

Integration of Social Determinants of Health Data into the Largest, Not-for-Profit Health System in South Florida

LOURDES M. ROJAS^{1,*}, GREGORY L. VINCENT¹, AND DON PARRIS¹

¹*Baptist Health South Florida – Center for Advanced Analytics, USA*

Abstract

Social determinants of health (SDOH) are the conditions in which people are born, grow, work, and live. Although evidence suggests that SDOH influence a range of health outcomes, health systems lack the infrastructure to access and act upon this information. The purpose of this manuscript is to explain the methodology that a health system used to: 1) identify and integrate publicly available SDOH data into the health systems' Data Warehouse, 2) integrate a HIPAA compliant geocoding software (via DeGAUSS), and 3) visualize data to inform SDOH projects (via Tableau). First, authors engaged key stakeholders across the health system to convey the implications of SDOH data for our patient population and identify variables of interest. As a result, fourteen publicly available data sets, accounting for >30,800 variables representing national, state, county, and census tract information over 2016–2019, were cleaned and integrated into our Data Warehouse. To pilot the data visualization, we created county and census tract level maps for our service areas and plotted common SDOH metrics (e.g., income, education, insurance status, etc.). This practical, methodological integration of SDOH data at a large health system demonstrated feasibility. Ultimately, we will repeat this process system wide to further understand the risk burden in our patient population and improve our prediction models – allowing us to become better partners with our community.

Keywords *data visualization; health system; social determinants of health*

1 Introduction

The United States, compared to other industrialized nations, struggles with high healthcare expenditures and suboptimal health outcomes (Solomon and Kanter, 2018). Researchers have posited that this is due to a misallocation of resources to sick care, rather than to an upstream approach to changing the social and behavioral causes of the disease (Marvasti and Stafford, 2012). Only 20% of health outcomes can be attributed to health care, whereas 80% of health outcomes can be attributed to health behaviors and social determinants of health (SDOH) (Hood et al., 2016). SDOH are “the conditions in which people are born, grow, live, work and age. These circumstances are shaped by the distribution of money, power and resources at global, national and local levels (WHO (World Health Organization), 2012).” The United States Office of Health Promotion and Disease Prevention has organized SDOH into five domains: economic stability, education, social and community context, health and healthcare, and neighborhood and built environment (Secretary’s Advisory Committee, 2010). Vast evidence supports the impact of social determinants of health on a variety of disease outcomes that plague our nation, including

* Email: LourdesRoja@baptisthealth.net.

cardiovascular disease, cancer, diabetes, HIV, and substance abuse disorders (Salina and Kones, 2018; Marmot, 2018; Walker et al., 2016; Geter et al., 2018; Dasgupta et al., 2018). Further, a deficiency in one SDOH domain propagates disparities in other domains. For example, if someone has low income and low education, they are less likely to have health insurance and less likely to access health care (Tolbert et al., 2018). This begs the question: Why are the majority of our health systems still *only focusing on the 20%*?

Fortunately, the understanding of how SDOH impact patient outcomes has been gaining momentum in health care. For example, the Aetna Foundation and the National Quality Forum collaborated with 60 healthcare stakeholders across the U.S. to discuss recommendations for how to address SDOH through “payment and quality innovation (Jent, 2019).” As another example, Kaiser Permanente deployed an SDOH initiative, which sought to address built environment factors affecting housing, obesity, and other health outcomes. Major organizations, such as the Center for Medicare and Medicaid Innovation, American Hospital Association, and Federally Qualified Health Centers, have been engaging in activities to identify and address SDOH needs (Alley et al., 2016; Jent, 2019; Centers for Medicare & Medicaid Services (CMS), 2017). It is necessary to invest in health outside of health care: it is fruitful for patients and for the healthcare system alike (Bradley et al., 2016). However, there are barriers to the measurement of SDOH in healthcare systems. Mainly, electronic medical records do not have SDOH related forms, do not prompt providers with SDOH questions, and most providers are not trained in how to address these questions (Gottlieb et al., 2016; Clark and Gurewich, 2017). One solution to these barriers is to utilize publicly available, census-tract level data sources geocoded to patient addresses as a proxy for SDOH characteristics (Gottlieb et al., 2016; Bazemore et al., 2015; Rivera and Hoffman, 2018; Hatem et al., 2019). Indeed, census-tract level variables that capture SDOH domains can paint an accurate picture of the patient’s living and working situation (Daniel et al., 2018).

The purpose of this manuscript is to describe the steps a healthcare team took to identify, integrate, and visualize SDOH data into a large, not-for profit health system in South Florida. By undertaking this pilot, we will provide a platform for future projects within the health system, with the ultimate goal to more adequately measure, meet, and predict our patients’ social determinants of health needs. This manuscript is written as a practical guide for other healthcare teams interested in including SDOH data from publicly available sources into their systems.

2 Methods

This project consisted of three general phases: identification, integration, and visualization of SDOH data. Our healthcare team was comprised of a Healthcare Data Scientist (project lead, 0.25 Full-Time Equivalent, “FTE”), Business Project Analyst (project coordinator, 0.15 FTE), Assistant Vice President (consultative role, 0.05 FTE), Data Manager (consultative role, 0.10 FTE), Senior Data Warehouse Specialist (Information Technology partner, 0.15 FTE), and an Analytic Intern (project support, 0.15 FTE) who was in a local graduate program. These FTE calculations are for the phases described below. However, an additional phase of “maintenance” would need to be considered for calculating percent effort. For maintenance, which would be comprised of bringing in new data as they are available on a yearly basis, we estimate 0.10 FTE for a project lead and 0.05 FTE for an Information Technology support.

In the identification phase, we conducted a thorough literature review to identify the most prevalent SDOH and their impact on population health. We then met with key clinical and executive stakeholders across the system to socialize the implications of SDOH data for our patient

population and identify variables of interest for specific stakeholders. In the integration phase, we downloaded data from the data sources' websites, cleaned the data, created data dictionaries, and worked closely with Information Technology to upload the data sources into the Data Warehouse. The data dictionaries consisted of two versions: a user-friendly version for healthcare system clients (Table 1) and a more detailed version for data managers (Table 2). Through github.com, we downloaded a HIPAA-compliant geocoding software, DeGAUSS ("Decentralized Geomarker Assessment for Multi-Site Studies", Brokamp, 2018), to be able to geocode patient and entity addresses. In the visualization phase, we focused on creating maps of the counties and census tracts within our health system on Tableau. The health system already had an active license with Tableau.

3 Results

3.1 Identification Phase

The key words of ("social determinants of health" OR "social determinants") AND ("healthcare system" OR "health care") were used to search PubMed and Google Scholar, with the filters of English language, peer-reviewed journals, and time range of 2014 to 2019. Within the search results, we focused in on articles that outlined processes for integration of SDOH into health care systems and application of SDOH data in healthcare systems. These articles pointed us to data sources, such as the American Community Survey from the U.S. Census Bureau, which we should be integrating.

From this review, we prepared a slide show for our clinical and executive stakeholder audience. The slide show covered: what are social determinants of health, a national focus on SDOH integration into health care, potential data sources and variables, potential applications of the data, and phases of our project. Potential applications included addition of SDOH variables into case management dashboards, inclusion of SDOH data into risk stratification algorithms, and mapping of SDOH resources to match patient needs within service areas of our hospital entities. The slide show ended with an open discussion with the stakeholders so our team could gather their variables of interest and include in the data pull. Stakeholders included executives and clinicians from different departments and entities across the system: Community Health Needs Assessment Committee, Pastoral Care Services, Addiction Treatment, Population Health Management, Marketing, Diversity and Inclusion Committee, and leaders from two of our hospital entities.

3.2 Integration Phase

Our team downloaded and cleaned over fourteen publicly available data sources, accounting for over 30,800 variables at the county and/or census tract level. To "clean" the data, we documented missing values the same throughout (i.e., ".") to provide uniformity in our visualization/analysis phase. We also documented all county or census level identifiers by the Federal Information Processing System (FIPS) code to be able to link all sources to the geocoding software. Variables were extracted from the following data sets: American Community Survey, Annual Homelessness Assessment Report, Florida Environmental Public Health Tracking, United States Small-Area Life Expectancy Estimates Project, Center for Disease Control and Prevention Wide-Ranging Online Data for Epidemiologic Research, Mapping Medicare Disparities Tool, Fatality Analysis Reporting System, Area Health Resource File from American Medical Association, National

Table 1: Sample of Data Dictionary File for Client. Each column on the Excel can be filtered as needed.

Column Name	Column Description
Table	Simplified name of data table, e.g., “Poverty Status in the past 12 months”
Social Determinants of Health Domain	Drop down menu: demographics, economic stability, education, health and healthcare, neighborhood and built environment, social and community context
U.S. Census Bureau American Community Survey, 2016, 1 year estimates: County Level	If the data can be found in a source, there would be an “x” in the column. For example, “poverty status in the past 12 months” would be “x” out for all US Census Bureau American Community Survey columns. This allows the client to see their options of where they can pull the data from.
U.S. Census Bureau American Community Survey, 2017, 1 year estimates: County Level	x
U.S. Census Bureau American Community Survey, 2018, 1 year estimates: County Level	x
U.S. Census Bureau American Community Survey, 2018, 5 year estimates: County Level, Census Tract	x
Annual Homeless Assessment Report (AHAR), 2016: County Level	x
Annual Homeless Assessment Report (AHAR), 2017: County Level	x
Annual Homeless Assessment Report (AHAR), 2018: County Level	x
Annual Homeless Assessment Report (AHAR), 2019: County Level	x
County Health Rankings Florida, 2019*: County Level	x
Florida Environmental Public Health Tracking, 2016, 1 year estimates: County Level	x
USALEEP, 2010-2015: Census Tract	x
CDC Wonder, 2011-2017: County Level	x
health system NPI, 2019: Health system locations and entities	x
Sample of Data Table: PDF or website link	Link to PDF or website that displays the data from its original source, e.g., Data\American Community Survey\Economic Stability\S1701_Poverty Status in the Past 12 Months.pdf

Table 2: Sample of Data Dictionary for Information Technology Team. Each column on the Excel can be filtered as needed. Supplement 2 contains the full data dictionary with links to download the publicly-available data.

Column Name	Column Description
Social Determinants of Health Domain	Drop down menu: demographics, economic stability, education, health and healthcare, neighborhood and built environment, social and community context
Year of Data	Options ranging from 2014–2019
Level of Data (State, County, and/or Census-Tract)	Drop down menu: state, county, and/or census-tract
Data Source	Name of data sources, e.g., “American Community Survey”
Link to Download Data	From the website, e.g., https://data.census.gov/cedsci/table?q=S1201&table=S1201&tid=ACSST1Y2018.S1201&lastDisplayedRow=27&hidePreview=true&g=0400000US12,12.050000
Table Name in Zip Folder	How data was downloaded, e.g., “ACSST1Y2018.S1201_data_with_overlays_2019-11-07T110629.csv”
Table Description	How data was described by data source, e.g., “S1201, Marital Status, 2018, 1 year estimate”
Field Name	Variable name, e.g., “S1201_C01_001E”
Field Description	Variable description, e.g., “Estimate!!Total!!Population 15 years and over”
How often data is released	Frequency (and month of year if applicable), e.g., “annually, October”
Date of data download	Date that health system downloaded the data into the data warehouse, e.g., “11/7/2019”

Center for Health Statistics, CDC Diabetes Interactive Atlas, USDA Map the Meal Gap, National Center for HIV/AIDS Viral Hepatitis STD and TB Prevention STI Atlas, National Center for Education Statistics, and Florida Behavioral Risk Factor Surveillance System.

We downloaded the most up to date versions, which in some cases was 2016, and in others 2019. For data sets that become available every year, such as the American Community Survey (ACS), we downloaded data from 2016 forward to align with the implementation of medical records across our health system. To ensure the use of relevant data, publicly available data sources will be monitored and updated within our system as new data is released to the public. The data dictionary outlines when new data sources are released. We also created a step-by-step guide for how to download data from these sources (one example is given in Supplementary

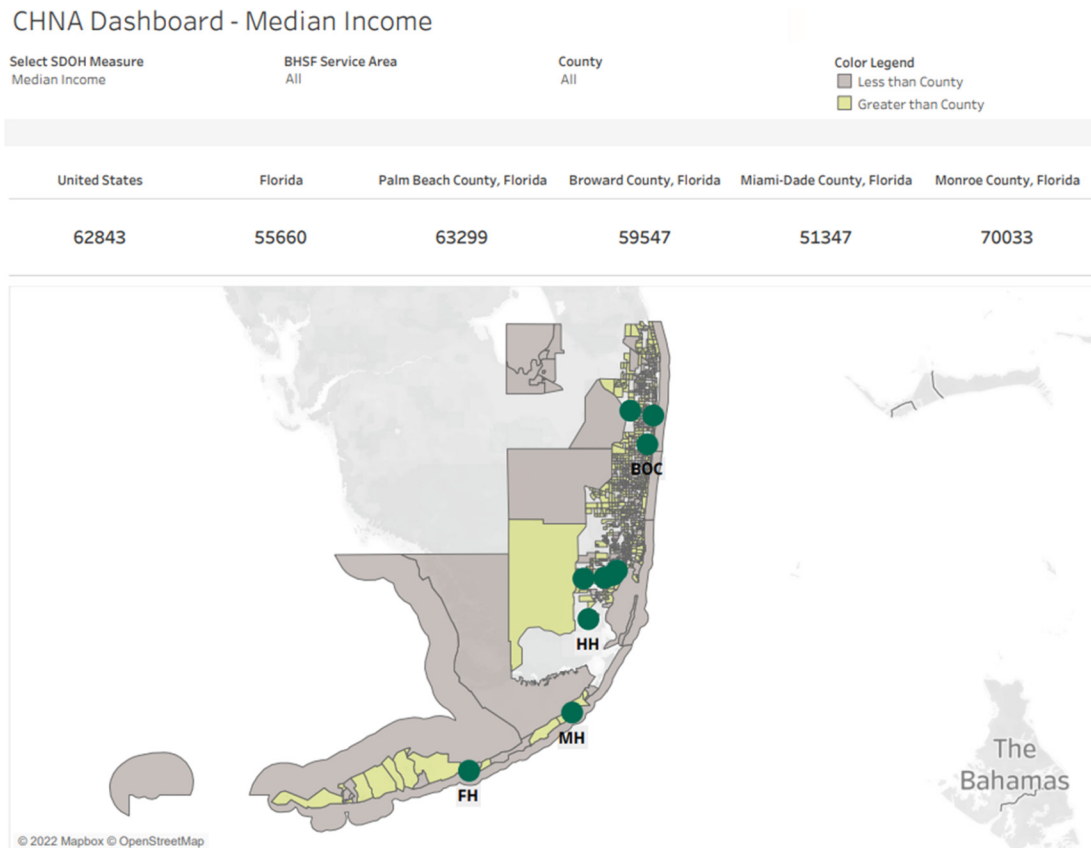


Figure 1: Screenshot of Tableau Dashboard. Filters can be selected for social determinant of health (SDOH) measure of interest, service area of interest, and county of interest.

Material). Once a year, the team meets with information technology team to assure the data has been uploaded and to go over any questions the information technology team may have.

To test the DeGAUSS software, a random subsample of patient addresses were geocoded. DeGAUSS has criteria set forth for usability of geocodes; the developers state that the assigned FIPS code must obtain a precision of “range” or “street” rather than “zip” or “city”, and must have a precision score of 0.5 or above. 98.5% of addresses obtained usability criteria. We also geocoded all health system hospital entities.

3.3 Visualization Phase

A pilot Tableau dashboard was created to support the community health’s department community health needs assessment report. This report was published on our health system’s website. In addition, the dashboard was used to guide community focus groups, where the facilitator would display the finding and ask the community about their thoughts. For this pilot, we utilized data points from the American Community survey: we analyzed the most up to date available 5-year estimate rather than the 1-year estimate because they have larger samples, smaller margins of error, and go down to the census-tract level (U.S. Census Bureau, 2018). The variables were median income, percent unemployment, percent uninsured, percent with bachelor’s degree or higher, and percent with high school diploma. The dashboard presented the estimates at the

national, state, and county level. Filters were in place to account for health system's service areas and counties. Figure 1 provides an example of one of the SDOH metrics.

4 Discussion

In the US, there is growing recognition that social, economic, and behavioral factors, commonly referred to as the social determinants of health (SDOH), significantly influence individual and population health outcomes (Marmot, 2005). Additionally, evidence suggests that these factors highly influence an individual's quality of care as well as their ability to attain and maintain their full health potential, leading them to consume more intensive ambulatory health care services (Katz et al., 2018). These health inequalities are striking between advantaged and marginalized groups in society and require partnerships with key stakeholders to adequately address the health needs of their community. Although evidence suggests the importance of systematically addressing such inequalities, health systems lack the capacity to integrate this data into their infrastructure and readily act upon such information. This manuscript outlined how one healthcare system approached SDOH data implementation through identification, integration, and visualization of linking publicly available SDOH data to patient service areas.

The novel approach outlined in this manuscript represents a simple and actionable integration of SDOH data into a large not-for-profit healthcare system. Although previous studies have successfully linked or integrated social and medical data into their EHR, these studies lack a sustainable approach to their integration. Studies in a systematic review of SDOH in EHRs did not mention the importance of establishing and disseminating a comprehensive data dictionary, cleaning and updating the data sets, and documenting hands-on feedback from stakeholders and hospital personnel (Chen et al., 2020; Heidari et al., 2022). On the contrary, this manuscript outlines a detailed summary of the identification and integration process. The systematic review also revealed that most studies focus on the American Community Survey as their neighborhood level measure, whereas we brought fourteen databases to match the needs of our stakeholders (Chen et al., 2020). Further, while frameworks exist for SDOH care coordination, this manuscript serves as a guide for the front-end data work that needs to be the foundation for future care coordination (Gibbins and Wickramasinghe, 2021).

While this manuscript outlines the process and a small pilot, there are many implications for SDOH data use within our healthcare system. Firstly, the socialization of the importance of SDOH to key stakeholders is critical to understanding the drivers and determinants of a patient's health at a community level. These needs are important in creating community level partnerships, identifying gaps in services and the allocation of resources. For example, based on a risk-adjustment model that incorporated community level variables from the American Community Survey, one health system was better able match their patients with health plans and allocate finances towards those health plans (Clark and Gurewich, 2017). As another example, based on an SDOH community assessment, an academic-practice-community partnership was formed for underinsured heart failure patients, in which these patients were referred to a specialty clinic rather than coming to the emergency department (Shirey et al., 2022). Secondly, we can define opportunities to improve care efficiency for patients by understanding their needs and the surrounding communities that we serve. For example, researchers have suggested that SDOH metrics available on a patient record could "identify all patients on a provider's panel who live in an area with a high proportion of fast food restaurants and send them information about where they can purchase fresh produce, along with recipes for quick, healthy meals" (Bazemore et al.,

2015). Thirdly, the insight gathered by such community-level as well as individual-level data may be beneficial at a system level by enhancing the performance of our predictive models. A recent systematic review that examined the use of SDOH in machine learning cardiovascular risk prediction models showed increased performance in the models that included social determinants (Zhao et al., 2021). In turn, risk stratification could lead to improved clinical decision support.

4.1 Limitations

While we had a dedicated team of scientists and analysts to complete this project, other health systems may not have the staff to do so. There are several other limitations to consider when a health system is interested in replicating this process. First, although data from the ACS is provided every year, the data are typically not up-to-date. For example, the ACS' data for 2021 will not be available until 2022. Additionally, the size of the community you are examining may have an impact on the data available. The ACS one year estimates represent data that have been collected over a year for geographical areas with a population of at least 65,000. For geographic areas with fewer than 65,000 residents, the Census Bureau combines five consecutive years of ACS data to produce multiyear estimates that represent data collected over a period of five years.

Second, while census-tract is an acceptable proxy for patient information, it will never be as accurate as individual-level data on the patient. Therefore, inferences made about an individual from aggregate or area-level data may suffer from the ecological fallacy. Statistical methods can be used to adjust area-level analyses to provide less biased estimates of individual-level parameters. Future studies should consider how we could validate census-tract information with individual-level patient data.

Lastly, collection of individual-level SDOH data on all patients may not even be feasible. For example, an emergency room patient may not be able to fill out a questionnaire due to their health status, or a clinician may not have the time to ask certain questions during a routine visit. Interestingly, researchers are exploring natural language processing as a method to extract individual level SDOH data (Patra et al., 2021). Further, most providers would not be able to act on the information even if collected as physicians and health systems oftentimes need to refer to community organizations or social service organizations to address SDOH issues. Moreover, patients do not always feel comfortable sharing this type of information, as they do not know how it will be used. Future studies should not only consider the patient perspective and privacy, but also work with the community to understand the implications of this data for standardized referral processes to be implemented.

Conflicts of Interest

No conflicts of interest to report.

Supplementary Material

Supplement 1 contains an example of how to download and clean the American Community Survey data.

Supplement 2 contains the full Data Dictionary for Information Technology, that was referenced in Table 2. With this Data Dictionary, the reader can find links to download the publicly-available data sets.

References

- Solomon S, Kanter M (2018). Health care steps up to social determinants of health: current context. *The Permanente Journal*.
- Marvasti FF, Stafford RS (2012). From “sick care” to health care: Reengineering prevention into the US system *The New England Journal of Medicine*, 367(10), 889.
- Hood CM, Gennuso KP, Swain GR, Catlin B (2016). County health rankings: relationships between determinant factors and health outcomes. *American Journal of Preventive Medicine*, 50(2): 129–135.
- WHO (World Health Organization) (2012). “What are the social determinants of health?”. http://www.who.int/social_determinants/sdh_definition/en/.
- Secretary’s Advisory Committee (2010). “Health Promotion and Disease Prevention Objectives for 2020. “Healthy People 2020: An Opportunity to Address the Societal Determinants of Health in the United States,” <http://www.healthypeople.gov/2010/hp2020/advisory/SocietalDeterminantsHealth.htm>. July 26.
- Salina AM, Kones R (2018). Cardiovascular risk management must always consider social determinants of health: Past lessons, data, proposals, and challenges. *Journal of Preventive Medicine*, 3(2): 14.
- Marmot M (2018). Health equity, cancer, and social determinants of health. *The Lancet Global Health*, 6(S29).
- Walker RJ, Williams JS, Egede LE (2016). Influence of race, ethnicity, and social determinants of health on diabetes outcomes. *The American Journal of the Medical Sciences*, 351(4): 366–373.
- Geter A, Sutton MY, Hubbard McCree D (2018). Social and structural determinants of HIV treatment and care among black women living with HIV infection: A systematic review: 2005–2016. *AIDS Care*, 30(4): 409–416.
- Dasgupta N, Beletsky L, Ciccarone D (2018). Opioid crisis: No easy fix to its social and economic determinants. *American Journal of Public Health*, 108(2): 182–186.
- Tolbert J, Orgera K, Damico A (2018). Key facts about the uninsured population. Kaiser Permanente. <https://www.kff.org/uninsured/fact-sheet/key-facts-about-the-uninsured-population/>. November 6.
- Jent K (2019). NQF to address social determinants of health with payment innovation, Patient Engagement Hit. <https://patientengagementhit.com/news/nqf-to-address-social-determinants-of-health-with-payment-innovation>.
- Alley DE, Asomugha CN, Conway PH, Sanghavi DM (2016). Accountable health communities—addressing social needs through medicare and medicaid. *The New England Journal of Medicine*, 374(1): 8–11.
- Centers for Medicare & Medicaid Services (CMS) (2017). Accountable Health Communities Model, <https://innovation.cms.gov/initiatives/ahcm/>. (Mayo 1)
- Bradley EH, Canavan M, Rogan E, Talbert-Slagle K, Ndumele C, Taylor L, et al. (2016). Variation in health outcomes: The role of spending on social services, public health, and health care, 2000–09. *Health Affairs*, 35(5): 760–768.
- Gottlieb L, Tobey R, Cantor J, Hessler D, Adler NE (2016). Integrating social and medical data to improve population health: Opportunities and barriers. *Health Affairs*, 35(11): 2116–2123.
- Clark MA, Gurewich D (2017). Integrating measures of social determinants of health into health care encounters: Opportunities and challenges. *Medical Care*, 55(9): 807–809.
- Bazemore AW, Cottrell EK, Gold R, Hughes LS, Phillips RL, Angier H, et al. (2015). “Community vital signs”: Incorporating geocoded social determinants into electronic records to

- promote patient and population health. *Journal of the American Medical Informatics Association*, 23(2): 407–412.
- Rivera B, Hoffman M (2018). Technical strategies for real-time geocoding in healthcare. In: *IEEE International Smart Cities Conference (ISC2)*. 1–5.
- Hatef E, Kharrazi H, Nelson K, Sylling P, Ma X, Lasser EC, et al. (2019). The association between neighborhood socioeconomic and housing characteristics with hospitalization: Results of a national study of veterans. *The Journal of the American Board of Family Medicine*, 32(6): 890–903.
- Daniel H, Bornstein SS, Kane GC (2018). Addressing social determinants to improve patient care and promote health equity: An American College of Physicians position paper. *Annals of Internal Medicine*, 168(8): 577–578.
- Brokamp C (2018). DeGAUSS: Decentralized Geomarker Assessment for Multi-Site Studies. *Journal of Open Source Software*.
- U.S. Census Bureau (2018). Understanding and using ACS Single Year and Multiyear Estimates. When to Use 1-year, 3-year, or 5-year Estimates. Understanding and Using American Community Survey Data. https://www.census.gov/content/dam/Census/library/publications/2018/acs/acs_general_handbook_2018_ch03.pdf.
- Marmot M (2005). Social determinants of health inequalities. *The Lancet*, 365(9464): 1099–1104.
- Katz A, Chateau D, Enns JE, Valdivia J, Taylor C, Walld R, et al. (2018). Association of the social determinants of health with quality of primary care. *Annals of Family Medicine*, 16(3): 217–224.
- 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.
- Heidari E, Brown C, Johnsrud M, Mastrangelo V, Spears A, Parikh M, et al. (2022). Procurement and use of social determinants of health data among key health care stakeholders. *Journal of Managed Care & Specialty Pharmacy*, 28(5): 538–543.
- Gibbins R, Wickramasinghe N (2021). Social determinants of health in the US: a framework to support superior care co-ordination and leverage digital health solutions. *Health Policy and Technology*, 10(2): 100523.
- Shirey MR, White-Williams C, Leading C (2022). Multisector partnerships to address social determinants of health and advance health equity for an underresourced heart failure patient population. *Nursing Administration Quarterly*, 46(3): 197–207.
- Zhao Y, Wood EP, Mirin N, Cook SH, Chunara R (2021). Social determinants in machine learning cardiovascular disease prediction models: A systematic review. *American Journal of Preventive Medicine*, 61(4): 596–605.
- Patra BG, Sharma MM, Vekaria V, Adekkanattu P, Patterson OV, Pathak J (2021). Extracting social determinants of health from electronic health records using natural language processing: A systematic review. *Journal of the American Medical Informatics Association*, 28(12): 2716–2727.