>2</sup> Note: The step-by-step protocol is based on Microsoft Excel. Other spreadsheet applications can be used. However, the step-by-step protocol will need to be adapted.

## Dataset Specifications

<b>Subject</b>	Public Health and Health Policy
<b>Specific subject area</b>	Infectious diseases and virology
<b>Type of dataset</b>	Tables
<b>Data Acquisition</b>	<p>This dataset will be generated from secondary data.</p> <p>This dataset will acquire the SOEM-IE data by way of extraction from the “SOEM_IND.xlsx” dataset. The following version of the file was used: September 5, 2021.</p> <p>Note, the SOEM-ID dataset is an extension of the SOEM. It was constructed using data extracted from the SOEM and Employed Labor Force Query System. How the data was extracted has been detailed elsewhere in the grey literature (Walsh 2021b).</p>
<b>Data format</b>	Analyzed
<b>Secondary source</b>	Walsh, Michael. Non-healthcare occupational exposure to SARS-CoV-2 across industries in the United States before March 2020: a dataset generating protocol (1.0). <i>Zenodo</i> . 2021. <a href="https://doi.org/10.5281/zenodo.5513598">https://doi.org/10.5281/zenodo.5513598</a> .
<b>Data accessibility</b>	The author commits to depositing this dataset with Mendeley Data upon acceptance of this protocol.

## **Step-by-step instructions**

The first step involves the creation of a master spreadsheet file:

1. Create: New Excel spreadsheet named “SOEM\_IND\_EDA”
  - a. Create: New tab named “SOEM\_ELF”
  - b. Create: New tab named “SOEM\_IND\_EMPLOYED”
  - c. Create: New tab named “SOEM\_IND\_OCCUPATIONS”

The second step involves the collection of source data on occupational exposure to SARS-CoV-2 across industries in the United States before March 2020:

1. Open: “SOEM\_IND.xlsx” worksheet
2. Go to: “SOEM\_ELF” tab
3. Copy: Columns A:K
4. Open: “SOEM\_IND\_EDA” worksheet
5. Go to: “SOEM\_ELF” tab
6. Paste Special: Copied Data
  - Paste as: Values

The third step involves summarizing the baseline measures of occupational exposure to SARS-CoV-2 across industries in the United States before March 2020 at the level of workers:

1. Open: “SOEM\_IND.xlsx” worksheet
2. Go to: “SOEM\_IND” tab
3. Copy: Columns A:B
4. Open: “SOEM\_IND\_EDA” worksheet
5. Go to: “SOEM\_IND\_EMPLOYED” tab
6. Paste: Copied Data
  - Columns: A:B
7. Insert: Row
  - Above: Row 1
8. Merge:
  - Cells: A1:A2
  - Cells: B1:B2
  - Cells: C1:C2
  - Cells: D1:D2
  - Cells: E1:E2
  - Cells: F1:F2
  - Cells: G1:H1
  - Cells: I1:L1
  - Cells: M1:P1
  - Cells: Q1:T1
  - Cells: U1:W1
  - Cells: X1:Z1
  - Cells: AA1:AB1
9. Label

## Working Draft | Do Not Cite Without Permission

- Cell C1 – “Total Employed (#)”
- Cell D1 – “NHCW Employed (#)”
- Cell E1 – “HCW Employed (#)”
- Cell F1 – “HCW Employed (%)”
- Cell G1 – “Employed Public Facing or Not (#)”
- Cell I1 – “Employed Indoor/Outdoor (#)”
- Cell M1 – “Employed Close Proximity (Measure 1, #)”
- Cell Q1: “Employed Close Proximity (Measure 2, #)”
- Cell U1: “Employed Combined (Proximity 1, #)”
- Cell X1: “Employed Combined (Proximity 2, #)”
- Cell AA1: “Employed Combined High (%)”
- Cell G2: “2”
- Cell H2: “1”
- Cell I2: “4”
- Cell J2: “3”
- Cell K2: “2”
- Cell L2: “1”
- Cell M2: “4”
- Cell N2: “3”
- Cell O2: “2”
- Cell Q2: “1”
- Cell R2: “4”
- Cell S2: “3”
- Cell T2: “2”
- Cell U2: “H”
- Cell V2: “M”
- Cell W2: “L”
- Cell X2: “H”
- Cell Y2: “M”
- Cell Z2: “L”
- Cell AA2: “Combined (Proximity 1)”
- Cell AB2: “Combined (Proximity 2)”

### 10. Run:

- Cell: C3
- Command:  
=SUMIF(SOEM\_ELF!\$A\$2:\$A\$37527,SOEM\_EMPLOYED!A3,SOEM\_ELF!\$C\$2:\$C\$37527)
- Apply: Entire Column

### 11. Run:

- Cell: D3
- Command:  
=SUMIFS(SOEM\_ELF!\$C\$2:\$C\$37527,SOEM\_ELF!\$A\$2:\$A\$37527,SOEM\_EMPLOYED!A3,SOEM\_ELF!\$D\$2:\$D\$37527,">0")
- Apply: Entire Column

### 12. Run:

- Cell: E3
- Command:  
=SUMIFS(SOEM\_ELF!\$C\$2:\$C\$37527,SOEM\_ELF!\$A\$2:\$A\$37527,SOEM\_EMPLOYED!A3,SOEM\_ELF!\$D\$2:\$D\$37527,"#N/A")

## Working Draft | Do Not Cite Without Permission

- Apply: Entire Column
- 13. Run:
  - Cell: F3
  - Command: =E3/C3
  - Number: Percentage
  - Round: One Decimal Place
  - Apply: Entire Column
- 14. Run:
  - Cell: G3
  - Command:  
=SUMIFS(SOEM\_elf!\$C\$2:\$C\$37527,SOEM\_elf!\$A\$2:\$A\$37527,SOEM\_EMPLOYED!A3,SOEM\_elf!\$D\$2:\$D\$37527,2)
  - Apply: Entire Column
- 15. Run:
  - Cell: H3
  - Command:  
=SUMIFS(SOEM\_elf!\$C\$2:\$C\$37527,SOEM\_elf!\$A\$2:\$A\$37527,SOEM\_EMPLOYED!A3,SOEM\_elf!\$D\$2:\$D\$37527,1)
  - Apply: Entire Column
- 16. Run:
  - Cell: I3
  - Command:  
=SUMIFS(SOEM\_elf!\$C\$2:\$C\$37527,SOEM\_elf!\$A\$2:\$A\$37527,SOEM\_EMPLOYED!A3,SOEM\_elf!\$E\$2:\$E\$37527,4)
  - Apply: Entire Column
- 17. Run:
  - Cell: J3
  - Command:  
=SUMIFS(SOEM\_elf!\$C\$2:\$C\$37527,SOEM\_elf!\$A\$2:\$A\$37527,SOEM\_EMPLOYED!A3,SOEM\_elf!\$E\$2:\$E\$37527,3)
  - Apply: Entire Column
- 18. Run:
  - Cell: K3
  - Command:  
=SUMIFS(SOEM\_elf!\$C\$2:\$C\$37527,SOEM\_elf!\$A\$2:\$A\$37527,SOEM\_EMPLOYED!A3,SOEM\_elf!\$E\$2:\$E\$37527,2)
  - Apply: Entire Column
- 19. Run:
  - Cell: L3
  - Command:  
=SUMIFS(SOEM\_elf!\$C\$2:\$C\$37527,SOEM\_elf!\$A\$2:\$A\$37527,SOEM\_EMPLOYED!A3,SOEM\_elf!\$E\$2:\$E\$37527,1)
  - Apply: Entire Column
- 20. Run:
  - Cell: M3
  - Command:  
=SUMIFS(SOEM\_elf!\$C\$2:\$C\$37527,SOEM\_elf!\$A\$2:\$A\$37527,SOEM\_EMPLOYED!A3,SOEM\_elf!\$F\$2:\$F\$37527,4)

## Working Draft | Do Not Cite Without Permission

- Apply: Entire Column
- 21. Run:
  - Cell: N3
  - Command:  
=SUMIFS(SOEM\_elf!\$C\$2:\$C\$37527,SOEM\_elf!\$A\$2:\$A\$37527,SOEM\_EMPLOYED!A3,SOEM\_elf!\$F\$2:\$F\$37527,3)
  - Apply: Entire Column
- 22. Run:
  - Cell: O3
  - Command:  
=SUMIFS(SOEM\_elf!\$C\$2:\$C\$37527,SOEM\_elf!\$A\$2:\$A\$37527,SOEM\_EMPLOYED!A3,SOEM\_elf!\$F\$2:\$F\$37527,2)
  - Apply: Entire Column
- 23. Run:
  - Cell: P3
  - Command:  
=SUMIFS(SOEM\_elf!\$C\$2:\$C\$37527,SOEM\_elf!\$A\$2:\$A\$37527,SOEM\_EMPLOYED!A3,SOEM\_elf!\$F\$2:\$F\$37527,1)
  - Apply: Entire Column
- 24. Run:
  - Cell: Q3
  - Command:  
=SUMIFS(SOEM\_elf!\$C\$2:\$C\$37527,SOEM\_elf!\$A\$2:\$A\$37527,SOEM\_EMPLOYED!A3,SOEM\_elf!\$G\$2:\$G\$37527,4)
  - Apply: Entire Column
- 25. Run:
  - Cell: R3
  - Command:  
=SUMIFS(SOEM\_elf!\$C\$2:\$C\$37527,SOEM\_elf!\$A\$2:\$A\$37527,SOEM\_EMPLOYED!A3,SOEM\_elf!\$G\$2:\$G\$37527,3)
  - Apply: Entire Column
- 26. Run:
  - Cell: S3
  - Command:  
=SUMIFS(SOEM\_elf!\$C\$2:\$C\$37527,SOEM\_elf!\$A\$2:\$A\$37527,SOEM\_EMPLOYED!A3,SOEM\_elf!\$G\$2:\$G\$37527,2)
  - Apply: Entire Column
- 27. Run:
  - Cell: T3
  - Command:  
=SUMIFS(SOEM\_elf!\$C\$2:\$C\$37527,SOEM\_elf!\$A\$2:\$A\$37527,SOEM\_EMPLOYED!A3,SOEM\_elf!\$G\$2:\$G\$37527,1)
  - Apply: Entire Column
- 28. Run:
  - Cell: U3
  - Command:  
=SUMIFS(SOEM\_elf!\$C\$2:\$C\$37527,SOEM\_elf!\$A\$2:\$A\$37527,SOEM\_EMPLOYED!A3,SOEM\_elf!\$H\$2:\$H\$37527,3)

## Working Draft | Do Not Cite Without Permission

- Apply: Entire Column
- 29. Run:
  - Cell: V3
  - Command:  
=SUMIFS(SOEM\_elf!\$C\$2:\$C\$37527,SOEM\_elf!\$A\$2:\$A\$37527,SOEM\_EMPLOYED!A3,SOEM\_elf!\$H\$2:\$H\$37527,2)
  - Apply: Entire Column
- 30. Run:
  - Cell: W3
  - Command:  
=SUMIFS(SOEM\_elf!\$C\$2:\$C\$37527,SOEM\_elf!\$A\$2:\$A\$37527,SOEM\_EMPLOYED!A3,SOEM\_elf!\$H\$2:\$H\$37527,1)
  - Apply: Entire Column
- 31. Run:
  - Cell: X3
  - Command:  
=SUMIFS(SOEM\_elf!\$C\$2:\$C\$37527,SOEM\_elf!\$A\$2:\$A\$37527,SOEM\_EMPLOYED!A3,SOEM\_elf!\$I\$2:\$I\$37527,3)
  - Apply: Entire Column
- 32. Run:
  - Cell: Y3
  - Command:  
=SUMIFS(SOEM\_elf!\$C\$2:\$C\$37527,SOEM\_elf!\$A\$2:\$A\$37527,SOEM\_EMPLOYED!A3,SOEM\_elf!\$I\$2:\$I\$37527,2)
  - Apply: Entire Column
- 33. Run:
  - Cell: Z3
  - Command:  
=SUMIFS(SOEM\_elf!\$C\$2:\$C\$37527,SOEM\_elf!\$A\$2:\$A\$37527,SOEM\_EMPLOYED!A3,SOEM\_elf!\$I\$2:\$I\$37527,1)
  - Apply: Entire Column
- 34. Run:
  - Cell: AA3
  - Command: =U3/(U3+V3+W3)
  - Number: Percentage
  - Round: One Decimal Place
  - Apply: Entire Column
- 35. Run:
  - Cell: AB3
  - Command: =X3/(X3+Y3+Z3)
  - Number: Percentage
  - Round: One Decimal Place
  - Apply: Entire Column

The fourth step involves summarizing the baseline measures of occupational exposure to SARS-CoV-2 across industries in the United States before March 2020 at the level of occupations:

1. Open: "SOEM\_IND.xlsx" worksheet

## Working Draft | Do Not Cite Without Permission

2. Go to: "SOEM\_IND" tab
3. Copy: Columns A:B
4. Open: "SOEM\_IND\_EDA" worksheet
5. Go to: "SOEM\_IND\_OCCUPATIONS" tab
6. Paste: Copied Data
  - a. Columns: A:B
7. Insert: Row
  - a. Above: Row 1
8. Merge:
  - a. Cells: A1:A2
  - b. Cells: B1:B2
  - c. Cells: C1:C2
  - d. Cells: D1:D2
  - e. Cells: E1:E2
  - f. Cells: F1:F2
  - g. Cells: G1:H1
  - h. Cells: I1:L1
  - i. Cells: M1:P1
  - j. Cells: Q1:T1
  - k. Cells: U1:W1
  - l. Cells: X1:Z1
  - m. Cells: AA1:AB1
9. Label
  - a. Cell C1 – "Total Occupations (#)"
  - b. Cell D1 – "NHCW Occupations (#)"
  - c. Cell E1 – "HCW Occupations (#)"
  - d. Cell F1 – "HCW Occupations (%)"
  - e. Cell G1 – "Occupations Public Facing or Not (#)"
  - f. Cell I1 – "Occupations Indoor/Outdoor (#)"
  - g. Cell M1 – "Occupations Close Proximity (Measure 1, #)"
  - h. Cell Q1: "Occupations Close Proximity (Measure 2, #)"
  - i. Cell U1: "Occupations Combined (Proximity 1, #)"
  - j. Cell X1: "Occupations Combined (Proximity 2, #)"
  - k. Cell AA1: "Occupations Combined High (%)"
  - l. Cell G2: "2"
  - m. Cell H2: "1"
  - n. Cell I2: "4"
  - o. Cell J2: "3"
  - p. Cell K2: "2"
  - q. Cell L2: "1"
  - r. Cell M2: "4"
  - s. Cell N2: "3"
  - t. Cell O2: "2"
  - u. Cell Q2: "1"
  - v. Cell R2: "4"
  - w. Cell S2: "3"
  - x. Cell T2: "2"
  - y. Cell U2: "H"



**Working Draft | Do Not Cite Without Permission**

- z. Cell V2: "M"
  - aa. Cell W2: "L"
  - bb. Cell X2: "H"
  - cc. Cell Y2: "M"
  - dd. Cell Z2: "L"
  - ee. Cell AA2: "Combined (Proximity 1)"
  - ff. Cell AB2: "Combined (Proximity 2)"
10. Run:
- a. Cell: C3
  - b. Command: =COUNTIF(SOEM\_elf!\$A\$2:\$A\$37527,SOEM\_Occupations!A3)
  - c. Apply: Entire Column
11. Run:
- a. Cell: D3
  - b. Command:  
=COUNTIFS(SOEM\_elf!\$A\$2:\$A\$37527,SOEM\_Occupations!A3,SOEM\_elf!\$D\$2:\$D\$37527,">0")
  - c. Apply: Entire Column
12. Run:
- a. Cell: E3
  - b. Command:  
=COUNTIFS(SOEM\_elf!\$A\$2:\$A\$37527,SOEM\_Occupations!A3,SOEM\_elf!\$D\$2:\$D\$37527,"#N/A")
  - c. Apply: Entire Column
13. Run:
- a. Cell: F3
  - b. Command: =E3/C3
  - c. Number: Percentage
  - d. Round: One Decimal Place
  - e. Apply: Entire Column
14. Run:
- a. Cell: G3
  - b. Command:  
=COUNTIFS(SOEM\_elf!\$A\$2:\$A\$37527,SOEM\_Occupations!A3,SOEM\_elf!\$D\$2:\$D\$37527,2)
  - c. Apply: Entire Column
15. Run:
- a. Cell: H3
  - b. Command:  
=COUNTIFS(SOEM\_elf!\$A\$2:\$A\$37527,SOEM\_Occupations!A3,SOEM\_elf!\$D\$2:\$D\$37527,1)
  - c. Apply: Entire Column
16. Run:
- a. Cell: I3
  - b. Command:  
=COUNTIFS(SOEM\_elf!\$A\$2:\$A\$37527,SOEM\_Occupations!A3,SOEM\_elf!\$E\$2:\$E\$37527,4)
  - c. Apply: Entire Column
17. Run:

**Working Draft | Do Not Cite Without Permission**

- a. Cell: J3
  - b. Command:  
=COUNTIFS(SOEM\_elf!\$A\$2:\$A\$37527,SOEM\_Occupations!A3,SOEM\_elf!\$E\$2:\$E\$37527,3)
  - c. Apply: Entire Column
18. Run:
  - a. Cell: K3
  - b. Command:  
=COUNTIFS(SOEM\_elf!\$A\$2:\$A\$37527,SOEM\_Occupations!A3,SOEM\_elf!\$E\$2:\$E\$37527,2)
  - c. Apply: Entire Column
19. Run:
  - a. Cell: L3
  - b. Command:  
=COUNTIFS(SOEM\_elf!\$A\$2:\$A\$37527,SOEM\_Occupations!A3,SOEM\_elf!\$E\$2:\$E\$37527,1)
  - c. Apply: Entire Column
20. Run:
  - a. Cell: M3
  - b. Command:  
=COUNTIFS(SOEM\_elf!\$A\$2:\$A\$37527,SOEM\_Occupations!A3,SOEM\_elf!\$F\$2:\$F\$37527,4)
  - c. Apply: Entire Column
21. Run:
  - a. Cell: N3
  - b. Command:  
=COUNTIFS(SOEM\_elf!\$A\$2:\$A\$37527,SOEM\_Occupations!A3,SOEM\_elf!\$F\$2:\$F\$37527,3)
  - c. Apply: Entire Column
22. Run:
  - a. Cell: O3
  - b. Command:  
=COUNTIFS(SOEM\_elf!\$A\$2:\$A\$37527,SOEM\_Occupations!A3,SOEM\_elf!\$F\$2:\$F\$37527,2)
  - c. Apply: Entire Column
23. Run:
  - a. Cell: P3
  - b. Command:  
=COUNTIFS(SOEM\_elf!\$A\$2:\$A\$37527,SOEM\_Occupations!A3,SOEM\_elf!\$F\$2:\$F\$37527,1)
  - c. Apply: Entire Column
24. Run:
  - a. Cell: Q3
  - b. Command:  
=COUNTIFS(SOEM\_elf!\$A\$2:\$A\$37527,SOEM\_Occupations!A3,SOEM\_elf!\$G\$2:\$G\$37527,4)
  - c. Apply: Entire Column
25. Run:

**Working Draft | Do Not Cite Without Permission**

- a. Cell: R3
  - b. Command:  
=COUNTIFS(SOEM\_elf!\$A\$2:\$A\$37527,SOEM\_Occupations!A3,SOEM\_elf!\$G\$2:\$G\$37527,3)
  - c. Apply: Entire Column
26. Run:
- a. Cell: S3
  - b. Command:  
=COUNTIFS(SOEM\_elf!\$A\$2:\$A\$37527,SOEM\_Occupations!A3,SOEM\_elf!\$G\$2:\$G\$37527,2)
  - c. Apply: Entire Column
27. Run:
- a. Cell: T3
  - b. Command:  
=COUNTIFS(SOEM\_elf!\$A\$2:\$A\$37527,SOEM\_Occupations!A3,SOEM\_elf!\$G\$2:\$G\$37527,1)
  - c. Apply: Entire Column
28. Run:
- a. Cell: U3
  - b. Command:  
=COUNTIFS(SOEM\_elf!\$A\$2:\$A\$37527,SOEM\_Occupations!A3,SOEM\_elf!\$H\$2:\$H\$37527,3)
  - c. Apply: Entire Column
29. Run:
- a. Cell: V3
  - b. Command:  
=COUNTIFS(SOEM\_elf!\$A\$2:\$A\$37527,SOEM\_Occupations!A3,SOEM\_elf!\$H\$2:\$H\$37527,2)
  - c. Apply: Entire Column
30. Run:
- a. Cell: W3
  - b. Command:  
=COUNTIFS(SOEM\_elf!\$A\$2:\$A\$37527,SOEM\_Occupations!A3,SOEM\_elf!\$H\$2:\$H\$37527,1)
  - c. Apply: Entire Column
31. Run:
- a. Cell: X3
  - b. Command:  
=COUNTIFS(SOEM\_elf!\$A\$2:\$A\$37527,SOEM\_Occupations!A3,SOEM\_elf!\$I\$2:\$I\$37527,3)
  - c. Apply: Entire Column
32. Run:
- a. Cell: Y3
  - b. Command:  
=COUNTIFS(SOEM\_elf!\$A\$2:\$A\$37527,SOEM\_Occupations!A3,SOEM\_elf!\$I\$2:\$I\$37527,2)
  - c. Apply: Entire Column
33. Run:

## Working Draft | Do Not Cite Without Permission

- a. Cell: Z3
  - b. Command:  
=COUNTIFS(SOEM\_ELF!\$A\$2:\$A\$37527,SOEM\_Occupations!A3,SOEM\_ELF!\$I\$2:\$I\$37527,1)
  - c. Apply: Entire Column
34. Run:
- a. Cell: C3
  - b. Command: =U3/(U3+V3+W3)
  - c. Number: Percentage
  - d. Round: One Decimal Place
  - e. Apply: Entire Column
35. Run:
- a. Cell: C3
  - b. Command: =X3/(X3+Y3+Z3)
  - c. Number: Percentage
  - d. Round: One Decimal Place
  - e. Apply: Entire Column

### Data description

This dataset is provided as a single Microsoft Excel file (.xlsx). The data elements are described in the following data dictionary:

Variable	Name	Definition	Source
-	HCW	Acronym for healthcare worker	
-	NHCW	Acronym for non-healthcare worker	
-	SOEM	Acronym for SARS-CoV-2 Occupational Exposure Matrix	See: SARS-CoV-2 Occupational Exposure Matrix. Council of State and Territorial Epidemiologists.
-	SOEM-IE	Acronym for SARS-CoV-2 Occupational Exposure Matrix Industry Extension	See: SOEM Industry Extension. Mendeley Data.
-	SARS-CoV-2	Acronym for Severe Acute Respiratory Syndrome Coronavirus 2	
-	SOEM-Rated Occupation	Any occupation rated for risk of SARS-CoV-2 exposure by the SOEM	

**Working Draft | Do Not Cite Without Permission**

1	Number of Workers	Number of workers per occupation per industry (2015-2019)	Employed Labor Force Database. Centers for Disease Control and Prevention.
2	Total Employed (#)	Aggregate of Variable 1 across all occupations within an industry	
3	HCW Employed (#)	Aggregate of Variable 1 across all non-SOEM-rated occupations within an industry	
4	NHCW Employed (#)	Aggregate of Variable 1 across all SOEM-rated occupations within an industry	
5	HCW Employed (%)	Variable 3 divided by Variable 2 expressed as a percentage rounded to one decimal place	
6	Total Occupations (#)	Count of Variable 1 across all occupations within an industry	
7	HCW Occupations (#)	Count of Variable 1 across all non-SOEM-rated occupations within an industry	
8	NHCW Occupations (#)	Count of Variable 1 across all SOEM-rated occupations within an industry	
9	HCW Occupations (%)	Variable 7 divided by Variable 2 expressed as a percentage rounded to one decimal place	
10	Public Facing Exposure Level	SOEM rating for public facing exposure level per occupation	SARS-CoV-2 Occupational Exposure Matrix - Appendix A. Council of State and Territorial Epidemiologists. 2021 May 27.
11	Working Indoors	SOEM rating for working indoors exposure level per occupation	SARS-CoV-2 Occupational Exposure Matrix - Appendix A.

**Working Draft | Do Not Cite Without Permission**

	Exposure Level		Council of State and Territorial Epidemiologists. 2021 May 27.
12	Close Proximity Exposure Level (M1)	SOEM rating for close proximity exposure level (measure one) per occupation *	SARS-CoV-2 Occupational Exposure Matrix - Appendix A. Council of State and Territorial Epidemiologists. 2021 May 27.
13	Close Proximity Exposure Level (M2)	SOEM rating for close proximity exposure level (measure two) per occupation	SARS-CoV-2 Occupational Exposure Matrix - Appendix A. Council of State and Territorial Epidemiologists. 2021 May 27.
14	Combined Exposure Group (M1)	SOEM rating for combined exposure group based on Variable 9 per occupation	SARS-CoV-2 Occupational Exposure Matrix - Appendix A. Council of State and Territorial Epidemiologists. 2021 May 27.
15	Combined Exposure Group (M2)	SOEM rating for combined exposure group based on Variable 10 per occupation	SARS-CoV-2 Occupational Exposure Matrix - Appendix A. Council of State and Territorial Epidemiologists. 2021 May 27.
16	Combined Exposure Level (M1)	Variable 14 converted to numerical ( i.e., Low -> 1, Medium -> 2, High ->3)	
17	Combined Exposure Level (M2)	Variable 15 converted to numerical ( i.e., Low -> 1, Medium -> 2, High ->3)	
18	Employed Public Facing or Not (#)	Aggregate of Variable 1 across all occupations within an industry that have a specific rating for Variable 10	
19	Employed Indoor/Outdoor (#)	Aggregate of Variable 1 across all occupations within an industry that have a specific rating for Variable 11	
20	Employed Close Proximity (Measure 1, #)	Aggregate of Variable 1 across all occupations within an industry that have a specific rating for Variable 12	

**Working Draft | Do Not Cite Without Permission**

21	Employed Close Proximity (Measure 2, #)	Aggregate of Variable 1 across all occupations within an industry that have a specific rating for Variable 13
22	Employed Combined (Proximity 1, #)	Aggregate of Variable 1 across all occupations within an industry that have a specific rating for Variable 16
23	Employed Combined (Proximity 2, #)	Aggregate of Variable 1 across all occupations within an industry that have a specific rating for Variable 17
-	Employed Combined High (%)	
25	Combined (Proximity 1)	Percentage of Variable 22 (High) to Variable 22 (All) rounded to one decimal place
26	Combined (Proximity 2)	Percentage of Variable 23 (High) to Variable 23 (All) rounded to one decimal place
27	Occupations Public Facing or Not (#)	Count of Variable 1 across all occupations within an industry that have a specific rating for Variable 10
28	Occupations Indoor/Outd oor (#)	Count of Variable 1 across all occupations within an industry that have a specific rating for Variable 11
29	Occupations Close Proximity (Measure 1, #)	Count of Variable 1 across all occupations within an industry that have a

**Working Draft | Do Not Cite Without Permission**

		specific rating for Variable 12
30	Occupations Close Proximity (Measure 2, #)	Count of Variable 1 across all occupations within an industry that have a specific rating for Variable 13
31	Occupations Combined (Proximity 1, #)	Count of Variable 1 across all occupations within an industry that have a specific rating for Variable 16
32	Occupations Combined (Proximity 2, #)	Count of Variable 1 across all occupations within an industry that have a specific rating for Variable 17
-	Occupations Combined High (%)	
33	Combined (Proximity 1)	Percentage of Variable 31 (High) to Variable 31 (All) rounded to one decimal place
34	Combined (Proximity 2)	Percentage of Variable 32 (High) to Variable 32 (All) rounded to one decimal place
<p>* Note: SOEM-IE made corrections made for the following errors reported by the authors of the SARS-CoV-2 Occupational Exposure Matrix: 1) 4850: Sales representatives, wholesale and manufacturing; 2) 5700: Secretaries and administrative assistants. Both values changed from 3 to 2.</p> <p>** Note: SOEM-IE rounded fractions to eleven decimal places.</p>		

**Value of the data**

- These data are useful because they provide new insights on occupational exposure to SARS-CoV-2 across industries in the United States before March 2020. They are particularly useful for exploring the distribution of the baseline measures on occupational exposure to SARS-CoV-2 across industries in the United States at the levels of workers and occupations.



- The direct beneficiaries of these data are policymakers, practitioners, and researchers engaged in occupational health and safety activities related to SARS-CoV-2. The indirect beneficiaries of these data are employers and workers who might benefit from a better understanding of occupational exposure to SARS-CoV-2 in the United States before March 2020.
- These data can be used: 1) to generate new hypotheses for research; 2) to inform policy, and prevention activities; 2) to introduce new concepts in public discourse about SARS-CoV-2; 3) to provide summary measures for program evaluation and monitoring.

### **Declaration of interests**

The author declares no known competing financial interests or personal relationships that could appear to influence his production of this dataset.

### **References**

Occupational Health Subcommittee Epidemiological Classification of COVID-19 Work-Relatedness and Documentation of Public-Facing Occupations. Council for State and Territorial Epidemiologists Website. Accessed on September 5, 2021. URL=[https://www.cste.org/resource/resmgr/occupationalhealth/publications/OH\\_Docs.zip](https://www.cste.org/resource/resmgr/occupationalhealth/publications/OH_Docs.zip).

About O\*NET. Occupational Information Network Website. Accessed on September 5, 2021. URL=<https://www.onetcenter.org/overview.html>.

Technical information. Employed Labor Force (ELF) query system website. Last reviewed on October 2, 2020. Accessed on September 5, 2021. URL=[https://wwwn.cdc.gov/Wisards/cps/cps\\_techinfo.aspx#tic2](https://wwwn.cdc.gov/Wisards/cps/cps_techinfo.aspx#tic2).

Walsh M. Measuring non-healthcare occupational exposure to SARS-CoV-2 across occupational groups in the United States: Version 2. *Protocols.io*. 2021. Doi=[dx.doi.org/10.17504/protocols.io.bw9gph3w](https://doi.org/10.17504/protocols.io.bw9gph3w).

Walsh, Michael.. Non-healthcare occupational exposure to SARS-CoV-2 across industries in the United States before March 2020: a dataset generating protocol (1.0). *Zenodo*. 2021. Doi=[10.5281/zenodo.5513598](https://doi.org/10.5281/zenodo.5513598).

Walsh, Michael. (2021). Non-healthcare occupational exposure to SARS-CoV-2 across industries in the United States before March 2020: a dataset (1.0). *Zenodo*. 2021. Accessed on September 17, 2021. URL=[https://zenodo.org/record/5513598/files/SOEM\\_IE.xlsx?download=1](https://zenodo.org/record/5513598/files/SOEM_IE.xlsx?download=1).