Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230351

Mapping Exposome Derived Phenotypes into SNOMED Codes

Christopher HAWTHORNE [a*], Luis MARCO RUIZ [b*] and
Guillermo LOPEZ CAMPOS [a,1]
[a] *Wellcome-Wolfson Institute for Experimental Medicine, Queen's University Belfast, Belfast,UK*
[b] *Norwegian Centre for E-health Research, University Hospital of North Norway, Tromsø, Norway*
ORCiD ID: Guillermo Lopez Campos https://orcid.org/0000-0003-3011-0940

Abstract. Human phenotypes define the healthy or diseased status of an individual and they arise from the complex interactions between environmental and genetic factors. The whole set of human exposures constitute the human exposome. These exposures have multiple sources including physical and socioeconomic factors. In this manuscript we have used text mining techniques to retrieve 1295 and 1903 Human Phenotype Ontology terms associated with these exposome factors and we have subsequently mapped 83% and 90% of the HPO terms respectively) into SNOMED as a clinically actionable code. We have developed a proof-of-concept approach to facilitate the integration of exposomic and clinical data

Keywords. Bioinformatics, Exposome, precision medicine, SNOMED-CT, Human Phenotype Ontology.

1. Introduction

Human phenotypes define the healthy or diseased status of an individual and they arise from the complex interactions between environmental and genetic factors. In the last couple of decades, great efforts have been made to improve the tailoring of medical practice to individuals, mostly through an improved understanding of the genome of an individual. More recently, with the development of precision medicine approaches, the exposome [1], defined as the whole set of life-through exposures of an individual, has also started to be considered. This "whole set" definition makes the exposome a broad and complex element that covers exposures to biological, chemical, physical and psychosocial agents. The contributions to exposomic research from each of these components is variable and so far, the focus has been mostly on understanding the chemical and biological domains, whereas an increasing interest is also noticeable in psychosocial aspects and the social determinants of health. Biomedical informatics play a key role in the development of efficient solutions that facilitate the analysis and integration of these data in multiple scenarios. In this manuscript we have combined computational exposomic research on physical and socioeconomic factors related

* Equal contributors

[1] Corresponding Author: Guillermo Lopez Campos, Wellcome-Wolfson Institute for Experimental Medicine, E-mail: g.lopezcampos@qub.ac.uk.

literature with SNOMED annotations as an approach for the future potential validation of the results and the integration of clinical data.

2. Methods

Bibliographic contents from the title and abstract of PubMed indexed literature around environmental 16 different physical factors (PF) ("lighting", "noise", "electromagnetic fields"…) and 7 different socioeconomic factors (SEF) (including "income", "poverty" and "educational status" among other) were annotated using ONASSIS tool to identify human phenotypes as previously described [2]. Phenotypes were then mapped to SNOMED-CT using the UMLS Terminology Service checking the available mappings in SNOMED-CT for each of the HPO codes representing the human phenotypes extracted from PubMed. The US 2022 SNOMED-CT version was used.

3. Results

A total of 5844 (4541 for PF + 1303 for SEF) different HPO terms were retrieved from the initial analyses. After filtering out phenotypes appearing in multiple factors 1294 (28%) and 909 (70%) unique HPO terms for physical and socioeconomic factors respectively remained in the dataset.

SNOMED mapping resulted in 1071 (83%) HPO terms mapped to different SNOMED codes for the physical factors dataset and 822 (90%) HPO terms mapped for the socioeconomic factors dataset. Interestingly, although initially a smaller set, SEF related HPO terms led to a more diverse set of SNOMED codes. A total of 8439 SNOMED codes (4523 for PF + 3916 for SEF) were mapped. The majority of the HPO terms were associated with more than one SNOMED code (min=0, average=3.4, max= 23 for the PF and min=0, average=4.3, max= 23 for the SF).

4. Conclusion

Socioeconomic related phenotypes are more easily mapped into a more diverse set of SNOMED-CT codes than those derived from those associated with the physical factors of the exposome. We have been able to develop a proof-of-concept approach to integrate exposomic related knowledge extracted from the literature with clinical vocabularies that might facilitate the validation of the research hypothesis derived from the literature using real world clinical data.

References

[1] Wild CP. Complementing the genome with an "exposome": the outstanding challenge of environmental exposure measurement in molecular epidemiology. Cancer Epidemiol Biomarkers Prev. 2005;14: 1847–50

[2] Hawthorne C, Quigley N, McClements C, Lopez-Campos GH. Construction of a Physical Factor Resource for Exposome Informatics Research. Stud Health Technol Inform. 2021 May; 281:1079-1080.