complicates attempts at combining data from the OSHA and SOII sources. Finally, we found that establishments identified in both sources tended to report similar information to each source.

Discussion: OSHA's electronic reporting requirement may continue to evolve. Employers' understanding of the new rules will presumably evolve as well. The nonrandom selection of employers into OSHA reporting status and other practical considerations could affect any BLS attempt to combine OSHA and SOII records, or substitute OSHA records for SOII survey information.

C4.2
Title: Efficiency of Autocoding Programs for Converting Job Descriptors into Standard Occupational Classification Codes
Authors: Bradley Evanoff, Skye Buckner-Petty, Ann Marie Dale

Background: Existing datasets often lack job exposure data. Standard Occupational Classification (SOC) codes, based on job information, can link publicly available work exposure data to individual health outcomes via a Job Exposure Matrix. However, job code assignment is a laborious process if done manually. Two recently developed automatic coding programs provide an efficient resource for job code assignment but the number of job titles that can be automatically assigned codes, and agreement with assignment from manual coding have not been tested other than during development of the tool.

Methods: Two publicly available programs, NIOSH Industry and Occupation Computerized Coding System (NIOCCS) version 2 and Standardized Occupation Coding for Computer-assisted Epidemiological Research (SOCcer) version 1, convert text based job descriptors into SOC codes and compute confidence scores based on the strength of those assignments. We entered industry and occupation descriptions from two existing cohorts into both autocoding programs. Independent manual SOC coding was also completed for each worker. Agreement between manual codes and autocodes was assessed at varying confidence scores. We also assessed agreement of several exposure values (from the Occupational Information Network, O*NET) linked by manual SOC codes versus those linked by autocodes, in order to examine how differences in coding might affect exposure assignments in general population cohort studies.

Results: NIOCCS produced SOC codes for the majority of subjects (Cohort A: 71%; Cohort B: 62%). The level of detail for these codes varied; detailed occupations (6-digit SOC codes) were available for 44% and 43% of cohorts A and B respectively. Comparison to manual codes showed strong agreement at the major group 2-digitlevel (kappa ≥ 0.8 and 0.6) and weaker agreement at the 6-digit level (kappa ≥ 0.6 and 0.3). SOCcer produced 6-digit SOC codes for all subjects with good agreement at the 2-digit level (kappa ≥ 0.7 and 0.6) and slightly lower at the 6-digit level (kappa ≥ 0.4 and 0.3). Agreement for O*NET exposures was very high for most comparisons within both cohorts for both programs (many ICCs > 0.8). Both programs produced better agreement as confidence scores increased.

Discussion: Both autocoding programs can be reliable tools to aid in assigning SOC codes that represent broad industry levels, with less agreement at finer levels of job codes. Given the availability of large public datasets with job information but no other work exposure data, autocoding of jobs provides exciting opportunities for analyzing work-related health outcomes in future studies. We are currently assessing the performance of NIOCCS version 3, which will be introduced later this year.

C4.3
Title: Enhancing Agriculture, Forestry, and Fishing Injury Surveillance Using Free Text Data
Authors: Erika Scott, Liane Hirabayashi

Background: Access to free text in existing administrative databases has proved useful in identifying and characterizing agricultural, forestry, and fishing (AFF) related injuries. Particularly, narratives from pre-hospital care reports (PCRs) provide specific details of the injury event directly from the scene and from interviewing the patient. These narratives, which are retained by a number of states, are systematically searched for AFF specific keywords and verified for AFF relatedness. The Occupational Injury and Illness Classification System (OIICS) is then applied to the dataset by a team of coders. To enhance this process, researchers are applying Bayesian methodologies to speed up text review and ultimately, reduce the cost of the surveillance system.

Methods: The process described above has been applied to PCRs from Maine and New Hampshire for a three-year period to create a confirmed injury dataset.

Agriculture, forestry and fishing records were identified by industry, and by the certainty of the injury report (e.g. true case, suspected case) This dataset was then split, along with non AFF records, into a training and validation datasets to build and test Bayesian algorithms for the determination of AFF records.

Results: Maine and New Hampshire had 767,060 pre-hospital care report records for 2008-2010. Of these, 28,341 contained one or more of 161 AFF keywords (searched either by character string or exact word). Of the keyword containing records, 1,203 were determined to be AFF related. Results of the Bayesian methodology are currently being testing and will be presented at the conference.

Discussion: Pre-hospital care reports are a rich source of occupational injury data, especially for agriculture, forestry and fishing. These injuries are able to be identified and coded using the OIICS classification scheme, making them comparable to other industries. Pre-hospital care reports have the potential to be a useful source of research data, beyond AFF, but for other industries and for public health in general. Conclusions on the success of using Bayesian methods to enhance coding of AFF cases will be discussed at the conference.

C4.4
Title: Household Survey of Occupational Injuries and Illnesses Pilot – Update and Discussion
Author: Elizabeth Rogers

Background: Research conducted by the Bureau of Labor Statistics (BLS) and others has identified under-reporting in the Survey of Occupational Injuries and Illnesses (SOII). In 2009 Congress charged BLS to develop a research program to examine the issue of under-reporting of occupational injuries and illnesses. OSHA and NIOSH were also asked to create similar research programs. In addition to internal research, BLS has also partnered with outside researchers on a variety of projects, including matching SOII and Workers' Compensation data, evaluating the use of multiple sources to identify nonfatal cases, and assessing the recordkeeping practices and knowledge of SOII respondents. While the scope of under-reporting to the SOII is not currently known, research points to filters that may reduce injury and illness reporting by employees or employers. These filters include employees not reporting injuries to their employers in the first place, insufficient employer

understanding of recordkeeping definitions and requirements, and disputes over the work-relatedness of an injury or illness. This may prevent employers from being aware of some occupational injuries and illnesses or from reporting known injuries and illnesses to the SOII. In light of this, BLS has examined various approaches to capturing a more complete count of workplace injuries and illnesses. One possible approach is the collection of occupational injury and illness data from workers directly. While a key strength of the SOII is the richness of the data published from it, comparing injury and illness data collected from households to SOII data may yield a greater understanding of the magnitude of workplace injuries and illnesses.

Findings/Discussion: To test this approach, in 2017 and 2018 BLS developed and piloted a Household Survey of Occupational Injuries and Illnesses (HSOII). The HSOII is a large-scale, nationally-representative telephone survey designed to yield responses from 3,520 workers in the United States. Responses are collected using random digit dial. Individuals are asked about occupational injuries and illnesses that they experienced in the past year, the effect those incidences have had on their work and pay, and their demographic characteristics. One goal of this pilot is to evaluate the feasibility of collecting occupational injury and illness data from workers directly. A second goal is to produce nationally representative top line estimates that can be compared to similar estimates from the SOII to better understand the magnitude of underreporting to the SOII. Data collection for the pilot HSOII will conclude in July, 2018.

BLS plans to present a summary of the research into SOII underreporting, a description of the HSOII pilot study, and the preliminary results of the HSOII pilot, including estimates from the pilot for available industries, occupations, case circumstances, and worker demographics and an assessment of data quality. The results will provide insight about the feasibility of collection of occupational injuries and illnesses from workers directly, a greater understanding of the potential magnitude of these cases, and challenges for any future collection of these data from households.
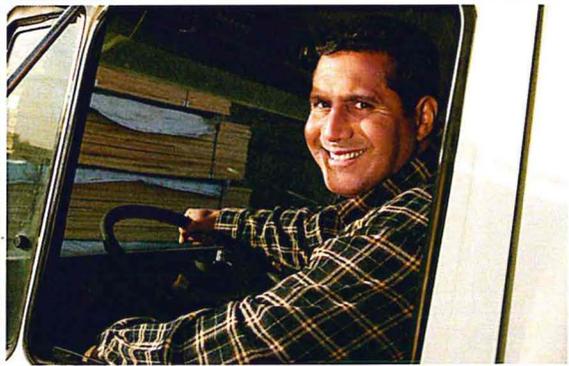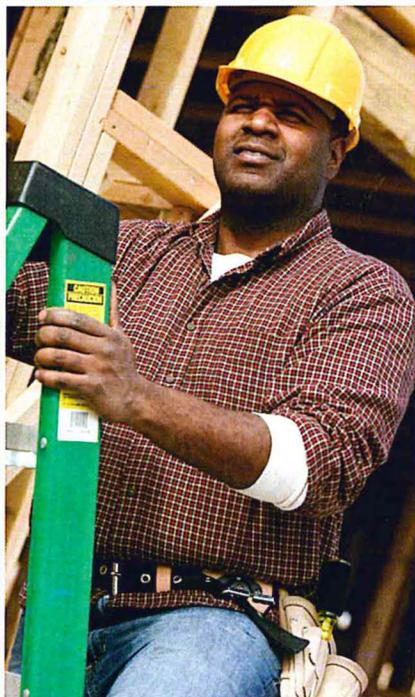
Session C5
Title: Economics of Worker Safety
Moderator: Tim Bushnell

NOIRS

National Occupational Injury Research Symposium

Advancing Worker Safety in the 21st Century Through Research & Practice

SYMPOSIUM PROGRAM

October 16-18, 2018 • Morgantown, WV