Identifying Social Needs Among Underserved Populations: Development of a Social Risk Score in the Electronic Health Record

Elham Hatef, Hsien-Yen Chang, Thomas Richards, Christopher Kitchen, Janya Budaraju, Iman Foroughmand, Elyse Lasser, Jonathan Weiner

Submitted to: JMIR Formative Research
on: November 20, 2023

Table of Contents

Identifying Social Needs Among Underserved Populations: Development of a Social Risk Score in the Electronic Health Record

Elham Hatef[1, 2] MD, MPH; Hsien-Yen Chang[2] PhD; Thomas Richards[2] MS; Christopher Kitchen[2] MIS; Janya Budaraju[2]; Iman Foroughmand[2] MD; Elyse Lasser[2] DrPH; Jonathan Weiner[2] DrPH

[1]Division of General Internal Medicine, Department of Medicine, Johns Hopkins School of Medicine, Baltimore, Maryland, USA Baltimore US
[2]Center for Population Health Information Technology, Department of Health Policy and Management, Johns Hopkins Bloomberg School of Public Health Baltimore US

Corresponding Author:
Elham Hatef MD, MPH
Division of General Internal Medicine, Department of Medicine, Johns Hopkins School of Medicine, Baltimore, Maryland, USA
624 N. Broadway
Room 502
Baltimore
US

Abstract

Background: Some existing predictive models have utilized the available data on social needs and/or social determinants of health (SDOH) challenges to predict health-related social needs or the need for various social service referrals. Despite these one-off research and pilot efforts, the work to date suggests that many technical and organizational challenges must be surmounted before SDOH-integrated solutions can be implemented on an ongoing, wide scale within most US healthcare organizations.

Objective: To retrieve available information in the electronic health record (EHR) relevant to the identification of persons with social needs, to link these with community-based measures, and to develop a social risk score for use within clinical practice to better identify patients at risk of having future social needs.

Methods: A retrospective study using EHR data (2016-2021) and data from the U.S. Census, American Community Survey. Predictors of interest included demographics, previous healthcare utilization, diagnostic/comorbidity, previously identified social needs, and neighborhood characteristics as reflected by the area deprivation index. The outcome variable was a binary indicator reflecting the likelihood of the presence of a patient having social needs. We applied a generalized estimating equation approach, adjusting for patient-level risk factors, the possible effect of geographically clustered data, and the effect of multiple visits for each patient.

Results: The model performance in predicting prospective social needs was acceptable (AUC:0.702, 95% CI: 0.699-0.705). Previous social needs (OR: 3.285, 95% CI: 3.237-3.335) and emergency department visits (OR: 1.659, 95% CI: 1.634-1.684) were the strongest predictors of future social needs.

Conclusions: Our model provides an opportunity to make use of available EHR data in conjunction with community-level data, to help identify patients with high social needs, for further assessment and or/referral to community-based organizations. Our proposed social risk score could help identify the subset of patients who would most benefit from further social needs screening and data collection to avoid potentially more burdensome primary data collection on all patients in a target population of interest.

(JMIR Preprints 20/11/2023:54732)
DOI: https://doi.org/10.2196/preprints.54732

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?
Please make my preprint PDF available to anyone at any time (recommended).
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
✓ **No, I do not wish to publish my submitted manuscript as a preprint.**

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✔ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Identifying Social Needs Among Underserved Populations: Development of a Social Risk Score in the Electronic Health Record

Elham Hatef, MD, MPH,[1,2], Hsien-Yen Chang, PhD[2], Thomas Richards, MS[2], Christopher Kitchen, MS,[2] Janya Budaraju[2], Iman Foroughmand, MD[2], Elyse Lasser, DrPH[2], Jonathan P. Weiner, DrPH[2]

[1]Division of General Internal Medicine, Department of Medicine, Johns Hopkins School of Medicine, Baltimore, Maryland, USA

[2]Center for Population Health Information Technology, Department of Health Policy and Management, Johns Hopkins Bloomberg School of Public Health

Corresponding Author

Elham Hatef, MD, MPH, FACPM

Center for Population Health IT, Department of Health Policy and Management Johns Hopkins Bloomberg School of Public Health

624 N. Broadway, Room 502, Baltimore, Maryland 21205

Office Phone: 443-287-2284

Email: ehatef1@jhu.edu

Word Count: 3347

ABSTRACT

Background: Patients with unmet social needs and social determinants of health (SDOH) challenges continue to face a disproportionate risk of increased prevalence of disease, healthcare utilization, higher healthcare costs, and worse outcomes. Thus, targeted interventions to address social needs and SDOH challenges are necessary to overcome widespread health disparities. Some existing predictive models have utilized the available data on social needs and/or SDOH challenges to predict health-related social needs or the need for various social service referrals. Despite these one-off research and pilot efforts, the work to date suggests that many technical and organizational challenges must be surmounted before SDOH-integrated solutions can be implemented on an ongoing, wide scale within most US healthcare organizations.

Objectives: We aimed to retrieve available information in the electronic health record (EHR) relevant to the identification of persons with social needs and to develop a social risk score for use within clinical practice to better identify patients at risk of having future social needs.

Methods: We conducted a retrospective study using EHR data (2016-2021) and data from the U.S. Census, American Community Survey. We developed a prospective model, using current year-1 risk factors to predict future year-2 outcomes, within four two-year cohorts (i.e., 2016-2017, 2017-2018, 2018-2019, and 2019-2020). Predictors of interest included demographics, previous healthcare utilization, diagnostic/comorbidity, previously identified social needs, and neighborhood characteristics as reflected by the area deprivation index. The outcome variable was a binary indicator reflecting the likelihood of the presence of a patient having social needs. We applied a generalized estimating equation approach, adjusting for patient-level risk factors, the possible effect of geographically clustered data, and the effect of multiple visits for each patient.

Results: The study population of 1,852,228 patients included middle-aged (mean age of 53.76 to 55.95 years old across study cohorts) white (63.49% to 64.79%) female (61.62% to 62.07%) patients from neighborhoods with high socio-economic status (mean area deprivation index percentile of 28.76 to 30.31). 8.28% to 11.55% of patients across the study cohorts had at least one social need documented in their EHR with safety issues and economic challenges (i.e., financial resource strain, employment, and food insecurity) being the most common documented social needs (4.71% and 3.14% of overall patients, respectively). The model had an area under the curve (AUC) of 0.702 (95% Confidence Interval (CI): 0.699-0.705) in predicting prospective social needs in the overall study population. Previous social needs (OR: 3.285, 95% CI: 3.237-3.335) and emergency department visits (OR: 1.659, 95% CI: 1.634-1.684) were the strongest predictors of future social needs.

Conclusion: Our model provides an opportunity to make use of available EHR data to help identify patients with high social needs, for further assessment and or/referral to community-based organizations. Our proposed social risk score could help identify the subset of patients who would most benefit from further social needs screening and data collection to avoid potentially more burdensome primary data collection on all patients in a target population of interest.

Keywords: Social Needs, Social Determinants of Health, Predictive Modeling, Electronic Health Records, Population Health.

INTRODUCTION

Addressing social needs and social determinants of health (SDOH) challenges in the healthcare systems has emerged as a key component of addressing health disparities.[1] Patients with unmet social needs continue to face a disproportionate risk of increased prevalence of disease, healthcare utilization, higher healthcare costs, and worse outcomes across a range of health-related domains.[2-5] Thus, health disparities cannot be resolved through traditional clinical interventions in the healthcare systems. Targeted interventions to address social needs and SDOH challenges, especially among minority populations are necessary to overcome widespread disparities.[6]

The use of coding systems such as International Classification of Diseases 10th Edition (ICD-10) codes for social needs (i.e., Z-codes) has increased in recent years suggesting that clinicians and provider organizations are increasingly aware of social needs and SDOH challenges and the importance of screening for and documenting such needs.[7] Social needs are also extensively documented in the unstructured electronic health records (EHR) such as free-text provider notes.[8-12] Moreover, the rapid adoption of EHRs nationwide and the creation of associated health information technology (HIT) tools have made it possible to use this growing body of data on social needs and SDOH challenges in risk prediction and adjustment models.[13-21]

Some existing predictive models utilize EHR and/or administrative claims data on social needs to predict patterns of healthcare utilization, cost, and health outcomes.[16,18-20,22] Population-level data on community characteristics is a key component of understanding and addressing SDOH challenges and their impact on healthcare utilization, cost, and outcomes.[23,24] Therefore, some EHR-based models have developed linkages to these population-level data to better account for community-level information in their risk predictions.[17,21,25-27] Some existing models have also

utilized the available data on social needs and/or SDOH challenges to predict health-related social needs[14,15] or the need for various social service referrals.[13] Despite these one-off research and pilot efforts, the work to date suggests that many technical and organizational challenges must be surmounted before SDOH-integrated HIT solutions can be implemented on an ongoing, wide scale within most US healthcare organizations.

Using both patient- and population-level data, we sought to develop a social predictive risk score based entirely on electronic information readily available within most healthcare delivery systems. Predictive models such as this could help providers to systematically identify patients at risk of having future social needs, who represent likely targets for further in-depth assessment of their social needs and ultimately potential referral to community-based organizations (CBOs) to address such needs. Using a systematic electronic case-finding screening approach such as this would help the healthcare systems to avoid burdensome and inefficient social needs assessment (e.g., primary data collection from every patient at every visit).

METHODS

Data Sources

This was a retrospective study using the Johns Hopkins Healthcare System (JHHS) EPIC-based EHR structured data from July 2016 to June 2021. Based on the patient's home address during the in-scope study periods, we linked community-level data (at the census block group level) from U.S. Census American Community Survey (ACS), 2018 5-year cohort.[28] We developed a prospective model (using current year-1 risk factors to predict future year-2 outcomes) within four such two-year cohorts (i.e., 2016-2017, 2017-2018, 2018-2019, and 2019-2020). Each cohort contained model predictors in the first year (2016 in the first, 2017 in the second, 2018 in the third, and 2019 in the fourth cohort) and model outcomes in the second year (2017 in the first, 2018 in the second, 2019 in the third and 2020 in the fourth cohort). The overall data was randomly split into training and validation datasets (80% of the data was used for model development while the remaining 20% was used for validation). The final model was applied to the 2020-2021 cohort to evaluate its accuracy.

Study Population

Adult patients who were at least 18 years old at the time of entering the observation period and were alive at the end of the observation, had at least 1 eligible encounter in the first and second year of each study cohort, and a valid address for linkage to population-level data were included in this study.

Variable Selection

We identified variables with the highest potential impact on the health and social well-being of minority populations through a review of the literature and consultation with minority health, population health, and social needs/SDOH experts at JHHS. We also sought input from primary care

providers and frontline workers such as social workers and care managers, representatives of community-based organizations, and patients/caregivers.

We identified a comprehensive list of predictors of interest available within the EHR's structured data including various patient and community-level characteristics as well as healthcare utilization measures (Table 1).[29] [30]

Table 1. Predictors of Interest Available Within the Electronic Health Record Structured Data for Inclusion in the Generalized Estimating Equation Model Predicting Prospective Social Needs for Patients at Johns Hopkins Health System Between 2016-2021.

Variable	Description
Demographics	
Age	
Gender	
Race	
Preferred Language	
Need for an Interpreter	
Previous Healthcare Utilization	
In-patient admissions	
Emergency Department Visits	
Previous Social Needs	
ICD-10 Codes	Documented using relevant ICD-10 codes
EPIC Wellness Registry	Documented in other structured social needs assessment fields, presented in the Wellness Registry Table
Clinical Characteristics (Derived from the Johns Hopkins ACG System Version 12.0 ,[29] a widely used population-based predictive modeling/case-finding methodology)	
Number of Chronic Conditions	
Medication Active Ingredients	
Resource Utilization Bands	Represents expected future utilization based on current morbidities.
Neighborhood Characteristics (Associated with the person's residence of longest duration)	
Area Deprivation Index	A composite measure allowing for ranking of neighborhoods across the country by their socioeconomic disadvantage, reported at census block group level[30]

ACG: Adjusted Clinical Group, ICD-10: International Classification of Diseases 10th Edition.

To develop the variable on previous social needs we obtained any ICD-10 codes presenting social needs using the *Compendium of Medical Terminology Codes for Social Risk Factors* developed by the *Social Interventions Research and Evaluation Network (SIREN)*[31] or any information on social

needs available in the JHHS-EHR Wellness Registry, a data mart table in EPIC storing information related to general patient health, consolidated from many subject areas including social history and risk scores. After reviewing the classification of the ICD-10 codes by *SIREN* we developed 13 subdomains and 5 domains of social needs (Figure 1).

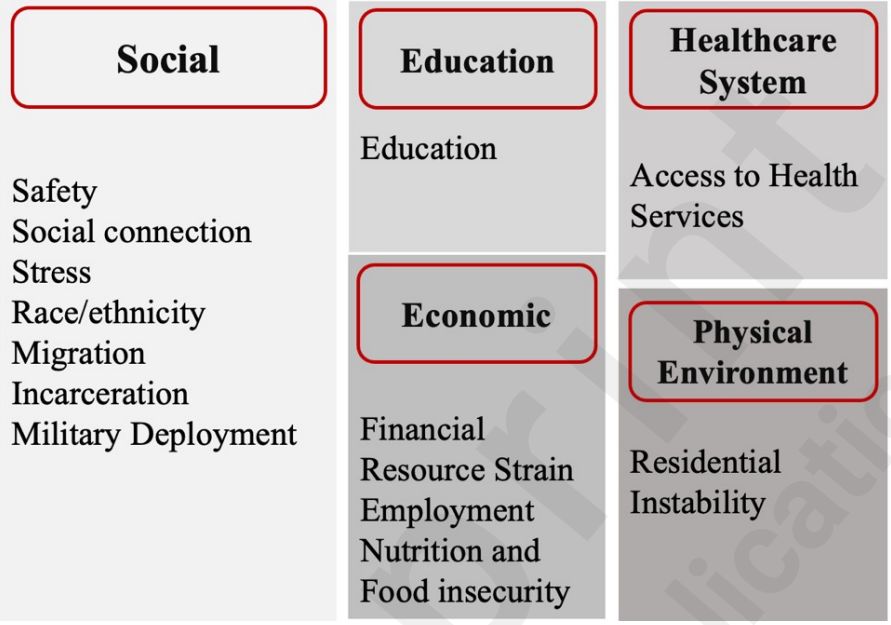

Figure 1: The Framework for Classification of Social Needs Identified in the Electronic Health Record Structured Data.

We reviewed the ICD-10 codes and mapped each to a unique social need subdomain. We also reviewed available information on social needs in EPIC Wellness Registry and selected variables corresponding to one of the 13 subdomains of social needs. We collapsed responses available for each variable to generate a binary variable (Yes/No indicator), suggesting the presence or absence of a social need. We defined previous social needs as a binary variable (Yes/No indicator), suggesting the presence or absence of any corresponding mapped ICD-10 codes or any corresponding social needs identified in the EPIC Wellness Registry to one or more of the 13 social needs subdomains. We defined the outcome as a binary indicator of having a social need in the second year of each cohort (using the same logic as for the development of the predictor of social needs).

The ACG Systems RUBs represent a simplified population segmentation system based on the overall

morbidity burden of each patient. Representing expected future utilization based on current morbidities, the measure is calculated using all available ICD-10 codes of a person in the EHR during year-1, ranked from low to high according to the expectations of resources used during year 2.

Statistical Analysis

We used a generalized estimating equation (GEE) model to predict prospective social needs, adjusting for the effect of the geographically clustered data as well as the effect of multiple visits for each patient (the records were clustered at the patient and 5-digit Zip Code level). The model selection was based on the goodness of fit test for GEE modeling, and the final risk score was composed using the variables identified as having the highest impact in the GEE model. We also validated the model using multiple denominators to ensure generalizability and retrained and tested the model for each subpopulation of interest (e.g., 65+ years old population, racial/ ethnic minority populations, and those living in the most and least disadvantaged neighborhoods).

The institutional review board of the Johns Hopkins Bloomberg School of Public Health reviewed and approved this study as exempt. The board approved the EHR data extraction for the secondary analysis of de-identified data.

RESULTS

Demographics

The final study population included 1,852,228 patients in total. To be included in the sample the patients had to be in at least one of four 2-year study cohorts (Table 2). Characteristics of patients across the study cohorts were comparable. Study cohorts included mostly middle-aged (mean age of 53.76 to 55.95 years old across study cohorts) white (63.49% to 64.79%) female (61.62% to 62.07%) patients from neighborhoods with high socio-economic status (mean ADI percentile of 28.76 to 30.31).

About 8.28% to 11.55% of patients across the study cohorts had at least one social need documented in the ICD-10 codes or EPIC Wellness Registry with safety issues and economic challenges (i.e., financial resource strain, employment, and food insecurity) being the most common documented social needs (4.71% and 3.14% of overall patients, respectively). (Refer to Supplement Table 1 for details on the social needs domains across study cohorts). About 18.67% to 21.18% of patients across the study cohorts had high or very high resource utilization bands, indicative of having a high disease burden, as reflected by many serious comorbidities.

Table 2. Characteristics of the Study Population for the Development of the Social Risk Score Using Electronic Health Record Data at Johns Hopkins Health System Between 2016-2021: Overall and by 2-Year Enrollment Cohorts.*

Variables	Study Cohort				Overall (no. 1,852,228)
	2016-2017 (no. 448,666)	2017-2018 (no. 442,366)	2018-2019 (no. 450,426)	2019-2020 (no. 510,770)	
Age – mean (standard deviation)					
	53.76 (17.65)	54.63 (17.43)	55.18 (17.25)	55.95 (17.34)	54.92 (17.43)
Gender – Female, no. (%)					
	278,488 (62.07%)	274,330 (62.01%)	278,574 (61.85 %)	314,741 (61.62%)	1,146,133 (61.88 %)
Race – no. (%)					
White	290,688 (64.79%)	285,558 (64.55%)	289,611 (64.30%)	324,279 (63.49%)	1,190,136 (64.25%)
Black	105,746 (23.57%)	103,879 (23.48%)	105,156 (23.35%)	118,989 (23.3%)	433,770 (23.42%)
Other	52,232 (11.64)	52,929 (11.97)	55,659 (12.35)	67,502 (13.21)	228,322 (12.33)
Preferred Language – English, no. (%)					
	430,331 (95.91%)	423,396 (95.71%)	430,024 (95.47%)	465,173 (91.07%)	1,748,924 (94.42%)
Interpreter Needed – Yes, no. (%)					
	9,356 (2.09%)	9,715 (2.20%)	10,632 (2.36%)	12,928 (2.53%)	42,631 (2.30%)
Area Deprivation Index National Rank – mean (standard deviation)†					
	30.31 (23.6)	29.93 (23.46)	29.65 (23.24)	28.76 (22.81)	29.63 (23.27)
Healthcare Utilization – no. (%)					
Any In-Patient Admission	39,364 (8.77%)	37,154 (8.4%)	38,407 (8.53%)	41,149 (8.06%)	156,074 (8.43%)
Any Emergency Department Visits	71,737 (15.99%)	69,436 (15.70%)	69,649 (15.46%)	77,308 (15.14%)	288,130 (15.56%)
Previous Social Needs – no. (%)					
Year 1 - ICD-10 Codes	37,137 (8.28%)	50,332 (11.38%)	52,037 (11.55%)	57,170 (10.21%)	191,693 (10.35%)
Year 1 - EPIC Wellness Registry	0 (0%)	0 (0%)	2 (0%)	6,663 (1.3%)	6,665 (0.36%)
Year 2 - ICD-10 Codes	45,272 (10.09%)	46,587 (10.53%)	42,475 (9.43%)	42,902 (8.40%)	177,236 (9.57%)
Year 2 - EPIC Wellness Registry	0 (0%)	3 (0%)	5,596 (1.24%)	11,079 (2.17%)	16,678 (0.90%)

Clinical Characteristics – mean (standard deviation)‡					
No. of Chronic Conditions	2.3 (2.7)	2.4 (2.8)	2.4 (2.8)	2.2 (2.7)	2.3 (2.7)
No. of Medication Active Ingredients	2.2 (6.1)	2.2 (6.1)	2.3 (6.3)	2.1 (5.9)	2.2 (6.1)
Resource Utilization Bands – no. (%)‡					
No or only invalid diagnosis	2,2507 (5.02%)	20,273 (4.58%)	16,897 (3.75%)	35,288 (6.91%)	94,965 (5.13%)
Healthy users	44,731 (9.97%)	43,408 (9.81%)	40,996 (9.1%)	53,142 (10.4%)	182,277 (9.84%)
Low resource utilization	59,400 (13.24%)	55,221 (12.48%)	57,667 (12.8%)	69,566 (13.62%)	241,854 (13.06%)
Moderate resource utilization	232,748 (51.88%)	231,680 (52.37%)	239,473 (53.17%)	257,424 (50.4%)	961,325 (51.9%)
High resource utilization	59,979 (13.37%)	60,841 (13.75%)	62,740 (13.93%)	63,338 (12.4%)	246,898 (13.33%)
Very high resource utilization	29,301 (6.53%)	30,943 (6.99%)	32,653 (7.25%)	32,012 (6.27%)	124,909 (6.74%)

ICD-10: International Classification of Diseases 10th Edition.
*All characteristics are reported based on the first year of each cohort unless otherwise indicated.
†Neighborhood characteristics for the person's residence of longest duration reported as a percentile of national rank.[30]
‡These clinical measures are derived from the Johns Hopkins Adjusted Clinical Group (ACG) System Version 12.0. Resource Utilization Band represents expected future utilization based on current morbidities.[29]

Generalized Estimating Equation Modeling

Details of the GEE models are presented in Table 3. The GEE model had an area under the curve (AUC) of 0.702 (95% Confidence Interval (CI): 0.699-0.705) in predicting prospective social needs in the overall study population. The strongest predictors of future social needs in the whole population in descending order were: social needs documented in the EHR during the previous year period (OR: 3.285, 95% CI: 3.237-3.335), 1+ ED visit in the previous periods (OR: 1.659, 95% CI: 1.634-1.684), and very high RUB measure indicative of a significant morbidity burden (OR: 1.371, 95% CI: 1.317-1.427).

To help assess bias and applicability to various sub-population, we identified comparable performance for models of select subgroups; with AUCs of 0.701 (95% CI: 0.696-0.706), 0.711 (95% CI: 0.706-0.716), and 0.711 (95% CI: 0.708-0.714) respectively for 65+ years old, black patients, and those living in the most disadvantaged neighborhoods. The strongest predictor of future social needs in the study subpopulations remained the previous social needs with OR of 3.043 (95% CI: 2.96-3.128), 2.9 (95% CI: 2.824-2.977), and 3.390 (95% CI: 3.334-3.447) among 65+ years old population, black patients, and those living in the most disadvantaged neighborhoods, respectively.

To ensure that most common social needs (i.e., safety and economic challenges) were not the main drivers of the model performance we performed a sensitivity analysis and ran the model after excluding patients with any social needs in subdomains of safety and/or economic challenges. This resulted in a slightly better-performing model with an AUC of 0.768 (95% CI: 0.763-0.773). In this instance, previous social needs were by far the most significant predictor

(OR: 11.857, 95% CI: 11.521-12.202), followed by ED visits (OR: 1.916, 95% CI: 1.865-1.969), need for an interpreter (OR:1.528, 95% CI: 1.397-1.672) and black race (OR: 1.307, 95% CI: 1.273-1.342). Contrary to the analyses performed with all social needs included, patients with higher RUBs had slightly less risk of increased social needs, and lower RUBs had slightly greater protective value regarding social needs (Refer to Supplement Table 2 for details of the GEE model).

Table 3. Generalized Estimating Equation Model Predicting Prospective Social Needs for Patients at Johns Hopkins Health System Using Electronic Health Record Data Between 2016-2021: Overall Model and Models for Selected Subpopulations.[*]

Variable	Overall Population	65+ years old Population	Racial Groups		Neighborhood Characteristics	
			White	Black	Most Disadvantaged	Least Disadvantaged
Area Under the Curve						
	0.702 (0.699-0.705)	0.701 (0.696-0.706)	0.689 (0.685-0.693)	0.711 (0.706-0.716)	0.711 (0.708-0.714)	0.667 (0.663-0.677)
Age – Years						
	0.994 (0.993-0.994)	1.002 (1.001-1.004)	0.994 (0.994-0.995)	0.993 (0.992-0.994)	0.991 (0.991-0.992)	1.002 (1.001-1.002)
Gender – Male (ref: female)						
	0.993 (0.981-1.004)	0.922 (0.903-0.941)	0.964 (0.95-0.979)	1.068 (1.046-1.092)	1.004 (0.991-1.017)	0.952 (0.927-0.977)
Race – Black (ref: White)						
	1.125 (1.11-1.141)	1.149 (1.118-1.182)	---	---	1.127 (1.111-1.144)	1.137 (1.087-1.190)
Preferred Language – English (ref: missing, others or sign language)						
	1.061 (1.026-1.097)	1.293 (1.221-1.371)	1.034 (0.991-1.08)	1.112 (1.043-1.187)	1.030 (0.993-1.069)	1.217 (1.125-1.317)
Interpreter Needed – Yes (ref: no or missing)						
	1.179 (1.122-1.238)	1.266 (1.156-1.386)	1.006 (0.908-1.114)	0.816 (0.679-0.981)	1.148 (1.089-1.211)	1.085 (0.944-1.247)
Area Deprivation Index National Rank – Percentile[†]						
	1.005 (1.005-1.005)	1.001 (1.001-1.002)	1.003 (1.003-1.004)	1.007 (1.006-1.007)	1.006 (1.006-1.006)	0.983 (0.979-0.988)
Healthcare Utilization						
Any In-Patient Admission	1.017 (0.993-1.042)	1.077 (1.03-1.126)	1.064 (1.03-1.1)	0.987 (0.947-1.029)	0.986 (0.960-1.013)	1.131 (1.071-1.194)
Any Emergency Department Visits	1.659 (1.634-1.684)	1.539 (1.495-1.584)	1.691 (1.656-1.728)	1.627 (1.587-1.668)	1.636 (1.608-1.664)	1.555 (1.503-1.608)
Previous Social Needs						
	3.285 (3.237-3.335)	3.043 (2.96-3.128)	3.459 (3.391-3.529)	2.9 (2.824-2.977)	3.390 (3.334-3.447)	2.775 (2.677-2.877)
Clinical Characteristics[‡]						
No. of Chronic Conditions	1.066 (1.064-1.069)	1.08 (1.07-1.08)	1.070 (1.067-1.074)	1.063 (1.058-1.068)	1.070 (1.067-1.073)	1.066 (1.058-1.073)
No. of Medication Active Ingredients	0.997 (0.996-0.998)	0.99 (0.99-1.00)	0.997 (0.995-0.998)	0.997 (0.995-0.999)	0.996 (0.995-0.997)	1.001 (0.999-1.004)

Resource Utilization Bands – (ref: no or only invalid diagnosis)[‡]						
Healthy users	0.838 (0.808-0.869)	0.792 (0.734-0.854)	0.836 (0.798-0.876)	0.785 (0.729-0.844)	0.837 (0.802-0.872)	0.787 (0.729-0.850)
Low resource utilization	0.855 (0.826-0.885)	0.837 (0.778-0.9)	0.856 (0.819-0.895)	0.846 (0.792-0.904)	0.849 (0.816-0.882)	0.847 (0.787-0.911)
Moderate resource utilization	0.964 (0.935-0.994)	0.935 (0.876-0.997)	0.942 (0.906-0.981)	1.008 (0.952-1.069)	0.978 (0.945-1.012)	0.908 (0.848-0.972)
High resource utilization	1.259 (1.217-1.302)	1.226 (1.144-1.314)	1.238 (1.184-1.294)	1.328 (1.246-1.415)	1.285 (1.238-1.335)	1.113 (1.030-1.204)
Very high resource utilization	1.371 (1.317-1.427)	1.218 (1.128-1.315)	1.291 (1.225-1.361)	1.514 (1.408-1.629)	1.420 (1.359-1.485)	1.136 (1.034-1.249)

[*] The first row reflects Area Under the Curve. And all subsequent rows reflect odds ratios and 95% confidence intervals.
[†] Neighborhood characteristics for the person's residence of longest duration reported as a percentile of national rank.[30]
[‡]These clinical measures are derived from the Johns Hopkins Adjusted Clinical Group (ACG) System Version 12.0. Resource Utilization Band represents expected future utilization based on current morbidities.[29]

DISCUSSION

Overview

Achieving a comprehensive assessment of a person's health and addressing health disparities goes beyond just documenting clinical diseases and medical interventions. We must also capture, standardize, analyze, and report reliable information on social needs and SDOH challenges within operational clinical decision support systems that are built into EHRs. Moreover, the rapid change towards "value-based" healthcare models[32] in the U.S. has required the incorporation of social needs and SDOH contexts and frameworks to ensure that the healthcare systems and health plans equitably address the needs of minority and disadvantaged communities.[33,34] For these value-based models to perform well it is critical that clinical and social interventions are aligned, and that no financial disincentives are imposed on providers who disproportionately serve minority and disadvantaged patients.[33,34]

To achieve these goals, applied research is needed to identify optimal solutions for the effective collection and application of social needs and SDOH information within EHRs; link provider-based data to community-level data describing the characteristics of patients' neighborhoods; and anchor such information to the providers' digital workflow. This approach will provide the vehicle for harnessing social needs and SDOH data to target interventions at the point of care (e.g., referrals of an individual); the health delivery system level (e.g., hiring a social worker in the clinic); or, the community (e.g., building or strengthening community-based initiatives).[35] To avoid burdensome and inefficient social needs assessment and data collection, it is essential to develop automated screening tools using EHR or community-based data to help identify the

subset of patients who would most benefit from social needs assessment and data collection.

Several EHR-based screening tools for social needs assessment have been piloted in recent years, and results have shown these tools to be effective in determining social needs and SDOH challenges.[36,37] However, the feasibility of such tools remains unclear, with healthcare systems needing to dedicate considerable time and budget to train and educate staff and manage workloads.[21] Our proposed social risk score aimed to reduce the burden of this process and increase its accuracy by identifying patients at high risk of having any social needs for more efficient screening.

Comparison with previous evidence

Our proposed model was based on a large and diverse dataset of patients in the JHHS-EHR. The AUC of our model in predicting prospective social needs was 0.702 (95% CI: 0.699-0.705) in the overall study population. This AUC may be the result of many instances of false negatives related to documentation of the social needs in structured EHR data. At the time of completing this study social needs screening and referral were not a common practice at our institutions. Thus, we expect many patients with social needs did not get a proper screening and documentation of such needs. The new mandate established by the Centers for Medicare and Medicaid Services requires the hospitals to report to the Inpatient Quality Reporting (IQR) program two brand new measures of social needs (i.e., number of patients screened for social needs and number of patients identified with selected social needs).[38] We expect the healthcare systems to establish more systematic and uniform processes for screening and documentation of social needs. This effort will increase the volume of data on social needs in the EHR, which

results in a better performance of our models in the future.

Our findings were comparable with those in Holcomb et al's study,[14] where they predicted health-related social needs using EHR and community-level data and machine learning modeling for Medicare and Medicaid beneficiaries participating in the Accountable Health Communities project. Their models performed relatively well with AUCs ranging from 0.59 to 0.68 for patients with different domains of social needs. Another notable mention was Byrne et al's study[15] where they used EHR data including responses to the Veteran's Health Administration's Homelessness Screening Clinical Reminder survey to develop and test predictive models of housing instability and homelessness. All their models performed well with the random forests models performing better than the logistic regression models for both the housing instability (85.4 vs 78.3) and homeless (91.6 vs 87.1) outcomes. Lastly, Kasthurirathne et al.[13] built random forest decision models to predict the need for social work referral using clinical and population-level data on SDOH challenges. The performance of the model ranged from an AUC of 0.713 for the model using both clinical and SDOH data to 0.731 for the model using clinical data.

Moreover, our results demonstrated that the most significant predictive factor for having prospective social needs was the documentation of prior social needs. Our study also found associations between prospective social needs with previous ED visits and morbidity-related high resource utilization presented in RUBs. Lastly, our model showed a minimal increased risk of social needs in association with characteristics of the neighborhood of residence, presented as the ADI measure. This finding was similar to the results of the Nguyen et al[39] study, where

they identified a small statistically significant association between the ADI and total score on social needs from the *Health Leads Social Needs Survey* among pediatric patient-families receiving primary care at a large academic institution were enrolled. A review by Chen et al[21] also indicated the low success of the integration of population-level data for predictive modeling and risk stratification purposes including the prediction for social-related service referrals,[13] in contrast to the performance of models using individual-level data in referrals to a social worker. [40] Overall, these findings indicated that individual characteristics played a more crucial role in predicting future social needs than neighborhood characteristics.

Clinical implications

The implementation of an EHR-based social risk score such as the one we developed would have many implications for clinicians and practice organizations. At the point of care, our social risk score could be integrated directly with EHR-derived data warehouses, thus the proposed risk score could be leveraged to allow clinicians to tailor more personalized care and modulate care coordination efforts. This personalized social risk score could also help support tools tailored to the needs of patients, which would empower the navigation of available social services.[35] At the health delivery system level, the linkages of geo-derived databases could improve the assessment of social needs and SDOH challenges for health systems and provide opportunities for longer-term plans to address those factors in current and future care management programs. Also, the merged clinical and non-clinical databases could enable providers to follow patients with social needs and SDOH challenges over time through their interaction with the health system. At the community level, the social risk score and social needs/SDOH data could enhance the care coordination efforts by integrating community-level data in the clinical decision support

tools and coordinated interventions. [41]

Limitations

Several limitations existed in our project. Firstly, EPIC wellness registry data was essentially non-existent between the years 2016 to 2018, which in addition to the possibility of ICD-10 codes being underutilized by providers might lead to an underrepresentation of social needs in our study population. Secondly, the Latino/Hispanic patients were under-represented in our study population, which may have impacted the generalizability of the proposed model. Furthermore, data on racial/ethnic minorities such as American Indians and Alaskan Natives, Native Hawaiian, and Other Pacific Islanders or multiracial individuals were limited in our dataset. Another important factor that might have affected our results was the potentially more frequent screening of social needs in females, ethnic and racial minority populations, those with higher disease burden, and super utilizers of healthcare services,[42] leading to biased results for these individuals. Misclassification and/or inconsistency in documenting social needs in the EHRs could influence our results.

Moreover, the use of EHR data as the sole source of information limited our data to services provided to the patients across the JHHS facilities and did not include other services outside JHHS. Also, we used the patient's home address to link the EHR data to the ACS community-level data. Thus, we did not include patients with a missed or invalid home address. This may have resulted in missing some patients with social needs such as residential instability. Finally, our final model was applied to the 2020-2021 cohort to evaluate its accuracy, which included

patient encounters during the first peak of the COVID-19 pandemic. While some sub-populations of patients experienced more social needs during the pandemic the challenges that the healthcare systems faced during this period impacted the screening and documentation of social needs and may have resulted in missed data on social needs for this cohort.

CONCLUSION

Screening for social needs and SDOH challenges using available secondary EHR data represents an important step toward addressing health disparities more efficiently and improving patient care and population health. Our proposed model integrated community-level data with patient-level to arrive at a social risk score that can be utilized to systematically identify patients at increased risk of having future social needs, and thus appropriate for in-depth assessment of their social needs and potential referral to CBOs to address these needs. Our model identified previous social needs and high morbidity levels as the strongest predictors of future social needs. Missing data in the EHR reflecting past clinician documentation of social needs may have impacted the performance of our proposed model and the predictive accuracy of models like ours will likely increase as capture of such information becomes more commonplace. Future studies should focus on developing EHR-integrated clinical decision support (CDS) tools to make this information available in the providers' digital workflow and at the point of care.

ACKNOWLADGEMENT

This work was supported by a grant from the National Institute on Minority Health and Health Disparities (NIMHD, Grant Number R01MD015844-01). Its contents are solely the responsibility of the authors and do not necessarily represent the official NIMHD views.

DATA AVAILABILITY

Data underlying this article were extracted from the electronic health record at the study site and cannot be shared publicly for the privacy of individuals who participated in the study.

CONFLICT OF INTEREST

Johns Hopkins University (JHU) holds the copyright to the ACG System and receives royalties from the global distribution of the ACG system. This revenue supports a portion of the authors' salary. The authors are members of a group of researchers who develop and maintain the ACG System with support from JHU. The authors report no other conflicts of interest in this work.

REFERENCES

1. Byhoff E, Freund KM, Garg A. Accelerating the Implementation of Social Determinants of Health Interventions in Internal Medicine. *J Gen Intern Med* 2018; **33**(2): 223-5.
2. Odonkor CA, Esparza R, Flores LE, Verduzco-Gutierrez M, Escalon MX, Solinsky R, Silver JK. Disparities in Health Care for Black Patients in Physical Medicine and Rehabilitation in the United States: A Narrative Review. *PM R* 2021; **13**(2): 180-203.
3. Bharmal N, Rennick A, Shideler A, Blazel M, Jones R, Wilson C, Pfoh ER. Health-Related Social Needs: Which Patients Respond to Screening and Who Receives Resources? *J Gen Intern Med* 2023.
4. Bensken WP, Alberti PM, Stange KC, Sajatovic M, Koroukian SM. ICD-10 Z-Code Health-Related Social Needs and Increased Healthcare Utilization. *Am J Prev Med* 2022; **62**(4): e232-e41.
5. Kepper MM, Walsh-Bailey C, Prusaczyk B, Zhao M, Herrick C, Foraker R. The adoption of social determinants of health documentation in clinical settings. *Health Serv Res* 2023; **58**(1): 67-77.
6. Thornton RL, Glover CM, Cene CW, Glik DC, Henderson JA, Williams DR. Evaluating Strategies For Reducing Health Disparities By Addressing The Social Determinants Of Health. *Health Aff (Millwood)* 2016; **35**(8): 1416-23.
7. Bensken WP, Alberti PM, Baker MC, Koroukian SM. An Increase in the Use of ICD-10 Z-Codes for Social Risks and Social Needs: 2015 to 2019. *Popul Health Manag* 2023.
8. Feller DJ, Bear Don't Walk Iv OJ, Zucker J, Yin MT, Gordon P, Elhadad N. Detecting Social and Behavioral Determinants of Health with Structured and Free-Text Clinical Data. *Appl Clin Inform* 2020; **11**(1): 172-81.
9. Han S, Zhang RF, Shi L, et al. Classifying social determinants of health from unstructured electronic health records using deep learning-based natural language processing. *J Biomed Inform* 2022; **127**: 103984.
10. Rouillard CJ, Nasser MA, Hu H, Roblin DW. Evaluation of a Natural Language Processing Approach to Identify Social Determinants of Health in Electronic Health Records in a Diverse Community Cohort. *Med Care* 2022; **60**(3): 248-55.
11. Bejan CA, Angiolillo J, Conway D, et al. Mining 100 million notes to find homelessness and adverse childhood experiences: 2 case studies of rare and severe social determinants of health in electronic health records. *J Am Med Inform Assoc* 2018; **25**(1): 61-71.
12. Chapman AB, Jones A, Kelley AT, et al. ReHouSED: A novel measurement of Veteran housing stability using natural language processing. *J Biomed Inform* 2021; **122**: 103903.
13. Kasthurirathne SN, Vest JR, Menachemi N, Halverson PK, Grannis SJ. Assessing the capacity of social determinants of health data to augment predictive models identifying patients in need of wraparound social services. *J Am Med Inform Assoc* 2018; **25**(1): 47-53.
14. Holcomb J, Oliveira LC, Highfield L, Hwang KO, Giancardo L, Bernstam EV. Predicting health-related social needs in Medicaid and Medicare populations using machine learning. *Sci Rep* 2022; **12**(1): 4554.
15. Byrne T, Montgomery AE, Fargo JD. Predictive modeling of housing instability and homelessness in the Veterans Health Administration. *Health Serv Res* 2019; **54**(1): 75-85.
16. Pandya CJ, Hatef E, Wu J, Richards T, Weiner JP, Kharrazi H. Impact of Social Needs in Electronic Health Records and Claims on Health Care Utilization and Costs Risk-Adjustment Models Within Medicaid Population. *Popul Health Manag* 2022; **25**(5): 658-68.

17. Bhavsar NA, Gao A, Phelan M, Pagidipati NJ, Goldstein BA. Value of Neighborhood Socioeconomic Status in Predicting Risk of Outcomes in Studies That Use Electronic Health Record Data. *JAMA Netw Open* 2018; **1**(5): e182716.
18. Hall AG, Davlyatov GK, Orewa GN, Mehta TS, Feldman SS. Multiple Electronic Health Record-Based Measures of Social Determinants of Health to Predict Return to the Emergency Department Following Discharge. *Popul Health Manag* 2022; **25**(6): 771-80.
19. Mosen DM, Banegas MP, Benuzillo JG, Hu WR, Brooks NB, Ertz-Berger BL. Association Between Social and Economic Needs With Future Healthcare Utilization. *Am J Prev Med* 2020; **58**(3): 457-60.
20. Seligman B, Tuljapurkar S, Rehkopf D. Machine learning approaches to the social determinants of health in the health and retirement study. *SSM Popul Health* 2018; **4**: 95-9.
21. Chen M, Tan X, Padman R. Social determinants of health in electronic health records and their impact on analysis and risk prediction: A systematic review. *J Am Med Inform Assoc* 2020; **27**(11): 1764-73.
22. Hao S, Wang Y, Jin B, et al. Development, Validation, and Deployment of a Real-Time 30 Day Hospital Readmission Risk Assessment Tool in the Maine Healthcare Information Exchange. *PLoS One* 2015; **10**(10): e0140271.
23. Comer KF, Grannis S, Dixon BE, Bodenhamer DJ, Wiehe SE. Incorporating geospatial capacity within clinical data systems to address social determinants of health. *Public Health Rep* 2011; **126 Suppl 3**(Suppl 3): 54-61.
24. Khoury MJ, Galea S. Will Precision Medicine Improve Population Health? *Jama* 2016; **316**(13): 1357-8.
25. Vest JR, Ben-Assuli O. Prediction of emergency department revisits using area-level social determinants of health measures and health information exchange information. *Int J Med Inform* 2019; **129**: 205-10.
26. Hatef E, Ma X, Rouhizadeh M, Singh G, Weiner JP, Kharrazi H. Assessing the Impact of Social Needs and Social Determinants of Health on Health Care Utilization: Using Patient- and Community-Level Data. *Popul Health Manag* 2021; **24**(2): 222-30.
27. Chang HY, Hatef E, Ma X, Weiner JP, Kharrazi H. Impact of Area Deprivation Index on the Performance of Claims-Based Risk-Adjustment Models in Predicting Health Care Costs and Utilization. *Popul Health Manag* 2021; **24**(3): 403-11.
28. The United States Census Bureau. American community survey (ACS). Available at: https://Www.census.gov/programs-surveys/acs/. Accessed on May 2, 2023.
29. The Johns Hopkins ACG System. 2018. Available at: https://www.hopkinsacg.org. Accessed on May 2, 2023.
30. University of Wisconsin School of Medicine and Public Health. Area Deprivation Index . https://www.neighborhoodatlas.medicine.wisc.edu/. Accessed on May 9, 2023.
31. Arons A, DeSilvey S, Fichtenberg C, Gottlieb L. Documenting social determinants of health-related clinical activities using standardized medical vocabularies. *JAMIA Open* 2019; **2**(1): 81-8.
32. Centers for Medicare and Medicaid Services (CMS). What are the value-based programs? Available at:
https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-
Programs/Value-Based- Programs.html. Accessed on May 9, 2023.
33. National Academies of Sciences E, Medicine. Integrating Social Care into the Delivery of

Health Care: Moving Upstream to Improve the Nation's Health. Washington, DC: The National Academies Press; 2019.
34. Centers for Medicare and Medicaid Services. MACRA: Delivery system reform, Medicare payment reform.
Available at: https://www.cms.gov/medicare/quality-initiatives-patient-assessment-instruments/value-basedprograms/macra-mips-and- apms/macra-mips-and-apms.html. Accessed on May 9, 2023.
35. Predmore Z, Hatef E, Weiner JP. Integrating Social and Behavioral Determinants of Health into Population Health Analytics: A Conceptual Framework and Suggested Road Map. *Popul Health Manag* 2019; **22**(6): 488-94.
36. Buitron de la Vega P, Losi S, Sprague Martinez L, et al. Implementing an EHR-based Screening and Referral System to Address Social Determinants of Health in Primary Care. *Med Care* 2019; **57 Suppl 6 Suppl 2**: S133-s9.
37. Gold R, Bunce A, Cowburn S, et al. Adoption of Social Determinants of Health EHR Tools by Community Health Centers. *Ann Fam Med* 2018; **16**(5): 399-407.
38. Centers for Medicare & Medicaid Services (CMS) DoHaHSH. Medicare Program; Hospital Inpatient Prospective Payment Systems for Acute Care Hospitals and the Long-Term Care Hospital Prospective Payment System and Policy Changes and Fiscal Year 2023 Rates; Quality Programs and Medicare Promoting Interoperability Program Requirements for Eligible Hospitals and Critical Access Hospitals; Costs Incurred for Qualified and Non-Qualified Deferred Compensation Plans; and Changes to Hospital and Critical Access Hospital Conditions of Participation. 2022. https://www.federalregister.gov/documents/2022/08/10/2022-16472/medicare-program-hospital-inpatient-prospective-payment-systems-for-acute-care-hospitals-and-the (accessed August 4, 2023 2023).
39. Nguyen C, Popescu F, Sommer EC, Adams LB, Barkin S. The Utility of the Brokamp Area Deprivation Index as a Prescreen for Social Risk in Primary Care. *J Pediatr* 2022; **249**: 43-9.
40. Vest JR, Menachemi N, Grannis SJ, et al. Impact of Risk Stratification on Referrals and Uptake of Wraparound Services That Address Social Determinants: A Stepped Wedged Trial. *Am J Prev Med* 2019; **56**(4): e125-e33.
41. Castrucci B, Auerbach J. Meeting Individual Social Needs Falls Short Of Addressing Social Determinants Of Health. Health Affairs Blog, January 16, 2019. https://www.healthaffairs.org/do/10.1377/hblog20190115.234942/full/. Accessed on May 10, 2023.
42. Nohria R, Xiao N, Guardado R, Drainoni ML, Smith C, Nokes K, Byhoff E. Implementing Health Related Social Needs Screening in an Outpatient Clinic. *J Prim Care Community Health* 2022; **13**: 21501319221118809.

Supplementary Files

Figures

The Framework for Classification of Social Needs Identified in the Electronic Health Record Structured Data.