Contents lists available at ScienceDirect

# SSM - Population Health

# Identifying individual social risk factors using unstructured data in electronic health records and their relationship with adverse clinical outcomes

S. Michaela Rikard [a], Bommae Kim [b], Jonathan D. Michel [b], Shayn M. Peirce [a,c], Laura E. Barnes [d], Michael D. Williams [c,e,*]

[a] *Department of Biomedical Engineering, University of Virginia, USA*
[b] *Department of Quality and Performance Improvement, University of Virginia Health System, USA*
[c] *School of Medicine, University of Virginia, USA*
[d] *Department of Systems and Information Engineering, University of Virginia, USA*
[e] *Frank Batten School of Leadership and Public Policy, University of Virginia, USA*

## ARTICLE INFO

## ABSTRACT

*Objective:* To determine the prevalence of individual-level social risk factors documented in unstructured data from electronic health records (EHRs) and the relationship between social risk factors and adverse clinical outcomes.

*Study setting:* Inpatient encounters for adults (≥18 years) at the University of Virginia Medical Center during a 12-month study period between July 2018 and June 2019. Inpatient encounters for labor and delivery patients were excluded, as well as encounters where the patient was discharged to hospice, left against medical advice, or expired in the hospital. The study population included 21,402 inpatient admissions, representing 15,116 unique patients who had at least one inpatient admission during the study period.

*Study design:* We identified measures related to individual social risk factors in EHRs through existing workflows, flowsheets, and clinical notes. Multivariate binomial logistic regression was performed to determine the association of individual social risk factors with unplanned inpatient readmissions, post-discharge emergency department (ED) visits, and extended length of stay (LOS). Other predictors included were age, sex, severity of illness, location of residence, and discharge destination.

*Results:* Predictors of 30-day unplanned readmissions included severity of illness (OR = 3.96), location of residence (OR = 1.31), social and community context (OR = 1.26), and economic stability (OR = 1.37). For 30-day post-discharge ED visits, significant predictors included location of residence (OR = 2.56), age (OR = 0.60), economic stability (OR = 1.39), education (OR = 1.38), social and community context (OR = 1.39), and neighborhood and built environment (OR = 1.61). For extended LOS, significant predictors were age (OR = 0.51), sex (OR = 1.18), severity of illness (OR = 2.14), discharge destination (OR = 2.42), location of residence (OR = 0.82), economic stability (OR = 1.14), neighborhood and built environment (OR = 1.31), and education (OR = 0.79).

*Conclusions:* Individual-level social risk factors are associated with increased risk for unplanned hospital readmissions, post-discharge ED visits, and extended LOS. While individual-level social risk factors are currently documented on an ad-hoc basis in EHRs, standardized SDoH screening tools using validated metrics could help eliminate bias in the collection of SDoH data and facilitate social risk screening.

<div style="border:1px solid">

# Callout box

**What is known on this topic:** Social determinants of health (SDoH) are modifiable factors that, when addressed through appropriate social policy and public health interventions, can improve patient health outcomes and reduce healthcare costs. However, these risk factors are not routinely or systematically captured by health systems at the individual level, limiting the understanding of the prevalence and impacts of these risk factors on adverse clinical outcomes.

**What this study adds:** This study demonstrates that individual social risk factors are associated with increased risk of unplanned inpatient readmissions, emergency department visits, and extended hospital length of stay. Furthermore, this study provides evidence that U.S. health systems are potentially already identifying many of these individual social risk factors through existing workflows and clinical notes in electronic health records (EHRs). Even in the absence of standardized SDoH screening tools, health systems can use existing EHR data to understand the prevalence of individual social risk factors and strategically prioritize further screening or interventions.

**Data availability**

The data that has been used is confidential.

</div>

## 1. Introduction

The U.S. spends far more on healthcare per capita than any other country – 17.7% of the nation's gross domestic product in 2019 (Martin, Hartman, Lassman, & Catlin, 2021; Papanicolas et al., 2019). Despite high healthcare expenditures, the U.S. ranks among the lowest in health outcomes and preventable mortality compared to other developed countries (Schneider et al., 2017). Clinical care, including access and quality of care, is estimated to account for only 20% of modifiable determinants of health. The remaining 80% of modifiable determinants of health include health behaviors (30%), physical environment (10%), and social and economic factors (40%) (Hood et al., 2016). Healthy People 2030 describes social determinants of health (SDoH) as "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." (Healthy People 2030) SDoH broadly encompass health behaviors (e.g., tobacco and alcohol use, diet, physical activity), socioeconomic factors (e.g., educational attainment, financial strain, social support systems, stress, interpersonal safety), and the physical environment (e.g., air and water quality, housing conditions, access to transportation) (Hood et al., 2016). Links between individual SDoH and population-level health outcomes (i.e., life expectancy and quality of life) have been widely documented (Bonomi et al., 2009; Dube et al., 2003; Getz et al., 2021; Ma et al., 2008; Telfair & Shelton, 2012; Turrentine et al., 2017).

The introduction of programs such as the Hospital Readmissions Reduction Program (HRRP) (Centers for Medicare & Medicaid Services, 2022), which reduces payments for hospitals that have excess 30-day readmission rates, has placed more focus on modifiable SDoH to improve health outcomes and thereby reduce healthcare costs. U.S. health systems are exploring ways to integrate SDoH data collection into EHRs to implement SDoH-related referrals for non-clinical care and improve quality performance (Chen et al., 2020). There is little consensus in the literature, or in clinical practice, about which SDoH measures should be captured, and social determinants are often conflated with social risk factors in the literature. Alderwick and

Gottlieb have outlined a SDoH lexicon for healthcare systems that defines social determinants as the conditions that shape health, but are neither positive or negative predictors of health by default (Alderwick & Gottlieb, 2019). Education is a social determinant of health, wherein higher formal educational attainment is associated with better health, and lower educational attainment with worse health. Social risk factors, however, are the specific adverse conditions associated with poor health outcomes, such as homelessness or food insecurity. A systematic review by Chen et al. on the integration of SDoH domains in EHRs found that only 16% of studies identified social risk factors as opposed to social or behavioral determinants of health (Chen et al., 2020). Furthermore, 57% of these studies report only neighborhood-level SDoH (e.g., median household income or neighborhood crime rates) and not individual-level SDoH (e.g., employment status or housing stability).

Social and behavioral determinants of health and associated risk factors have been a focus of research to improve predictive models for high-cost and high-risk patients (Bates et al., 2014). When integrating SDoH data with EHRs to improve risk assessments or prediction, the integration of neighborhood-level SDoH demonstrated no improvement in predictive performance (Bhavsar et al., 2018; Jamei et al., 2017; Vest & Ben-Assuli, 2019; Vutien et al., 2019). Alternatively, models that incorporated individual-level SDoH reported significant improvements in model performance when predicting outcomes ranging from medication adherence (Molfenter et al., 2012), to HIV risk (Feller et al., 2018), to risk for hospitalization (Chen et al., 2020; Takahashi et al., 2015). While many studies for risk prediction focus on inpatient readmissions due to programs like HRRP, this is not the only outcome of interest to health systems. Inappropriate ED visits contribute to high healthcare costs, ED overcrowding, and EDs are not equipped to provide the primary and preventative care that these patients often and preferentially need (Pines et al., 2013; Vest & Ben-Assuli, 2019). Some studies have estimated that up to 40% of all ED visits are clinically inappropriate (Naouri et al., 2020). Furthermore, extended hospital length of stay (LOS) contributes to high healthcare costs and increases the likelihood of hospital-acquired infections (Ward et al., 2021).

These is little consensus about the SDoH measures that should be captured in EHRs, or widespread use of standardized screening measures. Some researchers have used insurance claims data to identify a subset of social risk factors, but have found that these are often unreliably coded and that alternatively, text from clinical notes can be used to identify individual social risk factors with a much higher prevalence than administrative data (Navathe et al., 2018). In this study, our objective was to determine the prevalence with which individual social risk factors are documented through existing workflows and clinical notes as unstructured data in EHRs at the University of Virginia Medical Center, and to evaluate the association of those individual social risk factors with adverse clinical outcomes including unplanned hospital readmissions, ED revisits, and extended hospital LOS.

## 2. Methods

### 2.1. Study population

EHR data provided through EpicCare was queried for all inpatient admissions to the UVA Medical Center with a discharge date between July 1, 2018 and June 30, 2019. The study population was restricted to adults (age ≥18 years), excluded labor and delivery patients, and admissions where the patient expired in the hospital, was discharged to hospice, or was discharged against medical advice. The final study population included 21,402 inpatient admissions, representing 15,116 unique patients who had at least one inpatient admission during the study period.

## 2.2. Clinical outcomes

We investigated three outcomes in this study: 1) unplanned hospital readmissions, 2) ED revisits following discharge, and 3) extended hospital LOS. The primary endpoint for readmissions and ED revisits was 30 days following discharge from the index admission. Unplanned hospital readmissions exclude admissions for planned care, or follow-up treatment, such as scheduled chemotherapy and planned surgeries. ED revisits following discharge from the index admission include patients who visited the ED, but were not subsequently admitted to the hospital. If admitted through the ED, the outcome was considered an unplanned readmission. LOS was translated into a binary outcome by determining whether the actual LOS was greater than the expected LOS provided by 3M's All Patient Refined Diagnosis Related Groups (APR-DRG) model (McCormick et al., 2018).

## 2.3. Social risk factor measures

Despite efforts to develop standardized SDoH screening tools, there is currently no broad consensus on the specific questions, measures, or domains that should be captured. In a systematic review of studies that integrated SDoH measures with EHRs, Chen et al. employed a conceptual framework of SDoH domains and dimensions based on recommendations from the World Health Organization (Chen et al., 2020). This framework broadly categorizes SDoH into five domains – economic stability, education, healthcare access and quality, neighborhood and built environment, and social and community context. We utilized this framework for identifying social risk factors in EHR data. We queried EHRs for intake forms, discharge instructions, questionnaires, screening tools, social work and case management notes, and other existing note templates and flowsheets for any recorded responses and measures that contained information relevant to these SDoH domains and dimensions. This query identified a set of 41 unique fields currently being captured in UVA's EHR database (Supplemental Table 1) that were recorded by various healthcare providers during an inpatient admission. These measures spanned 11 different dimensions, and four of the five SDoH domains (Table 1). Next, we developed response criteria for each of these EHR fields to determine whether the recorded response indicated the presence of a social risk factor. Some fields had a preset list of

responses to choose from, while most were documented by providers with free text comments, and keywords were used to identify response criteria for those fields. Since these data were not systematically collected, it is possible that the absence of data was not random and subject to the bias of the health care providers documenting findings in EHRs. In order to reduce the impact of bias and missing data, we identified all encounters that a patient had with the health system (i.e., inpatient, outpatient, telehealth) during the 12 months preceding the index admission to determine a patient's exposure to social risk factors.

## 2.4. Statistical analysis

We developed a binomial generalized linear mixed model for each of the following clinical outcomes: 1) unplanned hospital readmissions, 2) post-discharge ED visits, and 3) extended LOS. All three clinical outcomes were modeled as binary outcomes. We included patients as random intercepts in these mixed models since a patient may have multiple admissions during the study period. The following predictor variables were included in all models: age, sex, severity of illness, location of residence, discharge destination, and risk factors for each of the SDoH domains identified in EHRs (economic stability, education, neighborhood and built environment, and social and community context). Severity of illness, which was obtained from 3M's APR-DRG model, was included as a predictor to account for the effect of severity of clinical conditions on the outcome measures. Location of residence was included as a predictor variable to account for the increased likelihood of readmissions and ED visits for patients living in proximity to the UVA Medical Center since admissions to other health systems could not be captured in this study. Discharge destination was included because discharge disposition could imply the patient's clinical condition at discharge and the post-discharge environment can impact readmissions and ED visits. Discharge to facilities can be delayed due to paperwork requirements and availability, potentially affecting the LOS outcome. All of these predictors were included to control for factors other than social risk factors that may impact the clinical outcomes measured.

SDoH domains were modeled as a binary variable, where a value of one indicated that the patient had an identified risk factor in that domain within the 12 months preceding the index admission. Categorical variables (i.e., sex, location of residence, and discharge destination) were factored into two levels. Location of residence was grouped by those who lived in the vicinity the of the hospital (Albemarle County and Charlottesville City) and all others. Discharge destination was grouped by those who were discharged to home (or home health) and those discharged to facilities (i.e., skilled nursing facilities, long term care, rehab). To aid in the interpretability of model coefficients, quantitative variables were rescaled to be on a similar scale as categorical variables. Severity of illness, which was originally classified on a scale of 1–4 (1 – minor, 2 – moderate, 3 – major, 4 – extreme) was normalized using mix-max scaling to a range of 0–1 (0 – minor, 0.33 – moderate, 0.66 – major, 1 – extreme). The age variable was divided by 100. All analysis was performed in R version 4.0.3 (R Core Team, 2021) using the package *lme4* (version 1.1.26) (Bates et al., 2015) and function *glmer* for implementation of the generalized linear mixed models.

## 3. Results

### 3.1. Prevalence of individual social risk factors documented in EHRs

A total of 30,924 inpatient admissions occurred during the study period and 21,402 admissions met the inclusion criteria, representing 15,116 unique patients treated at the UVA Medical Center over the 1-year study period (Fig. 1). We found that at least one measure within each of the four SDoH domains (i.e., economic stability, education, neighborhood and built environment, and social and community context) was documented within the 12-month period preceding the

**Table 1**

Framework for identifying SDoH domains and dimensions.

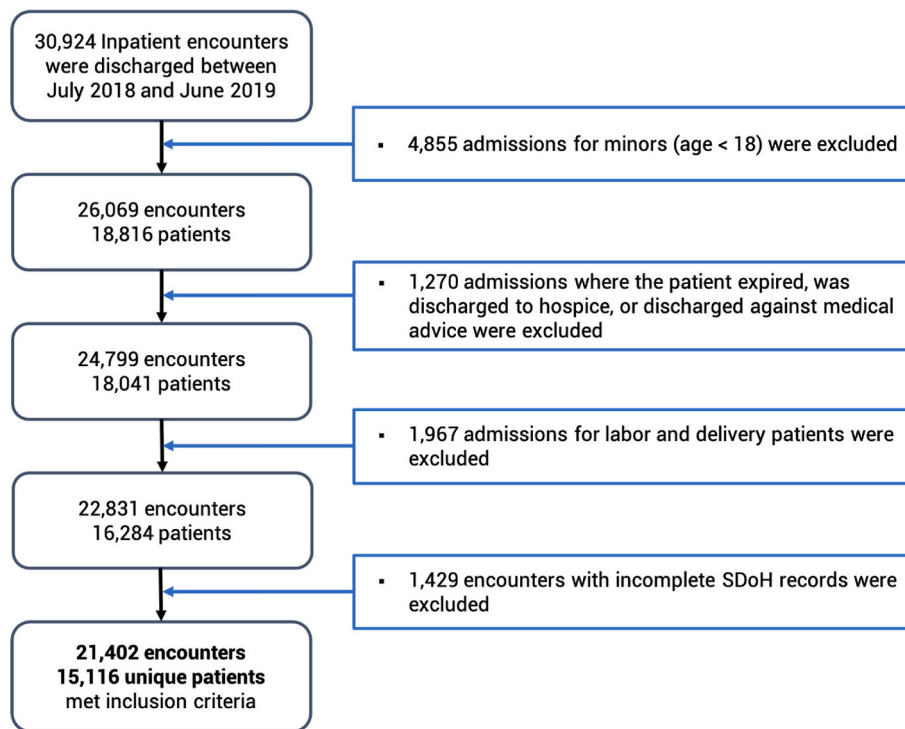| Domains | Dimensions | Identified in EHR unstructured data |
|---|---|---|
| Economic Stability | Employment | ✓ |
| | Financial resource strain | ✓ |
| | Housing instability | |
| | Food insecurity | |
| Education | High school graduation | ✓ |
| | Early childhood development and education | ✓ |
| | Language and literacy | ✓ |
| Healthcare Access and Quality | Access to health services | |
| | Access to primary care | |
| | Health literacy | |
| Neighborhood and Built Environment | Access to healthy foods | |
| | Neighborhood crime | |
| | Environmental conditions | |
| | Quality of housing | ✓ |
| | Transportation | ✓ |
| Social and Community Context | Family and community support | ✓ |
| | Interpersonal violence or abuse | ✓ |
| | Stress and depression | ✓ |
| | Marital status | ✓ |
| | Civic participation | |
| | Discrimination | |
| | Incarceration | |

**Fig. 1.** Defining the study population.

index admission for more than 94% of all inpatient admissions. Of the 21,402 admissions with complete SDoH records, 33.2% of admissions had risk factors identified in one SDoH domain, 16.3% had risk factors identified in two SDoH domains, 2.9% had risk factors identified in three domains, and less than 0.5% had risk factors identified in all four SDoH domains.

### 3.2. Patient and admission characteristics

We described the demographics (i.e., age, sex, race, ethnicity, primary language, and location of residence) for all patients in the study and for the patient subpopulations with identified risk factors in each of the SDoH domains (Table 2). Patient and admission characteristics for admissions that were excluded are provided in Supplemental Tables 2 and 3. Nearly half of patients in the study population were ages 55–74

**Table 2**
Patient characteristics for all patients and patient subpopulations with identified risk factors in each of the SDoH domains.

| Patient Characteristics | All patients, N (%) | Economic Stability, N (%) | Education, N (%) | Neighborhood and Built Environment, N (%) | Social and Community Context, N (%) |
|---|---|---|---|---|---|
| | 15,116 | 4,613 (30.5%[a]) | 850 (5.6%[a]) | 454 (3.0%[a]) | 4,591 (30.4%[a]) |
| **Age group (years)** | | | | | |
| 18–34 | 1,657 (11.0%) | 622 (13.5%) | 108 (12.7%) | 101 (22.2%) | 613 (13.4%) |
| 35–54 | 3,381 (22.4%) | 1,309 (28.4%) | 202 (23.8%) | 173 (38.1%) | 1,021 (22.2%) |
| 55–74 | 7,112 (47.0%) | 1,840 (39.9%) | 364 (42.8%) | 153 (33.7%) | 1,853 (40.4%) |
| ≥ 75 | 2,966 (19.6%) | 842 (18.3%) | 176 (20.7%) | 27 (5.9%) | 1,104 (24.0%) |
| **Sex** | | | | | |
| Male | 7,761 (51.3%) | 2,317 (50.2%) | 457 (53.8%) | 287 (63.2%) | 2,145 (46.7%) |
| Female | 7,355 (48.7%) | 2,296 (49.8%) | 393 (46.2%) | 167 (36.8%) | 2,446 (53.3%) |
| **Race** | | | | | |
| White | 12,102 (80.1%) | 3,371 (73.1%) | 539 (63.4%) | 346 (76.2%) | 3,580 (78.0%) |
| Black | 2,392 (15.8%) | 1,036 (22.5%) | 177 (20.8%) | 84 (18.5%) | 824 (17.9%) |
| Asian | 99 (0.7%) | 24 (0.5%) | 15 (1.8%) | 1 (0.2%) | 30 (0.7%) |
| Other | 485 (3.2%) | 176 (3.8%) | 115 (13.5%) | 23 (5.1%) | 145 (3.2%) |
| **Ethnicity** | | | | | |
| Non-Hispanic | 14,607 (96.6%) | 4,412 (95.6%) | 701 (82.5%) | 431 (94.9%) | 4,446 (96.8%) |
| Hispanic | 428 (2.8%) | 181 (3.9%) | 143 (16.8%) | 22 (4.8%) | 115 (2.5%) |
| **Primary Language** | | | | | |
| English | 14,727 (97.4%) | 4,429 (96.0%) | 669 (78.7%) | 436 (96.0%) | 4,497 (98.0%) |
| Non-English | 389 (2.6%) | 184 (4.0%) | 182 (21.4%) | 18 (4.0%) | 94 (2.0%) |
| **Location of Residence** | | | | | |
| Albemarle County[b] | 3,421 (22.6%) | 1,166 (25.3%) | 225 (26.5%) | 165 (36.3%) | 1,323 (28.8%) |
| Other | 11,695 (77.4%) | 3,447 (74.7%) | 625 (73.5%) | 289 (63.7%) | 3,268 (71.2%) |

Note: Column percentages reported unless otherwise indicated.
[a] Row percentages, as a percentage of all patients.
[b] Includes City of Charlottesville.

(47.0%) with slightly more male (51.3%) than female (48.7%) patients. The majority of patients were white (80.1%), non-Hispanic (96.6%), English was their primary language (97.4%), and resided outside of Albemarle County (77.4%). Overall, 30.5% of patients had a social risk factor identified in the economic stability SDoH domain, 5.6% had a risk factor in education, 3.0% had a risk factor in neighborhood and built environment, and 30.4% had a risk factor in social and community context. There were more patients in the 18–34 and 35–54 age groups with risk factors in the neighborhood and built environment domain (22.2% and 38.1%, respectively). In the ≥75 years age group, there were fewer patients with a risk factor in the neighborhood and built environment domain (5.9%) and more patients with a risk factor in the social and community context domain (24.0%). While females had a higher prevalence of risk factors in the social and community context domain (53.3%), males had a higher prevalence of risk factors in the neighborhood and built environment domain (63.2%). Black patients had a higher prevalence of risk factors across all SDoH domains, but most notably in the economic stability domain (22.5%) and education domain (20.8%). Hispanic patients and patients whose primary language was not English had a higher prevalence of risk factors in the education domain (16.8% and 21.4%, respectively).

We described the characteristics of admissions (i.e., length of stay, severity of illness, and discharge destination) for all admissions in the study and by clinical outcome (30-day readmissions, 30-day ED visit, and extended LOS) in Table 3. Among all inpatient admissions in the study population, 12.8% had a 30-day unplanned hospital readmission, 6.3% had a 30-day post-discharge ED visit, and 42.9% had an extended LOS. The median LOS for all admissions was 4.1 days with 22.6% of patients having a severity of illness classified as minor, 38.8% as moderate, 30.4% as major, and 8.2% as extreme. The majority of patients were discharged to home or home health (80.1%), while 19.9% were discharged to post-acute facilities (skilled nursing facilities, long-term care, or rehabilitation facilities). Admissions with a 30-day unplanned readmission or extended LOS had a higher median length of stay (5.1 and 6.7 days, respectively) and higher prevalence of major severity of illness (43.0% and 33.7%, respectively) or extreme severity of illness (13.3% and 10.4%, respectively). Discharge to post-acute care facilities from the hospital was more prevalent for encounters with a 30-day unplanned readmission (25.4%) and extended LOS (26.6%).

### 3.3. Individual social risk factors are associated with increased risk for adverse clinical outcomes

In this set of models, individual SDoH domains (economic stability, education, built neighborhood and environment, and social and community context) were included as independent predictors to determine whether certain SDoH domains were associated with higher risk for adverse clinical outcomes (Table 4). We found that 30-day unplanned readmissions were associated with higher severity of illness (OR = 3.96), location of residence in proximity to the UVA Medical Center (OR = 1.31), and social risk factors in the social and community context domain (OR = 1.26) and economic stability domain (OR = 1.37). Risk factors for neighborhood and built environment and education were not significant predictors of 30-day readmissions. Significant independent predictors of 30-day ED visits included age (OR = 0.60), location of residence in proximity to the UVA Medical Center (OR = 2.56) and risk factors across all of the SDoH domains: social and community context (OR = 1.39), economic stability (OR = 1.39), neighborhood and built environment (OR = 1.61), and education (OR = 1.38). All predictor variables with the exception of social and community context were significant predictors of extended LOS: age (OR = 0.51), sex (OR = 1.18), severity of illness (OR = 2.14), discharge to facilities (OR = 2.42), location of residence (OR = 0.82), economic stability (OR = 1.14), neighborhood and built environment (OR = 1.31), and education (OR = 0.79).

### 3.4. Aggregated SDoH risk score is associated with highest risk for 30-day post-discharge ED visits

In this set of models, risk factors for each SDoH domain were aggregated to reflect a total burden of SDoH risk on a range from 0 (no risk factors) to 4 (risk factors in all four domains). This aggregated SDoH risk score was normalized using min-max scaling to a range of 0–1, and included as a single predictor to compare the impact of other clinical and demographic variables with the presence of any SDoH risk factor on adverse clinical outcomes (Table 5). Severity of illness remained the most influential predictor of 30-day readmissions (OR = 3.98) with aggregated SDoH risk being the second most influential predictor (OR = 2.67). Location of residence remained a significant predictor of readmissions (OR = 1.32), and discharge to facilities was also a significant predictor in this model (OR = 1.13). Aggregated SDoH risk was associated with the highest risk (OR = 3.64) for 30-day ED visits among all independent predictors in this model. Other significant independent predictors of 30-day ED visits were age (OR = 0.58) and location of residence (OR = 2.65). All variables were significant independent predictors of extended LOS, with discharge to facilities having the highest associated risk (OR = 2.43), followed by severity of illness (OR = 2.21), SDoH risk (OR = 1.38), sex (OR = 1.17), location of residence (OR = 0.82), and age (OR = 0.60).

Furthermore, we compared 30-day (a commonly used industry standard) to 90-day endpoints for unplanned readmissions and ED revisits to see if there was a difference in the significance of predictors at a later end point for these outcomes (Supplemental Table 4). We found that discharge to facilities was a significant predictor of 30-day unplanned readmissions, but not a significant predictor for the 90-day endpoint. Significance of all other predictor variables across the two outcomes remained the same, but predictors had slightly higher odds ratios for 90-day outcomes since these outcomes would be more likely to occur within this endpoint.

### 4. Discussion

In this study we demonstrated that individual-level social risk factors are already widely documented in EHRs for an inpatient population treated at the UVA Medical Center in the absence of a specific SDoH screening tool. More than 94% of all inpatient encounters had a least one

**Table 3**
Admission characteristics by clinical outcome.

| Admission Characteristics | All Admissions, N (%) | 30-day readmission, N (%) | 30-day ED visit, N (%) | Extended LOS, N (%) |
|---|---|---|---|---|
| | 21,402 | 2,739 (12.8%[a]) | 1,350 (6.3%[a]) | 9,178 (42.9%[a]) |
| **Length of stay (days), median** | 4.1 | 5.1 | 4.2 | 6.7 |
| **Severity of Illness** | | | | |
| Minor | 4,840 (22.6%) | 280 (10.2%) | 275 (20.4%) | 1,715 (18.7%) |
| Moderate | 8,295 (38.8%) | 918 (33.5%) | 533 (39.5%) | 3,415 (37.2%) |
| Major | 6,510 (30.4%) | 1,177 (43.0%) | 442 (32.7%) | 3,089 (33.7%) |
| Extreme | 1,757 (8.2%) | 364 (13.3%) | 100 (7.4%) | 959 (10.4%) |
| **Discharge Destination** | | | | |
| Home[b] | 17,138 (80.1%) | 2,044 (74.6%) | 1,070 (79.3%) | 6,736 (73.4%) |
| Facilities[c] | 4,264 (19.9%) | 695 (25.4%) | 280 (20.7%) | 2,442 (26.6%) |

Note: Column percentages reported unless otherwise indicated.
[a] Row percentages, as a percentage of all admissions.
[b] Includes home health.
[c] Skilled nursing facilities, long-term care, and rehabilitation facilities.

**Table 4**

Individual SDoH domains and association with risk for adverse clinical outcomes.

| Predictors | 30-day unplanned readmission | | | 30-day ED visit without admission | | | Extended LOS | | |
|---|---|---|---|---|---|---|---|---|---|
| | Odds Ratio | 95% CI | p | Odds Ratio | 95% CI | p | Odds Ratio | 95% CI | p |
| Intercept | **0.04** | **0.03–0.04** | **<0.001** | **0.03** | **0.02–0.04** | **<0.001** | **0.58** | **0.50–0.67** | **<0.001** |
| Age | 1.09 | 0.81–1.47 | 0.570 | **0.60** | **0.41–0.88** | **0.009** | **0.51** | **0.41–0.64** | **<0.001** |
| Sex (Female) | 0.95 | 0.86–1.05 | 0.331 | 1.04 | 0.92–1.18 | 0.538 | **1.18** | **1.09–1.27** | **<0.001** |
| Severity of Illness | **3.96** | **3.36–4.67** | **<0.001** | 1.04 | 0.83–1.29 | 0.743 | **2.14** | **1.88–2.42** | **<0.001** |
| Discharge Destination (Facilities) | 1.12 | 1.00–1.26 | 0.050 | 0.95 | 0.80–1.11 | 0.506 | **2.42** | **2.20–2.67** | **<0.001** |
| Location of Residence (Albemarle County) | **1.31** | **1.17–1.46** | **<0.001** | **2.56** | **2.24–2.93** | **<0.001** | **0.82** | **0.75–0.90** | **<0.001** |
| Social and community context | **1.26** | **1.14–1.40** | **<0.001** | **1.39** | **1.21–1.59** | **<0.001** | 1.08 | 0.99–1.17 | 0.075 |
| Economic stability | **1.37** | **1.23–1.51** | **<0.001** | **1.39** | **1.22–1.59** | **<0.001** | **1.14** | **1.06–1.24** | **0.001** |
| Neighborhood and built environment | 1.13 | 0.87–1.46 | 0.359 | **1.61** | **1.21–2.16** | **0.001** | **1.31** | **1.06–1.62** | **0.013** |
| Education | 1.16 | 0.97–1.40 | 0.112 | **1.38** | **1.09–1.74** | **0.007** | **0.79** | **0.68–0.93** | **0.004** |
| *Random Effects* | | | | | | | | | |
| ICC | 0.21 | | | 0.25 | | | 0.28 | | |
| N | 15,116 | | | 15,116 | | | 15,116 | | |

**Table 5**

Aggregated SDoH risk factors and association with risk for adverse clinical outcomes.

| Predictors | 30-day unplanned readmission | | | 30-day ED visit without admission | | | Extended LOS | | |
|---|---|---|---|---|---|---|---|---|---|
| | Odds Ratio | 95% CI | p | Odds Ratio | 95% CI | p | Odds Ratio | 95% CI | p |
| Intercept | **0.04** | **0.03–0.05** | **<0.001** | **0.03** | **0.02–0.04** | **<0.001** | **0.60** | **0.51–0.69** | **<0.001** |
| Age | 1.06 | 0.79–1.43 | 0.699 | **0.58** | **0.40–0.85** | **0.005** | **0.49** | **0.39–0.62** | **<0.001** |
| Sex (Female) | 0.95 | 0.87–1.05 | 0.351 | 1.02 | 0.90–1.16 | 0.786 | **1.17** | **1.09–1.27** | **<0.001** |
| Severity of Illness | **3.98** | **3.38–4.69** | **<0.001** | 1.03 | 0.83–1.28 | 0.810 | **2.12** | **1.87–2.40** | **<0.001** |
| Discharge Destination (Facilities) | **1.13** | **1.01–1.27** | **0.039** | 0.95 | 0.80–1.12 | 0.574 | **2.43** | **2.20–2.68** | **<0.001** |
| Location of Residence (Albemarle County) | **1.32** | **1.18–1.47** | **<0.001** | **2.65** | **2.32–3.04** | **<0.001** | **0.82** | **0.75–0.90** | **<0.001** |
| SDoH | **2.67** | **2.13–3.34** | **<0.001** | **3.64** | **2.72–4.86** | **<0.001** | **1.38** | **1.15–1.65** | **0.001** |
| *Random Effects* | | | | | | | | | |
| ICC | 0.21 | | | 0.25 | | | 0.28 | | |
| N | 15,116 | | | 15,116 | | | 15,116 | | |

measure documented for each of the SDoH domains included in this study (economic stability, education, neighborhood and built environment, and social and community context). Admissions that were excluded from the study due to missing data for one or more of the SDoH domains had a significantly shorter hospital length of stay and were more likely to be classified as minor severity of illness than the study population (Supplemental Table 3), potentially suggesting that a thorough social history was not obtained due to limited time or less clinical severity. One-third of patients in the study population had a risk factor identified in at least one SDoH domain, and risk factors in the economic stability and social and community context domains were the most prevalent among this patient population. While risk factors for neighborhood and built environment were uncommon, they were a significant predictor of 30-day ED revisits and extended LOS, and had the highest odds ratio among the SDoH domains included in these models. Access to housing and transportation may create barriers to discharge and delay discharge from the hospital, and inhibit patients from obtaining medications or follow-up care (Syed et al., 2013). Prior research has shown that the rate of ED visits is significantly higher among persons experiencing homelessness (QuickStats, 2020).

The presence of individual-level social risk factors had significant effects on all clinical outcomes studied, and when social risk was included as an aggregated score it was associated with the highest risk of 30-day ED revisits among all predictors. The presence of risk factors in the economic stability SDoH domain were a significant predictor of all outcomes studied (30-day readmissions, 30-day ED revisits, and extended LOS). Social and community context was a significant predictor of post-discharge outcomes (readmissions and ED visits), but not extended LOS. The lack of family support or psychosocial care may increase the likelihood that a patient returns to the hospital, and indeed prior research has shown that the presence of high perceived social support among older adults significantly reduced hospital readmissions (Chan et al., 2019). Lack of family or community support may not pose

the same barriers to discharge from the hospital, but health systems should consider partnerships with community organizations to reduce the likelihood of costly readmissions for individuals with these risk factors. Notably for 30-day ED visits, clinical conditions (severity of illness and discharge destination) were not significant predictors of this outcome, while all SDoH domains were significant risk factors even when accounting for the patient's proximity to the UVA Medical Center. Social risk factors are potentially modifiable determinants of health that have significant impacts on ED revisits and should be a target for interventions among health systems seeking to impact this outcome. While all predictors except social and community context were associated with extended LOS, health systems who wish to reduce hospital LOS should consider social risk factors as potentially modifiable factors as opposed to other clinical conditions that are difficult to modify. Barriers to discharge including access to transportation and stable housing may be addressed through community partnerships to improve this outcome.

There are several limitations to this study, most notably the lack of standardized and validated SDoH screening questions or measures. Without a standardized screening tool and protocol there is likely implicit bias involved in the collection of SDoH data when clinicians make decisions about which questions are relevant to ask of individual patients (Atlus, 2017). The SDoH measures that we identified in EHRs do not cover all of the dimensions of social risk factors. For example, existing measures related to housing instability primarily capture those individuals experiencing chronic homelessness, and not individuals experiencing housing insecurity (e.g., inability to pay rent or utilities, or at risk for eviction). Similarly, transportation measures only assess an individual's need for transportation at the time of discharge from the hospital, but do not ask about access to transportation for follow-up appointments, or to obtain prescriptions. For these reasons, we anticipate that the prevalence of risk factors for many of these SDoH domains are potentially much higher in this patient population than is currently captured. Furthermore, EHR data was obtained from a single academic

medical center and cannot capture readmissions to other medical centers or emergency departments. Since the UVA Medical Center serves a wide geographic distribution and many patients reside in rural areas, this may artificially decrease readmissions and ED revisits for these populations.

A standardized system for the collection of individual-level social and behavioral risk factors will likely be necessary to improve upon current methods to accurately identify high-risk and high-cost patients and develop targeted interventions. There have been several proposed frameworks for integrating the collection of SDoH data into EHRs in parallel with calls for deployment of standardized SDoH screening tools. Multiple screening tools have been developed including the Accountable Health Communities Screening Tool (Billioux et al., 2017), PREPARE assessment tool, Health Leads Social Needs Assessment (Thomas-Henkel & Schulman, 2017), and others. Organizations such as the Institute of Medicine have developed recommendations for the relevant domains of social and behavioral health that should be captured in EHRs (Institute of Medicine, 2015). Others in the field have discussed an "opportunity index" for SDoH that could help health care systems identify priorities for quality improvement that would provide the greatest cost savings and benefit to patients and the community (Lee et al., 2016; Magnan, 2017). Important considerations for implementing SDoH screening tools include the format of a screening tool, who administers it, and the frequency of assessing needs. Electronic formats for self-disclosure, compared to in-person interviews, have been shown to have higher rates of disclosure for sensitive issues such as household violence and substance abuse (Gottlieb et al., 2014). An additional consideration is the appropriate frequency of risk assessment. Should SDoH screening be administered at every patient encounter, annually, or on some other frequency? The varied domains of social and behavioral health may change frequently and need to be assessed regularly, while others may be more stable especially in adulthood (Institute of Medicine, 2015).

These types of standardized screening tools for the structured collection of social risk factors are not widely used in clinical practice because they require the implementation of new data systems and widespread adoption from health care providers. Not all health systems will be poised to implement widespread standardized screening for individual social risk factors immediately, but this study demonstrates an opportunity for health systems to set priorities for interventions or further screening by identifying the risk factors that are already documented in EHRs. Despite the limitations of this study, we have demonstrated that individual-level social risk factors are already widely documented in EHRs and associated with increased risk for multiple adverse clinical outcomes. Unstructured EHR data may provide health systems with a useful starting point to understand the prevalence of social risk factors among their patient population, develop priorities for further screening, and design targeted interventions in partnership with community organizations.

## Ethical statement

This study was approved by the University of Virginia Institutional Review Board for Health Science Research with an exemption (IRB-HSR # 22081).

## Author statement

SMR, BK, and JDM contributed to the design of the study. SMR performed data collection and analysis. BK and LEB advised on data analysis and methodology. SMR drafted the initial manuscript. All authors contributed to manuscript revision and approved the final version submitted for publication.

## Funding

## Conflicts of interest

The authors declare that this research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Data availability

The data that has been used is confidential.

## Acknowledgements

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ssmph.2022.101210.

## References

Alderwick, H., & Gottlieb, L. M. (2019). Meanings and misunderstandings: A social determinants of health lexicon for health care systems. *The Milbank Quarterly, 97*(2), 407–419.

Atlus, A. (2017). *4 steps physicians can take to address implicit biases.* The DO.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48.

Bates, D. W., Saria, S., Ohno-Machado, L., Shah, A., & Escobar, G. (2014). Big data in health care: Using analytics to identify and manage high-risk and high-cost patients. *Health Affairs, 33*(7), 1123–1131.

Bhavsar, N. A., Gao, A., Phelan, M., Pagidipati, N. J., & Goldstein, B. A. (2018). Value of neighborhood socioeconomic status in predicting risk of outcomes in studies that use electronic health record data. *JAMA Network Open, 1*(5), Article e182716.

Billioux, A., Verlander, K., Anthony, S., & Alley, D. (2017). *Standardized screening for health-related social needs in clinical settings. The accountable health communities screening tool.* National Academy of Medicine.

Bonomi, A. E., Anderson, M. L., Rivara, F. P., & Thompson, R. S. (2009). Health care utilization and costs associated with physical and nonphysical-only intimate partner violence. *Health Services Research, 44*(3), 1052–1067.

Centers for Medicare & Medicaid Services. (2022). Hospital Readmissions Reduction Program (HRRP). https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/HRRP/Hospital-Readmission-Reduction-Program. (Accessed 19 April 2021).

Chan, B., Goldman, L. E., Sarkar, U., Guzman, D., Critchfield, J., Saha, S., & Kushel, M. (2019). High perceived social support and hospital readmissions in an older multi-ethnic, limited English proficiency, safety-net population. *BMC Health Services Research, 19*(1), 334.

Chen, M., Tan, X., & Padman, R. (2020). Social determinants of health in electronic health records and their impact on analysis and risk prediction: A systematic review. *Journal of the American Medical Informatics Association, 27*(11), 1764–1773.

Dube, S. R., Felitti, V. J., Dong, M., Giles, W. H., & Anda, R. F. (2003). The impact of adverse childhood experiences on health problems: Evidence from four birth cohorts dating back to 1900. *Preventive Medicine, 37*(3), 268–277.

Feller, D. J., Zucker, J., Yin, M. T., Gordon, P., & Elhadad, N. (2018). Using clinical notes and natural language processing for automated HIV risk assessment. *Journal of Acquired Immune Deficiency Syndromes, 77*(2), 160–166.

Getz, K. D., Lewey, J., Tam, V., Irizarry, O., Levine, L., Aplenc, R., & Arany, Z. (2021). Neighborhood education status drives racial disparities in clinical outcomes in PPCM. *American Heart Journal, AUGUST*(238), 27–32. https://doi.org/10.1016/j.ahj.2021.03.015

Gottlieb, L., Hessler, D., Long, D., Amaya, A., & Adler, N. (2014). A randomized trial on screening for social determinants of health: The iScreen study. *Pediatrics, 134*(6), e1611–1618.

Healthy People 2030, U.S. Department of Health and Human Services, Office of Disease Prevention and Health Promotion. Social Determinants of Health. https://health.gov/healthypeople/priority-areas/social-determinants-health.

Hood, C. M., Gennuso, K. P., Swain, G. R., & Catlin, B. B. (2016). County health rankings: Relationships between determinant factors and health outcomes. *American Journal of Preventive Medicine, 50*(2), 129–135.

*Capturing Social and Behavioral Domains and Measures in Electronic Health Records.* (2015). Washington, DC: Committee on the Recommended Social and Behavioral Domains and Measures for Electronic Health Records; Board on Population Health and Public Health Practice; Institute of Medicine.

Jamei, M., Nisnevich, A., Wetchler, E., Sudat, S., & Liu, E. (2017). Predicting all-cause risk of 30-day hospital readmission using artificial neural networks. *PLoS One, 12*(7), Article e0181173.

Lee, V. S., Kawamoto, K., Hess, R., Park, C., Young, J., Hunter, C., … Pendleton, R. (2016). Implementation of a value-driven outcomes program to identify high variability in clinical costs and outcomes and association with reduced cost and improved quality. *JAMA, 316*(10), 1061–1072.

Ma, C. T., Gee, L., & Kushel, M. B. (2008). Associations between housing instability and food insecurity with health care access in low-income children. *Ambulatory Pediatrics, 8*(1), 50–57.

Magnan, S. (2017). *Social Determinants of Health 101 for Health Care: Five Plus Five.* Washington, D.C.: NAM Perspectives.

Martin, A. B., Hartman, M., Lassman, D., & Catlin, A. (2021). National health care spending in 2019: Steady growth for the fourth consecutive year. *Health Affairs, 40* (1), 14–24.

McCormick, P. J., Lin, H. M., Deiner, S. G., & Levin, M. A. (2018). Validation of the all patient refined Diagnosis related group (APR-DRG) risk of mortality and severity of illness modifiers as a measure of perioperative risk. *Journal of Medical Systems, 42*(5), 81.

Molfenter, T. D., Bhattacharya, A., & Gustafson, D. H. (2012). The roles of past behavior and health beliefs in predicting medication adherence to a statin regimen. *Patient Preference and Adherence, 6*, 643–650.

Naouri, D., Ranchon, G., Vuagnat, A., Schmidt, J., & Yordanov, Y. (2020). Factors associated with inappropriate use of emergency departments: Findings from a cross-sectional national study in France. *BMJ Quality and Safety, 29*(6), 449–464.

Navathe, A. S., Zhong, F., Lei, V. J., Chang, F. Y., Sordo, M., Topaz, M., … Zhou, L. (2018). Hospital readmission and social risk factors identified from physician notes. *Health Services Research, 53*(2), 1110–1136.

Papanicolas, I., Woskie, L. R., Orlander, D., Orav, E. J., & Jha, A. K. (2019). The relationship between health spending and social spending in high-income countries: How does the US compare? *Health Affairs, 38*(9), 1567–1575.

Pines, J. M., Mullins, P. M., Cooper, J. K., Feng, L. B., & Roth, K. E. (2013). National trends in emergency department use, care patterns, and quality of care of older adults in the United States. *Journal of the American Geriatrics Society, 61*(1), 12–17.

QuickStats: Rate of emergency department (ED) visits, by homeless status and geographic region - National Hospital Ambulatory Medical Care Survey, United States, 2015-2018. (2020). *MMWR Morb Mortal Wkly Rep, 69*(50), 1931.

R Core Team. (2021). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.

Schneider, E. C., Sarnak, D. O., Squires, D., Shah, A., & Doty, M. M. (2017). *Mirror, mirror 2017: International comparison reflects flaws and opportunities for better U.S. health care.* The Commonwealth Fund.

Syed, S. T., Gerber, B. S., & Sharp, L. K. (2013). Traveling towards disease: Transportation barriers to health care access. *Journal of Community Health, 38*(5), 976–993.

Takahashi, P. Y., Ryu, E., Olson, J. E., Winkler, E., Hathcock, M., Gupta, R., … Cerhan, J. (2015). Health behaviors and quality of life predictors for risk of hospitalization in an electronic health record-linked biobank. *International Journal of General Medicine, 8*, 247–254.

Telfair, J., & Shelton, T. L. (2012). Educational attainment as a social determinant of health. *North Carolina Medical Journal, 73*(5), 358–365.

Thomas-Henkel, C., & Schulman, M. (2017). *Screening for Social Determinants of Health in Populations with Complex Needs: Implementation Considerations.* Center for Health Care Strategies.

Turrentine, F. E., Buckley, P. J., Sohn, M. W., & Williams, M. D. (2017). Travel time influences readmission risk: Geospatial mapping of surgical readmissions. *The American Surgeon, 83*(6), 573–582.

Vest, J. R., & Ben-Assuli, O. (2019). Prediction of emergency department revisits using area-level social determinants of health measures and health information exchange information. *International Journal of Medical Informatics, 129*, 205–210.

Vutien, P., Shah, R., Ma, K., Saleem, N., & Melson, J. (2019). Utilization of census tract-based neighborhood poverty rates to predict non-adherence to screening colonoscopy. *Digestive Diseases and Sciences, 64*(9), 2505–2513.

Ward, C., Patel, V., Elsaid, M. I., Jaisinghani, P., & Sharma, R. (2021). A case-control study of length of stay outliers. *American Journal of Managed Care, 27*(3), e66–e71.