

Improve Health Equity by analyzing SDOH from the EHR

Yanshan Wang, PhD, FAMIA
Department of Health Information Management
University of Pittsburgh

Motivation

- Social determinants of health (SDOH) have a major impact on health equity.
- Such as socioeconomic status, education, employment, lifestyles, social support networks, access to medical care, and neighborhood characteristics
- SDOH information is recorded in the structured Electronic Health Records (EHRs), such as diagnosis code, and documented in the free-text EHRs, such as clinical notes.
- There is a big **gap** in leveraging the heterogenous data within EHRs to infer a patient's SDOH status.

Project Description

- Aim 1: Curate SDOH concepts and collect EHRs.
 - In this project, we will focus on developing informatics methods to extract and analyze five SDOH categories that include: 1) housing status: stable housing, homeless, unstable housing, living with friends or relatives; 2) employment status: working part time, working full time, student, training, retired; 3) education status: post graduate, college educated, high school educated, 8th grade or less; 4) social support: family support, spiritual support, caregiver support, and lack of social support; and 5) lifestyles: physical inactivity, diet, sleep disorder.
 - We will retrieve both structured EHRs and clinical notes from the clinical data warehouse. We will create a gold standard dataset by manually annotating the SDOH concepts.
- Aim 2: Develop clinical natural language processing (NLP) and artificial intelligence (AI) models to identify SDOH status.
 - We will develop a novel lexicon of words and phrases related to the SDOH concepts and extract those concepts from clinical notes using clinical natural language processing (NLP).
 - We will combine the SDOH information from both clinical notes and structured EHRs and develop machine learning models to predict the patient-level SDOH status.



Extracting and analyzing social determinants of health embedded in the electronic health records using Artificial Intelligence and Natural Language Processing will help understand current health disparities and improve health equity.



Context

- SDOH is an important indicator for health equity since it indicates whether people have access to adequate diet, medical care, educational, and career opportunities, what are their healthy environmental conditions, and whether a person is exposed to physical or psychologic trauma. SDOH helps us develop comprehensive strategies to address potential risks for the population, particularly for the underserved populations.
- The majority acquisition methods to collect SDOH information for health equity research are based on self-reported interviews, surveys, or questionnaires and some lack biometric measures. Using such traditional ways to acquire SDOH information has become a bottleneck for epidemiologic and therapeutic investigations due to the following limitations: 1) non-scalable; 2) inefficient; 3) limited to patients' subjective experience; and 4) recall bias.
- EHRs offer large volumes of digitized information in real world clinical interventions and outcomes. However, EHRs remain underused in collecting SDOH information for research. Major barriers that hinder the use of EHRs are the heterogeneity of EHR and lots of SDOH information in clinical notes.
- Thus, clinical NLP to extract SDOH from clinical notes and AI models to combine SDOH from the heterogeneous EHR is needed.

Project Deliverables

- The clinical NLP and AI models of the project will be distributed publicly through the Open Health NLP (OHNLP) consortium, which is an open source consortium to promote NLP development efforts and to encourage participation in advancing future efforts.
- The research findings including the novel lexicon and NLP system will be published through informatics journals.
- We will work on a specific disease area and conduct a case-control study to compare the statistics of SDOH results between the case and control cohort to provide implications for health equity.

References and/or Acknowledgements

- The Principle Investigator would like to thank the School of Health and Rehabilitation Sciences (SHRS) for supporting the preliminary work.