Identifying Economic Insecurity Among Psoriasis Patients Using Natural Language Processing and Unstructured Clinical Notes

OMNY HEALTH

Vikas Kumar, Amanda Mummert, Lawrence Rasouliyan, Amanda G. Althoff, Stella Chang, Stacey Long
OMNY Health, Atlanta, GA, United States

BACKGROUND

- Social determinants of health (SDoH) are the conditions in the environments where people are born, live, learn, work, play, worship, and age that affect a wide range of health, functioning, and quality-of-life outcomes and risks[1].
- SDoH can be broadly categorized into subdomains pertaining to a particular aspect of the environment. Economic insecurity is a common subdomain which can be defined as financial instability to the point of adversely affecting patient health.
- While clinical codes for SDoH exist in the International Classification of Diseases, Tenth Revision, they are often underutilized in clinical practice resulting in a lack structured SDoH data in clinical records.
- Therefore, a method for identifying SDoH for patients in unstructured data such as clinical notes would be valuable for risk stratification and subsequent intervention.

OBJECTIVE

- The objective of this study was to train and apply a natural language processing (NLP) model on clinical notes to identify economic insecurity among psoriasis patients and to compare findings to those using structured clinical codes.

METHODS

- The OMNY Health Platform was used to access electronic health record data for patients with International Classification of Diseases, Tenth Revision (ICD-10) codes for economic insecurity (Z59.4 - Z59.7, Z59.86, Z59.87, Z59.9; Table 1).

Table 1: ICD-10 Codes Used in this Study and their Descriptions

ICD-10 Code	Description
Z59.4	Lack of adequate food and safe drinking water
Z59.5	Extreme poverty
Z59.6	Low income
Z59.7	Insufficient social insurance and welfare support
Z59.86	Financial insecurity
Z59.87	Material hardship due to limited financial resources, not elsewhere classified
Z59.9	Problem related to housing and economic circumstances, unspecified

- Deidentified unstructured clinical notes from patients with these diagnosis codes were split into sentences.
- Random sentence samples were annotated for presence of economic insecurity and was split into training (70%), validation (10%), and test (20%) sets.
- This data was used to fine-tune an open-source, transformer-based NLP model.[2]
- Probability thresholds for model predictions were chosen to maximize precision at an acceptable recall level.
- The model was applied to non-templated sentences of clinical notes from a broader psoriasis population derived from 5 specialty dermatology networks between 2017-2019.
- The number of patients having a sentence indicating economic insecurity was calculated and compared to the count of clinical-code-positive patients.
- A random sample of positively predicted sentences was checked for accuracy.

RESULTS

- A total of 2,007 sentences were sampled and annotated.
- At the chosen probability threshold of 0.91, precision/recall observed in the test set were 0.91/0.66.
- Application of the model to the psoriasis population (50,969 patients) yielded 686 patients positive for economic insecurity, compared to zero patients having corresponding clinical codes.
- Of the 100 positive sentences checked for accuracy, 60 (60%) were true positives.
- Table 2 shows examples of positively-predicted sentences using the NLP model and their ground-truth labels.

Table 2: Examples of positively-predicted sentences using the NLP model and their ground-truth labels

Sentence	Predicted Label	Actual Label
"Will consider a biological agent after the new year due to patient currently doesnt have health insurance."	Positive	Positive
"I called patient assistance asking if patient can fill out a new application with updated financial information."	Positive	Positive
"Will contact representative to compare cash prices."	Positive	Negative
"Patient states received letter of approval however can not afford 65 per month for medication."	Positive	Positive
"He now is working full time and has commercial insurance."	Positive	Negative

CONCLUSIONS

- NLP can be a useful tool to identify SDoH among patients not otherwise detectable using ICD-10 codes.
- Further model training and/or threshold adjustment is needed to improve accuracy.
- Similar analyses were performed for SDoH domains of undereducation, and housing insecurity; patient yields were significantly lower (not reported here), suggesting that some SDoH domains may be underreported in specialty clinical notes relative to others.
- Further research is needed to improve models for additional SDoH domains and obtain predictions for both primary care and specialty patients.

REFERENCES

1. Utilization of Z Codes for Social Determinants of Health among Medicare Fee-for-Service Beneficiaries, 2019. https://www.cms.gov/files/document/z-codes-data-highlight.pdf.
2. Gu Y, Tinn R, Cheng H, et al. (2020). Domain-Specific Language Model Pretraining for Biomedical Natural Language Processing. arXiv:2007.15779; https://huggingface.co/microsoft/BiomedNLP-PubMedBERT-base-uncased-abstract-fulltext

CONTACT INFORMATION

Vikas Kumar | Senior Data Scientist | **OMNY Health**
Email: vikas@omnyhealth.com | **Website:** omnyhealth.com

Presented at ISPOR Europe; 12-15 November 2023; Copenhagen, Denmark