

# Enriching Real-world Data with Social Determinants of Health for Health Outcomes and Health Equity: Successes, Challenges, and Opportunities

Zhe He<sup>1,2</sup>, Emily Pfaff<sup>3</sup>, Serena Jingchuan Guo<sup>4</sup>, Yi Guo<sup>5</sup>, Yonghui Wu<sup>5</sup>, Cui Tao<sup>6</sup>, Gregor Stiglic<sup>7,8,9</sup>, Jiang Bian<sup>5</sup>

<sup>1</sup> School of Information, Florida State University, United States

<sup>2</sup> Department of Behavioral Sciences and Social Medicine, College of Medicine, Florida State University, United States

<sup>3</sup> Department of Medicine, University of North Carolina at Chapel Hill School of Medicine, United States

<sup>4</sup> Department of Pharmaceutical Outcomes and Policy, College of Pharmacy, University of Florida, United States

<sup>5</sup> Department of Health Outcomes and Biomedical Informatics, College of Medicine, University of Florida, United States

<sup>6</sup> School of Biomedical Informatics, University of Texas Health Science Center at Houston, United States

<sup>7</sup> Faculty of Health Science, University of Maribor, Slovenia

<sup>8</sup> Faculty of Electrical Engineering and Computer Science, University of Maribor, Slovenia

<sup>9</sup> Usher Institute, University of Edinburgh, UK

## Summary

**Objective:** To summarize the recent methods and applications that leverage real-world data such as electronic health records (EHRs) with social determinants of health (SDoH) for public and population health and health equity and identify successes, challenges, and possible solutions.

**Methods:** In this opinion review, grounded on a social-ecological-model-based conceptual framework, we surveyed data sources and recent informatics approaches that enable leveraging SDoH along with real-world data to support public health and clinical health applications including helping design public health intervention, enhancing risk stratification, and enabling the prediction of unmet social needs.

**Results:** Besides summarizing data sources, we identified gaps in capturing SDoH data in existing EHR systems and opportunities to leverage informatics approaches to collect SDoH information either from structured and unstructured EHR data or through linking with public surveys and environmental data. We also surveyed recently developed ontologies for standardizing SDoH information and approaches that incorporate SDoH for disease risk stratification, public health crisis prediction, and development of tailored interventions.

**Conclusions:** To enable effective public health and clinical applications using real-world data with SDoH, it is necessary to develop both non-technical solutions involving incentives, policies, and training as well as technical solutions such as novel

social risk management tools that are integrated into clinical workflow. Ultimately, SDoH-powered social risk management, disease risk prediction, and development of SDoH tailored interventions for disease prevention and management have the potential to improve population health, reduce disparities, and improve health equity.

## Keywords

Social determinants of health, public health informatics, electronic health records, exposome

Yearb Med Inform 2023

<http://dx.doi.org/10.1055/s-0043-1768732>

## 1 Introduction

In the past decade, a rapidly growing body of literature has argued and demonstrated the important role of social determinants of health (SDoH) in shaping human health and well-being [1]. According to the World Health Organization (WHO), SDoH are the

conditions in which “*people are born, grow, live, work, and age*” [2]. These non-medical factors include social, societal, and environmental conditions such as income, education, employment, insurance, social relationships, physical environments, and more. Prior research has demonstrated that SDoH are major drivers of health

outcomes and more importantly, the main contributors to the widespread health inequities. It was estimated that, in the United States, SDoH could be responsible for up to 40% of all preventable deaths, significantly higher than the 10-15% for which better medical care can be accounted for [3–5]. Public health interventions that

target SDoH are instrumental for improving health and reducing the long-standing health inequities.

Recognizing the importance of SDoH on health, various professional societies and organizations, including the WHO [2], Healthy People 2030 [6], and the National Academy of Medicine (NAM) [7], have published frameworks that define SDoH and advocated for the collection of SDoH data. In particular, the NAM organized a committee on “*Capturing Social and Behavioral Domains and Measures in Electronic Health Records*”, which identified 12 SDoH measures to be included in patients’ electronic health records (EHRs) to inform the meaningful use of EHRs [8]. Internationally, the WHO European Health Equity Status Report initiative (HESRI) is being developed to promote policy making for health equity and well-being for the European Region [9, 10]. Canadian researchers also developed a rural-specific SDoH framework called “*Rural Community Health and Well-being Framework*”, which includes 13 categories of SDoH that are pertinent to rural residents [11]. SDoH influence health and well-being through a complex interplay between individual- and contextual-level factors. Individual-level SDoH are factors measured from an individual, such as education, occupation, and health behaviors, while contextual-level SDoH are factors measured from an individual’s surroundings, including both social and physical environments, such as built environment, healthcare quality, and community environment. At the individual level, collecting SDoH for a patient provides clinicians with a complete social context of a patient’s health status, facilitating shared decision-making and individualized treatment planning. At the community and societal levels, a successful public health intervention should simultaneously target individual- and contextual-level SDoH considering the powerful role and interacting nature of SDoH.

In recent years, there is an increasing number of studies that harness real-world data (RWD) with SDoH to support public and population health, with a particular emphasis on EHRs, claims and billing data, public health survey data (e.g., the National Health and Nutrition Examination Survey [12]), and other data such as exposome data.

In particular, to support precision prevention and treatment of diseases, SDoH is being incorporated in the models for disease screening and prediction to identify social risks. Efforts such as social prescribing would link patients with non-medical sources of support, such as community services, social services, and local organizations to address the underlying social and lifestyle factors that contribute to poor health and wellbeing, and to promote positive health outcomes [13]. Importantly, SDoH-enriched RWD can also support public health interventions by identifying populations at higher risk for certain health problems, allowing for targeted and more effective interventions for early prevention. By better understanding the impact of social and environmental factors on health and health care access, health systems and healthcare providers can work with communities and public health organizations to address these factors, reduce health disparities, and improve health equity, as SDoH are often the root causes of disparities and account for 80% of modifiable factors [14, 15]. Nonetheless, it is challenging to identify SDoH information and appropriately link it to clinical and public health data to enable these applications. Informatics approaches such as natural language processing, ontologies, spatiotemporal data integration offer promising solutions to tackle these challenges.

Given the prominent role of EHRs for disease prevention and treatment, and the increasing popularity of integrating SDoH into EHRs for health outcomes and health equity, this article reviews recent informatics approaches covering a wide range of methods and applications, including: (1) the collection of SDoH from both structured and unstructured EHR data, (2) linking public surveys and environmental data to EHR for measuring contextual-level SDoH, (3) the standardization of SDoH with ontologies, and (4) the utility of SDoH-enriched EHRs in public and population health applications including public health intervention, risk stratification, and prediction of unmet social needs.

We invited leading experts who have published extensively in these areas [16–29] to conceptualize this review article and co-author sections corresponding to their expertise. Figure 1 shows our conceptual framework, adopting the social-ecological

model [30] and the National Institute on Minority Health and Health Disparities (NIMHD) Research Framework [31], for integrating SDoH data with EHR data to support various health applications at the individual, family, community, and societal levels. The rest of this opinion review is organized as follows:

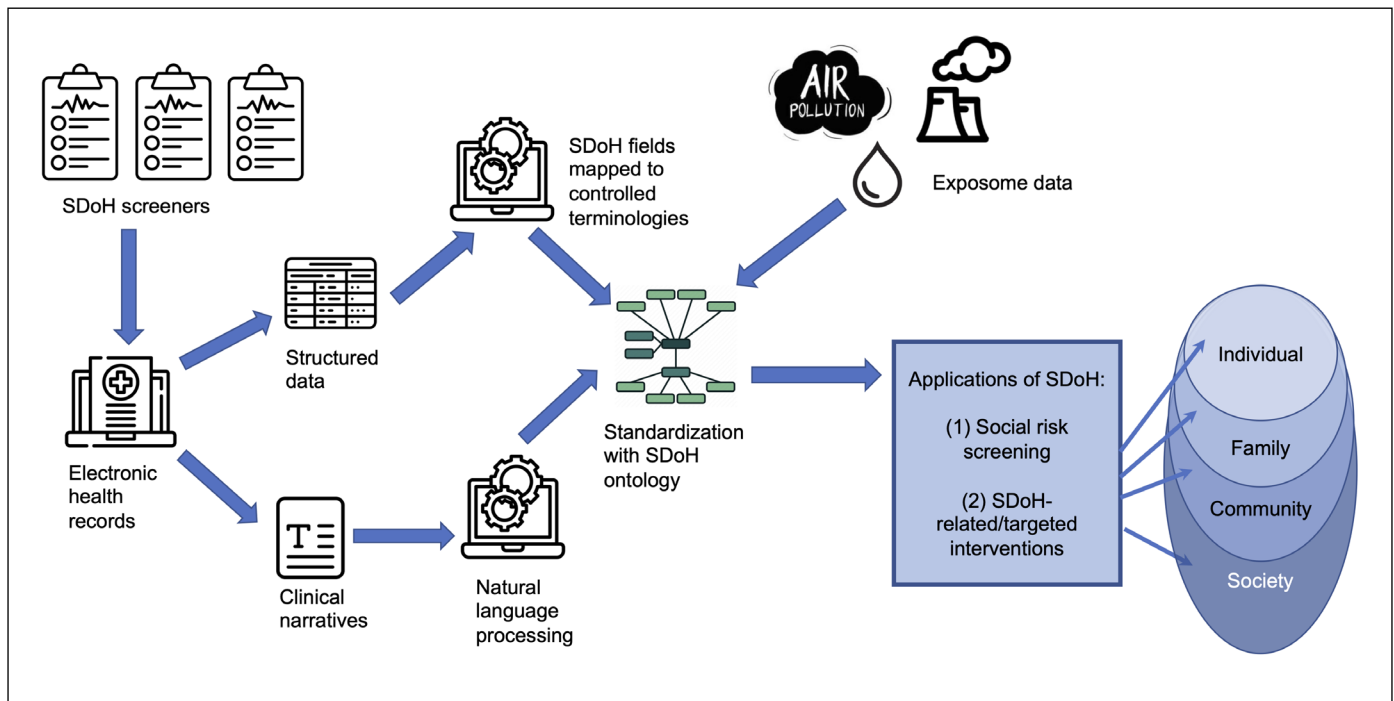
- In Section 2, we review the techniques for SDoH data capture and data engineering, as well as use cases;
- In Section 3, we review the techniques for creating SDoH ontologies;
- In Section 4, we review the public and population health applications using RWD enriched with SDoH;
- In Section 5, we summarize the challenges and promising pathways towards successful public and population health applications, leveraging RWD and SDoH.

## 2 SDoH Data Engineering

In this section, we will first review the standards for SDoH screening and discuss the challenges and opportunities for capturing SDoH information in structured EHR data (Section 2.1). Then, we will review natural language processing approaches for extracting SDoH information from clinical notes in EHRs and point out the low documentation rate of certain categories of SDoH in clinical notes (Section 2.2). Then, we will review recent efforts, challenges, and techniques of linking contextual SDoH data to EHR data (Section 2.3).

### 2.1 Structured Data and Tools in EHR Systems for Capturing SDoH

Given the increasing recognition of the importance of SDoH for patient care and population health, EHR vendors have started to implement structured SDoH fields to collect this information directly from patients during the course of care [32]. These structured fields typically cover commonly recognized SDoH domains including healthcare access, child care, financial strain, housing, transportation, food insecurity, education, and employment, among others [33].



**Fig. 1** A conceptual framework for integrating SDoH data with EHR data to support public and population health applications.

While the implementation of structured SDoH fields is a positive first step towards interoperability, there is still a lack of standardization among EHR vendors and health care systems regarding how and from whom SDoH information should be collected. This results in inconsistencies in the data collected and presents a challenge for the use and exchange of SDoH data across different systems. As identified by Arons *et al.* [33], the six most popular SDoH instruments and screening tools implemented in EHR systems are the NAM Recommended Social and Behavior Domains and Measures report [8]; the National Association of Community Health Center (NACHC)'s Protocol for Responding to and Assessing Patients' Assets, Risks, and Experiences (PRAPARE) survey [34]; the Centers for Medicare and Medicaid Services (CMS)'s Accountable Health Communities (AHC) survey [35]; the Health Leads questionnaire [36]; the Safe Environment for Every Kid (SEEK) questionnaire [37]; and the WE CARE survey instrument [38]. There is considerable variation among these tools and instruments in the questions that are

asked and the SDoH domains that are covered. This variation is compounded by the fact that many health care systems make additional customizations upon implementation, further limiting the opportunities for interoperability and standardization. In interviews, some top EHR vendors described the built-in flexibility of their SDoH data collection modules as a feature, noting that patient populations and reporting requirements vary from health system to health system. Those same vendors, however, also noted the disadvantages of this flexibility in terms of data sharing, interoperability, data aggregation, and analytics [39].

In the absence of a uniform standard for SDoH screening, mapping structured SDoH fields in EHRs to existing standard clinical terminologies, such as the International Classification of Diseases-Tenth Revision (ICD-10), Logical Observation Identifiers Names and Codes (LOINC), and the Systematized Nomenclature of Medicine (SNOMED), is a step toward greater interoperability. Both questions and answers have the potential to be mapped to these standard terminologies, as shown in Table 1.

EHR vendors have attempted to support this type of mapping with varying degrees of success. Unfortunately, many SDoH question/answer sets (particularly those designed to collect detailed information, such as “*In a typical week, how many times do you talk on the phone with family, friends, or neighbors?*”), do not have good matches within the standard terminologies [33,39]. In the absence of structured fields to collect SDoH data, the ICD-10 Z55-Z65 codes can be used to capture some SDoH in a standardized way in the EHR's Problem List (e.g., Z56, “*Problems related to employment and unemployment*”). However, while these codes have been available since 2016, a lack of clear guidelines for use, training, or incentives has led to slow and inconsistent uptake [41,42]. As of 2019, only 1.6% of Medicare beneficiaries had any Z-code in their records [43]. In a 2020 study [17], Guo *et al.* assessed the documentation of Z-codes in EHRs using data from a large clinical research network (the OneFlorida+ Clinical Research Consortium, covering ~15 million Floridians), and also found a

**Table 1** Mapping a SDoH question and the associated answer set to its LOINC equivalents [40].

SDoH item as shown in EHRs	Item type	LOINC equivalent
Has the electric, gas, oil, or water company threatened to shut off services in your home in the past 12 months?	Question	96779-4
No	Answer	LA32-8*
Yes	Answer	LA33-6*
Already shut off	Answer	LA32002-0

\*Note that LOINC has many different codes to signify common answers such as “No” and “Yes”; these codes are specific to a particular set of questions, including the one in this example

low utilization rate (270.61 per 100,000 at the encounter level and 2.03% at the patient level), although the utilization rate increased slightly from 255.62 to 292.79 per 100,000 since 2018.

Based on their finding of uneven use of structured SDoH fields at University of California San Francisco Health, Wang *et al.* suggest that it is not enough to simply add SDoH fields to EHRs and expect them to be used. Rather, those fields must be made an integral part of clinical workflow [44] and SDoH documentation must be incentivized with institutional policies and procedures [32]. Moreover, clinicians should be specifically trained to establish the empathy and trust necessary to collect this sensitive information from their patients [45]. Screening for social needs probes potentially stigmatizing aspects of individuals’ lives (e.g., poverty and racism), leading to potential harms through trauma, discrimination, or legal consequences. This concern is especially pronounced in existing survey-based SDoH screening without adequate face-to-face discussions [35]. Once the information is obtained, clinicians also lack training to use the SDoH information in clinical decision-making and formulating care plans accordingly [41]. Almost all existing SDoH screening tools were developed for universal screening but were not validated to predict specific outcomes, and there are often no actionable next steps even if certain SDoH issues were identified in the clinical settings. In other words, clinicians often do not know whether addressing the identified social

risks would lead to any specific health outcome improvements of patients at hand, nor do they necessarily have meaningful ways to address those identified social risks. Compounding these two issues, clinicians are less inclined to adopt SDoH screening tools in their routine care.

Incentives, policies, and training are non-technical gaps in the current methods of collecting structured SDoH information in EHRs, but must be addressed in order to set up the technical solutions for success. Nevertheless, as informaticians, we must also develop these technologies tailored to the clinician and patient needs.

## 2.2 SDoH Extraction from Clinical Narratives

Clinical notes and other free text fields offer a flexible and intuitive way for clinicians to document SDoH information. The informal nature of a clinical note allows for recording in-depth personal information such as a patient’s unstable housing situation or struggles with food insecurity. However, the lack of standardization for free text makes it challenging to analyze the data for both operational and research purposes and does not lend itself to interoperability. To better utilize SDoH information embedded in clinical notes, natural language processing (NLP) methods and tools have been developed to extract SDoH from clinical narratives.

Prior research has developed NLP systems to extract individual-level SDoH critical for public health studies, such as

substance use [46], homelessness and housing insecurity [47,48], employment status [49], and suicide attempt or ideation [18,50]. Both rule-based and machine learning-based methods have been applied. However, these systems can only extract a single SDoH at a time, and there is a lack of comprehensive NLP systems to extract multiple common SDoH from clinical narratives. Recent studies have developed clinical corpora with multiple common SDoH categories and applied more advanced deep learning-based NLP models. Feller *et al.* [51] developed a corpus of five SDoH categories and approached SDoH detection as a classification task using machine learning models. Lybarger *et al.* [52] developed a corpus containing 12 SDoH categories using clinical notes from the Medical Information Mart for Intensive Care (MIMIC)-III dataset and an existing dataset from the University of Washington and Harborview Medical Center. Han *et al.* [53] developed a corpus of 13 SDoH categories using MIMIC-III and approached SDoH detection as a classification task but used deep learning models. Similarly, Stemerman *et al.* [54] applied machine learning methods to detect 6 categories of SDoH through classification tasks. Yu *et al.* [19] developed a corpus of 19 SDoH categories using clinical notes from cancer patients at University of Florida Health and applied state-of-the-art transformer-based methods for SDoH extraction. Table 2 summarizes the detailed SDoH categories and data sources used in these studies. More recently, the well-known 2022 n2c2 NLP challenge organized an open challenge with a shared task focusing on SDoH [55].

While NLP methods based on the transformer models have shown promising results in extracting SDoH captured in clinical narratives, challenges remain. First, there is not an off-the-shelf comprehensive package for SDoH extraction, and the adoption of an NLP pipeline trained on one corpus often requires extensive fine-tuning when applied on a different corpus or at another institution. The accuracy of NLP methods depends on several factors, including the quality and consistency of the data, the choice of the NLP models, and the devel-

**Table 2** SDoH category and data sources in recently published NLP systems for SDoH extraction from clinical notes.

Study	SDoH category	Data source
Feller et al. [51]	Sexual history, Sexual orientation, Alcohol use, Substance use, Housing status	Clinical notes from Columbia University Medical Center
Lybarger et al. [52]	Substance use (alcohol, drug, and tobacco), Physical activity, Employment, Insurance, Living status, Sexual orientation, Gender identity, Country of origin, Race, and Environmental exposure	Clinical notes from MIMIC-III and University of Washington
Han et al. [53]	Economic, Education, Healthcare, Housing, Interaction with the legal system, Occupational, Sexual orientation, Social environment, Spiritual life, Substance abuse, Support circumstances and networks, Transportation, Other, Non-social	Clinical notes from MIMIC-III
Stemerman et al. [54]	General financial insecurity, Employment or income insecurity, Housing insecurity, Poor social support, Insurance insecurity, Food insecurity	Clinical notes from University of North Carolina at Chapel Hill Emergency Department
Yu et al. [19]	Abuse (physical or mental), Alcohol use, Drug use, Education, Ethnicity, Financial constraint, Gender, Language, Living condition, Living supply, Marital status, Occupation/Employment, Physical activity, Race, SDoH ICD, Sexual activity, Social cohesion, Tobacco use, Transportation	Clinical notes of cancer patients from the University of Florida Health

opment and training of the NLP algorithms. Furthermore, the complex and nuanced nature of SDoH information, as well as the challenges in standardizing free text, can make it difficult to extract and accurately categorize this information. Despite these challenges, NLP methods have the potential to greatly improve the analysis and utilization of SDoH information recorded in clinical narratives.

Second, the documentation of certain SDoH categories is poor in clinical notes. NLP is only useful when the SDoH are prevalent in clinical narratives. Yu *et al.* [56] reported that in the training corpus of 640 clinical notes from cancer patients, only 19 out of the 38 SDoH categories (based on a review of SDoH definitions from WHO, Healthy People 2030, and CDC) were observed. When the authors applied the trained NLP pipeline on a corpus of breast (n=7,971), lung (n=11,804), and colorectal cancer (n=6,240) patients from the University of Florida Health, among the 19 SDoH categories, 10 had an extraction rate of over 70%, including gender, race, tobacco use, alcohol use, drug use, education, living

supply, marital status, occupation, and sexual activity. The other 9 categories had a fairly low extraction rate, including abuse (physical and mental), ethnicity, financial constraint, language, living condition, physical activity, social cohesion, transportation, and ICD-10 Z codes of SDoH.

### 2.3 Identification of Contextual SDoH through Novel Data Linkage

Contextual SDoH are increasingly recognized as playing critical roles in not only population health but also disparities and structural inequities [57,58]. In environmental epidemiology, the exposome concept was coined to draw attention to a more comprehensive assessment of environmental exposures [59], where the internal exposome refers to “*exposures that impact the internal environment of the body*” such as metabolic factors and microbiota, while the external exposome refers to the “*social, cultural and ecological contexts in which the person lives their life*” such as climate factors and social capital, as well as “*the*

*specific external agents to which one is exposed*” such as specific contaminants, poor diet and lack of exercise [20]. In the United States, these external exposome data (or contextual-level SDoH) can be obtained from numerous publicly accessible data sources such as the American Community Survey (ACS) [60], County Health Rankings (CHR) [61], and Food Environment Atlas (FEA) [62]. Researchers have constructed comprehensive external exposome databases that include multiple domains of contextual-level SDoH. For example, Hu *et al.* have integrated external exposome data from multiple well-validated sources into a comprehensive set of variables of different spatial and temporal resolutions [63]. These contextual-level SDoH can be spatiotemporally linked using residential histories documented in EHR data to study a wide range of population health issues. Previous studies have documented that contextual SDoH have significant associations with health care access and various health outcomes. These associations can be uncovered by analyzing EHR data linked with exposome data [63,64], as well as the Exposome-Wide Association Study (ExWAS) approach (similar to the concept of Genome-Wide Association Study (GWAS) analyses), which enables to systematically screen the associations between thousands of contextual SDoH/environmental exposures and health outcomes based on an agnostic, untargeted, and hypothesis-generating approach [65]. A recently published Social and Environmental Determinants Address Enhancement toolkit (SEnDAE) [66] includes optional components for geocoding addresses that can extend the OMOP common data model. As OMOP operates on a global scale, this and similar initiatives should be seen as an essential step in the process of internationalizing the digitization of the SDoH.

Nevertheless, challenges remain in using contextual SDoH data. External exposome data sources are heterogenous and lack semantic standards [20]. Such heterogeneity also leads to methodological challenges with data engineering (*e.g.*, data source identification, variable selection, and data harmonization), spatiotemporal linkage (*e.g.*, geocoding of patient addresses and

spatiotemporal aggregation), and analyses and interpretation (e.g., ExWAS, prediction, and causation [67]). A very important caveat of ExWASs is the well-known idiom “*associations are not causations*” considering the ecological fallacy. Further, the old “*so what*” question still exists: what do we do with these statistically important contextual SDoH (even if causality was established)?

### 3 Semantic Standards of SDoH via Ontologies

Standardization of the measurement and management of SDoH for individuals, households, and communities as well as linking SDoH information to EHR data, is of utmost importance. Ontologies, usually defined as formal representations of a specific domain, can facilitate semantic interoperability across systems with formal definitions of concepts and their relationships. A few ontologies or terminologies exist that cover certain aspects of SDoH. For example, on the individual-level, the Ontology of Medically Related Social Entities (OMRSE) focuses on health-related social roles [68], while the Semantic Mining of Activity, Social, and Health (SMASH) data system ontology focuses on the interrelations of health, social activities, and daily physical activities [69]. Additionally, there are ontologies on the contextual-level such as the Environment Ontology (ENVO) [70], the Human Health Exposure Analysis Resource (HHEAR) ontology [71], the Child Health Exposure Analysis Resource (CHEAR) ontology [72], and the Environment Conditions, Treatments, and Exposures Ontology (ECTO) [73]. For ontologies related to contextual SDoH (or external exposome data), a previous review and assessment of existing semantic standards for external exposome data have detailed the current landscape, challenges, and future opportunities [20]. Despite the availability of ontologies that cover certain aspects of SDoH, they do not provide a comprehensive representation of SDoH, nor were they designed with the intention to link SDoH information to EHR data, among a number of other limitations [20].

In the recent two years, Rousseau *et al.* [74] developed an ontology-driven information model to integrate SDoH data with the EHRs for pediatric asthma. To achieve this, they identified a list of important environmental measures for pediatric asthma and then assessed existing SDoH frameworks, assessment tools, and terminologies to identify representative data standards for these measures. They found that even though there are LOINC and SNOMED CT concepts relevant to indoor and outdoor air quality measures, these terminologies do not align well with environmental exposure measurements and the concepts in these terminologies often lack the specificity with regards to the data elements from the air quality measurements and questionnaire. Kollapally *et al.* [75] prototyped the Social Determinant of Health Ontology (SOHO) aiming to cover terms related to negative societal phenomena that affect clinical outcomes. After a manual review of relevant publications, Healthy People 2030, and County Ranking models, the prototype of SOHO was developed with

189 classes among which 40% are covered by SNOMED CT, ICD-10-CM, or National Cancer Institute (NCI) Thesaurus with inconsistent coverage. SOHO only has IS-A relations and may not have the desired level of granularity for all the SDoH applications. In a more recent work, Dang *et al.* [21] developed a more comprehensive SDoH ontology called SDoHO whose category and topics were defined by incorporating mainstream sources including WHO, CDC, Healthy People 2020 & 2030, Kaiser Family Foundation, and NAM. Among others, SDoHO is a more formally defined ontology with 706 classes, 105 object properties, and 20 data properties, with 1,542 logical axioms and 966 declaration axioms. Their top-level classes include elements relevant to behavior and lifestyle, social and community context, health care, economic stability, neighborhood, food, and measures/indices/scores. SDoHO is aligned with standard terminologies including LOINC, SNOMED CT, and broadly the UMLS. Table 3 summarizes the recent SDoH ontologies.

**Table 3** Recent SDoH ontologies

Ontology Name	Purpose	Size	Formats
Ontology of Medically Related Social Entities (OMRSE) [68]	This ontology covers the domain of social entities that are related to health care, such as demographic information and the roles of various individuals and organizations.	644 classes	OWL
Semantic Mining of Activity, Social, and Health data (SMASH) [69]	The SMASH system primarily addresses the research topic of sustained weight loss with continued intervention with frequent social contacts.	189 classes	OWL/CSV/RDF
Environmental Ontology (ENVO) [70]	ENVO includes environmental entities such as ecosystems, environmental processes, and environmental qualities.	6,997 classes	OWL/CSV/RDF
Human Health Exposure Analysis Resource (HHEAR) [71]	HHEAR includes measures for exposures that cover the breadth of the “exposome,” which encompasses all environmental exposures including chemical, physical, and biological stressors, as well as lifestyle and social environments, from conception through adulthood.	45,347 classes	OWL/CSV/RDF
Environment Conditions, Treatments, and Exposures Ontology (ECTO) [73]	ECTO describes organismal exposures such as toxicological research, environmental variables, dietary features, and patient-reported data from surveys.	11,865 classes	OWL/CSV/RDF
Social Determinant of Health Ontology (SOHO) [75]	SOHO covers relevant concepts and IS-A relationships describing impacts and associations of social determinants.	189 classes	OWL
SDoHO [21]	SDoHO represents fundamental SDoH factors and their relationships in a standardized and measurable way. The ontology formally models classes, relationships, and constraints based on multiple SDoH-related resources.	706 classes	OWL

The challenges of developing and adopting SDoH ontologies for public and population health applications are multi-faceted. First, there is a lack of consensus on the information models and dimensions for SDoH. Leading organizations such as the WHO, NAM, CDC, Healthy People 2020 & 2030, all have developed their own frameworks for SDoH. It is challenging to consolidate the concepts and measures in these frameworks and define relationships between SDoH concepts. Second, even though existing standard terminologies such as ICD-10, SNOMED CT, and LOINC have added certain concepts for SDoH, their coverage is limited, the actual use of these codes in EHR is low [17], and the semantic alignment between these coding schemes are challenging [76]. Researchers often reported a lack of granularity in these terminologies. Small application ontologies that cover certain narrow aspects of SDoH or focus on a specific clinical domain continue to exist, and the categorization of the SDoH factors is not uniform across studies. Related to public and population health, there is a lack of alignment between exposome measures (such as air quality and water quality measures) and the representation of these measures in standard terminologies.

So far, we have not found studies that demonstrate the use of ontologies for linking SDoH to EHR data. To make effective use of SDoH ontologies for public health and epidemiology, a few questions await to be answered:

- What level of formalism is required for SDoH ontologies?
- How to use ontologies to standardize the measures of SDoH information?
- How granular should these ontologies be?
- How to effectively use ontologies to standardize SDoH data that can be integrated with EHR data at the patient level, neighborhood level, and regional level to model factors that impacts on health such as disease burden?
- How should SDoH ontologies be integrated with other ontologies to facilitate downstream use cases? For example, many SDoH-targeted interventions are closely related to behavioral changes, and lead to the needs to be linked to ontologies such as the Behavior Change Intervention Ontology (BCIO) [77] to guide intervention development.

In addition, besides recommending SDoH data elements, regulatory agencies should also recommend ontologies and provide a guideline on the standardization and integration of SDoH information.

## 4 Applications of SDoH

### 4.1 Incorporating SDoH in Disease Screening and Social Risk Prediction

To help guide precision prevention and treatment of diseases, it is critical to consider social risks and incorporate SDoH when developing disease screening and prediction models. Although there is overlap, individual-level and contextual-level SDoH approaches for assessing patient social risks are not equivalent [48] and it is important to consider both individual-level and contextual-level SDoH when developing prediction models. In fact, there have been some models of predicting social risks published, such as the polysocial risk score [78], polyexposure risk score [79], and polyexposomic risk score [80]. Taking the polysocial risk score as an example, it could help predict individual-level social risk of a disease or a particular health outcome with different combinations of social conditions without knowing the precise contribution of each social factor [78]. The social factors considered by polysocial risk scores include both individual SDoH factors such as income, education, religion, sex, race-ethnicity as well as contextual social, community, and physical environmental factors such as quality of housing, local crime level, and air and water quality. Although not explicitly stated as an approach for developing the polysocial risk score, Guo *et al.* examined both individual-level (extracted via NLP over clinical notes) and contextual-level (via spatiotemporal linkage) SDoH linked with EHRs and found novel SDoH associated with lower initiation of cardioprotective drugs in patients with type 2 diabetes (T2D) and varying effect across racial groups [81]. The polyexposure risk score, which combines multiple correlated nongenetic exposure and lifestyle factors, has been shown to provide modest incremental

prediction accuracy of predicting T2D over established clinical risk factors [79]. The polyexposomic risk score was initially developed for hypertensive disorders of pregnancy using external exposome-wide data consisting of 5,510 factors characterizing women's surrounding natural, built, and social environment during pregnancy [80]. The study found that neighborhood socioeconomic status, housing characteristics, meteorology factors, and air pollutants are predictive of hypertensive disorders of pregnancy.

SDoH data may also play a critical role in developing predictive models for critical public health crises such as opioid use crisis. Gao *et al.* found that Medicaid enrollees with a documented SDoH vulnerability had 26% higher odds of having an opioid use disorder than those without. However, the authors noted a high level of SDoH missingness in their data, suggesting that more consistent and thorough SDoH documentation may have major implications for such predictive models in the future [82]. In their study of factors leading to non-fatal overdose leading to intensive care unit admission, Mitra *et al.* addressed this missingness with NLP of clinical notes to fill gaps left by structured SDoH documentation and ultimately captured >99% of their SDoH variables from the free text [83].

### 4.2 Development of SDoH-related Interventions

To develop effective public health interventions that target population at higher risk for certain health problems and improve health equity, it is critical to identify effective social risk management strategies, particularly for marginalized groups. Advances in artificial intelligence combined with the increasing availability of RWD offer a unique opportunity to develop innovative approaches that improve both health outcomes and health equity by addressing SDoH. However, key data and methodologic barriers exist, some of which are extensively discussed above, such as the fact that RWD are not well-integrated with either contextual or individual-level SDoH data although factors from both levels are associated with T2D, with complex interplay among them.

Furthermore, from a modeling methodology perspective, although associations of multiple SDoH with health care and health outcomes are well documented [6], predictors may not be causally associated with outcomes. Therefore, there are critical gaps in understanding who may benefit from a given SDoH-targeted intervention (e.g., food pharmacy, transportation support for medical needs [84]). Machine learning (ML) has led to success in various RWD analysis tasks. However, RWD are observational in nature; thus, the causal inference framework needs to be incorporated with ML approaches (i.e., causal-principled ML models such as causal forest) to account for inherent biases (e.g., confounding and selection biases) when providing cause-and-effect estimates of potential SDoH interventions in RWD [85]. For example, Tang *et al.* used a causal ML method (i.e., doubly robust learning) to estimate the conditional average treatment effects and found a heterogeneous effect of SDoH on the risk of dementia [86]. Through causal-principled ML models, researchers can fill critical gaps in the causal effects of key actionable SDoH on healthcare and outcomes. Establishing causal effects of individual SDoH on the health outcomes of interest is critical as clinical practices are built on causality (e.g., via randomized controlled trials), so that we know exactly the potential benefits and harms of prescribing an intervention, regardless of whether it is a medical treatment or a SDoH intervention.

Nevertheless, knowing the causal effects of SDoH on health is not sufficient, as there are a number of other challenges beyond the data and methods, such as the lack of a social risk management tool in EHRs that can leverage existing rich data sources and consider the totality of both contextual and individual-level SDoH, to semi-automatically identify individuals at high social risk (e.g., social risk screening via polysocial risk scores) while limiting documentation burden. The field also lacks tools that can not only provide critical decision support information (e.g., prioritized key actionable SDoH and causal effect estimates), but also guide the next steps to address individual patients' unmet social needs (e.g., referral to community-based organizations for specific SDoH identified). From the informatics'

perspective, the usability (ease of use), acceptability (perceived usefulness), and how such tools are integrated in existing clinical workflows (considering the limited time providers already have with each patient) are critical. Addressing SDoH and unmet social needs is not necessarily a job of clinicians or even nurses, but an effort of the community with multiple stakeholders ranging from patients and caregivers to providers and health systems, to community organizations and government agencies. Informaticians play a critical role in providing novel technologies to support these activities ranging from data integration of heterogeneous sources to modeling with causal-principled methods to tool development via user-centered design considering human factors to EHR integration and implementation science via a learning health system framework.

## 5 Conclusions and Future Directions

SDoH factors affect people's health at the individual, family, community, and society levels. There is an increasing interest in examining the role of SDoH in public and population health, as well as health disparities using RWD. In this opinion review, we summarized data resources and recent informatics approaches to screen and harmonize different levels of SDoH from heterogeneous data sources and utilize SDoH with RWD. We also identified potential challenges and barriers to the low documenting rate of SDoH in EHR systems [87], including lack of integration into clinical workflows, lack of incentives for SDoH data collection, and lack of training and tools for clinicians to derive actionable insights for decision making. The informatics community has made strides in developing NLP methods to extract SDoH from clinical narratives, linking EHRs with public surveys and environmental data, creating SDoH ontologies for standardization, and developing SDoH-based social risk scores. To better leverage SDoH, future work should establish incentives, policies [88], quality measures, and training [42] to improve the collection and use of SDoH. Technical solutions such as social risk

management tools should follow user-centered design and be integrated into real-world clinical workflows to identify social risks and address unmet social needs. Note that even though the majority of recent studies on the methods and applications about linking EHRs with SDoH to improve public health were conducted in the United States, there are approaches for collecting SDoH data in the global context. A recent paper by Cossio [89] reviewed different approaches for digitally collecting or predicting SDoH such as quality of public transportation (Lisbon [90], Brazilian cities [91]), air quality (a city in Turkey [92]), and education (Sweden [93]). We believe that integrating SDoH into health care can improve public health, reduce healthcare disparities, and help inform public policies for effective interventions.

### Acknowledgments

This paper was partially supported by University of Florida-Florida State University Clinical and Translational Science Award (CTSA) funded by National Center for Advancing Translational Sciences (NCATS) under Award Number UL1TR001427, the University of North Carolina at Chapel Hill CTSA funded by NCATS under award number UL1TR002489. The authors are supported by the following awards from the National Institutes of Health (NIH), including R21LM013911, P01AA029547, R01DK133465, R01CA246418, R21ES032762, R21CA245858, and R01AG080624, a Patient-Centered Outcomes Research Institute (PCORI) award ME-2018C3-14754, a Centers for Disease Control and Prevention (CDC) award U18DP006512, and the following Slovenian Research Agency grants: ARRS N2-0101, ARRS P2-0057, ARRS N3-0307, ARRS BI-US/22-24-138.

### References

1. Braveman P, Gottlieb L. The social determinants of health: it's time to consider the causes of the causes. *Public Health Rep Wash DC* 1974 2014;129 Suppl 2:19–31. doi: 10.1177/00333549141291S206.
2. WHO. Social determinants of health n.d. <https://www.who.int/health-topics/social-determinants-of-health> (accessed January 29, 2023).
3. McGinnis JM, Williams-Russo P, Knickman JR. The case for more active policy attention to health



- promotion. *Health Aff Proj Hope* 2002;21:78–93. doi: 10.1377/hlthaff.21.2.78.
4. McGinnis JM, Foegen WH. Actual causes of death in the United States. *JAMA* 1993;270:2207–12.
  5. Danaei G, Ding EL, Mozaffarian D, Taylor B, Rehm J, Murray CJL, et al. The preventable causes of death in the United States: comparative risk assessment of dietary, lifestyle, and metabolic risk factors. *PLoS Med* 2009;6:e1000058. doi: 10.1371/journal.pmed.1000058.
  6. Social Determinants of Health - Healthy People 2030 n.d. <https://health.gov/healthypeople/priority-areas/social-determinants-health> (accessed February 5, 2023).
  7. What are the Social Determinants of Health? *Natl Acad Med* n.d. <https://nam.edu/programs/culture-of-health/young-leaders-visualize-health-equity/what-are-the-social-determinants-of-health/> (accessed February 5, 2023).
  8. 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. *Capturing Social and Behavioral Domains and Measures in Electronic Health Records: Phase 2*. Washington (DC): National Academies Press (US); 2015.
  9. WHO. Health Equity Status Report initiative n.d. <https://www.who.int/europe/initiatives/health-equity-status-report-initiative> (accessed April 25, 2023).
  10. Buzeti T, Madureira Lima J, Yang L, Brown C. Leaving no one behind: health equity as a catalyst for the sustainable development goals. *Eur J Public Health* 2020;30:i24–7. doi: 10.1093/eurpub/ckaa033.
  11. Annis R, Beattie M, Racher FF. *Rural Community Health and Well-Being: A Guide to Action*; 2004.
  12. Saydah S, Bullard KM, Chen Y, Ali MK, Gregg EW, Geiss L, et al. Trends in cardiovascular disease risk factors by obesity level in adults in the United States, NHANES 1999–2010. *Obes Silver Spring Md* 2014;22:1888–95. doi: 10.1002/oby.20761.
  13. Husk K, Elston J, Gradinger F, Callaghan L, Asthana S. Social prescribing: where is the evidence? *Br J Gen Pract* 2019;69:6–7. doi: 10.3399/bjgp19X700325.
  14. Hill-Briggs F, Ephraim PL, Vraney EA, Davidson KW, Pekmezaris R, Salas-Lopez D, et al. Social Determinants of Health, Race, and Diabetes Population Health Improvement: Black/African Americans as a Population Exemplar. *Curr Diab Rep* 2022;22:117–28. doi: 10.1007/s11892-022-01454-3.
  15. Ogunwole SM, Golden SH. Social Determinants of Health and Structural Inequities-Root Causes of Diabetes Disparities. *Diabetes Care* 2021;44:11–3. doi: 10.2337/dci20-0060.
  16. Guo Y, Bian J, Wang F. Editorial: Measuring and Analysing Social Determinants of Health in the Era of Big Data. *Front Public Health* 2022;10:902942. doi: 10.3389/fpubh.2022.902942.
  17. Guo Y, Chen Z, Xu K, George TJ, Wu Y, Hogan W, et al. International Classification of Diseases, Tenth Revision, Clinical Modification social determinants of health codes are poorly used in electronic health records. *Medicine (Baltimore)* 2020;99:e23818. doi: 10.1097/MD.00000000000023818.
  18. Patra BG, Sharma MM, Vekaria V, Adekkanattu P, Patterson OV, Glicksberg B, et al. Extracting social determinants of health from electronic health records using natural language processing: a systematic review. *J Am Med Inform Assoc* 2021;28:2716–27. doi: 10.1093/jamia/ocab170.
  19. Yu Z, Yang X, Dang C, Wu S, Adekkanattu P, Pathak J, et al. A Study of Social and Behavioral Determinants of Health in Lung Cancer Patients Using Transformers-based Natural Language Processing Models. *AMIA Annu Symp Proc* 2021:1225–33.
  20. Zhang H, Hu H, Diller M, Hogan WR, Proserpi M, Guo Y, et al. Semantic standards of external exposure data. *Environ Res* 2021;197:111185. doi: 10.1016/j.envres.2021.111185.
  21. Dang Y, Li F, Hu X, Keloth VK, Zhang M, Fu S, et al. Systematic Design and Evaluation of Social Determinants of Health Ontology (SDoHO) 2022. doi: 10.48550/arXiv.2212.01941.
  22. Guo SJ, Shao H. Growing global burden of type 1 diabetes needs multitiered precision public health interventions. *Lancet Diabetes Endocrinol* 2022;10:688–9. doi: 10.1016/S2213-8587(22)00257-1.
  23. He Z, Tao C, Bian J, Dumontier M, Hogan WR. Semantics-Powered Healthcare Engineering and Data Analytics. *J Healthc Eng* 2017;2017:e7983473. doi: 10.1155/2017/7983473.
  24. Amith M, He Z, Bian J, Lossio-Ventura JA, Tao C. Assessing the practice of biomedical ontology evaluation: Gaps and opportunities. *J Biomed Inform* 2018;80:1–13. doi: 10.1016/j.jbi.2018.02.010.
  25. Fecho K, Pfaff E, Xu H, Champion J, Cox S, Stillwell L, et al. A novel approach for integrating and sharing clinical data: the Translator Integrated Clinical and Environmental Exposures Service. *J Am Med Inform Assoc* 2019;26:1064–73. doi: 10.1093/jamia/ocz042.
  26. Kopitar L, Kocbek P, Cilar L, Sheikh A, Stiglic G. Early detection of type 2 diabetes mellitus using machine learning-based prediction models. *Sci Rep* 2020;10:11981. doi: 10.1038/s41598-020-68771-z.
  27. Fecho K, Ahalt SC, Knowles M, Krishnamurthy A, Leigh M, Morton K, et al. Leveraging Open Electronic Health Record Data and Environmental Exposures Data to Derive Insights Into Rare Pulmonary Disease. *Front Artif Intell* 2022;5:918888. doi: 10.3389/frai.2022.918888.
  28. Fecho K, Ahalt SC, Appold S, Arunachalam S, Pfaff E, Stillwell L, et al. Development and Application of an Open Tool for Sharing and Analyzing Integrated Clinical and Environmental Exposures Data: Asthma Use Case. *JMIR Form Res* 2022;6:e32357. doi: 10.2196/32357.
  29. Pfaff ER, Madlock-Brown C, Baratta JM, Bhatia A, Davis H, Girvin A, et al. Coding long COVID: characterizing a new disease through an ICD-10 lens. *BMC Med* 2023;21:58. doi: 10.1186/s12916-023-02737-6.
  30. CDC. The Social-Ecological Model: A Framework for Prevention 2022. <https://www.cdc.gov/violenceprevention/about/social-ecologicalmodel.html> (accessed February 12, 2023).
  31. NIMHD. The NIMHD Minority Health and Health Disparities Research Framework. NIMHD n.d. <https://www.nimhd.nih.gov/about/overview/research-framework/nimhd-framework.html> (accessed February 12, 2023).
  32. Wang M, Pantell MS, Gottlieb LM, Adler-Milstein J. Documentation and review of social determinants of health data in the EHR: measures and associated insights. *J Am Med Inform Assoc* 2021;28:2608–16. doi: 10.1093/jamia/ocab194.
  33. Arons A, DeSilvey S, Fichtenberg C, Gottlieb L. Documenting social determinants of health-related clinical activities using standardized medical vocabularies. *JAMIA Open* 2019;2:81–8. doi: 10.1093/jamiaopen/ooy051.
  34. National Association of Community Health Centers. PRAPARE: Protocol for Responding to and Assessing Patient Assets, Risks, and Experiences n.d. [https://www.in.gov/health/cdpc/files/PRAPARE\\_Assessment\\_Tool.pdf](https://www.in.gov/health/cdpc/files/PRAPARE_Assessment_Tool.pdf) (accessed January 16, 2023).
  35. Billioux A, Verlander K, Anthony S, Alley D. Standardized Screening for Health-Related Social Needs in Clinical Settings: The Accountable Health Communities Screening Tool. *NAM Perspect* 2017. doi: 10.31478/201705b.
  36. The Health Leads Screening Toolkit. Health Leads n.d. <https://healthleadsusa.org/resources/the-health-leads-screening-toolkit/> (accessed January 29, 2023).
  37. SEEK Parent Questionnaire n.d. <https://seekwell-being.org/wp-content/uploads/2022/10/SEEK-PQ-R-English-9-22.pdf> (accessed January 29, 2023).
  38. Garg A, Toy S, Tripodis Y, Silverstein M, Freeman E. Addressing social determinants of health at well child care visits: a cluster RCT. *Pediatrics* 2015;135:e296–304. doi: 10.1542/peds.2014-2888.
  39. Freij M, Dullabh P, Lewis S, Smith SR, Hovey L, Dhopeshwarkar R. Incorporating Social Determinants of Health in Electronic Health Records: Qualitative Study of Current Practices Among Top Vendors. *JMIR Med Inform* 2019;7:e13849. doi: 10.2196/13849.
  40. Walters K, Clark M, Dard S, et al. N3C data enhancements: A path for expanding common data models. *AMIA Summits Transl Sci Proc*. Forthcoming 2023.
  41. Camacho-Rivera M, Islam JY, Vidot DC, Espinoza J, Galiatsatos P, Sule A, et al. Social Determinants of Health During the COVID-19 Pandemic in the US: Precision Through Context. In: Hsueh P-YS, Wetter T, Zhu X, editors. *Personal Health Informatics. Patient Participation in Precision Health*. Cham: Springer International Publishing; 2022. p. 397–425. doi: 10.1007/978-3-031-07696-1\_19.
  42. Nour N, Stuckler D, Ajayi O, Abdalla ME. Effectiveness of alternative approaches to integrating SDOH into medical education: a scoping review. *BMC Med Educ* 2023;23:18. doi: 10.1186/s12909-022-03899-2.
  43. American Hospital Association. ICD-10-CM Coding for Social Determinants and Health n.d. <https://www.aha.org/system/files/2018-04/value-initiative-icd-10-code-social-determinants-of-health.pdf> (accessed January 16, 2023).
  44. Yan AF, Chen Z, Wang Y, Campbell JA, Xue Q-L, Williams MY, et al. Effectiveness of Social Needs

- Screening and Interventions in Clinical Settings on Utilization, Cost, and Clinical Outcomes: A Systematic Review. *Health Equity* 2022;6:454–75. doi: 10.1089/heap.2022.0010.
45. Abiri A, Evans DD, Hamilton JB. Strategies to Integrate the Practice of Social Emergency Medicine Into Routine Patient Care. *Adv Emerg Nurs J* 2022;44:78–83. doi: 10.1097/TME.0000000000000409.
  46. Yetisgen M, Vanderwende L. Automatic Identification of Substance Abuse from Social History in Clinical Text. In: ten Teije A, Popow C, Holmes JH, Sacchi L, editors. *Artif. Intell. Med.*, Cham: Springer International Publishing; 2017, p. 171–81. doi: 10.1007/978-3-319-59758-4\_18.
  47. Gundlapalli AV, Carter ME, Palmer M, Ginter T, Redd A, Pickard S, et al. Using natural language processing on the free text of clinical documents to screen for evidence of homelessness among US veterans. *AMIA Annu Symp Proc* 2013;2013:537–46.
  48. Hatf E, Rouhizadeh M, Nau C, Xie F, Rouillard C, Abu-Nasser M, et al. Development and assessment of a natural language processing model to identify residential instability in electronic health records' unstructured data: a comparison of 3 integrated healthcare delivery systems. *JAMIA Open* 2022;5:ooac006. doi: 10.1093/jamiaopen/ooac006.
  49. Dillahunt-Aspillaga C, Finch D, Massengale J, Kretzmer T, Luther SL, McCart JA. Using information from the electronic health record to improve measurement of unemployment in service members and veterans with mTBI and post-deployment stress. *PLoS One* 2014;9:e115873. doi: 10.1371/journal.pone.0115873.
  50. Fernandes AC, Dutta R, Velupillai S, Sanyal J, Stewart R, Chandran D. Identifying Suicide Ideation and Suicidal Attempts in a Psychiatric Clinical Research Database using Natural Language Processing. *Sci Rep* 2018;8:7426. doi: 10.1038/s41598-018-25773-2.
  51. 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;11:172–81. doi: 10.1055/s-0040-1702214.
  52. Lybarger K, Ostendorf M, Yetisgen M. Annotating social determinants of health using active learning, and characterizing determinants using neural event extraction. *J Biomed Inform* 2021;113:103631. doi: 10.1016/j.jbi.2020.103631.
  53. Han S, Zhang RF, Shi L, Richie R, Liu H, Tseng A, 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. doi: 10.1016/j.jbi.2021.103984.
  54. Sterman R, Arguello J, Brice J, Krishnamurthy A, Houston M, Kitzmiller R. Identification of social determinants of health using multi-label classification of electronic health record clinical notes. *JAMIA Open* 2021;4:ooaa069. doi: 10.1093/jamiaopen/ooaa069.
  55. National NLP Clinical Challenges (n2c2) Track 2 Extracting Social Determinants of Health n.d. <https://n2c2.dbmi.hms.harvard.edu/2022-track-2> (accessed February 5, 2023).
  56. Yu Z, Yang X, Dang C, Adekanlatu P, Patra BG, Peng Y, et al. SODA: A Natural Language Processing Package to Extract Social Determinants of Health for Cancer Studies 2022. doi: 10.48550/arXiv.2212.03000.
  57. Kolak M, Bhatt J, Park YH, Padrón NA, Molefe A. Quantification of Neighborhood-Level Social Determinants of Health in the Continental United States. *JAMA Netw Open* 2020;3:e1919928. doi: 10.1001/jamanetworkopen.2019.19928.
  58. World Health Organization. A conceptual framework for action on the social determinants of health. World Health Organization; 2010. <https://apps.who.int/iris/rest/bitstreams/52952/retrieve> (accessed May 2, 2023).
  59. Wild CP. Complementing the genome with an “exposome”: the outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol* 2005;14:1847–50. doi: 10.1158/1055-9965.EPI-05-0456.
  60. US Census Bureau. American Community Survey (ACS). *CensusGov* n.d. <https://www.census.gov/programs-surveys/acs> (accessed February 10, 2023).
  61. How Healthy is your County? | County Health Rankings. *Cty Health Rank Roadmaps* n.d. <https://www.countyhealthrankings.org/county-health-rankings-roadmaps> (accessed February 10, 2023).
  62. USDA ERS - Food Environment Atlas n.d. <https://www.ers.usda.gov/data-products/food-environment-atlas/> (accessed February 10, 2023).
  63. Hu H, Zheng Y, Wen X, Smith SS, Nizomov J, Fische J, et al. An external exposome-wide association study of COVID-19 mortality in the United States. *Sci Total Environ* 2021;768:144832. doi: 10.1016/j.scitotenv.2020.144832.
  64. Hu H, Zhao J, Savitz DA, Proserpi M, Zheng Y, Pearson TA. An external exposome-wide association study of hypertensive disorders of pregnancy. *Environ Int* 2020;141:105797. doi: 10.1016/j.envint.2020.105797.
  65. Juarez PD, Matthews-Juarez P. Applying an Exposome-Wide (ExWAS) Approach to Cancer Research. *Front Oncol* 2018;8:313. doi: 10.3389/fonc.2018.00313.
  66. Kingsbury P, Abajian H, Abajian M, Angyan P, Espinoza J, MacDonald B, et al. SENDAE: A Resource for Expanding Research into Social and Environmental Determinants of Health. *Comput Methods Programs Biomed* 2023:107542. doi: 10.1016/j.cmpb.2023.107542.
  67. Hu H, Liu X, Zheng Y, He X, Hart J, James P, et al. Methodological challenges in spatial and contextual exposome-health studies. *Crit Rev Environ Sci Technol* 2022;0:1–20. doi: 10.1080/10643389.2022.2093595.
  68. Hicks A, Hanna J, Welch D, Brochhausen M, Hogan WR. The ontology of medically related social entities: recent developments. *J Biomed Semant* 2016;7:47. doi: 10.1186/s13326-016-0087-8.
  69. Phan N, Dou D, Wang H, Kil D, Piniewski B. Ontology-based deep learning for human behavior prediction with explanations in health social networks. *Inf Sci* 2017;384:298–313. doi: 10.1016/j.ins.2016.08.038.
  70. Buttigieg PL, Pafilis E, Lewis SE, Schildhauer MP, Walls RL, Mungall CJ. The environment ontology in 2016: bridging domains with increased scope, semantic density, and interoperability. *J Biomed Semant* 2016;7:57. doi: 10.1186/s13326-016-0097-6.
  71. Viet SM, Falman JC, Merrill LS, Faustman EM, Savitz DA, Mervish N, et al. Human Health Exposure Analysis Resource (HHEAR): A model for incorporating the exposome into health studies. *Int J Hyg Environ Health* 2021;235:113768. doi: 10.1016/j.ijheh.2021.113768.
  72. Balshaw DM, Collman GW, Gray KA, Thompson CL. The Children's Health Exposure Analysis Resource: enabling research into the environmental influences on children's health outcomes. *Curr Opin Pediatr* 2017;29:385–9. doi: 10.1097/MOP.0000000000000491.
  73. Chan LE, Thessen AE, Duncan WD, Matentzoglou N, Schmitt C, Grondin CJ, et al. The Environmental Conditions, Treatments, and Exposures Ontology (ECTO): connecting toxicology and exposure to human health and beyond. *J Biomed Semantics* 2023;14(1):3. doi: 10.1186/s13326-023-00283-x.
  74. Rousseau JF, Oliveira E, Tierney WM, Khurshid A. Methods for development and application of data standards in an ontology-driven information model for measuring, managing, and computing social determinants of health for individuals, households, and communities evaluated through an example of asthma. *J Biomed Inform* 2022;136:104241. doi: 10.1016/j.jbi.2022.104241.
  75. Kollapally NM, Chen Y, Xu J, Geller J. An Ontology for the Social Determinants of Health Domain. *IEEE Computer Society*; 2022. p. 2403–10. doi: 10.1109/BIBM55620.2022.9995544.
  76. Hastings J. Achieving Inclusivity by Design: Social and Contextual Information in Medical Knowledge. *Yearb Med Inform* 2022;31:228–35. doi: 10.1055/s-0042-1742509.
  77. Michie S, West R, Finnerty AN, Norris E, Wright AJ, Marques MM, et al. Representation of behaviour change interventions and their evaluation: Development of the Upper Level of the Behaviour Change Intervention Ontology. *Wellcome Open Res* 2020;5:123. doi: 10.12688/wellcomeopenres.15902.2.
  78. Figueroa JF, Frakt AB, Jha AK. Addressing Social Determinants of Health: Time for a Polysocial Risk Score. *JAMA* 2020;323:1553–4. doi: 10.1001/jama.2020.2436.
  79. He Y, Lakhani CM, Rasooly D, Manrai AK, Tzoulaki I, Patel CJ. Comparisons of Polyexposure, Polygenic, and Clinical Risk Scores in Risk Prediction of Type 2 Diabetes. *Diabetes Care* 2021;44:935–43. doi: 10.2337/dc20-2049.
  80. Hu H, Zhao J, Bian J, Zheng Y, Pearson TA. Abstract P428: A Polyexposomic Risk Score for Hypertensive Disorders of Pregnancy Using External Exposome Data. *Circulation* 2020;141:AP428–AP428. doi: 10.1161/circ.141.suppl\_1.P428.
  81. Guo J, Hu H, Zheng Y, Guo Y, Chen A, Magnani JW, et al. Abstract 003: Contextual- And Personal-level Social Determinants Of Health And Real-world Adoption Of Novel Treatments For Improving Cardiovascular Outcomes In Type 2 Diabetes. *Circulation* 2022;145:A003–A003. doi: 10.1161/circ.145.suppl\_1.003.
  82. Gao W, Leighton C, Chen Y, Jones J, Mistry P.

- Predicting opioid use disorder and associated risk factors in a Medicaid managed care population. *Am J Manag Care* 2021;27:148–54. doi: 10.37765/ajmc.2021.88617.
83. Mitra A, Ahsan H, Li W, Liu W, Kerns RD, Tsai J, et al. Risk Factors Associated With Nonfatal Opioid Overdose Leading to Intensive Care Unit Admission: A Cross-sectional Study. *JMIR Med Inform* 2021;9:e32851. doi: 10.2196/32851.
84. Transportation to Support Rural Healthcare Overview - Rural Health Information Hub n.d. <https://www.ruralhealthinfo.org/topics/transportation> (accessed February 10, 2023).
85. Prosperi M, Guo Y, Sperrin M, Koopman JS, Min JS, He X, et al. Causal inference and counterfactual prediction in machine learning for actionable healthcare. *Nat Mach Intell* 2020;2:369–75. doi: 10.1038/s42256-020-0197-y.
86. Tang H, Wu Y, Brown J, Anton S, Hernandez I, Bian J, Guo J. Heterogeneous effect of social and behavioral determinants of health on the risk of dementia. *Alzheimers Dement* 2022;19:e063752.
87. Gold R, Bunce A, Cowburn S, Dambrun K, Dearing M, Middendorf M, et al. Adoption of Social Determinants of Health EHR Tools by Community Health Centers. *Ann Fam Med* 2018;16:399–407. doi: 10.1370/afm.2275.
88. 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;27:1764–73. doi: 10.1093/jamia/ocaa143.
89. Cossio M. Digital social determinants of health 2023. doi: 10.33774/coe-2023-xvh49. (working paper, non peer reviewed).
90. Foell S, Phithakkitmukoon S, Kortuem G, Veloso M, Bento C. Catch me if you can: Predicting mobility patterns of public transport users. Proceedings of the 17th International IEEE Conference on Intelligent Transportation Systems; 2014. p. 1995–2002. doi: 10.1109/ITSC.2014.6957997.
91. Costa C, Ha J, Lee S. Spatial disparity of income-weighted accessibility in Brazilian Cities: Application of a Google Maps API. *J Transp Geogr* 2021;90:102905. doi: 10.1016/j.jtrangeo.2020.102905.
92. Kök İ, Şimşek MU, Özdemir S. A deep learning model for air quality prediction in smart cities. Proceedings of the IEEE International Conference on Big Data; 2017. p. 1983–90. doi: 10.1109/BigData.2017.8258144.
93. Ludvigsson JF, Svedberg P, Olén O, Bruze G, Neovius M. The longitudinal integrated database for health insurance and labour market studies (LISA) and its use in medical research. *Eur J Epidemiol* 2019;34:423–37. doi: 10.1007/s10654-019-00511-8.

Correspondence to:  
Zhe He, PhD, FAMIA  
School of Information  
Florida State University  
142 Collegiate Loop  
Tallahassee, Florida, 32306-2100 United States  
Tel: +1 850 644 5775  
E-mail: zhe@fsu.edu