

Introduction In Germany more than 45,000 cases of COVID-19 are already recognized as Occupational Disease (OD) for workers in the health and welfare sector. Different long term effects of COVID-19 are described in literature, fatigue being one of the most common symptoms observed after the infection.

Objectives We assessed long term effects of COVID-19 three months and more after the SARS-CoV-2 infection in health and social workers (HSW).

Methods A self-administered questionnaire was sent to 3,800 HSW with COVID-19 recognized as OD in January 2021. The questionnaire was self-developed and assessed biographical data, symptoms at time of infection, quality of life (QoL) and workability as well as COVID-19 associated symptoms three months or more after infection. The study was approved by the Medical Ethic Committee in Hamburg, Germany.

Results A total of 2,018 HSW returned the questionnaire. Response rate was 53%. 82% were female and 52% were older than 50 years. 42% worked in hospitals and 30% in nursing homes, 9% were doctors and 56% nurses. Compared to the time before the infection, QoL was reduced by 72% and work ability was reduced by 83%. Severe fatigue was reported by 20%, severe concentration and memory problems by 14% and severe breathlessness by 9%. 4.5% had already taken part in a rehabilitation and 37% wished to take part in a rehabilitation.

Conclusion Long term fatigue symptoms are common in HSW more than three months after COVID-19. The need for rehabilitation in HSW with COVID-19 as OD is high. The further cause and the treatment options for long term effects and especially for fatigue need to be studied.

S-452 CHALLENGES OF INTEGRATING GENDER THEORETICAL CONCEPTS INTO EPIDEMIOLOGICAL HEALTH RESEARCH

¹Gabriele Bolte. ¹Institute of Public Health and Nursing Research, University of Bremen, Germany

10.1136/OEM-2021-EPI.441

In recent years awareness of the need to integrate sex/gender more comprehensively into epidemiological health research has increased. However, a discrepancy between gender theoretical concepts on one hand and epidemiological practice on the other hand still remains. A conceptualization of a static, individual characteristic sex/gender (often only dichotomous as male/female) does not capture variability, multidimensionality and contextuality of sex/gender. Moreover, interactions between sex-linked biology and gender relations, processes of embodiment as well as intersectionality in terms of power relations and processes of privilege or discrimination are hardly considered. This is especially true for environmental health research.

The presentation will draw on experiences gathered in two ongoing research projects: INGER and AdvanceDataAnalysis. The collaborative research project INGER (Integrating gender into environmental health research) developed a multidimensional sex/gender concept from an intersectionality perspective. This concept guided operationalizations for sex/gender-related data collection in a population-based study. Decision tree methods are currently applied to assess the relevance of several sex/gender dimensions when identifying subgroups with especially high environmental exposures. The subproject

AdvanceDataAnalysis as part of the collaborative research project AdvanceGender focuses on the more in-depth analysis of already existing data to support gender-sensitive health reporting. One approach is to define 'solution-linked' variables that indicate modifiable societal and contextual factors and help to explain heterogeneity across social dimensions.

With INGER and AdvanceDataAnalysis as examples, challenges in operationalization of sex/gender based on gender theoretical concepts, and in statistical analysis of the impact of multiple sex/gender dimensions on exposure variation and effect modification will be discussed.

Improvements in the integration of theoretically sound sex/gender concepts in epidemiological health research will enhance its validity and significance and, in perspective, contribute to more health equity.

S-464 AUTOMATED OCCUPATIONAL ENCODING TO THE CANADIAN NATIONAL OCCUPATION CLASSIFICATION USING AN ENSEMBLE CLASSIFIER FROM TF-IDF AND DOC2VEC EMBEDDINGS

¹Cesar Augusto Suarez Garcia, Anil Adishes, Christopher JO Baker. ¹University of New Brunswick, Canada

10.1136/OEM-2021-EPI.442

Introduction Occupational encoding is a technique that allows job titles provided by study participants to be categorized according to their role in the labor force. Encoding has primarily been a slow error-prone manual process which is ripe for automation.

Objectives Our goal was to design and test an automated coding prototype using machine learning techniques.

Methods The prototype classification system ENENOC (the ENsemble Encoder for the National Occupational Classification) is comprised of series of steps involving data cleaning, exact match search, multi classifier ensembling, hierarchical classification, and multiple output selection. In the absence of exact matching between job title input and NOC category descriptions, the input data is embedded using the TF-IDF algorithm and Doc2Vec. The embeddings are fed into a hierarchical, ensemble classifier that uses classical machine learning techniques: Random Forests, Support Vector Machine and K-Nearest Neighbour. Ensemble encoding is achieved using a majority-voting system. The hierarchical two tier classification methodology first predicts the first digit of the NOC code followed while the second tier predicts the second third and fourth digit of the NOC code for the input data. The combined approach produces a single, 4-digit code as a top choice, as well as four alternate NOC codes, that serve as additional ranked choice based on the Doc2Vec model.

Results The prototype was benchmarked on a manually annotated data set comprising of 64,000 records. It produced a top-1 Per-Digit Macro F1-Score of 0.65 and a top-5 Per-Digit Macro F1-Score of 0.76, both of which are highly within published accuracy ranges for manual coding (44% to 89% inter-annotator agreement). ENENOC coded 30,000 job titles in 3 hours.

Conclusion The ENENOC prototype is a sophisticated ENsemble Encoder for the National Occupational Classification which has state of the art performance accuracy with significant speed improvements over manual coding.