

Do social determinants of health documented in clinical notes improve hospital prediction in home healthcare?Mollie Hobensack¹, Anahita Davoudi², Maxim Topaz^{1,2}¹ Columbia University School of Nursing ² VNS Health**Background.**

One in five patients in home healthcare are hospitalized per year in the United States¹. Social determinants of health (SDOH) influence more than 50% of health outcomes such as hospitalization². Home healthcare clinicians provide care within a patient's home, allowing them to assess the SDOH that influence a patient's daily life. Home healthcare clinicians document SDOH in their clinical notes that contribute to increased risk for hospitalization³. However, few studies have investigated if SDOH documented in clinical notes improve the identification of patients at risk for hospitalization. Thus, this study investigates if SDOH documented in clinical notes improves machine learning prediction of patients who may be hospitalized in the home health setting.

Methods.

This study uses data from VNS Health, a non-profit home healthcare agency located in New York, NY. Patients received home healthcare services from January 2015-December 2017. These data contain 86,866 episodes of home healthcare reflective of 65,593 unique patients. Structured data from the Outcome and Assessment Information Set (i.e., required Medicare document) and unstructured clinical notes were analyzed in this study.

In a previous study, we extracted the following SDOH using natural language processing from the home healthcare clinical notes: Social environment, Physical environment, Education and literacy, Food insecurity, Access to care, Housing and economic circumstances. We then created two datasets: a) a dataset with only structured data, and b) a dataset with structured data + SDOH extracted from clinical notes. Our outcome of interest was hospitalization. We added 5000 keywords using Term Frequency - Inverse Document Frequency (TF-IDF) to enhance model performance. We split the model into 80% training and 20% testing. We applied four machine learning models: Logistic Regression, LightGBM, and AutoGluon. We used 5 fold cross validation and evaluated model performance using the following metrics: F-score, Precision, Recall, and Accuracy.

Results.

	Logistic Regression		LightGBM		AutoGluon	
	Structured	+SDOH	Structured	+SDOH	Structured	+SDOH
Precision	0.32	0.32	0.53	0.55	0.45	0.44
Recall	0.57	0.58	0.68	0.63	0.60	0.62
Accuracy	0.77	0.77	0.87	0.87	0.84	0.83
F-score	0.41	0.41	0.59	0.59	0.51	0.51

Conclusion.

This is one of the first studies to examine the influence of SDOH documented in clinical notes on machine learning models that predict hospitalization in the home healthcare setting⁴. Our results suggest that incorporating SDOH from clinical notes shows almost no performance difference. However, we believe this is an important finding because it prompts further informatics research to discover how best to leverage SDOH documentation. The clinical value of SDOH information may not be in prediction but in identification that heightens awareness to social risk factors documented in previous home healthcare clinical notes. We encourage future researchers to engage home healthcare clinicians to identify how to leverage SDOH to support their clinical practice.

References.

1. *Report to Congress: Medicare Payment Policy.*; 2021. www.medpac.gov.
2. Braveman P, Gottlieb L. The Social Determinants of Health: It's Time to Consider the Causes of the Causes. *Public Health Reports.* 2014;129:19.
3. Hobensack M, Ojo M, Barron Y, et al. Documentation of hospitalization risk factors in electronic health records (EHRs): a qualitative study with home healthcare clinicians. 2022;ocac023. doi:<https://doi.org/10.1093/jamia/ocac023>
4. Hobensack M, Song J, Scharp D, Bowles KH, Topaz M. Machine learning applied to electronic health record data in home healthcare: A scoping review. *Int J Med Inform.* 2023;170:104978. doi:10.1016/J.IJMEDINF.2022.104978