

Employment Data in Washington's Rapid Health Information Network (RHINO)

A comparison of two methods for coding employer and occupation free text data collected among Emergency Department visits for Heat Related Illness, 2020-2022.

SHARP TECHNICAL REPORT 80-22-2024

JANUARY 2024

Contents

EXECUTIVE SUMMARY 1

 Key Findings 1

 Conclusion..... 2

INTRODUCTION..... 3

METHODS..... 4

 Emergency department data 4

 Employment Security Department data 5

 Record linkage..... 8

 Free text coding 9

 Data analysis 9

RESULTS 11

 Match results 12

 Comparison of IO Codes Assigned: ESD vs. Text-based codes..... 14

 Industry codes..... 14

 Occupation Codes 20

DISCUSSION..... 24

 Limitations 24

 Conclusion..... 25

Appendix 26

Table of Tables

Table 1. Examples of ESD establishment data.....	7
Table 2. Examples of records linked at each step in the matching algorithm.....	10
Table 3. Employer name and occupation free text data in RHINO ED visit data by patient age.....	11
Table 4. Expected inclusion in ESD data, based on manual review of RHINO ED data.	11
Table 5. Number of RHINO ED records linked at each step in the linkage process.....	12
Table 6. Number of matches identified and confirmed through manual review by linkage step.....	12
Table 7. Agreement in industry classification between ESD linkage and text-based assignment	15
Table 8. Distribution of text-based codes by ESD-based NAICS codes.....	16
Table 9. Industry Sector by coding method.....	17
Table 10. HRI record rates, counts, and Prevention Index rank by industry and method of coding	18
Table 11. Agreement in occupation classification between ESD linkage and text-based assignment.....	20
Table 12. Occupation by coding method.....	21
Table 13. HRI record rates, counts, and Prevention Index rank by occupation and method of coding.....	22

Table of Figures

Figure 1. Record linkage between RHINO ED visit data and ESD employment data.....	4
Figure 2. Flowchart of HRI ED records and methods used to assign industry and occupation codes.....	13
Figure 3. Prevention Index rank by industry and method of assigning industry code.	19
Figure 4. Prevention Index rank by occupation and method of assigning industry code.	23

Abbreviations

ED	Emergency Department
ESD	Washington State Employment Security Department
HRI	Heat-related illness
IO	Industry and Occupation
L&I	Washington State Department of Labor & Industries
NAICS	North American Industry Classification System
NIOCCS	NIOSH Industry and Occupation Computerized Coding System
NIOSH	National Institute for Occupational Safety and Health
PI	Prevention Index
RHINO	Washington State Department of Health’s Rapid Health Information NetwOrk
SHARP	L&I’s Safety and Health Assessment and Research for Prevention Program
SOC	Standard Occupation Classification system
UI	Unemployment Insurance

Acknowledgements

This work was supported in part by Cooperative Agreement OH008487 from the National Institute for Occupational Safety and Health (NIOSH). The contents are solely the responsibility of the authors and do not necessarily represent the official views of NIOSH and the Center for Disease Control and Prevention.

Many thanks to our colleagues at the Washington State Department of Health’s Rapid Health Information NetwOrk who helped us obtain and understand the ED visit data.

EXECUTIVE SUMMARY

The Washington State Department of Health's Rapid Health Information NetWOrk (RHINO) collects near real-time non-federal emergency department (ED) visit data. The ED visit data has the potential to be a useful source for occupational injury and illness surveillance, especially if the occupation and employer name information, recorded in the ED visit data as free text fields, can be coded according to standard industry and occupation classification systems.

We sought to assign industry codes to ED visits for heat-related illness (HRI) by linking the RHINO ED data to Washington Employment Security (ESD) data, and then compare those codes to the industry codes assigned based solely on the free text employer and occupation data captured in the ED data. Here we describe the work data reported among ED visit data, the protocol developed to link the records, the success of the linkage, and the results of the two coding methods.

Key Findings

- Employer name was reported more frequently than occupation in the RHINO ED visit data.
 - Among patients age 20-65, 23% of records included employer data, 11% reported occupation data, and 6% reported both employer and occupation data.
- The majority of records with employer data were able to be assigned an industry code by linking to ESD data, based on similarity of worker and employer names.
 - 80% of records included in the linkage attempt were assigned an industry code through the ESD linkage.
 - 50% of records included in the linkage attempt linked to ESD records based on the most stringent linkage criteria – requiring high similarity on worker and employer names.
- Manual record review was a crucial step in the ESD linkage process.
 - 19% of all matches were determined to be false matches (mismatched worker or employer name) based on manual record review.
- Assigning industry based on the ESD linkage resulted in more coded records (fewer records considered non-classifiable) compared with coding based on free text alone. The increase was greater for industry than occupation.
 - Industry: 85% classified via ESD linkage vs 75% classified based on free text
 - Occupation: 41% classified via ESD linkage vs 36% classified based on free text
- Agreement between coding methods varied
 - By industry:
 - High (>80% of records in agreement) for Educational Services and Public Administration
 - Low (<45% of records in agreement) for Agriculture, Forestry, Fishing and Hunting; Administrative and Support and Waste Management and Remediation Services; and Wholesale Trade
 - And by occupation:
 - High (>80%) for Protective Service Occupations; Food preparation and serving related occupations; and Management, Business, Science (other than Healthcare), and Arts Occupations

- Low (<45%): Construction and Extraction Occupations; Building, Grounds Cleaning, and Maintenance Occupations; and Farming, Fishing, and Forestry Occupations
- Prevention Index (PI) rankings differed by coding method.
 - The ESD linkage resulted in higher PI rankings for six industry sectors, including three sectors characterized by substantial outdoor work activity: Agriculture, Forestry, Fishing and Hunting, Construction, and Administrative and Support and Waste Management and Remediation Services
 - The ESD linkage resulted in higher PI rankings for two occupation major groups: Farming, Fishing, and Forestry occupations (ranked 3rd based on ESD codes vs 9th based on text only codes), and Construction and Extraction occupations (1st based on ESD codes vs 2nd based on text only codes)

Conclusion

Because many of the industries and occupations involving outdoor work (and thus potentially at increased risk of HRI), are disproportionately missed by relying on free text coding alone, the linkage attempt to ESD is recommended. Further refinement of the linking algorithm may result in an increase in the number of records linked and a decrease in the percent of false matches, potentially reducing the amount of manual record review needed to confirm true matches and increasing the efficiency of the linkage process.

INTRODUCTION

The Washington State Department of Health's Rapid Health Information NetwOrk (RHINO) collects near real-time healthcare visit data from non-federal hospitals and clinics, and is a key source of emergency department (ED) visit data. The ED visit data has the potential to be a useful source for occupational injury and illness surveillance, especially if the occupation and employer name information, recorded in the RHINO ED visit data as free text fields, can be coded according to standard industry and occupation classification systems.

The CDC's National Institute for Occupational Safety and Health (NIOSH) developed a program to assign industry and occupation codes to free text descriptions of each data element. The NIOSH Industry and Occupation Computerized Classification System (NIOCCS), which utilizes machine learning, is able to assign industry codes when employer names are provided instead of descriptions of industry, although generally it is more successful for large employers or when the employer name includes a key word that describes the industry.

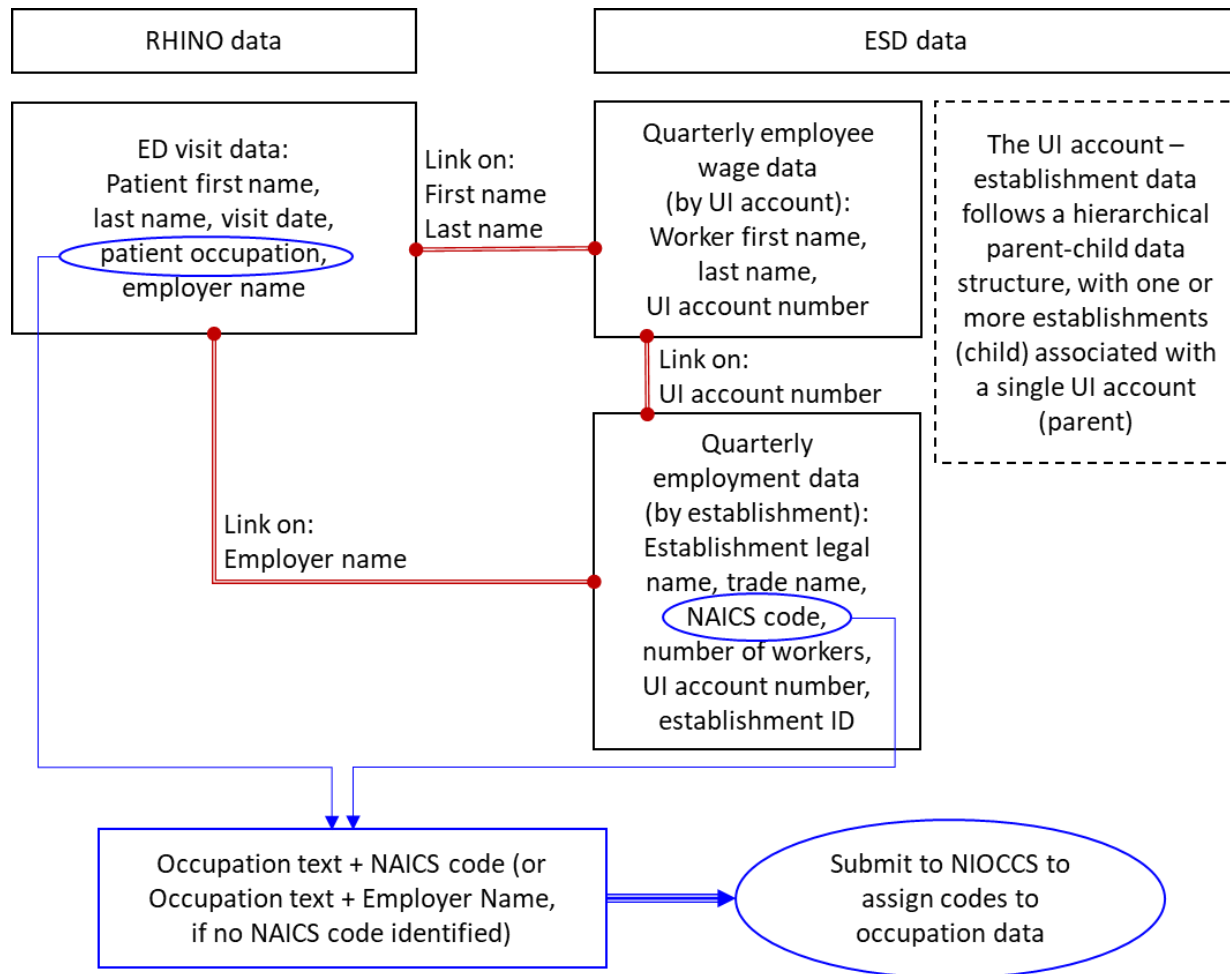
Washington's Employment Security Department (ESD) maintains data on all employers covered by the state's unemployment insurance laws. The data include legal and trade names of business establishments, North American Industry Classification System (NAICS) codes assigned to establishments, and the first and last names of individual workers.

We sought to assign industry codes to ED visits for heat-related illness (HRI) by linking the RHINO ED visit data to the ESD data, and then compare those codes to the codes assigned by NIOCCS based solely on the free text data. Here we describe the protocol developed to link the records, the success of the linkage, and compare the results of the two coding methods.

METHODS

To classify industry and occupation among ED visits for HRI, RHINO ED visit data was first linked to ESD Unemployment Insurance (UI) data to identify the industry code associated with the patient's employer. Then, we used the NIOSH Industry and Occupation Computerized Classification System (NIOCCS) to assign occupation codes to all records, and industry codes to records unable to be linked to ESD data (Figure 1).

Figure 1. Record linkage between RHINO ED visit data and ESD employment data.



Emergency department data

Emergency department visits for HRI that occurred between May 1 and September 30, in the years 2020 – 2022 were identified using the query definition¹ developed by the National Syndromic Surveillance Program (NSSP) and state partners (see Appendix), and extracted from the NSSP ESSENCE platform

¹ The query definition is inclusive of all HRI and not limited to work-related HRI.

(Electronic Surveillance System for the Early Notification of Community-Based Epidemics) and then linked to the WA DOH RHINO database to collect patient occupation and employer data that are not included in ESSENCE. Data were shared with the LNI SHARP program in May 2023. The RHINO ED visit data included two employment variables: occupation and employer name, each reported as a separate, free text field.

The complete list of responses recorded in the occupation and employer fields were reviewed to identify responses that indicated something other than an occupation, industry, or employer (e.g., “disabled”, “unemployed”, “unknown”, values that were numbers or dates). These responses were excluded from attempts to assign industry and occupation codes.

Employer names were cleaned by removing punctuation and special characters, store numbers, business structure nomenclature (“LLC”, “INC”), and “THE” at the beginning of text strings. Ampersands, “NORTHWEST”, and “SERVICES” were replaced with “AND”, “NW”, and “SVC”, respectively. Lastly, all spaces were removed.

Patient name were cleaned by removing punctuation, special characters, and the suffixes JR, SR, III, IV.

Employment Security Department data

With few exceptions, nearly all employers operating in Washington, and all employees working for those employers are captured in data files maintained by ESD. Two data files maintained by ESD were used in the record linkage:

The first is quarterly employee wage data, which include worker first and last name, and the employer unemployment insurance account number (UI account number). Quarterly employee wage data is reported by employers for an entire UI account (vs. the more granular establishment unit).

The second is quarterly establishment employment data, which include the establishment identification number, legal name, trade name, NAICS code, the number of workers employed at the establishment, and the UI account number. Note that multiple establishments may be associated with a single UI account.

While the quarterly establishment data include federal agencies (e.g., Department of Defense, National Park Service, United States Postal Service, Veterans Administration Hospitals, various military accounts), individual workers for these agencies are not reported, and are thus not included in the employee quarterly wage records. Self-employed workers are not included in either the establishment data or the employee wage records. Washington residents who work in another state are not included in the Washington employment data (and Washington residents who sought care in an ED in another state are not included in the Washington ED data).

We defined employer name as trade name, unless trade name was blank, in which case legal name was used. All NAICS codes were converted to the 2017 version of NAICS, converting 2022 NAICS codes when necessary, using the Census concordance reference. After cleaning employer names according to the same procedures used for the RHINO ED data, we summed the number of workers by employer name and NAICS code. For each unique employer name, we then retained the NAICS code associated with the greatest number of workers. This decision rule, to retain the NAICS code associated with the greatest number of workers, was necessary to ensure that only one NAICS code was assigned to each ED record, and avoid assigning multiple NAICS codes per ED record. See Table 1 for examples of establishment

records included vs. excluded in linkage attempts to RHINO ED data, based on employer name and number of workers. For example, in Table 1, Account 1111 contains three records, two of which share the same name. Applying the study's decision resulted in retaining the record with the greatest employment (the first record, having 24 workers), and excluding the record with the same employer name and fewer workers (second record, having 14 workers). Because the third record in account 1111 has a unique employer name, that record was also retained for the linkage attempt. Account 5555 contains five records, but only four unique employer names. The second record was excluded from the linkage attempt because it shares the same employer name as the first record (HOSPITALMEDICALCENTER), and has fewer workers than the first record (6 vs. 1037 workers). The remaining three records in account 5555 have unique employer names (HOSPITALCLINIC, HOSPITALCLINICEASTSIDE, CAREHOUSE) and, thus, all are included in the linkage attempt.

ESD employer names and worker names were cleaned according to the same procedures used for the RHINO ED data. RHINO records indicating employment with an entity not included in the ESD employee wage file (e.g., military, self-employed) were flagged and excluded from attempts to match to ESD employment data.

Table 1. Examples of ESD establishment data. For each employer name (as it appears post cleaning), the NAICS code associated with the greatest number of workers was included in the linkage attempt to the RHINO ED data. Bold font indicates ESD records included in the linkage attempt. Synthetic data presented.

Account	Employer name	NAICS code	NAICS description	Number of workers	Record linkage
1111	DAIRYCO	311511	Fluid Milk Manufacturing	24	Included
1111	DAIRYCO	445299	All Other Specialty Food Stores	14	Excluded
1111	WASHINGTONFARMFOOD	311511	Fluid Milk Manufacturing	46	Included
2222	SILVIASPIZZA	722513	Limited-Service Restaurants	99	Included
2222	SILVIASPIZZA	551114	Corporate, Subsidiary, and Regional Managing Offices	1	Excluded
3333	COMPUTERCOMPANYNAME	423430	Computer and [...] Software Merchant Wholesalers	97	Included
3333	CCN	423430	Computer and [...]Software Merchant Wholesalers	3	Included
4444	INSURANCEGROUP	524126	Direct Property and Casualty Insurance Carriers	26	Included
4444	INSURANCEGROUP	524113	Direct Life Insurance Carriers	23	Excluded
4444	INSURANCENEWWORLDLIFE	524126	Direct Property and Casualty Insurance Carriers	50	Included
5555	HOSPITALMEDICALCENTER	622110	General Medical and Surgical Hospitals	1037	Included
5555	HOSPITALMEDICALCENTER	621111	Offices of Physicians (except Mental Health Specialists)	6	Excluded
5555	HOSPITALCLINIC	621111	Offices of Physicians (except Mental Health Specialists)	79	Included
5555	HOSPITALCLINICEASTSIDE	621111	Offices of Physicians (except Mental Health Specialists)	8	Included
5555	CAREHOUSE	623110	Nursing Care Facilities (Skilled Nursing Facilities)	13	Included

Record linkage

RHINO ED records with work responses recorded in the employer name field (excluding responses such as 'unemployed', 'unknown') were linked to ESD records through a multiple-step process. To allow for fuzzy matching (linking records with similar but not identical responses), SAS's `compge` function was used, which calculates the generalized edit distance (GED) between two strings, with lower scores indicating greater similarity between the two strings. The multi-step process that follows was developed to maximize true links and minimize false links (links between mismatched records). After each step, records were reviewed to identify false links. Records left unlinked after a particular step were included in the next step, in an attempt to eventually identify a link (i.e., records unlinked after step 1 advanced to step 2; records unlinked after step 2 advanced to step 3, etc.).

In step 1, records were linked based on first three letters of person first name or first initial if only an initial was provided, exact match on last name, and first initial of employer name, plus a similarity in employer names as measured by a GED score no greater than 500 (in each step of the multi-step process, GED cutoff scores were determined by manually reviewing records resulting from a range of cutoff scores.) Because this approach potentially links one RHINO ED record to multiple employment records, for each RHINO ED record, we retained the link to the ESD record with most similar employer name (lowest GED score). When multiple ESD records had identical GED scores, we used employment as the tie breaker, retaining the link to the ESD record with the greatest employment.

In step 2, we allowed for slight differences in both person and employer names, linking remaining RHINO ED records (unlinked after step 1) to ESD records on the first initial of person first name, and requiring similarity between person last names (GED score no greater than 200), and similarity in employer names (GED score under 350). When RHINO ED records linked to multiple ESD records, we retained the link with the most similar last name and the most similar employer name. Again, employment was used as a tie-breaker, retaining the link to the ESD record with the greatest employment.

In step 3, we linked on employer name only, requiring a match on the first letter of the employer name and a similarity in employer names (GED scores no greater than 350). In this step, the ESD data was limited to employers reporting an average of 0.5 workers per quarter during the study period. Restricting the linkage by employment removed 182,648 of ESD records, and reduced the potential for false matches to establishments with scant employment. Among records linked on employer name, for each RHINO ED record we first retained the link to the ESD record with the most similar employer name and greatest employment. When no link reflected both the most similar employer name and greatest employment, we retained the link to the most similar employer name, among links with GED scores no greater than 50.

In step 4, we linked on worker name only, requiring the first three letters of the first name to match (or first initial if only initial was provided), similarity in last name (GED score no greater than 200), and employment in year-quarter of the ED visit. This step allowed us to link records where the employer name was recorded under the occupation field in the RHINO ED data, and where there were large differences in employer name between the two data sources (e.g., Monster vs. Hansen Beverage). Although we initially allowed slight differences in last name, final linkages were limited to records with identical matches on first and last person names. Among these, we prioritized links to ESD records with the greatest employment, followed by links among records that matched on the first two letters of the

employer name. All links resulting from this step were reviewed to ensure that the ESD employer information was appropriate given the information reported in the RHINO ED record.

In the final step, step 5, we linked remaining RHINO ED records to RHINO ED records already linked to ESD data, based on the cleaned employer name data. Links were required to have a GED score for employer name similarity of 200 or less, and were manually reviewed to identify false matches.

Table 2 includes examples of records matched at each step.

Free text coding

Two files were prepared for inputting into NIOCCS. The first file, consisting of RHINO ED records linked to ESD data, included the occupation free text data recorded in RHINO and the NAICS industry code identified from the linkage to the ESD data. The second file included all RHINO ED records having employment data, with the employer and the occupation free text data entered into NIOCCS as industry and occupation data, respectively. This allowed for a comparison between the industry codes assigned via the ESD linkage and those assigned by NIOCCS based on the information recorded in the text fields. Data were uploaded to NIOCCS on June 27, 2023.

Data analysis

Industry codes were aggregated to the NAICS sector level for analysis, with the exception of Information, Finance and Insurance, Real Estate, Rental and Leasing, Professional, Scientific, and Technical Services, and Management of Companies and Enterprises which was grouped into a single super-sector.

Generally, occupation codes were aggregated to the SOC major group level. Three super groups were created for Management, Business, Science, and Arts occupations (SOC codes 11-0000 – 27-0000); Healthcare occupations (SOC codes 29-0000 – 31-0000); and Sales and Office occupations (41-0000 – 43-0000).

We calculated the percent of records assigned the same codes (industry sector or occupation major group) by both coding systems, the percent coded a different code, and the percent classified as having Insufficient Information.

For each coding method, we estimated rates of ED visits for HRI by industry and occupation, using 2021 5-yr ACS employment estimates for Washington, and calculated the Prevention Index (PI)², defined as the average of the visit count rank and visit rate rank for each industry and occupation. We then compared the PI rankings produced by each coding method.

² Prevention Index = (visit count rank + visit rate rank)/2. Prevention Index estimates are then ranked.

Table 2. Examples of records linked at each step in the matching algorithm. Synthetic data presented.

Link Step	Worker name, RHINO ED data	Worker name, ESD data	Occupation, RHINO ED data	Employer name, RHINO ED data	Employer name, ESD data	NAICS description, ESD data
1	JOHN BROWN	JOHN BROWN		TOGLINS	TOGNLINS	Industrial Building Construction
2	JANE MILLNER	J MILNER		VONCOY	VONCOYSUPPLY	Roofing, Siding, and Insulation Material Merchant Wholesalers
3				SCOTTPJOHNSON	SCOTTPJOHNSON	Offices of Real Estate Agents and Brokers
4	TYLER DAVIS	TYLER DAVIS		PEBBLEROOFINGSUPPLY	PRSACQUISITION	Other Building Material Dealers
5			ROOFING TECH	WHITTLEANDSNOW	WETTLEANDSNOW ^a	Roofing Contractors

^aRHINO ED data employer name linked in preceding step to ESD data with employer name of WETTLEANDSNOWROOFING

RESULTS

Of the 5446 RHINO HRI ED records, 1091 (20%) had employer name or occupation data reported in the free text fields. Twice as many records included employer name data compared to records with occupation data (16% vs. 8%), and only 4% of records reported both employer name and occupation data (Table 3). The percent of records with employer or occupation data was higher among working age adults (29% of records among patients age 20-64 included employer or occupation data).

Table 3. Employer name and occupation free text data recorded in RHINO ED visit data by patient age, among the 5446 HRI ED visits.

Work data reported in ED visit data	≤19 yrs (n=620)	20-64 yrs (n=3265)	≥65 yrs (n=1561)	Total (n=5446)
Employer name	41 (7%)	774 (24%)	43 (3%)	858 (16%)
Occupation	s	363 (11%)	s	443 (8%)
Both employer name and occupation	s	194 (6%)	s	210 (4%)
Employer name and/or occupation	44 (7%)	943 (29%)	104 (7%)	1091 (20%)
Neither	576 (93%)	2322 (71%)	1457 (93%)	4355 (80%)

s=suppressed due to small numbers.

Of the 858 RHINO ED records with employer data reported, 8% were not expected to be captured in the ESD worker data file, given the type of employer reported (Table 4), leaving 789 RHINO ED records included in the linkage attempt with ESD data.

Table 4. Expected inclusion in ESD data, based on manual review of RHINO ED employer data.

Employer category	Records (%)
Potentially in ESD worker and establishment data	789 (92%)
Potentially in ESD establishment data, excluded from ESD worker data	
US Postal Service or Military	35 (4%)
Excluded from both ESD establishment data and ESD worker data	
Self-employed	34 (4%)
TOTAL RHINO ED records with work info reported	858 (100%)

Match results

Of the 789 RHINO ED records included in the linkage attempt with ESD data, 629 records (80%) were linked to ESD records (Table 5). The remaining 20% (160 records) were dependent on NIOCCS for assignment of an industry code.

Step 1 accounted for the greatest number of linked records (linking half of all records attempted) and the greatest portion of links confirmed through manual review as true matches (95% confirmed as true matches). The lowest percent of links confirmed as true matches resulted from step 4 (attempted match on worker name alone), with 44% of matches confirmed as true (Table 6).

Figure 2 presents the HRI ED records and the method used to assign industry and occupation codes.

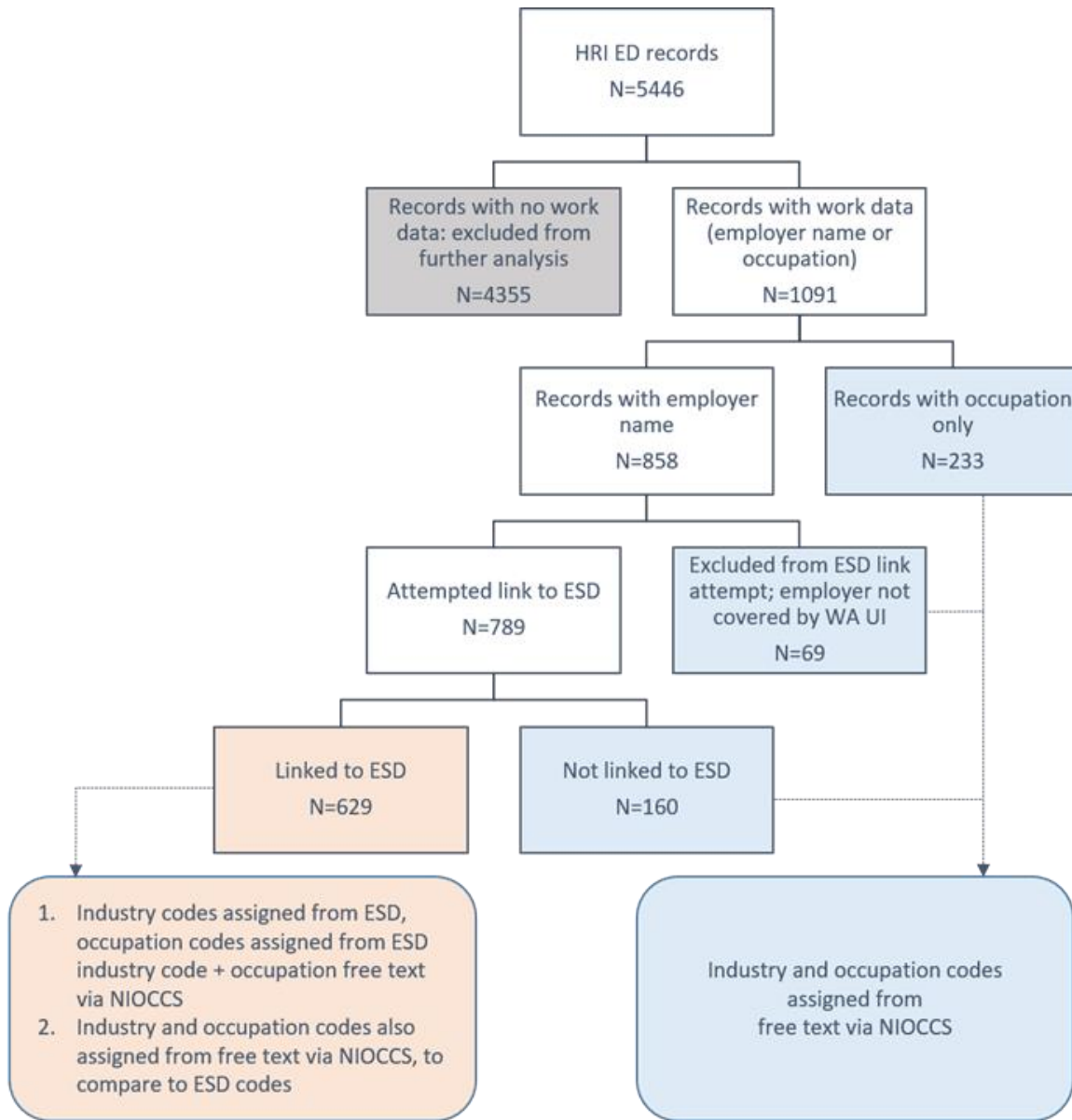
Table 5. Number of RHINO ED records linked at each step in the linkage process.

Link results	Records
Total records linked to ESD	629 (80%)
Step 1	395 (50%)
Step 2	54 (7%)
Step 3	101 (13%)
Step 4	67 (8%)
Step 5	12 (2%)
Unlinked	160 (20%)
Total records attempted	789 (100%)

Table 6. Number of matches identified in the linking process and confirmed through manual review by linkage step.

Step	Description	Matches identified	False matches	True matches	% True
1	Exact match on person name, fuzzy match on employer name	417	22	395	95%
2	Fuzzy match on person name, fuzzy employer name	64	10	54	84%
3	Match on employer name only (fuzzy match)	125	24	101	81%
4	Match on person name only (exact match)	154	87	67	44%
5	Match on RHINO employer names already linked to ESD data (fuzzy match)	16	4	12	75%

Figure 2. Flowchart of HRI ED records and methods used to assign industry and occupation codes



Comparison of IO Codes Assigned: ESD vs. Text-based codes

Industry codes

Comparing industry sector codes assigned through the ESD linkage to those assigned by NIOCCS based on the employer and occupation free text fields, among the 629 records linked to ESD, the percent of records assigned the same codes differed by industry. Agreement was greatest among utilities, where 100% of records coded through the ESD linkage were assigned the same sector code by NIOCCS based on the employer and occupation free text data, although the total number of records was small (Table 7). Wholesale trade saw the least amount of agreement, where only 16% of records coded by ESD were assigned the same sector code by NIOCCS. There were five groups (four sectors plus one super-sector) where 50% or less of the records were assigned the same codes by both methods.

Overall, 17% of records assigned a NAICS code through the ESD linkage lacked sufficient information in the text fields for NIOCCS to assign an industry code. The percent of records deemed by NIOCCS to have insufficient information varied by sector, and was greatest for records coded by ESD as administrative and support and waste management and remediation services, where more than one-third were unable to be assigned an industry code by NIOCCS.

Table 8 displays the distribution of NIOCCS-assigned codes by ESD-based codes among the 629 records linked to ESD. Among the 37 records classified through the ESD linkage as agriculture, forestry, fishing and hunting, 43% were assigned the same industry sector code based on the information recorded in the free text employer and occupation data, while 8% were classified as manufacturing, 8% as retail trade, and 14% as some other sector. Twenty-seven percent of records coded by ESD as agriculture, forestry, fishing and hunting were not classifiable by NIOCCS based on the text data reported. Among the 25 records coded by ESD as wholesale trade, more were coded by NIOCCS as manufacturing (25%) than wholesale trade.

Slight differences were observed in the distribution of industry codes by method of coding for all 1091 ED records with employment data (ESD NAICS for 629 linked records plus text-based codes for remaining 462 records vs. 1091 records coded solely from employer and occupation free text data) (Table 9). Based on the ESD plus text method of assigning codes, the greatest number of records occurred among the construction industry. However, based on text-only codes, construction ranked fourth while the manufacturing sector accounted for the most records.

For eight sectors plus military, the two methods resulted in the most similar estimates ($\pm 9\%$). For six sectors, the ESD plus text method resulted in record numbers that were 16-170% greater than the number of records produced by the text-only method. The text-only method resulted in 67% more uncoded records (presented in the table 9 as “Insufficient information”).

Table 7. Agreement in industry classification between ESD-based coding method and text-based assignment, among RHINO ED records linked to ESD (n=629).

Industry sector (NAICS code)	Number of records, based on ESD coding	% assigned same sector based on free text	% assigned different Sector based on free text	% classified as insufficient information based on free text
Utilities (22)	s	100%	0%	0%
Educational services (61)	27	89%	11%	0%
Public administration (92)	33	85%	12%	3%
Arts, entertainment, and recreation (71)	s	75%	8%	17%
Retail trade (44-45)	77	73%	14%	13%
Accommodation and food services (72)	58	66%	19%	16%
Health care and social assistance (62)	45	67%	20%	13%
Transportation and warehousing (48-49)	38	61%	24%	16%
Manufacturing (31-33)	61	57%	28%	15%
Construction (23)	94	52%	29%	19%
Other services (except public administration) (81)	18	50%	28%	22%
Information, finance, insurance, real estate, professional, management services (51-55)	47	49%	28%	23%
Agriculture, forestry, fishing and hunting (11)	37	43%	30%	27%
Admin and support and waste mgmt svc (56)	54	41%	26%	33%
Wholesale trade (42)	25	16%	60%	24%
Total	629	59%	24%	17%

s=Suppressed due to small numbers.

Table 8. Distribution of text-based NAICS codes by ESD-based NAICS codes (column percentages), among RHINO ED records linked to ESD (n=629).

Text-based codes	ESD-based codes														
	11 (n=37)	22 (n=s)	23 (n=94)	31-33 (n=61)	42 (n=25)	44-45 (n=77)	48-49 (n=38)	51-55 (n=47)	56 (n=54)	61 (n=27)	62 (n=45)	71 (n=s)	72 (n=58)	81 (n=18)	92 (n=33)
11	43%	0%	0%	8%	4%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%
21	0%	0%	1%	3%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
22	0%	100%	2%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	3%
23	0%	0%	52%	2%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	3%
31-33	8%	0%	7%	57%	24%	9%	5%	2%	4%	0%	0%	0%	9%	6%	0%
42	3%	0%	1%	0%	16%	1%	0%	2%	0%	0%	0%	0%	0%	0%	0%
44-45	8%	0%	1%	8%	16%	73%	0%	4%	4%	0%	0%	0%	2%	6%	0%
48-49	0%	0%	0%	2%	4%	0%	61%	0%	2%	4%	0%	0%	5%	6%	0%
51-55	5%	0%	5%	0%	0%	0%	5%	49%	6%	0%	2%	0%	0%	0%	0%
56	0%	0%	1%	0%	8%	0%	0%	0%	41%	0%	0%	0%	0%	0%	0%
61	0%	0%	0%	0%	0%	0%	0%	0%	0%	89%	4%	0%	0%	11%	3%
62	3%	0%	3%	0%	4%	1%	3%	11%	11%	0%	67%	0%	0%	0%	0%
71	0%	0%	1%	0%	0%	0%	0%	0%	0%	4%	0%	75%	2%	0%	3%
72	3%	0%	1%	5%	0%	0%	5%	6%	0%	0%	0%	0%	66%	0%	0%
81	0%	0%	0%	0%	0%	0%	0%	2%	0%	0%	9%	0%	2%	50%	0%
92	0%	0%	4%	0%	0%	0%	5%	0%	0%	4%	2%	8%	0%	0%	85%
98^a	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	2%	0%	0%	0%	0%
99	27%	0%	19%	15%	24%	13%	16%	23%	33%	0%	13%	17%	16%	22%	3%

^aNIOCCS code for retired and non-paid workers.

s=Suppressed due to small numbers.

Table 9. Industry Sector by coding method, among records with employer or occupation data reported in RHINO ED data (n=1091).

Industry Sector	ESD plus text codes ^a n (%)	Text-based codes n (%)	Difference in record counts ^b
Construction (23)	126 (12%)	84 (8%)	+50%
Transportation and warehousing (48-49)	96 (9%)	89 (8%)	+8%
Retail trade (44-45)	94 (9%)	92 (8%)	+2%
Manufacturing (31-33)	86 (8%)	94 (9%)	-9%
Accommodation and food services (72)	72 (7%)	62 (6%)	+16%
Information, finance, real estate, professional, mgmt svc (51-55)	71 (7%)	60 (5%)	+18%
Admin, support and waste mgmt and remediation svc (56)	70 (6%)	41 (4%)	+71%
Health care and social assistance (62)	64 (6%)	67 (6%)	-4%
Public administration (92)	57 (5%)	61 (6%)	-7%
Agriculture, forestry, fishing and hunting (11)	50 (5%)	36 (3%)	+39%
Educational services (61)	39 (4%)	41 (4%)	-5%
Other services (except public admin) (81)	38 (3%)	35 (3%)	+9%
Wholesale trade (42)	27 (2%)	10 (1%)	+170%
Arts, entertainment, and recreation (71)	19 (2%)	20 (2%)	-5%
Utilities (22)	s	s	-50%
Mining, quarrying, oil & gas extraction (21)	s	s	-100%
Military	12 (1%)	12 (1%)	0%
Non-worker (98)	s	s	-20%
Insufficient information (99)	163 (15%)	273 (25%)	-40%
TOTAL	1091 (100%)	1091 (100%)	

^aCodes based on linkage to ESD data for 629 records, NIOCCS-assigned codes based on text to remaining 462 records.

^bThe percent difference in the number of records coded from ESD plus text compared to the number of records coded from text alone.

s=Suppressed due to small numbers.

Based on the ESD plus text method of coding, agriculture, forestry, fishing and hunting had the highest rate of ED visits for HRI by industry, followed closely by transportation and warehousing (Table 10). The rate rankings for those two sectors were reversed, based on the text-only method of coding. Both methods ranked transportation and warehousing highest on the Prevention Index. The greatest difference in Prevention Index ranking was found for administrative and support, waste management and remediation services and wholesale trade; both sectors ranked 5 places higher based on ESD plus text codes compared with text-only codes. Manufacturing ranked four places higher based on text-only codes. Figure 3 shows the industry sectors by Prevention Index rank for each coding method.

Table 10. Comparison of industry-specific HRI record rates, counts, and Prevention Index rank by method of coding^a.

Industry Sector	ESD Rate ^b	ESD Rate Rank	Text Rate ^b	Text Rate Rank	ESD Count	ESD Count Rank	Text Count	Text Count Rank	ESD PI ^c Rank	Text PI ^c Rank	PI rank diff ^d
Agriculture, forestry, fishing and hunting (11)	58.5	1	42.1	2	50	10	36	11	4	6	2
Transportation and warehousing (48-49)	53.5	2	49.6	1	96	2	89	3	1	1	0
Admin, support and waste mgmt and remediation svc (56)	49.0	3	29.1	5	69	7	41	9	3	8	5
Construction (23)	47.7	4	31.8	4	126	1	84	4	2	3	1
Public administration (92)	31.0	5	33.2	3	57	9	61	7	7	4	3
Accommodation and food services (72)	29.6	6	25.5	7	72	5	62	6	5	7	2
Wholesale trade (42)	27.7	7	10.3	13	27	13	10	14	9	14	5
Manufacturing (31-33)	24.7	8	27.0	6	86	4	94	1	6	2	4
Arts, entertainment, and recreation (71)	23.7	9	25.0	8	19	14	20	13	13	10	3
Other services (except public admin) (81)	23.0	10	20.6	10	38	11	34	12	12	11	1
Retail trade (44-45)	21.9	11	21.4	9	94	3	92	2	8	5	3
Health care and social assistance (62)	12.9	12	13.7	11	63	8	67	5	10	9	1
Educational services (61)	12.3	13	13.0	12	38	11	40	10	14	12	2
Information, finance, real estate, prof, mgmt svc (51-55)	10.8	14	9.2	14	71	6	60	8	11	13	2

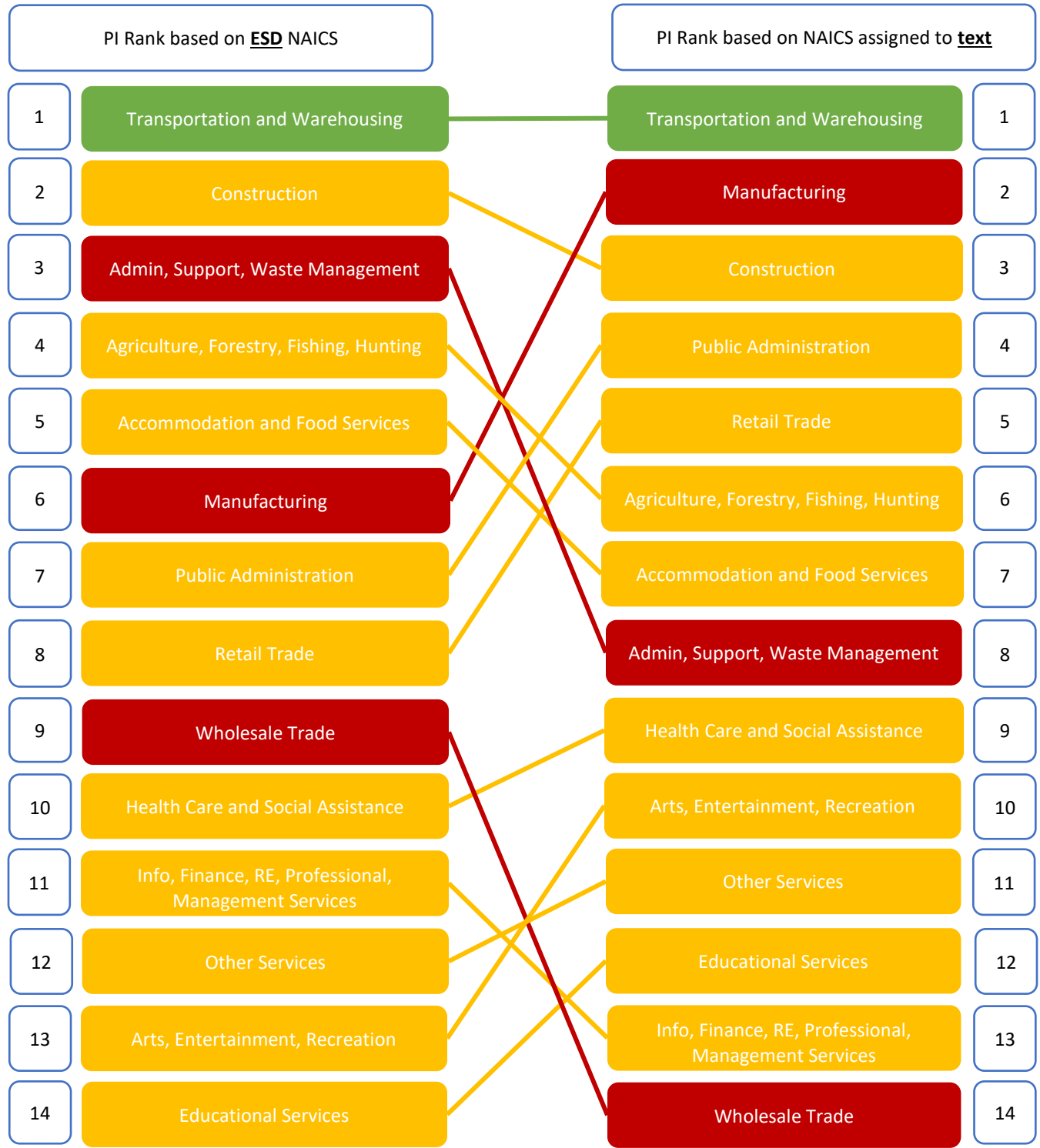
^aExcludes military, Insufficient information, sectors with fewer than 10 visits, non-worker categories, and visits among patients under age 16 years.

^bRecords per 100,000 workers, based on 2021 5-yr ACS estimates.

^cPI rank is the rank of the Prevention Index, which was calculated by averaging the record rank and the rate rank. Ties were broken by awarding the higher PI rank to the group with the higher rate.

^dThe absolute difference between the PI rankings of the two coding methods.

Figure 3. Prevention Index (PI) rank by industry and method of assigning industry code. Green shading indicates no difference in rankings between the two coding methods, yellow indicates ranking difference of 1 to 3 places, red indicates a difference in ranking of 4 or more places between the two coding methods.



Occupation Codes

Most records lacked sufficient information for assigning an occupation code. Of the 629 records linked to ESD, 203 were assigned an occupation code (by NIOCCS) based on the data reported in the occupation free text field or the NAICS code identified in ESD. Of the 10 records coded via the ESD linkage method as protective service occupations, all were assigned the same occupation major group based on the employer and occupation free text data (Table 11). Agreement was also high for food preparation and serving related occupations. In five occupation major groups, 50% or less of the records assigned codes using the ESD method were assigned the same codes based on text data. Generally, text-based codes were more likely to result in no code (“Insufficient information”) rather than assignment of a different occupation.

Table 11. Agreement in occupation classification between ESD linkage and text-based assignment, among RHINO ED records linked to ESD (n=629).

Occupation grouping (SOC code)	Number of records, based on ESD coding method ^a	% assigned same group based on free text	% assigned different group based on free text	% classified as insufficient information based on free text
Protective service occupations (33)	10	100%	0%	0%
Food preparation and serving related occ. (35)	10	90%	0%	10%
Management, business, science, arts occ. (11-27)	24	83%	13%	4%
Sales and office occupations (41-43)	19	79%	5%	16%
Healthcare occupations (29-31)	14	79%	0%	21%
Transportation and material moving occ. (53)	29	72%	10%	17%
Production occupations (51)	s	60%	20%	20%
Installation, maintenance, and repair occ. (49)	14	50%	7%	43%
Construction and extraction occupations (47)	38	42%	16%	42%
Building, grounds cleaning, maintenance occ. (37)	s	25%	13%	63%
Farming, fishing, and forestry occupations (45)	22	18%	5%	77%
Military	10	0%	0%	100%
Insufficient Information	426	-	3%	97%
Total	629	4%	20%	76%
Total, excluding ESD-based codes classified as Insufficient Information	203	58%	8%	33%

^aNAICS industry codes identified through ESD linkage were uploaded along with the occupation free text data for the NIOCCS auto-coder to assign an occupation code.

s=Suppressed due to small numbers.

Of the 1091 records with employer or occupation data reported in RHINO ED data, the majority lacked sufficient information to be assigned an occupation code using either coding method. Both methods assigned the highest number of records to transportation and material moving occupations, and the record counts were similar (Table 12). The greatest difference in record counts was found among farming, fishing, and forestry occupations, with the ESD plus text method classifying over 200% more records as the text-only approach.

Table 12. Occupation by coding method, among records with employer or occupation data reported in RHINO ED data (n=1091).

Occupation Grouping	ESD plus text codes ^a n (%)	Text-based codes n (%)	Difference in record counts ^b
Transportation and material moving occupations (53)	72 (7%)	69 (6%)	4%
Management, business, science, arts occupations (11-27)	67 (6%)	66 (6%)	2%
Construction and extraction occupations (47)	61 (6%)	39 (4%)	56%
Sales and office occupations (41-43)	56 (5%)	57 (5%)	-2%
Healthcare occupations (29-31)	30 (3%)	32 (3%)	-6%
Farming, fishing, and forestry occupations (45)	25 (2%)	s	>200%
Installation, maintenance, and repair occupations (49)	24 (2%)	18 (2%)	33%
Military (98)	23 (2%)	13 (1%)	77%
Food preparation and serving related occupations (35)	22 (2%)	23 (2%)	-4%
Protective service occupations (33)	20 (2%)	23 (2%)	-13%
Building, grounds cleaning, maintenance occupations (37)	19 (2%)	14 (1%)	36%
Production occupations (51)	14 (1%)	16 (1%)	-13%
Personal care and service occupations (39)	s	s	-17%
Non-worker (90-91)	s	s	0%
Insufficient information (99)	649 (59%)	703 (64%)	-8%

^a Codes based on occupation free text and NAICS codes assigned through linkage to ESD data for 629 records, NIOCCS-assigned codes based on occupation and employer text to remaining 462 records.

^bThe percent difference in records coded from ESD plus text compared to records coded from text alone
s=Suppressed due to small numbers.

Based on the ESD plus text method of coding, farming, fishing, and forestry occupations had the highest rate of ED visits for HRI by occupation (Table 13). The text-only method suggested that the highest rate was among protective service occupations, although the rate was lower than the ESD plus text-based rate for farming, fishing, and forestry occupations. Both methods ranked construction and extraction occupations and transportation and material moving occupations as the top two occupations based on the Prevention Index, although the exact ranking were switched between the two methods. The greatest difference in Prevention Index ranking was found for farming, fishing, and forestry occupations (ranked 6 places higher based on ESD plus text codes compared with text-only codes) and food preparation and related serving occupations (4 places higher based on text-only codes). Figure 4 shows the occupation groupings by Prevention Index rank for each coding method.

Table 13. Comparison of occupation-specific HRI visit rates, counts, and Prevention Index rank by method of coding^a.

Occupation Grouping	ESD Rate ^b	ESD Rate Rank	Text Rate ^b	Text Rate Rank	ESD Count	ESD Count Rank	Text Count	Text Count Rank	ESD PI ^c Rank	Text PI ^c Rank	PI rank diff ^d
Farming, Fishing, and Forestry Occupations (45)	45.6	1	14.6	5	25	6	<10	11	3	9	6
Construction and Extraction Occupations (47)	33.3	2	21.3	3	61	3	39	4	1	2	1
Protective service occupations (33)	29.7	3	34.2	1	20	9	23	7	4	3	1
Transportation and Material Moving Occupations (53)	26.3	4	25.2	2	72	1	69	1	2	1	1
Installation, Maintenance, and Repair Occupations (49)	22.1	5	16.6	4	24	7	18	8	5	4	1
Building and grounds cleaning and maintenance occupations (37)	17.0	6	12.5	6	19	10	14	10	10	10	0
Food preparation and serving related occupations (35)	11.4	7	11.9	7	22	8	23	6	9	5	4
Healthcare occupations (29-31)	8.9	8	9.5	8	30	5	32	5	6	6	0
Production Occupations (51)	8.2	9	9.4	9	14	11	16	9	11	11	0
Sales and office occupations (41-43)	7.9	10	8.1	10	56	4	57	3	7	7	0
Management, business, science, and arts occupations (11-27)	4.8	11	4.7	11	67	2	66	2	8	8	0

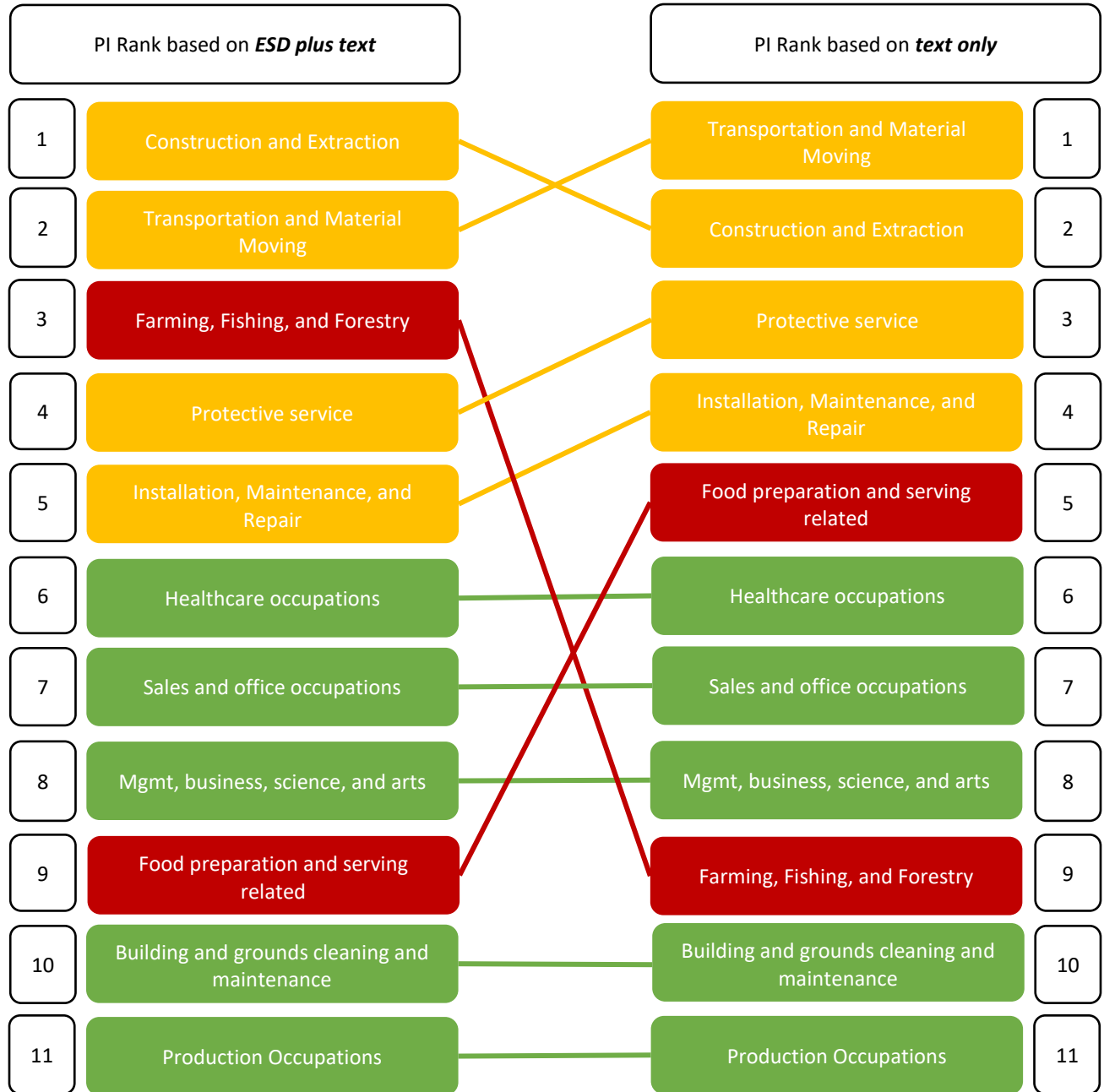
^aExcludes military, Insufficient information, occupations with fewer than 10 visits using both coding methods, non-worker categories, and visits among patients under age 16 years.

^bRecords per 100,000 workers, based on 2021 5-yr ACS estimates.

^cPI rank is the rank of the Prevention Index, which was calculated by averaging the visit rate rank and count rank. Ties were broken by awarding the higher PI rank to the group with the higher rate.

^dThe absolute difference between the PI rankings of the two coding methods.

Figure 4. Prevention Index (PI) rank by occupation and method of assigning industry code. Green shading indicates no difference in rankings between the two coding methods, yellow indicates ranking difference of 1 to 3 places, red indicates a difference in ranking of 4 or more places between the two coding methods.



DISCUSSION

The data available in ESD is a near-comprehensive list of all employers in Washington subject to state unemployment insurance laws, and their associated industry codes. While the current version of NIOCCS is sometimes able to assign industry codes based on employer name data, those instances are generally limited to records among large employers, employer names that include keywords suggestive of an industry, or clues provided in the occupation free text field.

The ESD linkage method of assigning industry codes is more resource intensive than relying solely on NIOCCS, requiring cleaning of both the RHINO ED data and the ESD data prior to linking, and manual record review after each linkage step to identify false matches. However, linking to ESD allowed for a more complete characterization of industry and – to a lesser extent – occupation among ED visits for HRI.

Not only did the ESD linkage method allow for the classification of industry and occupation for more records (i.e., fewer records determined to be of “Insufficient Information”), the classification resulting from the ESD linkage shifted the distribution of records, suggesting a different prioritization for HRI prevention. Among the industries more likely to be classified through the ESD linkage method versus free text alone were several sectors traditionally engaged in outdoor work, notably: construction, administrative and support and waste management and remediation services (which includes, among other industries, landscaping services and temporary help services), and agriculture, forestry, fishing and hunting. The ESD coding method increased visit counts and rates among these three industry sectors, and placed them at a higher priority for prevention, compared with the text-based codes. Of the top five industry sectors ranked by prevention index using the ESD classification, only two were also in the top five based on free text classification (namely, transportation and warehousing and construction).

Occupation classification showed similar differences between the two coding methods (both methods used NIOCCS to code occupation, but where the text-only method used both employer name and occupation free text fields, the ESD linkage method supplied NIOCCS with the occupation free text field and the NAICS code identified from the ESD linkage instead of employer free text field). Other than military, the biggest increases in record classification using the ESD linkage occurred among construction and extraction occupations, farming, fishing, and forestry occupations, and building, grounds cleaning, and maintenance occupations. Ranking the occupation major groups according to the Prevention Index, the biggest difference between the two coding methods was found for farming, fishing, and forestry occupations, which was ranked 3rd based on the ESD linkage method and 9th based on the text only method.

Limitations

Employer and occupation data is likely lacking for a large portion of ED visits for HRI. Based on census data, approximately 75% of Washington adults age 20-64 are employed. In comparison, work data was captured for 29% of ED visits among patients age 20-64. While the comparison between the two is imperfect – since ED data is visit-based rather than person-based, potentially capturing multiple visits per person – the low percent of ED records among working age adults with employment data reported is likely due, at least in part, to missing data. The true distribution of industry and occupation may be different than presented here if the industry and occupation data is not missing at random but instead,

more likely to be missing among select industries and occupations. Increasing collection of industry and occupation among all employed patients would help mitigate the issue of missing data.

The linkage process favored assignment of industry codes associated with the greatest employment within an employer name or UI account. It is possible that differences between the text-based codes and the ESD-based codes arose from selecting a single industry code from among multiple associated with a multi-establishment employer.

The few variables present in both the ESD and RHINO ED data limit the ability to link records. Lacking additional variables like date of birth (absent from the ESD data) or social security number (absent from the RHINO ED data), forces increased reliance on employer name data, and impedes attempts to link records with missing work data.

Conclusion

Because many of the industries and occupations involving outdoor work (and thus potentially at increased risk of HRI), are disproportionately missed by relying on free text coding alone, the linkage attempt to ESD is recommended. Further refinement of the linking algorithm may result an increase in the number of records linked and a decrease in the percent of false matches, potentially reducing the amount of manual record review needed to confirm true matches and increasing the efficiency of the linkage process.

Appendix

Heat Related Illness v2 definition uses ICD-10-CM codes and free text terms in the Chief Complaint History, Admit Reason Combo, and Discharge Diagnosis fields:

```
(,^[;/ ]992[0-9];^,or,^[;/ ]992.[0-9];^,OR,^[;/ ]E900;^,OR,^[;/ ]E900.[09];^,OR,^[;/ ]E900[09];^,OR,^[;/ ]T67^,OR,^[;/ ]X30.X^,OR,^[;/ ]X30X^,),OR,(, (^HEAT,OR,^HEAT[ -/\.:;]^,OR,^HEATCRAMP^,OR,^HEATEX^,OR,^HEATST^,OR,^HYPERTHERM^,OR,^SUNSTR^,OR,^SUNSTROKE^,OR,^SUN-STR^,OR,^ TO HOT^,OR,^ TOO HOT^,OR,(, (^HEET^,OR,^HOT^,)),AND,(, ^EXCESSIVE^,OR,^EXHAUST^,OR,^EXPOS^,OR,^FATIGUE^,OR,^CRAMP^,OR,^STRESS^,OR,^IN CAR^,OR,^OUTSIDE^,OR,^PROSTRATION^,)),),ANDNOT,(, ^ALLERG^,OR,^FEELING HOT^,OR,^FEELS HOT^,OR,^FELT HOT^,OR,(, ^HOT^,AND,^SENSATION^,)),OR,^HEAT SENSATION^,OR,^ INFLAM^,OR,(, ^PAIN^,AND,(, ^LIMB^,OR,^ARM^,OR,^SHOULDER^,OR,^ELBOW^,OR,^WRIST^,OR,^HAND^,OR,^LEG^,OR,^HIP^,OR,^GROIN^,OR,^THIGH^,OR,^KNEE^,OR,^ANKLE^,OR,^FOOT^,OR,^FEET^,OR,^BACK^,OR,^NECK^,OR,^FLANK^,OR,^RED^,OR,^JAW^,OR,^MOUTH^,OR,^TEETH^,OR,^TOOTH^,)),OR,^RADIAT^,OR,^REDNESS^,OR,^SWELL^,OR,^SWOLLEN^,OR,^SURG^,OR,^POST OP^,OR,^IBUPROFEN^,OR,^IBUPROPHEN^,OR,^ALIEVE^,OR,^MOTRIN^,OR,^TYLENOL^,OR,^INJUR^,OR,^TRAUMA^,OR,(, (^HEAT,OR,^HEAT[ -/\.:;]^,)),AND,(, ^ICE^,OR,^APPLIED^,OR,^APPLY^,OR,^APPLYING^,OR,^TRIED^,OR,^USED^,OR,^USING^,OR,^COLD^,OR,^RASH^,)),OR,^HEAT PACK^,OR,^HEATING PAD^,OR,^LUMBAGO^,OR,^RELIEF^,OR,^RESOLVE^,OR,^RELIEVE^,OR,^RELEIVE^,OR,^DENTAL^,OR,(, ^HOT^,AND,(, ^COLD^,OR,^COFF^,OR,^SHOWER^,)),),OR,(, ^ORAL^,AND,^SURG^,)),OR,(, ^SENSITIV^,AND,(, ^HEAT^,OR,^HOT^,)),OR,^HOT DOG^,OR,^HOT GREASE^,OR,^HOT EPPERS^,OR,^HOT TEA^,OR,^HEAT ACHE^,OR,^HEAT CONDITION^,OR,^HEATACHE^,OR,^HEAT ATTACK^,OR,^HEAT BEAT^,OR,^HEATBEAT^,OR,^HEAT FAILURE^,OR,^HEAT BURN^,OR,^HEATBURN^,OR,^HEAT FLUTTER^,OR,^HEAT RACING^,OR,^HEAT RATE^,OR,^HEATRATE^,OR,^HEATLH^,OR,^HEATH^,OR,^HEATTH^,OR,^HITTING HEAT^,OR,^PALPITATION^,OR,^CHEAT^,OR,^WHEAT^,OR,^HEATER^,OR,^HEATHER^,OR,^HEATING^,OR,^HOTEL^,OR,^LITHOTR^,OR,^METHOTR^,OR,^PHOTO^,OR,^PSYCHOTIC^,OR,^SHEATH^,OR,^SHEET^,OR,^SHOT^,OR,^SUNDAY^,OR,^THEAT^,OR,^WHEAT^,OR,^ACCIDENT^,OR,^ALCOHOL^,OR,^ETOH^,OR,(, ^BURN^,AND,^MOUTH^,)),OR,^DISTRESS^,OR,^FEVER^,OR,^GETS HOT^,OR,^HEAT FLASH^,OR,^HOT FLASH^,OR,^HIVES^,OR,^HOT TUB^,OR,^NO HEAT^,OR,^OVEN^,OR,^SUICID^,OR,^HEAT OF THE MOMENT^,OR,^CONTACT WITH OTHER HEAT AND HOT^,OR,^W92^,))
```