

Meaningful Information in the Age of Big Data: A Scoping Review on Social Determinants of Health Data Collection for Electronic Health Records

Kelsey Berg (✉ kelsey.berg@uleth.ca)

University of Lethbridge Faculty of Health Sciences <https://orcid.org/0000-0002-8871-5048>

Chelsea Doktorchik

University of Calgary

Hude Quan

University of Calgary

Vineet Saini

Alberta Health Services

Research article

Keywords: Social Determinants of Health, Electronic Health Records, Information Systems, Data Collection, Big Data, Population health, Health equity, Social constructivism, Health Management.

Posted Date: October 24th, 2019

DOI: <https://doi.org/10.21203/rs.2.16433/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background: Electronic Health Records (EHRs) are key tools for integrating patient data into health information systems (IS). Advances in automated data collection methodology, particularly the collection of social determinants of health (SDOH), provide opportunities to advance health promotion and illness prevention through advanced analytics (i.e. “Big Data” techniques). We ask how current data collection processes in EHRs permit SDOH data to flow throughout health systems.

Methods: Using a scoping review framework, we searched through medical literature to identify current practices in SDOH data collection within EHR systems. We extracted relevant information on data collection methodology, specifically focusing on uses of automated technology. We discuss our findings in the context of research methodology and potential for health equity.

Results: Practitioners collect a variety of SDOH data at point of care through EHR, predominantly via embedded screening tools and clinical notes, and primarily capturing data on financial security, housing status, and social support. Health systems are increasingly using digital technology in data collection, including natural language processing algorithms. However overall use of automated technology is limited to date. End uses of data pertain to improving system efficiency, patient care-coordination, and addressing health disparities.

Discussion & Conclusion: EHRs can realistically promote collection and meaningful use of SDOH data, although EHRs have not extensively been used to collect and manage this type of information. Future applied research on systems-level application of SDOH data is necessary, and should incorporate a range of stakeholders and interdisciplinary teams of researchers and practitioners in fields of health, computing, and social sciences.

Background

Social determinants of health refer to determinants that interact with health through social mechanisms. They are described as “the conditions in which people are born, grow, live, work and age”(1) They arise from the patterns of relationships which influence social, political and economic position, which in turn influence “life chances”(2). Distal social determinants include financial or social status, which impact health through proximal factors, such as access to nutritious food and health services, and other distal factors, such as migration patterns, political participation, etc.

Although the mechanisms connecting SDOH and health outcomes are complex, research from health sciences show increasing evidence that a significant burden of disease is associated with SDOH (3-9), with population data revealing predictable associations between specific SDOH factors and health status (10). For example, morbidity and mortality due to for a wide range of chronic and communicable health conditions are consistently and negatively, with few exceptions, associated with income level (11). As SDOH are highly predictive of health outcomes, population level SDOH data would enable health systems to address SDOH inequalities in order to better manage or prevent harmful health outcomes.

Key tools for capturing health information are Electronic Health Records (EHRs). These records store and manage patient health data, and can be connected to a broader information system (IS) in order to communicate a range of health information through diverse channels. As EHR systems become more widely used within health systems (12, 13), greater quantity of patient data will be available in digitized and analysis-ready format. Health information systems can currently incorporate SDOH measurements. For example, social conditions are quantified and qualified based on individual attributes, such as annual income, level of education, occupation status; as well as the attributes of a population to which a patient or research subject belongs, such as median income level, density of services, proportion of immigrants, etc. (4). While SDOH measurement are used in health research, and these data are relevant for health profiles, SDOH in EHRs have been limited to date (5, 12-14).

Nevertheless, the health care sector is increasingly incorporating data-driven approaches into operations. While new health data streams entails governance challenges, the potential for meaningful public health applications is vast (15). “Big Data” is considered to have considerable potential in core functions of Public Health, including surveillance, hypothesis-generating research, causal inference (16), and promoting health equity (17). While technological changes have contributed to medical research through increased data acquisition in fields such as genomics (18), health profiles remain incomplete without SDOH data. Advanced analytic techniques with large data sets have been shown to accurately identify and categorize SDOH, such as the structure of social networks (19) or risk of food insecurity (20). Unfortunately, and as mentioned above, clinical interactions do not in general include screening for SDOH, meaning that highly influential data pieces are not consistently considered in health analyses.

The reasons for inconsistent SDOH screening are multi-factorial (12, 21). There is wide variability in data collection and management standards (14), leading to EHRs often lacking well-designed documentation tools and mechanisms for managing SDOH data (13). Peer-reviewed knowledge on best practice in SDOH collection through EHRs is also still relatively nascent (7), and there is limited evidence that SDOH data can be used to effectively minimize health harms arising from social conditions (14, 22). As such, this review aims to clarify the current potential for Big Data in advancing health equity by focusing on SDOH data collection methodology and use of automated collection techniques in EHRs.

Methods

The following research question guided the scoping review: *How are SDOH factors are currently being integrated into Electronic Health Records?*

The search strategy incorporated the three step approach recommended by the Joanna Briggs Institute (23) including initial database searching, an iteratively-revised comprehensive database search after increased familiarity with search terms, and a final search of the reference lists of the included studies. We reviewed a comprehensive medical database (PubMed) using iterative search strings based on the

search terms found in table 1.1. We included the MeSH descriptor data to clarify the search concepts for the purpose of this review. Our final search was conducted in May, 2019, and included articles retrievable through PubMed up until that date.

Table 1.1: Search terms to identify articles on SDOH data collection in EHRs from PubMed database.

MeSH term	MeSH Scope Note
Social Determinants of Health	The circumstances in which people are born, grow up, live, work, and age, as well as the systems put in place to deal with illness. These circumstances are in turn shaped by a wider set of forces: economics, social policies, and politics (http://www.cdc.gov/socialdeterminants/).
Electronic Health Record	Media that facilitate transportability of pertinent information concerning patient's illness across varied providers and geographic locations. Some versions include direct linkages to online consumer health information that is relevant to the health conditions and treatments related to a specific patient.
Data collection	Systematic gathering of data for a particular purpose from various sources, including questionnaires, interviews, observation, existing records, and electronic devices. The process is usually preliminary to statistical analysis of the data.

Our search inclusion criteria were based on Population, Concept, and Context, as per Joanna Briggs Institute's scoping review methodology guidelines (23). We included research articles which targeted EHR stakeholders and addressed SDOH data collection in the context of information flow into health systems. Detailed criteria can be found in table 1.2.

Table 1.2: Inclusion and exclusion criteria for articles pertaining to SDOH data collection in EHRs.

Component	Inclusion	Exclusion
Population	- Stakeholders in EHR or clinical information system (CIS) data, including patients, health care practitioners, and individuals in health service delivery systems.	- Not EHR or CIS users. - Data warehouse subjects for research purposes only.
Concept	- Research on SDOH data collection, including SDOH variables such as demographic data, social and economic risk factors, cultural/political identity.	- Not research, such as perspective or guidelines. - Collection process does not include SDOH data, i.e, exclusively biological, ecological, psychological, or behavioural data.
Context	- Data collection processes within an EHR system, including automated or otherwise technologically enhanced processes.	- Did not use an EHR system, e.g. paper record only, or screening tool not connected to EHR system. - Not data collection, e.g. data management or analysis only.

We extracted data from the included articles on key aspects of data collection methodology, such as research design, operational definitions, collection tools, data type, and data use. We also extracted information on current or potential use of automated data collection tools in order to understand the current state of automated technology in SDOH data collection. We also included the EHR and EHR context (country of origin, details of CIS, type of health service organization, etc.).

Table 1.3 Data extraction categories from studies on SDOH data collection in EHRs.

Study population	Operational terms for SDOH data	SDOH Data categories	Electronic System	EHR context (name, health sector, etc.)	Data Collection Methods	Data collection rationale	Link to Automation	Outcomes of implementation	Comments

Results

Figure 1.1 PRISMA flow diagram for research articles on SDOH data collection in EHRs.

We identified 3 articles in PubMed using all three search concepts. We eliminated the third concept, data collection, from our search strategy (table 1.1), which then yielded 168 results. We retained the concept as part of the selection strategy during full-text review. Search results are presented in figure 1.1. A summary of extracted findings are presented in table 2.1. All extracted data are available in the supplementary data file.

Table 2.1 Extracted findings from review of SDOH data collection in EHRs.

Data input format

Social history or social screening sections of the EHR were commonly used to operationalize SDOH data collection, providing structured and unstructured digital data. Screening tools included binary screening questions (e.g. “Are you having problems with housing conditions?” (12), likert scales (e.g. “How would you rate [the client’s participation in] social network (family, work, friends)?” (24), and categorical or interval survey items (e.g. “What is your sexual orientation”, “What was your total family income before taxes last year?” (25) for self- or practitioner-administered questionnaires accessible through EHR portals (5-7, 9, 21, 24-28). Free-text documentation in clinical notes also contained SDOH data in the form of common social history topics (29, 30) such as the impact of “monetary assets, occupational level/security, and housing of housing stability and social support” on health (30). Some studies specified that a combination of both structured and unstructured SDOH data were contained in EHRs (12, 31, 32).

Data collection frameworks

Screening questions were developed based on national or professional guidelines, such as the National Academy of Medicine recommendations for SD[O]H (7), Institute of Medicine (IOM) social and behavioural domains (27, 31) and Medical Legal Partnership I-HELP categories (33). Other evidence-informed tools and frameworks included the Patient Centered Assessment Method (24) or the Wilkinson and Marmot SD[O]H categories (30). The theoretical underpinnings of data categories were not always provided, but evidence for basis in local population contexts was present in several descriptions of the collection methods (7, 28).

Automation in data collection tools

In all the cases identified in this review, a health care professional provided the data entry point into the health system. Physicians, social workers, or other health care staff were primarily responsible for identifying and coding relevant social information. Several SDOH data collection tools were self-administered, but only at the point-of-care (25, 28). We identified no studies in which patient provided free-text data were analysed and screened for SDOH data. This review revealed that verbal or written forms of collection were used exclusively, and no documentation of other input (e.g.: visual or audio) was recorded in the studies present in our review.

No health system had a fully automated collection process, although automation was relevant to collection processes through the use of digital devices for gathering patient information. While this review did not specifically focus on practices beyond collection, our collected articles addressed health care staff’s preference for free-text notes over questionnaires (12), as well as the possibility for Natural Language Processing (NLP) algorithms to efficiently and reliably identify SDOH data (29, 32, 34). One study revealed that data extracted from clinical notes were more comprehensive than data extracted from screening questionnaires (32).

Types of data and end uses

The most commonly collected SDOH information pertained to housing conditions, financial security, and social support. Other specific SDOH data items included access to social services (5, 12, 24), domestic violence (7, 12, 27) education (5, 7, 27), health literacy or navigation needs (24, 27, 35), family relationships and needs (29, 32, 33), childcare needs (5), occupation status and employment needs (29, 30, 33), race and ethnicity (9, 25, 27, 32), acculturation (27), language (25, 27, 28), religiosity (7, 25), sexual orientation (25, 28), legal needs (33, 35) addictions (30), and access to transportation (30).

The research articles did not always explicitly mention the end uses of the data. However rationale generally pertained to system improvement (7, 21, 24, 34, 36), such as determining efficient practices in collecting information on social support (37); preventative health (27, 29, 30, 36), such as understanding cancer screening practices within a population (9); and promoting health equity (25, 27, 34) for instance by improving level of assistance with SDOH needs (7). Other studies reported that SDOH data collection at the care site was a response to specific mandates, such as the Veteran’s Health Administration (VHA)’s call to end homelessness among users (21) or the American Academy of Nursing (AAN) call to action on social and behavioural determinants of health (31).

Limitations

One limitation of this research is that the majority of articles which met inclusion criteria were from the United States (n=15). No studies outside of North America were identified. This may influence the generalizability of our findings, as unique features of American political economy, such as health policies and specific population diversity and inequalities, may shape SDOH data collection methodology. One of the two articles from Canadian study sites provides an example of how regional information management contexts influence SDOH screening. The authors stated that their survey design did not initially follow the principles of Ownership, Control, Access, and Possession (OCAP®), which governs research concerning Canada’s Indigenous populations (38), and thus influenced how the authors reported study findings (6). This observation points to the need for additional research on contextually driven data collection methods, as well as further investigation into how recommendations from leading health organizations (e.g.: the Institute of Medicine) or guidelines (e.g.: OCAP®) influence practices in SDOH surveillance.

The search strategy, particularly the selection of key terms, also shaped the article pool. We searched for “Social Determinants of Health”, rather than individual terms associated with SDOH. This was considered more feasible as broad categories of SDOH exist (e.g.: income, education, occupation) but lead to a broad range of variables. For instance, income could be conceptualised as ‘family income’, ‘after tax income’, ‘accumulated wealth’, ‘ability to “make ends meet”’, etc. Furthermore, a complete set of all social determinants of health is not possible. Such a list could conceivably contain such factors as ‘uneven sidewalks’, ‘political corruption’, ‘density of green spaces’, etc. We therefore limited the search to ‘Social Determinants of Health’ as the concept of interest. Although this permitted us to further clarify how this concept is operationalized in contemporary research with EHRs, this search strategy would not have fully captured the breadth of research articles pertaining to data collection of SDOH which may not have been conceptualized as such. For example, information on ‘early childhood experiences’ were not collected in the EHRs described in this review, although this is a well-established SDOH (11). The full spectrum of SDOH data collection methodologies may therefore not be represented here.

We also searched exclusively in the published medical literature. It is possible that studies relevant to our research question may be present in the grey literature or in non-medical health databases and therefore were not included in this review. While this limits the breadth of our findings, per scoping review protocol, we do not intend to reveal all relevant information on the concept but rather map out the current and dominant themes in a new and emerging area of research.

Discussion

This review demonstrates that several health systems have been able to integrate SDOH data into EHRs. As data collection methods have apparently been designed to avoid disrupting workflow (25, 27), EHRs with embedded SDOH data collection tools could conceivably be scaled and expanded into other health systems. Further research into the information technology should improve efficiency and accuracy of data entry. However, before technical processes of SDOH data collection are developed, it is worthwhile to consider the meaningfulness of these data points and their potential to impact health equity.

Current SDOH data collection methodologies in EHRs and implications for health equity

Care access barriers

The first point to consider in population analyses using EHR data is that data collected through EHRs represent only a subset of the population: those with access and ability to navigate electronic health portals, or individuals with access to a health care provider who has access and ability to navigate electronic health portals. The parameters of EHR-derived datasets are therefore limited by technology infrastructure, meaning they are already defined by the privilege of access to health services and information technology. Evidence further suggests that SDOH (specifically intersections of age, income, and race) influence likelihood of using internet technology in health contexts (39), creating complex intersections of healthy inequities. Indeed, inequalities in access to information communication technology remains “one of the biggest hurdles” to enhancing well-being through digital tools (40).

Separate to population level analyses, the addition of SDOH data into EHRs is expected to advance precision health for individual patients by supporting clinical decision-making (29). However given the access barriers mentioned above, and our finding that most processes for SDOH data collection for EHRs take place within primary care settings, where “persistent health and access inequities” are still actualized (41); caution must be taken to avoid undermining equity by developing technology which excludes certain subsets of the population from health advancements.

Situated ontologies

Technological advancement in language processing and data management suggest that SDOH data collection methods may become efficient to the point where ‘Big Data’ analytics are possible. Even so, the ontological nature of SDOH data and social knowledge paradigms shed light on the complexity of data flow. As identified in this review, SDOH variables used in EHRs varied depending on the local context. Health practitioners may screen for only a handful of loosely defined variables, while others may collect precisely formatted information in as many as 108 domains (31). Decisions around which data to collect are determined by the priorities, and subsequent informational needs, of the health system in question. For example, considerable research is devoted to which data are necessary to improve care continuity and chance of recovery (24), reduce rate of readmission (32), or improve population rates of participation preventative practices (6). The landscape of SDOH metrics and valid associations therefore changes across environments based on situated needs. Consequently, individuals and systems derive knowledge on health risks associated with SDOH from subjective perspectives, or the knowledge produced through a specific position in time and place (42). This points to the fact that SDOH are social constructs (43) defined and created through narrative discourse (44). By shaping the representation of the data, text-based tools for identifying and capturing SDOH data also contribute to developing the construct itself.

This has both direct and indirect effects on the patient. For example, a patient may feel that the provided SDOH information may further extend power differentials through stigmatization (40). To clarify, physicians may exert authority over the patient via their professionalization; the act of introducing new perceived differentials, such as through income or education “class”, can further distance the patient from their provider. This can lead to discomfort in the interaction and/or low response rates to SDOH screening (25). Extreme lack of cultural safety, or awareness and deconstruction of cultural power imbalances, can even lead to care avoidance (45), with serious consequences on health. Indirectly, the act of classifying the patient determines their representation elsewhere in the IS. While the continuous and accurate representation of a patient is an ideal target for health systems, this includes the ability to represent the patient throughout the changing life course of a patient. There is therefore a need to engage with SDOH data as dynamic rather than static information, and recognize that health professionals produce these data via relational discourse.

Potential for equity-oriented IS design

Relationality in data collection methods to disperse concentration of social power

Prioritizing input and collaboration from various data users throughout the local information network in IS design may clarify necessary and situated methodological considerations for SDOH data collection. Roadmaps created through ‘stakeholder engagement’ (46) are examples of how health system administrators can incorporate local experience into the design of effective SDOH data collection systems. Further integrating discourse from multiple perspectives into SDOH data collection also “troubles” the current constructive narratives, to borrow a sociological term. By redistributing and sharing control over language, IS design which follows a participatory approach to development can dislodge the privilege of certain positions (e.g.: that of medical professionals) and create a more ethical and equitable process of documenting and interpreting social reality.

The reviewed literature showed an apparent gap in input from patients in IS design. The specific collection methods can shape the patient’s experience with the health system, which is an SDOH variable in and of itself (47), as well as determine how they are represented through health data. Far from being a socially neutral process, we are cautioned that if the field of data science does not provide space to address how “scientific practices themselves inadvertently legitimate and further disseminate political and cultural values and interest” (for example

'institutional erasure' of non-gender conforming individuals in the health system (48)), it may "end up complicit" in perpetuating social inequality (49). As methodological decisions for capturing SDOH data require critical thought and direct experience with social power structures, as well as consideration for mechanical and professional feasibility, interdisciplinary and participatory research is a fundamental aspect of future work.

Refocusing data sources and collection technologies for transformational information flow

While EHRs are designed to capture information on individual patients, patients are not the sole source for understanding the nature of a given environment as it pertains to health. Linkage with data from other sources, such as government or other public data sets or direct observations from the care setting, also appear to be necessary to create a picture of the social context surrounding the health delivery system. In addition to permitting entry of secondary data into the IS, these connections would allow health evidence to flow into policy decision-making mechanisms. As policy-level decisions are necessary to effectively modify the social structures which contribute to disease (50), and clinical level treatment recommendations or service recommendations alone are likely to be inefficient in addressing health disparities (51), an information infrastructure which connects health data to policy decisions could lead to greater impact. We noted that several EHRs systems were connected to broader data sets through governmental or academic partnerships (e.g. (13)). While data linkage was beyond the scope of this review, interoperability between data sources should be considered in future research in IS design.

In addition to data linkage, emergent technologies permit innovation in the type of SDOH data admissible into an IS. Systems could incorporate non-text based data formats, such as geo-spatial data to determine neighbourhood 'walkability' (52), social network data on levels of social support (53), as well as a variety of input from the clinical setting, where interactions can be considered SDOHs in and of themselves. Although modern technology provides the capacity to "track, synthesize, and visualize" features of a patient's social context, EHRs have largely not capitalized on these capabilities (54). Research on technologically enhanced data collection tools for SDOH is a current gap in the literature and their use in collecting and integrating data into EHRs is a potential avenue for exploration.

Future directions in applied research

We identified multiple health information systems which currently incorporate SDOH data from EHRs into their operations. In spite of similar software platforms and collection tools, SDOH variables were not identical across contexts. Applied research should therefore consider local context in EHR design, namely the unique care pathways and social determinants characterising a given population. Situated knowledge on SDOH also serves to promote accountability and impact by generating locally usable and relevant information. Future contextual research, particularly research using transdisciplinary participatory methodologies, may further refine best practices in data collection as well as clarify (and avoid, in so far as possible) how data streams can perpetuate health inequities.

It is also important to note that 'Big Data' refers not only to the amount of data, but the proportion of collected data relative to all available information (55). As the data economy has shown; patients, or rather human beings, are nearly unlimited sources of data. While a complete profile of all relevant health information for every individual patient is beyond possible, the comprehensiveness of SDOH data could be enhanced by incorporating relevant SDOH information from outside the health sector, such as public domain or community data (56-58). Further research on data linkage is a next step in developing SDOH data frameworks in EHRs. Conversely, in comparison to the patient, the health service delivery system may be a more appropriate unit of analysis. In such a scenario, data access barriers are reduced, as the system is monitoring itself, and the potential for impact would be greater, as the health system has a higher degree of control over its own decisions and 'behaviours' compared to the control it exerts over patients. As factors within the health system greatly influence the experience of care, a shift in surveillance towards the system of care rather than the individual members of a population, may be a promising avenue for impactful research.

Finally, in addition to the equity implications described above, the use of 'Big Data' in the public sector creates additional challenges with respect to management, quality, ethical and privacy concerns (15, 59). Key challenges around information dissemination, such as maintaining privacy and information security, directing relevant information to maximize impact and prevent information overload, and fully understanding the ethical implications of data collection and use, should all be further explored in research and policy discussions on data management.

Conclusion

This review clarified methods of collecting SDOH data for EHRs, which are increasingly relevant inputs for effective health planning and promotion. Current practices predominantly involve embedded and structured SDOH screening tools in the EHR, although the use of free-text data may increase as NLP algorithms become available to health systems. As the comprehensive range of SDOH variables tend to be specific to given populations, applying SDOH data collection tools will need to take local context into consideration. This also speaks to the paradigmatic issue of engaging with SDOH data as dynamic constructs in the experience of care. Although there is considerable perceived potential for automating SDOH data collection in order to enhance health analytics, researchers and practitioners must attend to the implications for the stated health equity goals. Evidence-informed systems-level changes based on situated knowledge should be considered an end goal of SDOH data collection methodology, rather than a sole focus on individual or behaviour driven health promotion strategies. In conclusion, mobilizing EHRs to promote SDOH data collection is a step towards facilitating 'Big Data' analytics in health information systems; however, further interdisciplinary and participatory research is necessary in order to capitalize on SDOH data for equity-oriented health promotion.

References

1. Organization WH. Closing the gap: policy into practice on social determinants of health: discussion paper. 2011.
2. The distribution of power within the community: Classes, Stände, Parties by Max Weber. *Journal of Classical Sociology*. 2010;10(2):137-52.
3. Marmot M. Social determinants of health inequalities. *The Lancet*. 2005;365(9464):1099-104.
4. Ratcliff KS. *The social determinants of health: looking upstream*. Cambridge, UK;Malden, MA;: Polity; 2017.

5. Gottlieb LM, Tirozzi KJ, Manchanda R, Burns AR, Sandel MT. Moving electronic medical records upstream: incorporating social determinants of health. *Am J Prev Med.* 2015;48(2):215-8.
6. Lofters AK, Schuler A, Slater M, Baxter NN, Persaud N, Pinto AD, et al. Using self-reported data on the social determinants of health in primary care to identify cancer screening disparities: opportunities and challenges. *BMC family practice.* 2017;18(1):31.
7. 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(5):399-407.
8. Conway M, Keyhani S, Christensen L, South BR, Vali M, Walter LC, et al. Moonstone: a novel natural language processing system for inferring social risk from clinical narratives. *Journal of biomedical semantics.* 2019;10(1):6.
9. Lofters AK, Schuler A, Slater M, Baxter NN, Persaud N, Pinto AD, et al. Using self-reported data on the social determinants of health in primary care to identify cancer screening disparities: opportunities and challenges. *BMC family practice.* 2017;18(1):31.
10. Mahamoud A, Roche B, Homer J. Modelling the social determinants of health and simulating short-term and long-term intervention impacts for the city of Toronto, Canada. *Social science & medicine.* 2013;93:247-55.
11. Mikkonen J, Raphael D. Social determinants of health: The Canadian facts: York University, School of Health Policy and Management; 2010.
12. Beck AF, Klein MD, Kahn RS. Identifying social risk via a clinical social history embedded in the electronic health record. *Clin Pediatr (Phila).* 2012;51(10):972-7.
13. Winden TJ, Chen ES, Monsen KA, Wang Y, Melton GB. Evaluation of Flowsheet Documentation in the Electronic Health Record for Residence, Living Situation, and Living Conditions. *AMIA Joint Summits on Translational Science proceedings AMIA Joint Summits on Translational Science.* 2018;2017:236-45.
14. Cantor MN, Thorpe L. Integrating data on social determinants of health into electronic health records. *Health Affairs.* 2018;37(4):585-90.
15. Perakslis E, Coravos A. Is health-care data the new blood? *The Lancet Digital Health.* 2019;1(1):e8-e9.
16. Mooney S, Pejaver V. Big Data in Public Health: Terminology, Machine Learning, and Privacy. *Annual Review of Public Health.* 2018:95-112.
17. Zhang X, Perez-Stable EJ, Bourne PE, Peprah E, Duru OK, Breen N, et al. Big Data Science: Opportunities and Challenges to Address Minority Health and Health Disparities in the 21st Century. *Ethnicity & disease.* 2017;27(2):95-106.
18. Biesecker LG. Hypothesis-generating research and predictive medicine. *Genome Res.* 2013;23(7):1051-3.
19. Hunter RF MH, Davis M, Tully MA, Valente TW, Kee F. "Hidden" social networks in behavior change interventions. *Am J Public Health.* 2015(105):513-6.
20. Barbosa RM, Nelson DR. The Use of Support Vector Machine to Analyze Food Security in a Region of Brazil. *Applied Artificial Intelligence.* 2016;30(4):318-30.
21. Chhabra M, Sorrentino AE, Cusack M, Dichter ME, Montgomery AE, True G. Screening for Housing Instability: Providers' Reflections on Addressing a Social Determinant of Health. *J Gen Intern Med.* 2019.
22. Braveman P, Egerter S, Williams DR. The social determinants of health: coming of age. *Annual review of public health.* 2011;32:381-98.
23. Peters M, Godfrey C, McInerney P, Soares CB, Khalil H, Parker D. Methodology for JBI scoping reviews. *The Joanna Briggs Institute Reviewers manual 2015: The Joanna Briggs Institute; 2015.* p. 3-24.
24. Hewner S, Casucci S, Sullivan S, Mistretta F, Xue Y, Johnson B, et al. Integrating Social Determinants of Health into Primary Care Clinical and Informational Workflow during Care Transitions. *EGEMS (Wash DC).* 2017;5(2):2.
25. Pinto AD, Glattstein-Young G, Mohamed A, Bloch G, Leung FH, Glazier RH. Building a Foundation to Reduce Health Inequities: Routine Collection of Sociodemographic Data in Primary Care. *J Am Board Fam Med.* 2016;29(3):348-55.
26. Lewis JH, Whelihan K, Navarro I, Boyle KR. Community health center provider ability to identify, treat and account for the social determinants of health: a card study. *BMC family practice.* 2016;17:121.
27. Palacio A, Suarez M, Tamariz L, Seo D. A Road Map to Integrate Social Determinants of Health into Electronic Health Records. *Population health management.* 2017;20(6):424-6.
28. Tan-McGrory A, Bennett-AbuAyyash C, Gee S, Dabney K, Cowden JD, Williams L, et al. A patient and family data domain collection framework for identifying disparities in pediatrics: results from the pediatric health equity collaborative. *BMC pediatrics.* 2018;18(1):18.
29. Lindemann EA, Chen ES, Wang Y, Skube SJ, Melton GB. Representation of Social History Factors Across Age Groups: A Topic Analysis of Free-Text Social Documentation. *AMIA Annu Symp Proc.* 2017;2017:1169-78.
30. Rabovsky AJ, Rothberg MB, Rose SL, Brateanu A, Kou L, Misra-Hebert AD. Content and Outcomes of Social Work Consultation for Patients with Diabetes in Primary Care. *Journal of the American Board of Family Medicine : JABFM.* 2017;30(1):35-43.
31. Monsen KA, Kapinos N, Rudenick JM, Warmbold K, McMahon SK, Schorr EN. Social Determinants Documentation in Electronic Health Records With and Without Standardized Terminologies. *West J Nurs Res.* 2016;38(10):1399-400.
32. Navathe AS, Zhong F, Lei VJ, Chang FY, Sordo M, Topaz M, et al. Hospital Readmission and Social Risk Factors Identified from Physician Notes. *Health Serv Res.* 2018;53(2):1110-36.
33. Theiss J, Regenstein M. Facing the Need: Screening Practices for the Social Determinants of Health. *The Journal of Law, Medicine & Ethics.* 2017;45(3):431-41.
34. Vest JR, Grannis SJ, Haut DP, Halverson PK, Menachemi N. Using structured and unstructured data to identify patients' need for services that address the social determinants of health. *International journal of medical informatics.* 2017;107:101-6.
35. Vest JR, Grannis SJ, Haut DP, Halverson PK, Menachemi N. Using structured and unstructured data to identify patients' need for services that address the social determinants of health. *International journal of medical informatics.* 2017;107:101-6.
36. Winden TJ, Chen ES, Wang Y, Lindemann E, Melton GB. Residence, Living Situation, and Living Conditions Information Documentation in Clinical Practice. *AMIA Annual Symposium proceedings AMIA Symposium.* 2017;2017:1783-92.
37. Zhu VJ, Lenert LA, Bunnell BE, Obeid JS, Jefferson M, Hughes-Halbert CA. Automatically identifying social isolation from clinical narratives for patients with prostate Cancer. *BMC medical informatics and decision making.* 2019;19(1):43.

38. Mecredy G, Sutherland R, Jones C. First Nations Data Governance, Privacy, and the Importance of the OCAP® principles. *International Journal of Population Data Science* U6 - ctx_ver=Z3988-2004&ctx_enc=info%3Aofi%2Fenc%3AUTF-8&rft_id=info%3Aid%2Fsummonserialssolutionscom&rft_val_fmt=info%3Aofi%2Ffmt%3Akev%3Amtx%3Ajournal&rftgenre=article&rfttitle=F09-01&rftpub=Swansea+University&rftissn=2399-4908&rftvolume=3&rftissue=4&rft_id=info:doi/10.23889%2Fijpds.v3i4.911&rftexternalDBID=DOA&rftexternalDocID=oai_doaj_org_article_1951 US U7 - Journal Article. 2018;3(4).
39. Choi NG, DiNitto DM. The Digital Divide Among Low-Income Homebound Older Adults: Internet Use Patterns, eHealth Literacy, and Attitudes Toward Computer/Internet Use. *Journal of medical Internet research*. 2013;15(5):e93.
40. Burr C, Taddeo M, Floridi L. The Ethics of Digital Well-Being: A Thematic Review. Available at SSRN 3338441. 2019.
41. Browne AJ, Varcoe C, Ford-Gilboe M, Wathen CN, Team ER, Team ER, et al. EQUIP Healthcare: An overview of a multi-component intervention to enhance equity-oriented care in primary health care settings. *INTERNATIONAL JOURNAL FOR EQUITY IN HEALTH*. 2015;14(1):152-.
42. Haraway D. Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. *Feminist Studies*. 1988;14(3):575-99.
43. Catalyst N. Social Determinants of Health (SDOH). *NEJM Catalyst*. 2017;1.
44. Berger PL, Luckmann T. *THE SOCIAL CONSTRUCTION OF REALITY*1966.
45. Berg K, McLane P, Eshkakogan N, Mantha J, Lee T, Crowshoe C, et al. Perspectives on Indigenous cultural competency and safety in Canadian hospital emergency departments: A scoping review. *International emergency nursing*. 2019.
46. Palacio A, Seo D, Medina H, Singh V, Suarez M, Tamariz L. Provider Perspectives on the Collection of Social Determinants of Health. *Population health management*. 2018;21(6):501-8.
47. Gilson L, Doherty, J., Loewenson, R. & Francis, V. . *CHALLENGING INEQUITY THROUGH HEALTH SYSTEMS* Final Report Knowledge Network on Health Systems. . World Health Organization, Health CoSDo; 2007.
48. Bauer GR, Hammond R, Travers R, Kaay M, Hohenadel KM, Boyce M. "I Don't Think This Is Theoretical; This Is Our Lives": How Erasure Impacts Health Care for Transgender People. *JANAC-JOURNAL OF THE ASSOCIATION OF NURSES IN AIDS CARE*. 2009;20(5):348-61.
49. Harding SG. *Science and social inequality: feminist and postcolonial issues*. Urbana and Chicago: University of Illinois Press; 2006.
50. Organization WH. *CLOSING THE GAP: Policy into practice on social determinants of health*. . (2011).
51. Bird Y, Lemstra M, Rogers M, Moraros J. The relationship between socioeconomic status/income and prevalence of diabetes and associated conditions: A cross-sectional population-based study in Saskatchewan, Canada. *INTERNATIONAL JOURNAL FOR EQUITY IN HEALTH*. 2015;14(1):93-.
52. Zandieh R, Flacke J, Martínez-Martín JA, Jones P, Van Maarseveen M. Do Inequalities in Neighborhood Walkability Drive Disparities in Older Adults' Outdoor Walking? *International journal of environmental research and public health*. 2017;14(7):740.
53. Grosberg D, Grinvald H, Reuveni H, Magnezi R. Frequent Surfing on Social Health Networks is Associated With Increased Knowledge and Patient Health Activation. *Journal of medical Internet research*. 2016;18(8).
54. Zulman DM, Shah NH, Verghese A. Evolutionary Pressures on the Electronic Health Record: Caring for ComplexityEvolution of the Electronic Health RecordEvolution of the Electronic Health Record. *JAMA*. 2016;316(9):923-4.
55. Mayer-Schönberger V, Cukier K. *Big data: A revolution that will transform how we live, work, and think*: Houghton Mifflin Harcourt; 2013.
56. Bazemore AW, Cottrell EK, Gold R, Hughes LS, Phillips RL, Angier H, et al. "Community vital signs": incorporating geocoded social determinants into electronic records to promote patient and population health. *Journal of the American Medical Informatics Association : JAMIA*. 2016;23(2):407-12.
57. Jamei M, Nisnevich A, Wetchler E, Sudat S, Liu E. Predicting all-cause risk of 30-day hospital readmission using artificial neural networks. *PLoS One*. 2017;12(7):e0181173.
58. Ye C, Fu T, Hao S, Zhang Y, Wang O, Jin B, et al. Prediction of Incident Hypertension Within the Next Year: Prospective Study Using Statewide Electronic Health Records and Machine Learning. *Journal of medical Internet research*. 2018;20(1):e22.
59. Fredriksson C, Mubarak, F., Tuohimaa, M. and Zhan, M. Big Data in the Public Sector: A Systematic. *Scandinavian Journal of Public Administration*. 2017;21(3):39-61.

Declarations

Ethics Approval: Not applicable

Consent for publication: Not applicable

Availability of data and materials: All data generated or analysed during this study are included in this published article and its supplementary information files.

Competing interests: The authors declare that they have no competing interests

Funding: This work was supported by CIHR Planning and Dissemination grant 397886.

Authors' contributions: All authors contributed to the design of the search strategy. Researchers KB and CD reviewed articles. Researcher KB was principally responsible for data extraction and analysis. All authors reviewed the manuscript before publication.

Acknowledgements: The authors acknowledge and thank Dr. Mingkai Peng for his contribution to the search strategy, as well Dr. Julia Brassolotto for her input on the framing of the analysis.

List of Abbreviations

EHR: Electronic Health Record

IS: Information System

IOM: Institute of Medicine

MeSH: Medical Subject Heading

NLP: Natural Language Processing

OCAP®: Ownership, Control, Access, Possession

SDOH: Social Determinants of Health

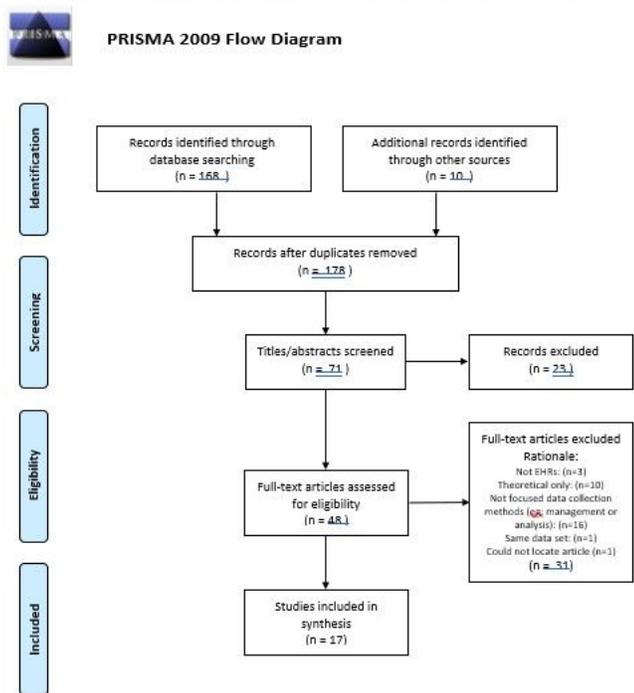
Table 2.1 Results

First Author Year	Operational terms for SDOH data	SDOH variables	Electronic System	Data Collection Methods	Link to Automation
Beck 2012	Social history questions.	Financial ("Are you doing okay making ends meet?"), Food security ("In the past month, did anyone in your family go hungry because there was not enough money?"), Housing ("Are you having problems with housing conditions?"), Access to services ("Are you having problems receiving WIC [Women, Infants, and Children], food stamps, day care vouchers, [or] health insurance"), Domestic safety, ("Do you feel that you and your children are safe in your relationships?"), unspecified questions on depression and anhedonia	Centricity Physician Office EHR (General Electric Medical Systems Information Technology, 2003)	EHR-embedded social history template, with structured questions and free text box	None specified
Chhabra 2019	Homelessness Screening Clinical Reminder (HSCR)	Homelessness (Y/N))	Corporal Michael J. Crescenz VA Medical Center (CMCVAMC) (Philadelphia, PA) EHR system	Homelessness Screening Clinical Reminder (HSCR) embedded into the EMR in outpatient clinics.	None specified
Gold 2018	Social Determinants of Health (SDH)	14 screening questions: Education level, learning style, housing condition, access to necessities, food security, domestic violence, physical activity, social connectedness, support, isolation, stress levels, religious connection . Screening list included as supplementary data to article: http://www.annfammed.org/content/suppl/2018/09/06/16.5.399.DC1/Gold_Supp_Apps.pdf	Epic EHR	SDH Screening Questions Included in OCHIN's EHR Tool, based on PRAPARE and National Academy of Medicine recommendations	"we added reportable text shortcuts to help document domains for when the patient was given SDH-related information or to document if assistance with SDH needs was declined."
Gottlieb 2015	1. Basic Resource Needs 2. Utility shut-off protection for low-income patients. 3. Homelessness among Veterans	1. Patient resident status, family resident status, household income; CPS involvement, CPS outcome, Concerns regarding basic needs: education, health insurance, job resources, food, childcare, housing, utilities, other; Day care/Education/Employment: family caretaker during week (categ, family child care provider, pre-school, child care centre, head start, pre-kindergarten. 2. Utility shut-off protection letter, auto-populated with patient data. 3. "housing needs, including where they slept the preceding night and whether they have housing stability."	1. Epic EHR 2. GE Centricity EMR. 3. Veteran's Affairs EMR	1. Social needs are captured as structured data within physician note templates. Screening categories in Epic which parallel Health Leads' [referral agency] scope of services. 2. legally formatted shut-off protection letter that is now integrated into their EMR. 3. Adapted assessment tool developed by the National Center on Homelessness Among Veterans to identify Veterans presenting with low-acuity symptoms to ED during clinic hours	"By triaging and automating some referrals, EMRs may help maximize clinical professional efficiency to meet the demand for SDH interventions and follow-up."
Hewner 2017	Assessment of social determinants of health using the Patient Centered Assessment Method (PCAM)	12 item, 4-point Likert scales ranking home security and stability, self-determination, social network, financial resources, health literacy, access to coordinated services.	Elmwood Health Center, a Patient-Centered Medical Home (PCMH)	After completing a post discharge telephone call with the patient or caregiver, the nurse selects the response that most closely matches his or her perception for each question. The PCAM is administered by professional staff without extensive psychiatric background.	None specified
Lindemann 2017	Common social history topics by age	28 unique factors were identified in manual analysis, and the 10 most common displayed for automated analysis. Several topics were consistent through all age groups across both methods of analysis: Family, Living Condition, Living Situation, Living Situation Exposure, Marital Status, Occupation, Residence, and SH Other	EHR of Fairview Health Services (FHS).	Free-text Social History Documentation section within the enterprise EHR	Social history documents extracted from the CDR were pre-processed with an open source biomedical Natural Language Processing (NLP) pipeline to extract sections and split sections into statements
Lofters 2017	Socio-demographic characteristics	Income level: i) What was your total family income before taxes last year? (where response categories were in \$30,000 increments), and ii) How many people does your income support?; Housing status: 9 categories (excluding "prefer not to answer" and "do not	EMR of Family Health Team, as a common medical home	Sociodemographic survey (developed by a multi-organizational steering committee) is completed either	Option to complete survey on tablet, but

		know"), available in supplementary file of article; Race & Ethnicity: 16 potential responses; Age, sex, postal code (to determine area income level)	model consisting of 6 clinics inn downtown Toronto.	on an electronic tablet or on paper and then transferred to the electronic medical record (EMR) by clerical staff. Responses to the survey questions are stored directly in the patient's EMR and are immediately available for viewing	manual data transfer.
Monsen 2016	IOM social and behavioural domains	107 unique SBDH items	Nine EHRs (six acute/ambulatory care and three community care)	SBDH items were documented using free text, structured text, and clinical terminologies in diverse screens and by multiple clinicians and others.	Free text analysis potential
Navathe 2018	Social factors	Structured EHR data: Race or ethnicity, marital status. Unstructured (clinical notes): tobacco use, alcohol abuse, drug abuse, depression, housing instability, fall risk, and poor social support.	EHR of multihospital academic health system.	Clinical notes and structured EHR data	To extract social factor information from physician notes, we utilized MTERMS, an NLP system validated for identifying clinical terms within medical record text
Palacio 2017	IOM Domains, but tailored to population	IOM domains: Race/Ethnicity; Education; Financial Resource-Strain; Stress; Depression; Physical Activity; Tobacco Use; Alcohol Use; Social Connection or Isolation; Intimate-partner violence; Residential Violence; Census-tract median income. Removed depression, alcohol and DV screening due to particulars of collections methods. Added country of origin, years living in the United States, language of preference, acculturation, health literacy, food insecurity, living arrangements, and transportation. We pilot tested the survey in a sample of patients and fine-tuned the questionnaire based on feedback.	EHR of the University of Miami Health System	Principles of collection and integration include: Collect SDH across the system without disrupting existing clinical workflows; Providers should have access to SDH data to improve patient care; Analyze SDH data to identify and test system-wide strategies that improve processes of care and reduce disparities among patients with high social risks; Develop community partnerships to tackle social barriers to care.	" We worked with the IT department to create smart data elements from the survey, integrate geocoding data, and define relevant EHR data elements that would be merged into the survey data automatically and create an environment for ad hoc merges."
Pinto 2016	Sociodemographic variables	14 survey questions: English as a first language/ability to speak english, Immigration status, race/ethnicity, disabilities, gender, sexual orientation, religion/spirituality, family income, family size, housing, self-rated health. The options "Prefer not to answer" and "Do not know" were available for each question.	EMRs of Four major health organizations in Toronto	Patient questionnaires in waiting room, self-administered using an electronic tablet. Question domains were identified based on studies that identified variables that are consistently tied to differences in access to health services, the quality of health services, and health outcomes. The wording of questions was informed by a literature review and refined through an iterative process, with numerous meetings and consultations involving staff and physicians at all 4 organizations over 4 years. Data collection was integrated into the standard workflow at registration	One site used electronic tablet for survey collection
Rabovsky 2017	SDH categories established by Wilkinson and Marmot.	(1) the social gradient, encompassing the impact that monetary assets, occupational level/security, and housing have on health; (2) stress, or the effects of continuous anxiety, insecurity, and low self-esteem; (3) early life, or the impact that previous emotional and developmental experiences of childhood and adolescence have on adult health; (4) social exclusion, which is the effects of being treated less than equal as a result of discrimination, debilitation, racism, or stigmatization (such as is experienced by ex-convicts, the homeless, and those who are mentally ill); (5) work, or the specific contribution of stress at work to overall health; (6) unemployment, or the increased risk of premature death experienced by the unemployed and their families; (7) social support, including the impact that both emotional and tangible support systems have on health; (8) addiction, which encompasses the effects of alcohol, nicotine, and drug dependence both as a result of social inequality as well as a means of increasing its impact; (9) food, or how access to healthy foods can	ERHs in Cleveland Clinic Health System	Social Worker encounter and notes	None specified

		influence chronic disease management and progression; and (10) transport, encompassing both the ability to arrive at appointments and walk/exercise in safe environments.			
Tan-McGrory 2018	IOM SDOH domains, but adapted to local context and resources	1) caregivers, 2) race and ethnicity, 3) language, 4) sexual orientation and gender identity, 5) disability, and 6) social determinants of health.	No specific EHR	Predominantly questionnaire-based at POC, including pad technologies, but also described the use of EHR home portals.	Incorporated electronic collection devices
Theiss 2017	"I-HELP" categories	income and insurance needs, housing and utilities needs, education and employment needs, legal status needs, and personal and family needs	EHRs of health care and legal partners that collaborate in MLPs.	Verbal and paper screenings, sometimes responses entered into EHR	"A referral process that is linked to an electronic health record could enable an automatic referral when a patient screens positive for a particular need"
Vest 2017	SDOH needs, as determined by health care professional	SDH services needed: social work, behavioral health, nutrition counseling, respiratory therapy, financial planning, medical-legal partnership assistance, patient navigation, and pharmacist consultation	Eskenazi Health's home grown EHR	Queried diagnosis and billing codes from the EHR and the HIE for ICD-9, ICD-10, and CPT procedure codes associated with behavioral health, nutritionist, respiratory therapy, and pharmacist consultation. Reviewed the EHR's unstructured data (i.e. orders and notes) for additional documentation of SDH services need.	Used natural language processing to identify instances of social worker contact with patients
Winden 2017	Residence, Living Situation, and Living Conditions	Residence: dwelling types, physical residence, and geographic location and includes safety considerations such as railings or number of floors and steps. Living Situation: with whom the patient lives such as roommates, family members, multi-resident dwelling as well as how many others they live with. Living Conditions: environmental cleanliness and precautions against infection and disease and includes such things as animals, and presence of mold or an unclean living space	University of Minnesota-affiliated Fairview Health Services (FHS) electronic health record system	a single free text field that can be documented on by any EHR clinical user. Study collected 200 random progress notes authored by social workers, 200 progress notes authored by physical therapists, and 200 progress notes authored by occupational therapists, as well as 1,200 social documentation notes	Sentence-level annotation using General Architecture for Text Engineering (GATE) to identify and classify sentences related to the three topic areas. Statement-level annotation using the brat rapid annotation tool to identify elements, attributes and relationships.
Winden 2018	Residence, Living Situation, and Living Conditions	Flowsheet prompts: "home", "house", "housing", "residence", "live", "living", "lives", "people", "mold", "insect", "rodent", "water", "heat", "social", "density" "Stairs", "railings", "safety", "safe", "facility", "group home", "skilled nursing facility", "assisted living facility", "support system", "family", "support", "housing conditions", "caregiver", "bathroom", "community support", "rehab", "assistive device", "social/environment", "equipment", "social support", "household", "transitional care", "social connectedness", "live alone"	Fairview Health System (FHS) EHR system	Flowsheets and unstructured text	Not in data collection phases

Figures



From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *BMC Med* 8(7): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit www.prisma-statement.org.

Figure 1

Flow Diagram

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementaryfile.xlsx](#)