

extracted odds ratios/relative risks (and standard errors) comparing workers in each sector to within study reference category. To avoid complications due to study heterogeneity, we simply produced descriptive forest plots where possible utilising the odds ratios that have been adjusted for demographics.

Results Eight studies were identified from Norway (one), and California (one), and the UK (five). Three studies at time of writing were peer reviewed, and five of the eight were cohort studies. Food production was the least well investigated with limited evidence of increased infection, severe infection or mortality compared to the 'other' groups. Healthcare and transport show wide variation in the odds ratios reported. Public facing roles did indicate greater infection risk, specifically in the first wave.

Conclusions Considerable study heterogeneity is present, particularly with respect to the chosen reference groups, meaning objective comparisons are limited here. However, as would be expected public facing roles, especially in the first wave, did appear to experience increased infection. Further prospective work with subject level data is needed to better understand the occupational risks.

S-498 **METHODOLOGICAL ISSUES IN STUDYING OCCUPATIONAL DIFFERENCES IN RISK OF COVID-19**

¹Neil Pearce. ¹*London School of Hygiene and Tropical Medicine, United Kingdom*

10.1136/OEM-2021-EPI.452

It is well-established that certain occupations that have frequent contact with patients (e.g. health care workers) or the public (e.g. bus drivers), may have increased risks of COVID-19. However, estimating these risks involves a number of important methodological problems. Not everyone who has a SARS-CoV-2 infection gets symptoms; not everyone with symptoms gets tested; not all tests yield valid results. Therefore, even just estimating the incidence or prevalence of COVID-19 in particular occupational groups is difficult. Moreover, these selection pressures may be different for different occupational groups (e.g. health care workers may be more likely to be tested than some other occupations). In addition, unlike most other occupational exposures, you can get infected at work, or get the same infection at home, potentially yielding quite severe confounding by lifestyle and living conditions. These methodological issues are not insurmountable, but require careful study design and data analysis. In particular: (i) when not everyone is being tested, then comparisons between and within occupations can be biased, but this bias can be minimised using the test-negative design; and (ii) comparisons between occupations require careful and rigorous adjustment for work-related 'living conditions', i.e. the fact that workers in insecure low-paid jobs may also have a high risk of contracting SARS-CoV-2 infections outside of the workplace.

S-499 **THE APPLICATION OF ARTIFICIAL INTELLIGENCE IN THE CODING OF OCCUPATIONAL INFORMATION**

¹Anil Adisesh, Christopher JO Baker. ¹*University of Calgary, Canada*

10.1136/OEM-2021-EPI.453

Many research studies seek to identify the social determinants of health and occupation is an important predictor, both at the level of the individual as well as for populations. Whereas job titles are usually solicited during interviews or by questionnaire, before being able to use this information the responses need to be categorized using a coding system, such as the Canadian National Occupational Classification (NOC).

Manual coding is the usual method, which is a time-consuming and error-prone activity with variable or inconsistent outcomes from teams of coders. In recent work the ACA-NOC algorithm¹ was developed to perform automated coding based on matching job title text with the NOC's job titles and textual descriptions. This algorithm was benchmarked on a small sample manually coded data set with subject matter experts subsequent review of coding discrepancies to facilitate functional improvements to the algorithm. Performance levels achieved illustrated the viability of the approach albeit larger benchmarking data sets were required.

CanPATH² has collected data from approximately 330,000 volunteer Canadians, including information about health, lifestyle, occupation, environment and behavior. We report on the further benchmarking and further development of this algorithm in CanPATH funded project using over 60,000 manually coded job titles from the constituent Alberta Tomorrow Project. The algorithm was also applied to over 100,000 un-coded job titles from Atlantic PATH, including the Core questionnaire and occupational history data.

The core outcome of the project identified that auto-coding results are comparable to manual coding in accuracy and superior in speed e.g. 2 years of manual coding (64,000 records) can be auto coded in 72 hours. The algorithm was considered ready for deployment in operational settings: point of care, decision support for manual coders.

Additional insights gained during the project revealed that (i) NOC and ATP data sets have a distribution bias where some NOC categories were over or under-represented and numerous non-standard lexical features were found in job titles and NOC job descriptions, (ii) benchmarking datasets from ATP included coding errors that were corrected by expert coders leading to the creation of gold standard test sets for further algorithm improvement studies, (iii) a study on 17 categories of occupations initially difficult to code, identified some job categories with near 90% coding accuracy.

Automated coding of job titles to the NOC has been shown to be both practicable to good levels of accuracy and shown to significantly accelerate manual coding efforts from years to autocoding in a matter of hours without decrease in accuracy. Autocoding can replace costly, error prone manual labor with accurate point-of-care auto-coding such that patient occupation information during healthcare encounters could now supplement existing administrative data sets in electronic health record systems. This data can be used better to understand the socioeconomic consequences of health conditions, advise patients about returning to work with a health condition, recognizing occupations at risk of disease e.g. as in the COVID-19 pandemic.

REFERENCES

1. Bao H, Baker CJO, Adisesh A. Occupation coding of job titles: iterative development of an automated coding algorithm for the Canadian national occupation classification (ACA-NOC). *JMIR Form Res* 2020 Aug 5;4(8):e16422. doi:10.2196/16422
2. CanPath the Canadian Partnership for Tomorrow Project. <https://canpath.ca/> accessed: 01.09.2021