

# **BioHealth-OKN: A Dynamically-Updated Open Knowledge Network for Health: Integrating Biomedical Insights With Social Determinants of Health**

*Aidong Zhang, UVA; Rahmatollah Beheshti, Stefan Bekiranov, Kishlay Jha, Cathy Wu*

## **Use case description and societal challenge being addressed**

Substantial research demonstrates that social determinants of health (SDoH) influence patient care outcomes, including mental health outcomes. However, incorporating a patient's social determinants into risk prediction remains a significant challenge, at both the patient and population levels. We plan to utilize knowledge graphs to understand how SDoH factors contribute to risk for suicidality and post-traumatic stress disorder (PTSD) within the Veterans Health Administration (VHA). Utilizing data on social determinants from structured VHA electronic record health data and unstructured clinical documentation data, we will extract SDoH data from patient records and incorporate into knowledge graphs predicting suicide outcomes (positive Columbia suicide screens, suicide attempts, completed suicides) as well as PTSD outcomes (positive initial PTSD screens, PCL-5 scores). We anticipate both potential patient and population use cases. The patient-level use case is to predict individual patient risk to provide clinical decision support to primary care and mental health teams, which may affect plans for specialty referral, case management escalation, and treatment escalation to prevent adverse outcomes such as suicide and mental health-related hospitalization. The population level use case is to gather insights at the VA facility, VISN (regional), or national VHA level on suicide and PTSD risk factors, which may be used to inform decision making on staffing and mental health resource allocation. End users will include mental health providers, primary care providers, and organizational leaders (e.g. facility suicide coordinators, mental health service leaders, institutional leaders). Insights may be translatable into clinical decision support tools that interface with the electronic health record system and offer patient-specific insights or aggregator summary data.

## **Knowledge graph source datasets**

Datasets we used to extract the entities and relationships include the PubMed dataset, which contains the citations and abstracts of biomedical literature from several NLM literature resources; we utilized the publicly available titles and abstracts from PubMed. We also utilized the Medical Information Mart for Intensive Care-III (MIMIC-III) dataset. For the MIMIC-III dataset, we utilized the discharge summary in the NOTEEVENT.csv. The current knowledge graph has 248,333 unique entities and 18,201,480 unique relationships, in the future we will include the articles of the biomedical literature, and the size will be a huge increase.

## **User queries / competency queries for the use case**

Predicate represents the relationship between two nodes in the knowledge graph. Subject node and object node is the start and the end node of a pair of relationships between two concepts we extracted from the datasets listed above. Predicate, subject node, and object node are pre-defined by Unified Medical Language System.

- Find the concepts that cause 'Homelessness'
- Find the concepts that 'occurs in' 'Homeless Youth'
- Find the concepts that affect the 'Homeless family'
- Retrieve nodes connected by a 'METHOD\_OF' relationship to nodes named 'Suicide prevention' - identify various methods of interventions that are aimed at preventing suicide
- Find nodes that coexist with 'Suicide prevention' - look for other interventions, conditions, or factors that are typically associated with suicide prevention efforts.
- Retrieve nodes that have an 'AFFECTS' relationship with 'Homeless persons'. This aims to discover factors that have a positive or negative impact on the existence or wellbeing of homeless individuals.
- Identify what prevents 'Post-Traumatic' conditions. It is focused on interventions or factors that could help in preventing post-traumatic events.
- Finds nodes that have a 'COEXISTS\_WITH' relationship with 'Housing problem'. This could help in identifying other issues or conditions that often occur alongside housing problems.
- Identify nodes that have a 'CAUSES' relationship with 'Unemployment'. The goal is to explore different factors or conditions that may lead to or cause unemployment.