noise induced hearing loss {NIHL}. Standard guidelines were followed and PTA was performed using a NIOSH approved audiogram machine. For this analysis, we selected HEOs between the ages of 45 and 57 years, with 10 or more years of service. These results were compared with 106 age matched employees, who were involved in non-mining activity in the same industry with the same amount of noise exposure but from different sources. Results were validated using standard statistical tools.

Result Results revealed 60% (64/108) of the HEOs suffered NIHL. Of these the impact on the left side was found to be more pronounced for 90% (58/64) versus 10% (6/64) in the right ear. In the comparative group of 106 non-mining employees 55% (59/106) had abnormal audiogram reports but with no specific inclination to the right or left side.

Discussion NIHL is common in the mining industry. The observed laterality in hearing loss among HEOs is likely multi-factorial and can be attributed to the presence of the engine on the left hand side in the Indian vehicles, the failure to sound proof the operator's cabin and/or comply with the issued personal protective equipment {PPE}. Our results underscore the importance of the need for compliance with the use of PPEs and vehicular upgrade and maintenance.

Methodological Issues

CROSS-WALKING COUNTRIES' INDUSTRY CLASSIFICATIONS USING CONCORDANCE FILES COMPARED TO FUZZY DATA MATCHING, TO UTILISE AN INTERNATIONAL EXPOSURE DATASET

Ganesh Selvaraj*, Michael Butchard. *Worksafe New Zealand, Wellington, New Zealand*

Background Carcinogen exposure data can potentially guide the work of health and safety (H and S) regulators. This project aims to use CAREX Canada data to estimate carcinogen exposures in New Zealand industries. This requires the creation of a cross-walk between the countries' industry classifications.

Methods Agile and big-data-science methodologies were used to construct two versions of an industry classification cross-walk from the 2006 Australian and New Zealand Standard Industrial Classification (ANZSIC06) to the Canadian version of the 2002 North American Industrial Classification (NAICS2002), used by CAREX Canada.

Firstly, concordance files from government statistics bureaus cross-walked the path: ANZSIC06 ->International Standard Industrial Classification of All Economic Activities Rev4 ->NAICS2017->NAICS2012->NAICS2007->NAICS2002. The cross-walk accounted for 'one-to-many-to-one', non-machine formats, and missing/erroneous values.

Secondly, a fuzzy data matching pipeline was designed. Data preparation removed redundant, stop, and common domain words, and lemmatised using morphological analysis (e.g. fishing to fish). Data matching used a hybrid algorithm combining 'JaroWinkler-distance' and a token-sort approach (i. e. ignoring the positional occurrence of words in a sentence) to match descriptions. A trial-and-error approach was used to assign weightings and concatenate the hierarchical industry

classification levels to improve match accuracy. Python language was used for implementation.

For each method, random samples of 50 matches were manually classified as either poor or sufficient by two people. Disagreements were discussed and consensus reached.

Results The concordance cross-walk sample had 52% (95% C. I. 38%–66%) sufficient matches compared to 84% (95% C.I. 74%–94%) for the fuzzy data matching pipeline cross-walk sample.

Conclusions Cross-walking countries' industry classifications using a fuzzy data matching pipeline was more accurate than using a concordance cross-walk. The pipeline is modular enough to easily include more components. This work is part of a vision to design a semantic big-data lake, enabling integration of any data relevant to H and S.

EFFICIENCY OF AUTOCODING PROGRAMS FOR CONVERTING JOB DESCRIPTORS INTO STANDARD OCCUPATIONAL CLASSIFICATION CODES

Ann Dale*, Skye Buckner-Petty, Bradley Evanoff. *Washington University School of Medicine – General Medical Sciences, St. Louis, USA*

Introduction Standard Occupational Classification (SOC) codes can link work exposure data to individual health outcomes, but manually assigning job codes is laborious. We tested two recently developed automatic coding programs.

Methods We entered self-reported job titles and industry from two existing cohorts into two publicly available autocoding programs, the NIOSH Industry and Occupation Computerized Coding System (NIOCCS) and the Standardized Occupation Coding for Computer-assisted Epidemiological Research (SOCcer), and assessed agreement between autocodes and manual coding. 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 1: 85%; Cohort 2: 79%). The level of detail for these codes varied slightly; 6-digit SOC codes (detailed occupations) were available for 84% and 76% of cohorts A and B respectively. Comparison to manual codes showed strong agreement at the major group level (kappa=0.8 for both cohorts) and weaker agreement at the 6-digit level (kappa $\geq$0.4 and 0.6). SOCcer produced 6-digit SOC codes for all subjects with good agreement at the 2-digit level (kappa $\geq$0.6 and 0.7) and slightly lower at the 6-digit level (kappa $\geq$0.3 and 0.4). Agreement for O*NET exposures was very high for most comparisons within both cohorts for both programs (many ICCs>0.8).

Conclusion 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.