

Natural language processing to extract social risk factors influencing health

August 21 2023



Credit: CC0 Public Domain

Social risk factors such as financial instability and housing insecurity are increasingly recognized as influencing health. But unlike diagnosis codes, prescription information, lab or other test reports, social risk

factors do not adhere to standardized, controlled terminology in a patient's electronic medical record, making this information difficult to extract from the clinical notes where they typically are found.

A new study has found that a [natural language processing](#) (NLP) system developed by Regenstrief Institute and Indiana University Richard M. Fairbanks School of Public Health informaticians showed excellent performance when ported to a new health system and tested on more than six million clinical notes of patients seen in Florida. Performance was evaluated for generalizability and portability, defined as ease and accuracy when deploying the software in a new environment and of updating its use to meet the needs of new data.

"Social factors have a great impact on our health. It's not just the [medical care](#) that we receive, but it's also the places where we live, the places where we work and our access to food and transportation and other resources that have a major influence on our health," said Chris Harle, Ph.D., the Regenstrief and IU Fairbanks School faculty member who is senior author on the study.

"It's important for the clinicians and health systems providing medical care to know about people's social risk factors so when prescribing medications, ordering tests or planning to perform a procedure, they can better treat the whole person—perhaps with lower cost drugs or alternative sources for tests—and can also link them to services that help address their needs for a safe place to live and healthy food to eat."

In this study, the researchers' NLP rule-based model searched through text that physicians or other clinicians had written in the [clinical notes](#) of patients' [electronic health records](#), looking for key words or phrases that were likely to indicate difficulty with housing (for example: lack of permanent address) or financial needs (for example: inability to afford follow-up care) of patients at a health care system in a new and quite

different geographic area.

In spite of challenges (for example: name of a homeless shelter without indication of the facility's function or regional variation or local nuances in language), the research scientists verified that the NLP models, with relatively simple modifications, could deliver highly accurate performance as compared to the gold standard of human review.

"Is a patient diagnosed with diabetes? It's relatively easy to find that information in an electronic health record because the same words and codes are more likely to be used in [health systems](#) in central Indiana as are used in Florida or elsewhere in the U.S. But social risk factors don't have nearly as established and widely used words, phrases or codes to identify them. Therefore, it's harder to search through and determine a patient has a financial need than it is to say a patient has diabetes," said Dr. Harle.

"Our work is important for patients because ultimately their health is related to a variety of factors in their life, including [social factors](#). For example, are clinicians incorporating in their decision making a patient's ability to recover from a surgery as it's going to be different if they have stable housing versus unstable housing?"

"The more that we can disseminate and adapt natural language processing and other artificial intelligence methods that fully describe a patient to give clinicians a full 360 understanding of patients' needs, the better. If we can extract social information more efficiently, it's less costly. Then we can start to take what we'd call a population health perspective."

"So, if a health system can efficiently identify the patients who have housing instability—the population of patients who have this need—then the health care system may be able to employ a more proactive

population-based intervention to serve that whole group of people, connecting them, for example, to the housing services in the community or financial resources that might be available."

Dr. Harle, an information scientist and health services researcher who focuses on the design, adoption, use and value of health information systems, notes that this study was a team effort across multiple institutions of professionals who work in the clinical arena (including individuals who study how patients access and use care), public health, population health and health care administration as well as technically knowledgeable and skilled systems specialists.

"Bringing people together who have that diversity of understanding leads to pragmatically useful studies like this one," he said.

"Generalizability and portability of [natural language](#) processing system to extract individual social risk factors" is published in the *International Journal of Medical Informatics*.

More information: Tanja Magoc et al, Generalizability and portability of natural language processing system to extract individual social risk factors, *International Journal of Medical Informatics* (2023). [DOI: 10.1016/j.ijmedinf.2023.105115](#)

Provided by Regenstrief Institute

Citation: Natural language processing to extract social risk factors influencing health (2023, August 21) retrieved 28 April 2024 from <https://medicalxpress.com/news/2023-08-natural-language-social-factors-health.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private

study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.