

# An Ontology-Guided Analysis to Unveil Disparities of Social History Documentation in Discharge Notes

Zhaoyi Sun, MS, Sitong Liu, MS, Wenyu Zeng, MS, John H. Gennari, PhD\*  
University of Washington, Seattle, WA

## Introduction

Documenting social history in discharge notes is a critical aspect of patient care. This insight is essential for integrating social and medical care effectively, from research, such as analyzing the impact of adverse childhood experiences on clinical outcomes, to care delivery, such as connecting patients experiencing homelessness to shelters. (1) Despite its significance, a preliminary extraction from the MIMIC-IV database reveals a deficiency, with only 9,128 out of 331,793 discharge notes containing a social history section. The lengths and contents of these social history texts also vary significantly, as shown in Figure 1. Such disparities, especially when analyzed across different racial groups, remain largely unexplored. To bridge this gap, we leveraged the hierarchical structure of the Social Determinants of Health Ontology (SDoHO) (2) as a standard to manually annotate 200 examples of social history documents with terms from the SDoHO. Next, we used logistic regression to analyze these 200 labeled notes by bootstrapping 10,000 samples for the disparity in the amount and style of social history notes across races, particularly in the categories of “behavior and lifestyle”, “economic stability”, and “social and community context”. This study sheds light on the disparities of social history documentation in clinical settings across races, as well as setting the stage for future research aimed at improving health outcomes through a better understanding of the social determinants of health.

## Methods

Our annotation guidelines provided definitions for each category based on the hierarchical structure of SDoHO. The “behavior and lifestyle” category includes physical ability, substance use, and sexual orientation; the “economic stability” category includes employment status and housing instability; the “social and community context” category includes adverse childhood experience, safety concern, and social support. All three categories are labeled as a boolean variable. “0” means this category is not documented in social history, while “1” means this category is

### Social History 1:

Currently homeless. In the recent past has been living in "crack houses" and half-way houses. The patient is a retired \_\_, now on disability. She had previously been living with her aunt, also an IV drug abuser, and her 3 children. She broke up with a same-sex partner in \_\_. Pt born in \_\_, moved into \_\_ years ago and \_\_ daughter living with the grandmother. She was incarcerated for \_\_ years for grand theft auto, armed robbery, and DWI. She was recently released from prison. She has a 25-pack-year tobacco history, still smoking 1 pack a day and frequent (approximately 2 drinks per day) alcohol use. She has a history of abuse, especially crack cocaine and heroin (sniffing). She denies any IV drug abuse x years. Approximately 20 detox attempts, most recently in (quit because she was discriminated against). The longest sobriety was \_\_ years while in prison. Notably, she denies any significant alcohol use - beers per week.

### Social History 2:

Married, has 3 children

### Social History 3:

non-smoker

Figure 1. Examples of social history in MIMIC-IV discharge notes

documented. The Inter-Annotator Agreement (IAA) on 20 notes showed a Fleiss’ kappa of 0.907, indicating an almost perfect agreement among 3 annotators. We annotated 200 discharge notes as the gold standard, each with a social history section. Using these 200 labeled cases, we analyzed the association between race and the documentation of three social determinant categories using logistic regression. These cases were enhanced by bootstrapping 10,000 samples to ensure robust statistical analysis. When calculating odds ratios (OR), White was set as the control group. 95% confidence intervals (95% CI) of OR were calculated and reported.

## Results

Here, we present a selection of our results, which generally show that underrepresented groups had less documentation. American Indian/Alaska Native (AI/AN) (OR = 0.27, 95% CI = 0.22-0.33), Asian (OR = 0.54, 95% CI = 0.40-0.73), and Black (OR = 0.48, 95% CI = 0.41-0.56) were found to have less “behavior and lifestyle” documentation in their social history than their White counterparts. AI/AN (OR = 0.78, 95% CI = 0.65-0.93), Asian (OR = 0.51, 95% CI = 0.41-0.63), and Black (OR = 0.58, 95% CI = 0.52-0.64) were found to have less “economic stability” documentation in their social history than their White counterparts.

## Discussion and Conclusions

Our findings reveal significant disparities in the documentation of social history across different racial groups. Black and Asian patients had “behavioral and lifestyle” comments about half as frequently as white patients (and AI/AN even less frequently). These disparities underscore the need for improved documentation practices to ensure equitable consideration of social determinants of health in patient care. By building on existing NLP frameworks, we can further analyze these disparities on the whole sample of 9,128 notes with social history and contribute to more equitable healthcare outcomes.

## References

1. 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 Nov 25;28(12):2608–16.
2. Dang Y, Li F, Hu X, et al. Systematic design and data-driven evaluation of social determinants of health ontology (SDoHO). J Am Med Inform Assoc. 2023 Aug 18;30(9):1465–73.