

International Journal of Population Data Science

Journal Website: www.ijpds.org



Swansea University
Prifysgol Abertawe

How Well are the Social Determinants of Health Documented in the Electronic Medical Records in Calgary, Alberta?

Namneet Sandhu^{1*,†}, Elliot A. Martin^{1,2†}, Hude Quan^{1,3}, Eshleen Grewal³, and Karen L. Tang^{3,4,5}

Submission History

Submitted:	14/01/2026
Accepted:	03/06/2026
Published:	16/10/2026

¹ Centre for Health Informatics, Cumming School of Medicine, University of Calgary, Alberta, Canada

² Provincial Research Data Services, Health Shared Services, Alberta, Canada

³ Department of Community Health Sciences, Cumming School of Medicine, University of Calgary, Alberta, Canada

⁴ Department of Medicine, Cumming School of Medicine, University of Calgary, Alberta, Canada

⁵ O'Brien Institute for Public Health, University of Calgary, Alberta, Canada

[†] indicates co-first authorship

Abstract

Introduction

Social determinants of health (SDOH), including economic status, education level, and housing, have significant impact on an individual's health. SDOH data are therefore important not only for health equity purposes, but also for evaluations of health systems' performance, making it pertinent for hospitals to collect patient-level SDOH data.

Objective

Our study assessed how often SDOH data are documented for hospitalised patients in Calgary, Alberta, Canada.

Methods

We used chart review data of 10,659 randomly selected patients discharged from four acute care hospitals in Calgary between 2017 and 2022 and linked these data to the administrative Discharge Abstract Database. Descriptive statistics were used to measure documentation rates of nine SDOH variables: immigration status, refugee status, ethnicity, homelessness, marital status, living alone, financial strain, employment status and gender identity. Results were stratified by patient age, sex, length of stay (LOS) and number of Charlson comorbidities. Adjusted odds ratios were calculated using logistic regression models to measure the associations between covariates and documentation of each SDOH variable.

Results

In this cohort, marital status had the highest documentation (87.9% of charts) followed by homelessness (81.9%) and living alone status (71.8%). Refugee status (0.3%), immigration status (5.6%) and ethnicity (7.1%) were among the least documented SDOH variables. Complex patients with a higher number of Charlson comorbidities and longer LOS had higher rates of SDOH documentation. For example, the odds of documentation of ethnicity increased by 15% for each additional Charlson comorbidity recorded (aOR 1.15, 95% CI: 1.05-1.25). These results were statistically significant ($p < 0.01$).

Conclusion

Our study found wide variation in the documentation of different SDOH variables in hospital records. Low rates of several SDOH variables indicate the need for improving documentation in patient records.

Keywords

social determinants of health, social determinants, SDOH, electronic medical record, missing documentation

*Corresponding Author:

Email Address: namneet.sandhu@ucalgary.ca (Namneet Sandhu)

Introduction

Social determinants of health (SDOH) are the conditions in which individuals live, grow, work and age that shape their health. These conditions include, but are not limited to, education, employment, income, and housing [1, 2]. Stratification based on differences in these conditions leads to unfair and avoidable differences in health, known as health inequities [3–5]. For example, populations that are disadvantaged based on race, ethnicity, immigration, or education have been demonstrated to have higher mortality [1, 6, 7]. Thus, the ability of a healthcare system to identify and address health inequities depends on the availability and accessibility of SDOH data. At a systems level, SDOH data are important not only for health equity but also for evaluations of health systems performance, through risk adjustment. For example, in-hospital mortality is a widely used indicator of health system performance and for benchmarking but is affected by patient-level SDOH such as poverty [8–10]. Not adequately accounting for these factors may result in inaccurate assessments of hospital performance and quality [10]. Therefore, there has been increasing attention to adequate collection and documentation of patient-level SDOH in hospitals, particularly within electronic medical records (EMR), which are longitudinal records of clinical encounters [11–13].

In the EMR, SDOH can exist within structured fields (i.e., where users select the appropriate value or characteristic from a list of options), or in unstructured textual form, such as within clinical notes [14–17]. Despite the potential for an EMR to be a comprehensive repository of patient-level SDOH, there appears to be a significant underreporting of many SDOH in hospitals. For example, we conducted a recent study to develop Natural Language Processing (NLP) methods to extract certain SDOH data, such as education, presence of a language barrier, and living alone and employment status, from an inpatient EMR [18]. We found that within structured fields, while marital status was nearly universally documented, there were zero entries for occupation. Such variation across different SDOH variables is consistent with the literature, primarily from the United States, suggesting that while core demographics such as legal sex are generally highly complete in EMRs, other data such as gender identity, social support and housing status are under-documented [19–22]. In addition, SDOH tend to be more commonly captured in notes in the EMR (as unstructured data, requiring extraction such as through NLP) rather than as structured data [21–24]. Current hospital-based studies on SDOH documentation, however, tend to focus on a narrow set of SDOH variables and elements, and/or a specific disease cohorts and populations, limiting their generalizability [23–25].

What remains unknown in the current evidence base is how reporting rates of SDOH data vary across a wide range of variables especially within the hospital context, and the extent to which patient factors influence these rates of capture. This information is important to ascertain for several reasons. First, knowledge of which SDOH variables are poorly documented in EMR is a necessary antecedent to improving the documentation of SDOH. Second, as the use of NLP grows in extracting SDOH from unstructured notes from EMR, efforts to improve these technologies should

focus on the variables that are highly prevalent (i.e., there is little value in refining the extraction of data when the data do not exist). Lastly, the exploration of factors that influence SDOH documentation can inform current and future interventions and quality improvement initiatives to address barriers related to SDOH data collection and documentation. We therefore undertook a cross-sectional study of 10,659 inpatients in Calgary, Alberta, Canada between 2017 and 2022 to determine the prevalence of documentation of nine SDOH variables (immigrant status, refugee status, ethnicity, marital status, employment status, presence of financial strain, homelessness, living alone status, and gender identity), and the influence of patient and hospitalisation factors that influence these documentation rates.

Methods

Study Cohort

The study cohort included 10,659 randomly selected inpatients who were discharged from any of the four acute care hospitals, which used the same EMR, in Calgary, Alberta, Canada from 2017 to 2022. Inclusion criteria included: 1) age between 18 and 100 years old; and 2) presence of a discharge summary within the hospitalisation EMR record. For patients with multiple admissions, only the most recent hospitalisation was included. Obstetric and psychiatric admissions were excluded.

Data Sources

The EMR records for each patient underwent manual chart review to identify whether the following SDOH variables were documented: immigration status, refugee status, ethnicity, marital status, employment status, presence of financial strain, whether the patient was experiencing homelessness, living alone status, and gender identity. Sunrise Clinical Manager (SCM) was used by all four acute care hospitals [26]; it was not used in the outpatient setting (e.g. in primary care). Only the most recent hospital visit was reviewed, though tables containing structured demographic information about patients are persistent and may have been entered during an earlier encounter. All information contained in unstructured data such as clinical notes would only come from the index hospital visit.

Data were extracted using chart review conducted as part of a previous study [27]. Definitions and/or response categories for the SDOH variables are provided in the Supplemental File 1, which was the standardised form completed by each reviewer when extracting patient demographics. Financial strain was defined as the self-reported difficulty in meeting basic or health-related needs due to finances and would include broader terms such as documented food insecurity in the chart. Biologic sex was not obtained through chart review – but rather, through linkage to administrative data (Discharge Abstract Database, see below). Gender identity was captured through gendered terms such as men, women, genderqueer, genderfluid, non-binary, and two-spirit. Chart review was conducted by six registered nurses, each with a minimum of three years of practice in a clinical setting within surgery, general internal medicine, intensive care, oncology, and/or cardiology [27]. They were provided with four rounds

Table 1: Baseline Characteristics of Study Cohort from DAD administrative data

Variable	Characteristic
N	10659
Female Sex (n, %)	5566 (52.2%)
Age (median, IQR)	62 years (46-76 years)
Length of Stay (median, IQR)	4 days (2-8 days)
Charlson Comorbidities (n, %)	
Number of Charlson Comorbidities = 0	6013 (56.4%)
Number of Charlson Comorbidities = 1	3012 (28.3%)
Number of Charlson Comorbidities ≥ 2	1634 (15.3%)
Physician Specialty	
Surgical (n, %)	3640 (34.2%)
Non-surgical (n, %)	7019 (65.8%)

of training for EMR chart review with periodic monitoring to ensure data quality. An instruction manual was also provided to the chart reviewers, guiding the chart review process (Supplemental File 2). Both structured fields and unstructured text (e.g., consultation reports, discharge summaries, and multi-disciplinary daily progress notes) were reviewed from the most recent hospital visit in the patient's chart. Charts were manually reviewed and reviewers had no ability to automate the review through search terms. If a patient declined to answer a question, no data related to that SDOH variable would be extracted; that is, chart reviewers did not delineate between missing SDOH information versus whether a patient declined to answer. The overall agreement across chart reviewers was substantial (Conger's Kappa 0.80 [95% CI: 0.78-0.82]) [27, 28]. Inter-rater reliability scores ranged from 0.70 to 0.87 throughout each phase of periodic monitoring [27, 28]. SDOH data obtained through chart review were linked to the administrative Discharge Abstract Database (DAD), which captures administrative, clinical, and demographic information from each hospital stay [29]. Diagnoses are coded using the International Classification of Diseases, 10th version, Canadian Coding Standard (ICD-10-CA) [30]. Data extracted from DAD included baseline patient and visit characteristics such as age at admission, biologic sex, length of stay (LOS), physician specialty (i.e. surgical versus non-surgical), and diagnostic codes to calculate the number of Charlson comorbidities [31].

Data Analysis

Summary statistics were calculated to determine the proportion of all charts that had each of the nine SDOH variables documented. These were then stratified by patient and hospitalisation factors including age, number of Charlson comorbidities, LOS, and physician specialty. For stratified analysis, age and LOS were divided into quartiles, with each stratum containing 25% of admissions.

Adjusted odds ratios were separately computed for each SDOH variable by fitting a logistic regression model using the generalised linear model library in R [32]. In each case, the documentation status of the SDOH variable of interest was used as the outcome, and the above patient and hospitalisation variables were used as covariates. This allowed us to determine

how the covariates of interest impacted the documentation of each SDOH variable separately. The decision to run separate logistic regression models for each SDOH variable, rather than a single model with multiple outcomes, was made because documentation rates and patient-level predictors were expected to vary across SDOH variables. Different SDOH variables are likely subject to different clinical workflows and documentation practices.

Results

The study cohort consisted of a total of 10,659 patients, of whom 52.2% were females (see Table 1 for sample characteristics). The median age of the study cohort was 62 years, and the median length of stay was 4 days. Most patients (56.4%) had no Charlson comorbidities, while 15.3% had 2 or more Charlson comorbidities. The study cohort was stratified into two physician specialties, surgical and non-surgical, and 65.8% of the patients were from the non-surgical group.

Table 2 shows the documentation rates of the different SDOH variables. Marital status was the most frequently documented (87.44%) SDOH variable, followed by homelessness status (81.93%) and living alone status (71.83%). Amongst the least captured SDOH variables were refugee status (0.34%), immigrant status (5.62%) and ethnicity (7.14%). In stratified analysis (Table 2), patients who had longer hospital stays (≥ 8 days) were observed to have higher documentation rates of each SDOH variable compared to those who had shorter hospital stays. Similarly, older patients (≥ 76 years) had higher documentation rates of all SDOH variables, with the exception of employment status. Patients who had multiple comorbidities (≥ 2) were also found to have higher documentation rates of SDOH variables compared to those with only 1 or no Charlson comorbidities (Table 2).

Table 3 presents the adjusted odds ratio demonstrating associations between documentation of SDOH variables and the patient and hospitalisation variables. Odds of documentation of SDOH variables amongst surgical cases were lower than non-surgical cases. For instance, surgical cases had decreased odds of documentation of ethnicity (aOR 0.39, 95% CI: 0.31-0.47), immigrant status (aOR 0.33, 95% CI:

Table 2: Documentation Rates of SDOH Variables – Overall and Stratified by Patient and Hospitalisation Factors

	Immigrant (%)	Refugee (%)	Ethnicity (%)	Marital status (%)	Employment status (%)	Living alone (%)	Homeless (%)	Financial strain (%)	Gender identity ^a (%)
All	5.62	0.34	7.14	87.44	32.09	71.83	81.93	11.75	33.38
Sex^a									
Female	5.44	0.23	7.17	87.32	25.08	71.31	81.94	10.15	35.20
Male	5.82	0.45	7.11	87.56	39.78	72.42	81.93	13.50	31.36
Age (years)^b									
<46	5.28	0.19	6.98	85.51	32.54	58.37	74.40	10.10	34.70
46 to <62	4.70	0.39	5.78	86.47	34.43	66.42	77.91	13.15	33.77
62 to <76	4.82	0.36	5.82	86.64	33.57	74.95	83.00	11.72	30.45
≥76	7.66	0.41	9.97	88.99	27.89	86.80	91.97	12.01	34.77
LOS (days)^b									
<2	3.14	0.16	3.92	86.41	19.79	58.20	72.44	4.71	33.52
2 to <4	5.48	0.09	7.04	87.30	32.89	70.31	80.72	10.51	32.02
4 to <8	5.86	0.28	8.24	87.72	36.48	77.25	86.37	14.10	32.99
≥8	9.33	0.88	11.22	88.90	46.47	89.38	93.71	21.59	34.70
Specialty									
Surgical	2.25	0.11	3.32	86.76	22.58	62.12	75.77	6.26	32.53
Non-surgical	7.37	0.46	9.12	87.79	37.03	76.86	85.13	14.59	33.82
Charlson comorbidities									
0	4.32	0.20	5.60	87.13	28.31	65.81	78.30	8.93	33.99
1	6.54	0.46	8.53	87.28	35.56	75.93	84.66	14.48	31.64
≥2	8.69	0.61	10.22	88.86	39.66	86.41	90.27	17.07	34.33

^aThe variables sex and gender identity are from DAD administrative data and EMR chart review respectively.

^bCategories for age and LOS variables represent quartiles in the cohort.

0.26-0.42) and financial strain (aOR 0.43, 95% CI: 0.37-0.51). Odds of documentation of most SDOH variables were observed to increase with an increase in the number of Charlson comorbidities. Similarly, the odds of documentation increased with an increase in the length of hospital stay. For example, the odds of documenting ethnicity increased by 1% with an increase of one day in the LOS (aOR 1.01, 95% CI: 1.01-1.01), whereas the odds increased by 15% for each additional Charlson comorbidity recorded (aOR 1.15, 95% CI: 1.05-1.25). These findings were statistically significant ($p < 0.001$).

Having recorded female sex in DAD was associated with lower odds of documentation of employment status (aOR 0.48, 95% CI: 0.44-0.53) and financial strain (aOR 0.73, 95% CI: 0.65-0.83). These findings were statistically significant ($p < 0.001$). These logistic regression results are congruent with the higher rates of documentation of employment status in males compared to females, as reported in Table 2. Gender identity was more likely to be documented in patients recorded as female sex compared to patients recorded as male sex (aOR 1.18, 95% CI: 1.09-1.29). The association of female sex with other SDOH variables was not found to be statistically significant. Age was not observed to be associated with documentation rates of most SDOH variables, with the exception of employment status. Marital status, as reported in Table 2, was well documented and not impacted by any of the patient variables – possibly due to the existence of a known structured field in the EMR for this data element

[18]. Refugee status, which had a very low documentation rate, did not have any significant association with any of the patient/hospitalisation variables studied.

Finally, Table 4 shows how the documentation rates have changed over time. There was a gradual increase in documentation rates over time for most variables, with the largest occurring in living alone and employment status.

Discussion

In this study of 10,659 inpatients from four acute care facilities in Calgary, Alberta, Canada, we assessed rates of documentation of nine SDOH variables. We found substantial variability in the rates of documentation across different SDOH, with marital status as the most frequently documented (with 87.44% of all charts having this recorded), followed by homelessness (81.93%) and living alone status (71.83%). Refugee status, immigration status, ethnicity, and financial strain were amongst the least documented SDOH variables. Complex patients with higher numbers of comorbidities and longer hospital stays had significantly higher odds of documentation of SDOH, whereas admission to surgical services had lower rates of SDOH documentation. We also observed an increase in the documentation rate of most SDOH variables over the study period, with the largest changes seen in employment status going from a documentation rate of 25.35% in 2017 to 40.62% in 2022 and living alone going from

Table 3: Adjusted Odds Ratio of Documentation of SDOH Variables

	Immigrant aOR (95%CI)	Refugee aOR (95%CI)	Ethnicity aOR (95%CI)	Marital status aOR (95%CI)	Employment status aOR (95%CI)	Living alone aOR (95%CI)	Homeless aOR (95%CI)	Financial strain aOR (95%CI)	Gender Identity ^a aOR (95%CI)
Female ^a	0.93 (0.79–1.11)	0.54 (0.26–1.05)	1.01 (0.87–1.18)	0.98 (0.87–1.10)	0.48 (0.44–0.53)	1.00 (0.91–1.09)	1.05 (0.95–1.16)	0.73 (0.65–0.83)	1.18 (1.09–1.29)
Age (years)	1.00 (1.00–1.00)	1.00 (0.98–1.02)	1.00 (1.00–1.00)	1.00 (1.00–1.01)	0.99 (0.98–0.99)	1.02 (1.02–1.02)	1.01 (1.01–1.02)	0.99 (0.99–1.00)	1.00 (1.00–1.00)
Length of stay (days)	1.01 (1.01–1.01)	1.01 (1.00–1.02)	1.01 (1.01–1.01)	1.00 (1.00–1.00)	1.03 (1.02–1.03)	1.06 (1.05–1.07)	1.07 (1.05–1.08)	1.02 (1.02–1.03)	1.00 (1.00–1.01)
Surgical	0.33 (0.26–0.42)	0.29 (0.085–0.76)	0.39 (0.31–0.47)	0.96 (0.85–1.09)	0.47 (0.43–0.52)	0.66 (0.60–0.72)	0.70 (0.63–0.78)	0.43 (0.37–0.51)	0.95 (0.87–1.04)
Charlson comorbidities (n)	1.15 (1.05–1.27)	1.33 (0.94–1.83)	1.15 (1.05–1.25)	1.00 (0.93–1.08)	1.08 (1.02–1.14)	1.13 (1.06–1.21)	1.07 (0.99–1.16)	1.16 (1.08–1.25)	0.99 (0.94–1.05)

^aFemale refers to sex obtained from the DAD administrative data. Gender identity was obtained from the EMR chart review.

^bAdjusted odds ratios with p-values <0.01 are in bold text.

67.25% to 78.37% in that period. This finding may reflect the growing recognition of the importance of SDOH collection in the healthcare system.

Our findings on the variation of documentation rates across different SDOH are consistent with existing evidence, which is based almost entirely within the United States. These studies show that identity and demographic fields – such as race and legal sex – are nearly universally complete in EMR because they are mandated at registration, billing, or for reporting purposes [19–21, 33]. Unlike the United States, however, Canada does not mandate the capture of race or ethnicity. The result of this is, as expected, lower documentation of ethnicity – we found that only 7.14% of charts contain this information. Existing literature also suggests that socioeconomic and social risk elements are severely under-documented in the inpatient setting; they are virtually absent within structured fields [17, 21], but even when considering notes and unstructured text, they remain significantly under-documented [21–24, 34]. For example, in one study, the use of text mining techniques detected the presence of unstructured variables such as housing, financial strain, and social isolation in only 1–3% of nearly 1.2 million charts [21]. Our study similarly shows the near-complete absence of certain SDOH variables (such as refugee status); however, our estimates of documentation for financial strain, living alone status, and homelessness status are higher than those reported in the literature. Therefore, while it is true that documentation rates across different SDOH variables do vary widely, which variables are highly documented versus not and the absolute rates of their completion depend considerably on regional and national contexts.

There is a substantial body of evidence that explores barriers to the collection of SDOH data. At the patient level, while patients are generally open and receptive to the collection of these data, there remain concerns about confidentiality, perceived benefit, and potential stigmatisation [35–38]. At the provider level, barriers to SDOH data collection include workload and time demands, perceived inability to act on such data, comfort with asking patients about SDOH, and differences in opinion about who is responsible for SDOH collection [36, 37, 39–42]. With increasing attention to, and calls for, universal collection of SDOH data in hospital [12, 13], there is a need not only to understand these perspectives, experiences, and individual-level barriers (to inform considerations of how SDOH data should be collected) but also systematic gaps in the current collection and documentation of SDOH data (to inform targeted areas of focus for future intervention). Our study provides this important exploratory information. For example, we found that patients who have fewer comorbidities and shorter hospital stays have SDOH less commonly recorded in their EMR charts. Similarly, despite the importance of SDOH data in predicting surgical and post-surgical outcomes [43–45], patients admitted to surgical services have lower odds of having their SDOH documented. These populations are therefore important targets for interventions aimed at improving inpatient SDOH collection. These findings also shed light on the subpopulations that might be systematically “missing” when using SDOH and social barrier data extracted from EMR to inform program development and policy planning to identify and address health inequity.

Table 4: Documentation Rates of SDOH Variables Over Time

Year	Immigrant (%)	Refugee (%)	Ethnicity (%)	Marital Status (%)	Employment Status (%)	Living Alone (%)	Homeless (%)	Financial Strain (%)	Gender Identity ^a (%)
2017	4.77	0.2	6.32	86.55	25.35	67.25	78.28	10.69	32.89
2018	5.88	0.42	6.95	87.47	28.74	67.52	78.27	11.28	33.14
2019	5.93	0.43	6.47	87.22	30.03	69.76	80	10.73	34.34
2020	5.33	0.15	6.9	87.23	31.07	72.55	83.76	11.89	33.66
2021	5.77	0.51	7.94	88.24	37.32	75.77	84.93	12.1	32.21
2022	6.1	0.12	8.4	87.34	40.62	78.37	85.73	15.19	35.67

Despite the presence of these barriers, there remains interest in, and calls for, the widespread and systematic collection of SDOH data across entire patient populations. This interest stems from the recognition that SDOH data could be used not only to personalise and improve patient-centred care at the individual level, but is also vital at a systems level for health systems evaluation, program planning, and identification and mitigation of health inequities. A 2022 systematic review examined the integration of SDOH screening in EMR and the effects of interventions and quality improvement initiatives targeting SDOH. This review found these SDOH interventions consistently demonstrated improved process outcomes as well as improved access to resources and resolution of patient-identified concerns [46]. However, studies reporting outcomes on healthcare utilisation and cost outcomes were both limited in number and reported mixed results [46]. More recently, the Centers for Medicaid and Medicare Services in the United States implemented and tested the Accountable Health Communities Model, with 28 participating organisations [47]. Beneficiaries residing in the community were comprehensively screened for social needs, referred to community services, with a subset being provided with navigation services, while simultaneously strengthening the partnerships between clinical partners and community organisations. This intervention resulted in reduced readmissions, emergency department visits, and ambulatory care visits, generating an estimated \$200 million in net savings [47]. Ultimately, the scale of health inequity attributable to unmet social needs, combined with this emerging evidence suggests that systematic identification and documentation of SDOH remains a necessary foundation and prerequisite for the development and coordination of clinical, community, and systems-level interventions to produce downstream benefits.

The strengths of our study include an extensive, large-scale, rigorously implemented chart review of over 10,000 charts to obtain SDOH documentation rates. We therefore provide a comprehensive view of the prevalence of SDOH documentation in the EMR, not limited to structured fields (which tend to have significant under-documentation of SDOH) and not limited by the performance of the approaches used (such as with natural language processing) to extract SDOH from unstructured notes. This makes the study unique and notable in the field. Another strength is the broad scope, across multiple centres and admission services, without limiting to any specific disease cohort. Despite these strengths, we acknowledge certain limitations of this study. First, while we included a broad population across multiple centres (i.e.,

all four hospitals in the city of Calgary), all centres were located in one city in Canada, and all used the same single EMR system. This may limit generalizability to other jurisdictions, which may have different mandates, workflows, processes, and EMR platforms that may influence whether, which, and how SDOH are collected. Further, since we only included records from inpatient stays at acute care hospitals, generalizability of our findings is limited to other settings such as urgent care facilities and outpatient and/or primary care clinics. Second, our documentation rate estimates do not differentiate between structured and unstructured data. There have been numerous studies on this topic, consistently showing higher documentation of SDOH in unstructured notes compared to structured fields, so this line of inquiry was not further pursued; such information would have been helpful in data interpretation (e.g., whether the existence of structured fields like marital status was associated with improved documentation of certain SDOH variables). Third, the SDOH variables of interest in this study were selected based on study team discussion and consensus and informed by our team's prior work [18]. While established SDOH frameworks, such as those proposed by the World Health Organization [4], identify a broader range of structural and social determinants than those examined in this study, the variables selected represent individual-level factors commonly documented in clinical records and relevant to the inpatient setting. The absence of a formal SDOH framework to guide variable selection may have resulted in the omission of relevant SDOH variables. Fourth, the low documentation rates of certain SDOH variables, such as refugee status, result in few documented cases especially within certain patient subgroups (e.g., patients under 46 years, or surgical patients). Effect estimates and significance tests for these variables should be interpreted with caution. Lastly, separate logistic regression models were run for each SDOH variable to assess the associations between patient factors and SDOH documentation rates. While this approach is aligned with our research objective (i.e. to assess documentation rates across different SDOH variables), which acknowledges that documentation rates and the factors that influence them are likely to vary across different SDOH variables, we also recognise that this approach would not capture co-documentation patterns or correlations between SDOH variables. In particular, failing to account for co-documentation patterns means that we cannot determine whether certain patient factors are associated with systematically lower documentation across multiple SDOH domains simultaneously, which may have different clinical and health equity implications. Approaches that jointly model

multiple correlated outcomes or identify subgroups of patients based on overall documentation could be addressed in future work.

Conclusion

There is a need for comprehensive patient-level SDOH data in hospital-based EMR. These are essential for program development, health systems evaluations, and improving equity of outcomes and delivery of care. Despite their importance, we found that there is wide variation in documentation across different SDOH variables. Unless this documentation can be improved, the intensive efforts aimed at developing and improving technologies to extract certain SDOH data from EMR notes may be of limited utility [16, 48, 49]. Our findings also inform future interventions and initiatives to improve SDOH documentation by highlighting systematic gaps, such as severely under-documented SDOH variables and populations that are less likely to have SDOH documented, in current practice. Addressing these gaps and strengthening SDOH collection and documentation is essential to guide equitable care and system improvement.

Acknowledgements

The authors acknowledge that the administrative data were extracted from the Health Shared Services Enterprise Data Warehouse with support provided by the Alberta Strategy for Patient Oriented Research Support Unit (AbSPORU), which is funded by CIHR, Alberta Innovates, University Hospital Foundation, University of Alberta, University of Calgary and Health Shared Services. This study is based in part on data provided by Alberta Health and Health Shared Services. The interpretation and conclusions contained herein are those of the researchers, and neither the Government of Alberta nor Health Shared Services expressed any opinion in relation to this study.

Dr. Hude Quan passed away during the publication of this work, and the remaining authors would like to express their gratitude for his wisdom and guidance here and elsewhere.

Statement on Conflicts of Interest

The authors declare that there are no conflicts of interest.

Ethics Statement

This study is approved by the Conjoint Health Research Ethics Board at the University of Calgary, REB21-0416.

Data Availability Statement

The chart review data used in this study are available upon reasonable request to the PI. The administrative data were obtained from the Alberta Strategy for Patient Oriented Research Unit (AbSPORU). Due to data sharing policies of the data custodians, the administrative dataset cannot be

made publicly available. Researchers in Alberta may apply for access with approval from the data custodians here: (www.absporu.ca).

AI Disclosure Statement

The authors declare that no generative AI tools were used in the preparation of this manuscript.

References

1. World Health Organization. World report on social determinants of health equity, 2025. Geneva: World Health Organization; 2025.
2. Government of Canada. Social determinants of health and health inequalities. [Available from: <https://www.canada.ca/en/public-health/services/health-promotion/population-health/what-determines-health.html>].
3. Braveman P, Gottlieb L. The social determinants of health: it's time to consider the causes of the causes. Public Health Rep. 2014;129 Suppl 2(Suppl 2):19-31. <https://doi.org/10.1177/00333549141291S206>
4. Solar O, Irwin A. A conceptual framework for action on the social determinants of health. Social Determinants of Health Discussion Paper 2. Policy and Practice. Geneva, Switzerland: World Health Organization; 2010.
5. World Health Organization. Operational framework for monitoring social determinants of health equity. Geneva2024.
6. Statistics Canada. Health of Canadians: Health determinants [Available from: <https://www150.statcan.gc.ca/n1/pub/82-570-x/2024001/section1-eng.htm>].
7. Statistics Canada. Health of Canadians: Health outcomes [Available from: <https://www150.statcan.gc.ca/n1/pub/82-570-x/2024001/section2-eng.htm>].
8. Canadian Institute for Health Information. Hospital Deaths (HSMR) 2025 [Available from: <https://www.cihi.ca/en/indicators/hospital-deaths-hsmr>].
9. Phelos HM, Kass NM, Deeb AP, Brown JB. Social determinants of health and patient-level mortality prediction after trauma. J Trauma Acute Care Surg. 2022;92(2):287–95. <https://doi.org/10.1097/TA.0000000000003454>
10. Tran LD. Social Risk Adjustment in Health Care Performance Measures. JAMA Netw Open. 2020;3(6):e208020. <https://doi.org/10.1001/jamanetworkopen.2020.8020>
11. Davis VH, Dainty KN, Dhalla IA, Sheehan KA, Wong BM, Pinto AD. "Addressing the bigger picture": A qualitative study of internal medicine patients' perspectives on social needs data collection and

- use. PLoS One. 2023;18(6):e0285795. <https://doi.org/10.1371/journal.pone.0285795>
12. Lasser EC, Kim JM, Hatef E, Kharrazi H, Marsteller JA, DeCamp LR. Social and Behavioral Variables in the Electronic Health Record: A Path Forward to Increase Data Quality and Utility. *Acad Med*. 2021;96(7):1050-6. <https://doi.org/10.1097/ACM.0000000000004071>
13. Institute of Medicine. Capturing Social and Behavioral Domains and Measures in Electronic Health Records: Phase 2. Washington (DC): National Academies Press (US); 2014. <https://doi.org/10.17226/18951>
14. Häyrynen K, Saranto K, Nykänen P. Definition, structure, content, use and impacts of electronic health records: a review of the research literature. *Int J Med Inform*. 2008;77(5):291-304. <https://doi.org/10.1016/j.ijmedinf.2007.09.001>
15. Menachemi N, Collum TH. Benefits and drawbacks of electronic health record systems. *Risk Manag Healthc Policy*. 2011;4:47-55. <https://doi.org/10.2147/RMHP.S12985>
16. Patra BG, Sharma MM, Vekaria V, Adekanattu P, Patterson OV, Glicksberg B, et al. Extracting social determinants of health from electronic health records using natural language processing: a systematic review. *J Am Med Inform Assoc*. 2021;28(12):2716-27. <https://doi.org/10.1093/jamia/ocab170>
17. Wang M, Pantell MS, Gottlieb LM, Adler-Milstein J. Documentation and review of social determinants of health data in the EHR: measures and associated insights. *J Am Med Inform Assoc*. 2021;28(12):2608-16. <https://doi.org/10.1093/jamia/ocab194>
18. Martin EA, D'Souza AG, Saini V, Tang K, Quan H, Eastwood CA. Extracting social determinants of health from inpatient electronic medical records using natural language processing. *J Epidemiol Popul Health*. 2024;72(6):202791. <https://doi.org/10.1016/j.jep.2024.202791>
19. Foer D, Rubins DM, Nguyen V, McDowell A, Quint M, Kellaway M, et al. Utilization of electronic health record sex and gender demographic fields: a metadata and mixed methods analysis. *J Am Med Inform Assoc*. 2024;31(4):910-8. <https://doi.org/10.1093/jamia/ocae016>
20. McDowell A, Fung V, Bates DW, Foer D. Factors Associated with Completeness of Sex and Gender Fields in Electronic Health Records. *LGBT Health*. 2025;12(3):212-9. <https://doi.org/10.1089/lgbt.2023.0359>
21. Hatef E, Rouhizadeh M, Tia I, Lasser E, Hill-Briggs F, Marsteller J, et al. Assessing the Availability of Data on Social and Behavioral Determinants in Structured and Unstructured Electronic Health Records: A Retrospective Analysis of a Multilevel Health Care System. *JMIR Med Inform*. 2019;7(3):e13802. <https://doi.org/10.2196/13802>
22. Hatef E, Ma X, Rouhizadeh M, Singh G, Weiner JP, Kharrazi H. Assessing the Impact of Social Needs and Social Determinants of Health on Health Care Utilization: Using Patient- and Community-Level Data. *Popul Health Manag*. 2021;24(2):222-30. <https://doi.org/10.1089/pop.2020.0043>
23. 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. <https://doi.org/10.1111/1475-6773.12670>
24. Wray CM, Vali M, Walter LC, Christensen L, Abdelrahman S, Chapman W, et al. Examining the Interfacility Variation of Social Determinants of Health in the Veterans Health Administration. *Fed Pract*. 2021;38(1):15-9. <https://doi.org/10.12788/fp.0080>
25. Thompson HM, Kronk CA, Feasley K, Pachwicewicz P, Karnik NS. Implementation of Gender Identity and Assigned Sex at Birth Data Collection in Electronic Health Records: Where Are We Now? *Int J Environ Res Public Health*. 2021;18(12). <https://doi.org/10.3390/ijerph18126599>
26. Lee S, Xu Y, AG DAS, Martin EA, Doktorchik C, Zhang Z, et al. Unlocking the Potential of Electronic Health Records for Health Research. *Int J Popul Data Sci*. 2020;5(1):1123. <https://doi.org/10.23889/ijpds.v5i1.1123>
27. Wu G, Eastwood C, Sapiro N, Cheliger C, Southern DA, Quan H, et al. Achieving high inter-rater reliability in establishing data labels: a retrospective chart review study. *BMJ Open Qual*. 2024;13(2). <https://doi.org/10.1136/bmjopen-2023-002722>
28. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-74. <https://doi.org/10.2307/2529310>
29. Canadian Institute for Health Information. Discharge Abstract Database (DAD) metadata. [Available from: <https://www.cihi.ca/en/discharge-abstract-database-dad-metadata>].
30. Canadian Institute for Health Information. Codes, Classifications, and Terminologies [Available from: <https://www.cihi.ca/en/submit-data-and-view-standards/codes-classifications-and-terminologies>].
31. Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi JC, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med Care*. 2005;43(11):1130-9. <https://doi.org/10.1097/01.mlr.0000182534.19832.83>
32. Team RC. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2025.
33. Craven CK, Highfield L, Basit M, Bernstam EV, Choi BY, Ferrer RL, et al. Toward standardization, harmonization,

- and integration of social determinants of health data: A Texas Clinical and Translational Science Award institutions collaboration. *J Clin Transl Sci*. 2024;8(1):e17. <https://doi.org/10.1017/cts.2024.2>
34. Harris DR, Anthony N, Quesinberry D, Delcher C. Evidence of housing instability identified by addresses, clinical notes, and diagnostic codes in a real-world population with substance use disorders. *J Clin Transl Sci*. 2023;7(1):e196. <https://doi.org/10.1017/cts.2023.626>
 35. Davis VH, Rodger L, Pinto AD. Collection and Use of Social Determinants of Health Data in Inpatient General Internal Medicine Wards: A Scoping Review. *J Gen Intern Med*. 2023;38(2):480-9. <https://doi.org/10.1007/s11606-022-07937-z>
 36. Trochez RJ, Sharma S, Stollendorf DP, Mixon AS, Novak LL, Rajmane A, et al. Screening Health-Related Social Needs in Hospitals: A Systematic Review of Health Care Professional and Patient Perspectives. *Popul Health Manag*. 2023;26(3):157-67. <https://doi.org/10.1089/pop.2022.0279>
 37. DesRoches CM, Wachenheim D, Garcia A, Harcourt K, Henry J, Shah R, et al. Clinician and Patient Perspectives on the Exchange of Sensitive Social Determinants of Health Information. *JAMA Netw Open*. 2024;7(10):e2444376. <https://doi.org/10.1001/jamanetworkopen.2024.44376>
 38. What does social isolation have to do with my arthritis? Patient and provider views of the implementation of an electronic Social Determinants of Health risk assessment tool. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*: Sage Publications Inc; 2021. p. 707-10. <https://doi.org/10.1177/1071181320641164>
 39. Kepper MM, Walsh-Bailey C, Prusaczyk B, Zhao M, Herrick C, Foraker R. The adoption of social determinants of health documentation in clinical settings. *Health Serv Res*. 2023;58(1):67-77. <https://doi.org/10.1111/1475-6773.14039>
 40. Sood N, Stetter C, Kunselman A, Jasani S. The relationship between perceptions of electronic health record usability and clinical importance of social and environmental determinants of health on provider documentation. *PLOS Digit Health*. 2024;3(1):e0000428. <https://doi.org/10.1371/journal.pdig.0000428>
 41. Kostelanetz S, Pettapiece-Phillips M, Weems J, Spalding T, Roumie C, Wilkins CH, et al. Health Care Professionals' Perspectives on Universal Screening of Social Determinants of Health: A Mixed-Methods Study. *Popul Health Manag*. 2022;25(3):367-74. <https://doi.org/10.1089/pop.2021.0176>
 42. Schickedanz A, Hamity C, Rogers A, Sharp AL, Jackson A. Clinician Experiences and Attitudes Regarding Screening for Social Determinants of Health in a Large Integrated Health System. *Med Care*. 2019;57 Suppl 6 Suppl 2(Suppl 6 2):S197-s201. <https://doi.org/10.1097/MLR.0000000000001051>
 43. Sullivan GA, Krishnan V, Silver C, Smith C, Raval MV, Gulack BC, et al. Association of social determinants of health-related diagnosis codes with postoperative outcomes. *World J Surg*. 2024;48(5):1004-13. <https://doi.org/10.1002/wjs.12135>
 44. Pollak YLE, Lee JY, Khalid SI, Aquina CT, Hayden DM, Becerra AZ. Social determinants of health Z-codes and postoperative outcomes after colorectal surgery: A national population-based study. *Am J Surg*. 2022;224(5):1301-7. <https://doi.org/10.1016/j.amjsurg.2022.06.012>
 45. Abdelhack M, Tripathi S, Chen Y, Avidan MS, King CR. Social Vulnerability and Surgery Outcomes: A Cross-sectional Analysis. *Res Sq*. 2023. <https://doi.org/10.1186/s12889-024-19418-5>
 46. Yan AF, Chen Z, Wang Y, Campbell JA, Xue QL, Williams MY, et al. Effectiveness of Social Needs Screening and Interventions in Clinical Settings on Utilization, Cost, and Clinical Outcomes: A Systematic Review. *Health Equity*. 2022;6(1):454-75. <https://doi.org/10.1089/heq.2022.0010>
 47. Centers for Medicare & Medicaid Services. Accountable Health Communities Model – Final Evaluation Report Executive Summary. U.S. Department of Health and Human Services; 2026.
 48. Guevara M, Chen S, Thomas S, Chaunzwa TL, Franco I, Kann BH, et al. Large language models to identify social determinants of health in electronic health records. *NPJ Digit Med*. 2024;7(1):6. <https://doi.org/10.1038/s41746-023-00970-0>
 49. Gabriel RA, Litake O, Simpson S, Burton BN, Waterman RS, Macias AA. On the development and validation of large language model-based classifiers for identifying social determinants of health. *Proc Natl Acad Sci U S A*. 2024;121(39):e2320716121. <https://doi.org/10.1073/pnas.2320716121>

Abbreviations

AbSPORU:	Alberta Strategy for Patient Oriented Research Unit
CI:	Confidence Interval
CIHR:	Canadian Institute for Health Research
DAD:	Discharge Abstract Database
EMR:	Electronic Medical Record
ICD-10-	International Classification of Diseases,10 th
CA:	Version, Canadian Coding Standard
LOS:	Length of Stay
NLP:	Natural Language Processing
OR:	Odds Ratio
SCM:	Sunrise Clinical Manager
SDOH:	Social Determinants of Health