Perspective

From Genes to Geography, from Cells to Community, from Biomolecules to Behaviors: The Importance of Social Determinants of Health

Jaysón Davidson [1,2,*], Rohit Vashisht [2] and Atul J. Butte [2]

1   Pharmaceutical Science and Pharmacogenomics Graduate Program, University of California San Francisco, San Francisco, CA 94143, USA
2   Bakar Computational Health Sciences Institute, University of California San Francisco, San Francisco, CA 94143, USA
*   Correspondence: jayso'n.davidson@ucsf.edu

Abstract: Much scientific work over the past few decades has linked health outcomes and disease risk to genomics, to derive a better understanding of disease mechanisms at the genetic and molecular level. However, genomics alone does not quite capture the full picture of one's overall health. Modern computational biomedical research is moving in the direction of including social/environmental factors that ultimately affect quality of life and health outcomes at both the population and individual level. The future of studying disease now lies at the hands of the social determinants of health (SDOH) to answer pressing clinical questions and address healthcare disparities across population groups through its integration into electronic health records (EHRs). In this perspective article, we argue that the SDOH are the future of disease risk and health outcomes studies due to their vast coverage of a patient's overall health. SDOH data availability in EHRs has improved tremendously over the years with EHR toolkits, diagnosis codes, wearable devices, and census tract information to study disease risk. We discuss the availability of SDOH data, challenges in SDOH implementation, its future in real-world evidence studies, and the next steps to report study outcomes in an equitable and actionable way.

Keywords: social determinants of health; electronic health records; real-world evidence; census tract; data science

1. Introduction

Understanding disease at the molecular level has dominated the field of genetics, which has been the major basis for studying disease risk over the past two decades. Researchers have presented considerable evidence that disease risk is generally conferred through genetic inheritance, and now more recently, through specific rare and common mutations [1,2]. Using tools in molecular and cellular biology, researchers and medical providers can investigate many diseases and conditions. However, the results of previous investigations have shown that disease risk is too complex to model using genetics or molecules alone. Indeed, genetic, social, and environmental factors including socioeconomic status, geolocation, and age, as well as racial and ethnic background play a role in disease risk across different population groups [3]. Growing evidence increasingly indicates the importance of accounting for the social and environmental factors that are likely to affect health outcomes. While Dr. Phil Bourne, whom this Special Issue honors, is certainly known for his work in computational methodologies and structural biology, he also understood the importance of external influences on health and called for better methods to measure and "describe individuals' activity spaces and exposure to the built, natural, social, and economic environments that influence behaviors and health outcomes" [4].

2. Social Determinants of Health

Social determinants of health (SDOH) are one of the ways to capture, represent, and assess the impact of social and environmental factors in clinical research, thus improving patient care. SDOH are the conditions in which people are born, live, work, play, worship, and age, which affect a wide range of health, functioning, quality of life outcomes, and risks [5]. A patient's SDOH can be used to estimate their access to healthcare and treatments, their positive or negative health outcomes, and to assess comorbidities by using information related to an individual's health including alcohol and tobacco usage, socioeconomic status, insurance status, living situation, access to healthy foods, access to health literacy, and access to quality of care [5]. The main components of the SDOH commonly gathered in medicine are grouped into five domains: economic stability, education access and quality, healthcare access and equity, social and community context, and neighborhood and built environment [5]. Though noted separately, each domain is interconnected to match the complexity of SDOH variables and represent SDOH at both the population and individual levels [6,7].

Population-level SDOH measures are heavily reliant on census tract information derived from the United States Census Bureau (U.S. Census). Census tracts are indicative of geographical areas, which are defined as small, relatively permanent statistical subdivisions of a county providing information on demographic and housing estimates, occupation codes, industry codes, product and service codes, and material/fuel codes [8]. Census tracts have surveys such as the American community survey, decennial census, economic surveys, population estimates, public sector census, and economic censuses that can be leveraged to assess the overall impact of socioeconomic parameters on the health and wellbeing of patients in a given healthcare system at a given geographical location. Census tract information is gathered by assigning each person, household, housing unit, institution, farm, business establishment, or other responding entity to a specific location, and then assigning that location to a zip code tabulation area appropriate to the census or sample survey by way of geocoding [8,9]. The geocoding process ensures that the Census Bureau can provide correct counts for small geographic entities and that both the Census Bureau and data users can accumulate the data for small entities to provide totals for larger geographic entities such as zip code areas. Census tract information has been used to develop indices that directly explain the SDOH of people by using their zip code location to develop the area deprivation index, social vulnerability index, and modified retail food index [10–13]. Indices that use census tract information often categorize data by socioeconomic status, location, and education to calculate the deprivation or vulnerability of people residing in a location.

SDOH are utilized in clinical care and research studies by way of electronic health records (EHRs) which are the primary way to capture real-world data from providers on patient encounters in a health system [14]. EHRs provide a unique opportunity to study the relationship between SDOH and the management and outcomes of clinical diseases through real-world data (RWD). RWD captured in EHRs are used to develop real-world evidence (RWE) studies that analyze data and inform providers about the causes of different treatment strategies, disease risk, quality of life, and outcomes for different patients and populations. RWE studies often contain diverse patient populations that are representative of real patients' health where common SDOH are collected. Prior to EHRs, the SDOH were primarily captured by population-level questionnaires administered by the U.S. Census or through direct questionnaires administered in clinical trials. However, the innovation of EHRs has provided us with patient-derived data to help us understand the social and lifestyle factors of patients. SDOH data coupled with questionnaires and clinical data in EHRs could be used to enable precision medical studies on healthcare access and health outcomes, by linking with data about treatments, disease conditions, drug response, insurance status, and demographics.

Although the classification of SDOH at the individual or patient level is becoming increasingly standardized for operational and clinical research purposes, a current challenge

in the wide adoption of SDOH in RWE studies is that of missing data, HIPAA regulations, and quality control issues that severely limit the amount of data available to answer clinical questions with high precision [15–18]. Therefore, the roles that the SDOH play in various chronic illnesses and diseases are ill-defined but have the potential to address population- and person-specific questions in the future. Research shows that public health goals cannot be realized without addressing the underlying SDOH that contribute to disparities and outcomes [19,20]. Therefore, healthcare research should strive to include SDOH in addition to race/ethnicity in RWE studies. A plethora of research reveals numerous socioeconomic parameters potentially accelerating disease risk, especially among minorities [20]. We must improve our understanding of the impact of SDOH on disease risk by investigating the different roles that SDOH play for patients, population groups, healthcare providers, healthcare access, and health outcomes (Figure 1).
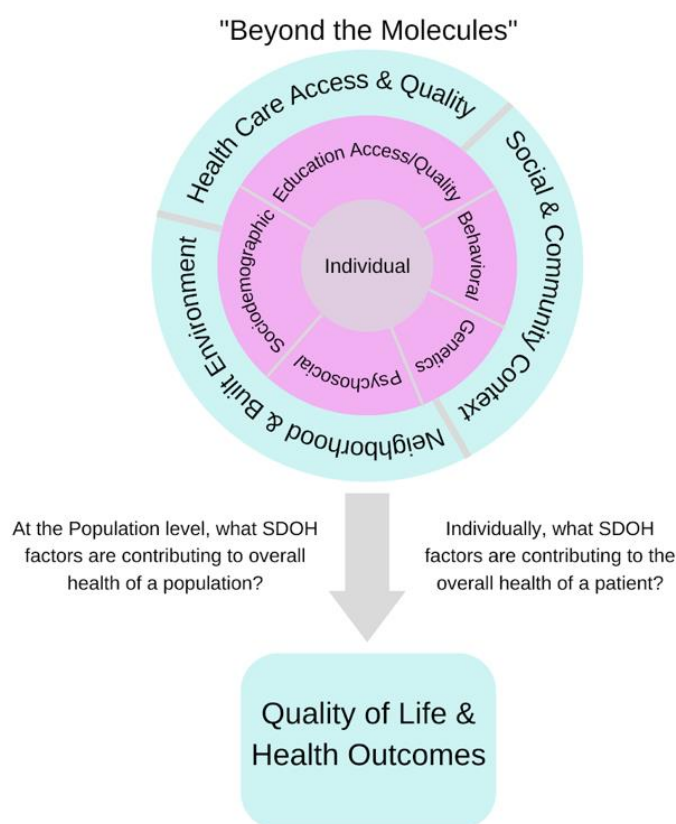

Figure 1. Grouped SDOH factors are categorized at the population and individual levels. At each level, we define the SDOH factors that contribute to the overall health of a population and the overall health of a patient, which mark the difference between a good outcome and a poor outcome.

3. SDOH Integration into Electronic Health Records

In EHR databases across the country, there is incompleteness of SDOH data, which has led previous RWE studies to use only race/ethnicity, sex, and age as measures of SDOH. In theory, those demographics can provide context, but cannot capture the full picture of one's overall health. In an attempt to capture SDOH effectively in EHRs, efforts to map de-identified patients' information to census tracts have been extremely important in providing researchers the ability to use evidence-based SDOH to answer clinical questions. However, the SDOH are often only captured in clinical notes, without structured coding, and we need better methods to obtain SDOH data trapped in notes. Currently, social aspects found in clinical notes vary across EHR databases in the country; however, the Institute of Medicine (IOM) has worked aggressively on identifying SDOH domains to be suggested for use in EHRs for academic research purposes [21]. The type of data suggested

includes sociodemographic domains, psychological domains, behavioral domains, and individual-level social relationship and living condition domains. Our field is still in the early days of extracting specific SDOH information and mapping such data to and from EHRs, but alongside social indices, other structured data elements, such as insurance status, can now be used to understand a patient's socioeconomic status. The indices can be used to answer questions related to SDOH and in cooperation with EHRs to understand surgical outcomes, drug distribution, health outcomes, and hospital readmissions. As time progresses, we will effectively utilize more SDOH data in EHRs and RWE studies.

EHR toolkits offer precise categorization of SDOH captured from census tracts for use in RWE studies. SDOH-standardized vocabularies are offered through these toolkits to map data to census tracts in EHR databases, but these are not yet widely adopted. A popular toolkit called the PhenX toolkit offers ontologies such as health insurance coverage, food insecurity, air quality index, wealth, job insecurity, food swamp, and more to increase the SDOH measures used in studies [22]. The SDOH ontologies were precisely chosen to enable highly qualitative measures that will increase the statistical power of studies [22]. The International Classification of Diseases (ICD) coding system has incorporated specific ICD-Z codes for SDOH that refer to problems related to education and literacy, housing, economic circumstances, social environments, upbringing, primary support groups, psychosocial circumstances, and occupational exposures to risk factors (Table 1). Although important, the Z codes are listed as non-diagnosis codes rather than disease-specific codes. Yet, they all play a major part in understanding the causes of health disparities in communities. At this point, there are several SDOH-standardized vocabularies and ontologies to use, but challenges occur with the utilization of these codes and capturing these codes and related data from patients in an effective way to use in research. Until EHR databases increase their SDOH data availability for SDOH-specific studies, navigating this field will remain difficult.

Table 1. International Classification of Disease (ICD) Z codes for SDOH [23].

ICD-10 CM Code	SDoH Categories
Z55	Problems related to education and literacy
Z56	Problems related to employment and unemployment
Z57	Occupational exposure to risk factors
Z58	Problems related to physical environment
Z59	Problems related to housing and economic circumstances
Z60	Problems related to social environment
Z62	Problems related to upbringing
Z63	Other problems related to primary support group, including family circumstances
Z64	Problems related to certain psychosocial circumstances
Z65	Problems related to other psychosocial circumstances

When SDOH are made available for research, one can start to model and predict occurrences across a wide range of diseases, relating SDOH elements to numerical measurement tests, diagnostics, and other health outcomes. It will be important to choose diseases that are easy to define in terms of severity, prevalence, and incidence of the condition, as well as diseases that have been known to be prominent in lower socioeconomic areas. With SDOH mapped to and from EHRs, we can understand what causes differences in the prevalence and incidence of a disease between different population groups. To do this, we can use descriptive biostatistical methods such as ordinal logistic regression and multivariate logistic regression that describe the relationship between categorical variables

predominantly found in EHR datasets, alongside other effective analysis techniques for EHRs as previously described [24].

4. The Future of SDOH in Real-World Evidence Studies

The next step in biomedical informatics is to develop a streamlined method for mapping census tract information to EHRs that can be implemented by each institution. In structured data, there needs to be a unified goal to create more diverse databases by providing healthcare access to underserved populations, thus ensuring that SDOH studies are well-balanced and produce answers backed by realistic outcomes. In the future, we could increase the versatility of SDOH for medical studies by developing a universal "medical grade social index" that could contain specific SDOH information that is most directly related to clinical outcomes, to best answer research questions. The power of this data will be to create knowledge that fills current gaps in healthcare, such as the need to incorporate brick-and-mortar needs-based healthcare services into deprived communities.

The future looks bright for even newer sources of data on SDOH. Such patient-derived data have allowed us to obtain information on SDOH beyond that in EHRs thanks to people using health measures including wearable fitness monitors, smartphones, step trackers, food trackers, and telemedicine. Tracking information from patients allows us to understand patients' daily activities that can lead to chronic illnesses and diseases, and thus understand how to potentially eliminate them. Currently, however, patient-derived data offer limited information across larger populations. Luckily, the adoption of mHealth has increased SDOH measures by utilizing mobile technology to improve health goals in communities. mHealth is a medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants, and other wireless devices, as defined by the World Health Organization [25]. Whilst mHealth is not widely adopted yet, patient-derived data can be used in the future to estimate a patient's health or risk of clinical diseases and provide better patient-centered health approaches to improve overall individual health. How mHealth and digital devices will feed into SDOH is still an active area of research. Other newer sources of SDOH-related data will also include genomic information for large populations uploaded to EHR databases, which is already being implemented by various institutions around the country [26].

Thus far, questionnaires, population surveys, and patient-reported outcomes are heavily used in SDOH and epidemiological research studies because they allow researchers to answer clinical research questions when structured data elements of the EHR database system themselves do not give a full picture of SDOH data. Due to inconsistencies in SDOH data availability, the field has been using questionnaire-based data to answer in-depth questions related to SDOH, along with EHR data, to understand surgical outcomes, drug distribution, health outcomes, and hospital readmissions [27–30].

5. Conclusions

The communication of RWE studies, especially related to disparities in health, has the potential to create political conflict, medical mistrust, and harm to marginalized communities. Therefore, it is important to effectively communicate results in a manner that is understood by both clinicians and patients from all paths of life so that the knowledge of health outcomes is more accessible. More importantly, including a diverse set of researchers to harness the power of this data is important for leveraging different mindsets to understand SDOH clinical questions in a different context. Since we know that a large portion of health disparities occur in communities under-represented in medicine, the best way to relate to these communities is by having people who understand the vernacular, culture, and the patience needed to communicate risk, health literacy, and health accessibility to the most affected communities.

All of this is an acknowledgment and perhaps a frustration that studying SDOH is likely to be harder than studying genes, molecules, and cells. However, the more we finetune the idea and execution of SDOH in EHR databases across the country, the more we

will increase the opportunities to use precision medicine to target clinical diseases. Precision medicine can be used to accurately prescribe patients based on RWE of health outcomes and treatment patterns respective to different population groups. More importantly, it will give clinicians the ability to prescribe patients using patient-centered approaches derived from research. Nevertheless, it is critical to remember that precision medicine means more than just genes, molecules, and cells. The future implementation of SDOH will provide greater precision of treatments based on an array of demographics, lifestyle factors, and environmental factors, all of which are likely to make a greater difference for our patients than any given measured cell or base-pair in their genome.

References

1. Baptista, P.V. Principles in genetic risk assessment. *Ther. Clin. Risk Manag.* **2005**, *1*, 15–20. [CrossRef]
2. Hall, W.D.; Morley, K.I.; Lucke, J.C. The prediction of disease risk in genomic medicine. *EMBO Rep.* **2004**, *5*, S22–S26. [CrossRef] [PubMed]
3. Fraiman, Y.S.; Wojcik, M.H. The influence of social determinants of health on the genetic diagnostic odyssey: Who remains undiagnosed, why, and to what effect? *Pediatr. Res.* **2020**, *89*, 295–300. [CrossRef] [PubMed]
4. Breen, N.; Berrigan, D.; Jackson, J.S.; Wong, D.W.; Wood, F.B.; Denny, J.C.; Zhang, X.; Bourne, P.E. Translational Health Disparities Research in a Data-Rich World. *Health Equity* **2019**, *3*, 588–600. [CrossRef] [PubMed]
5. Healthy People 2030, U.S. Department of Health and Human Services, Office of Disease Prevention and Health Promotion. 2018. Available online: https://health.gov/healthypeople/objectives-and-data/social-determinants-health (accessed on 11 March 2022).
6. Haire-Joshu, D.; Hill-Briggs, F. The Next Generation of Diabetes Translation: A Path to Health Equity. *Annu. Rev. Public Health* **2019**, *40*, 391–410. [CrossRef] [PubMed]
7. Kind, A.J.; Buckingham, W.R. Making Neighborhood-Disadvantage Metrics Accessible—The Neighborhood Atlas. *N. Engl. J. Med.* **2018**, *378*, 2456–2458. [CrossRef]
8. U.S. Census Bureau. 1994. Census Tracts and Block Numbering Areas. Available online: https://www2.census.gov/geo/pdfs/reference/GARM/Ch10GARM.pdf (accessed on 11 November 2021).

9.   Kolak, M.; Bhatt, J.; Park, Y.H.; Padrón, N.A.; Molefe, A. Quantification of Neighborhood-Level Social Determinants of Health in the Continental United States. *JAMA Netw. Open* **2020**, *3*, e1919928. [CrossRef]

10.   Singh, G.K. Area Deprivation and Widening Inequalities in US Mortality, 1969–1998. *Am. J. Public Health* **2003**, *93*, 1137–1143. [CrossRef]

11.   Flanagan, B.E.; Gregory, E.W.; Hallisey, E.J.; Heitgerd, J.L.; Lewis, B. A Social Vulnerability Index for Disaster Management. *J. Homel. Secur. Emerg. Manag.* **2020**, 4–6. [CrossRef]

12.   Nagano, I.; Wu, Z.; Asano, K.; Takahashi, Y. Census Tract Level State Maps of the Modified Retail Food Environment Index (mRFEI). *Natl. Cent. Chronic Dis. Prev. Health Promot. Div.* **2011**, *178*, 134–142.

13.   Knighton, A.J.; Savitz, L.; Belnap, T.; Stephenson, B.; VanDerslice, J. Introduction of an Area Deprivation Index Measuring Patient Socio-economic Status in an Integrated Health System: Implications for Population Health. *EGEMs (Gener. Evid. Methods Improv. Patient Outcomes)* **2016**, *4*, 9. [CrossRef] [PubMed]

14.   Nelson, C.A.; Butte, A.J.; Baranzini, S.E. Integrating biomedical research and electronic health records to create knowledge-based biologically meaningful machine-readable embeddings. *Nat. Commun.* **2019**, *10*, 1–7. [CrossRef] [PubMed]

15.   The Office for Civil Rights (OCR); Malin, B. Guidance Regarding Methods for de-identification of protected health information in accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule. *Health Inf. Priv.* **2012**, 1–32. Available online: https://privacysecurityacademy.com/wp-content/uploads/2021/03/HHS-OCR-Guidance-on-De-Identification-of-PHI-2012.pdf (accessed on 25 August 2022).

16.   Shortreed, S.M.; Cook, A.J.; Coley, R.Y.; Bobb, J.F.; Nelson, J.C. Challenges and opportunities for using big health care data to advance medical science and public health. *Am. J. Epidemiol.* **2019**, *188*, 851–861. [CrossRef] [PubMed]

17.   Rudrapatna, V.A.; Butte, A.J. Opportunities and challenges in using real-world data for health care. *J. Clin. Investig.* **2020**, *130*, 565–574. [CrossRef] [PubMed]

18.   Cantor, M.N.; Thorpe, L. Integrating data on social determinants of health into electronic health records. *Health Aff.* **2018**, *37*, 585–590. [CrossRef] [PubMed]

19.   Marmot, M. Social determinants of health inequalities. *Lancet* **2005**, *365*, 1099–1104. [CrossRef]

20.   Thornton, P.L.; Kumanyika, S.K.; Gregg, E.W.; Araneta, M.R.; Baskin, M.L.; Chin, M.H.; Mangione, C.M. New research directions on disparities in obesity and type 2 diabetes. *Ann. N. Y. Acad. Sci.* **2020**, *1461*, 5–24. [CrossRef]

21.   IOM (Institute of Medicine). *Capturing Social and Behavioral Domains and Measures in Electronic Health Records: Phase 2*; National Academies Press: Washington, DC, USA, 2015; pp. 1–351. [CrossRef]

22.   Hamilton, C.M.; Strader, L.C.; Pratt, J.G.; Maiese, D.; Hendershot, T.; Kwok, R.K.; Hammond, J.A.; Huggins, W.; Jackman, D.; Pan, H.; et al. The PhenX Toolkit: Get the Most From Your Measures. *Am. J. Epidemiol.* **2011**, *174*, 253–260. [CrossRef]

23.   Brämer, G.R. *International Statistical Classification of Diseases and Related Health Problems: Tenth Revision*, 2nd ed.; World Health Organization: Geneva, Switzerland, 2004.

24.   MIT Critical Data. Secondary Analysis of Electronic Health Records. In *Secondary Analysis of Electronic Health Records*; Springer International Publishing: Berlin/Heidelberg, Germany, 2016; pp. 1–427. [CrossRef]

25.   World Health Organization. *mHealth: New Horizons for Health through Mobile Technologies: Second Global Survey on Ehealth*; World Health Organization: Geneva, Switzerland, 2011.

26.   Akbari, P.; Sosina, O.A.; Bovijn, J.; Landheer, K.; Nielsen, J.B.; Kim, M.; Aykul, S.; De, T.; Haas, M.E.; Hindy, G.; et al. Multiancestry exome sequencing reveals INHBE mutations associated with favorable fat distribution and protection from diabetes. *Nat. Commun.* **2022**, *13*, 4844. [CrossRef]

27.   Hughes, M.M.; Wang, A.; Grossman, M.K.; Pun, E.; Whiteman, A.; Deng, L.; Hallisey, E.; Sharpe, J.D.; Ussery, E.N.; Stokley, S.; et al. County-Level COVID-19 Vaccination Coverage and Social Vulnerability—United States, 14 December 2020–1 March 2021. *MMWR. Morb. Mortal. Wkly. Rep.* **2021**, *70*, 431–436. [CrossRef] [PubMed]

28.   Azap, R.A.; Paredes, A.Z.; Diaz, A.; Hyer, J.M.; Pawlik, T.M. The association of neighborhood social vulnerability with surgical textbook outcomes among patients undergoing hepatopancreatic surgery. *Surgery* **2020**, *168*, 868–875. [CrossRef] [PubMed]

29.   Ghirimoldi, F.M.; Schmidt, S.; Simon, R.C.; Wang, C.-P.; Wang, Z.; Brimhall, B.B.; Damien, P.; Moffett, E.E.; Manuel, L.S.; Sarwar, Z.U.; et al. Association of Socioeconomic Area Deprivation Index with Hospital Readmissions After Colon and Rectal Surgery. *J. Gastrointest. Surg.* **2020**, *25*, 795–808. [CrossRef] [PubMed]

30.   Rosenzweig, M.Q.; Althouse, A.D.; Sabik, L.; Arnold, R.; Chu, E.; Smith, T.J.; Smith, K.; White, D.; Schenker, Y. The Association Between Area Deprivation Index and Patient-Reported Outcomes in Patients with Advanced Cancer. *Health Equity* **2021**, *5*, 8–16. [CrossRef]