Title

Artificial Intelligence Approaches to Identify Social Determinants of Health in Pediatric and Opioid-Related Care: Analyses of Transformer-Based Models, Structured Data, and Unstructured Clinical Notes

Author

Mehta, Shivani Gaurang

Publication Date

2025

Peer reviewed|Thesis/dissertation

Artificial Intelligence Approaches to Identify Social Determinants of Health in Pediatric and Opioid-Related Care: Analyses of Transformer-Based Models, Structured Data, and Unstructured Clinical Notes

by

Shivani Mehta

DISSERTATION

Submitted in partial satisfaction of the requirements for degree of

DOCTOR OF PHILOSOPHY

in

Epidemiology and Translational Science

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

Signed by:

William Brown III ______________________ William Brown III

4618B0EAD7FC452...

Chair

DocuSigned by:

Matthew Pantell ______________________ Matthew Pantell

DocuSigned by:

Yulin Hswen ______________________ Yulin Hswen

8AB285E7A20A4AC...

____

____

Committee Members

Dedication

This work is lovingly dedicated to my parents. Your endless love, unwavering guidance, and countless sacrifices have been the bedrock of my journey. Not only have you been my greatest supporters, but also my best friends—offering unconditional love and a constant source of encouragement. Your wisdom has served as my guiding light, and the life lessons you've taught me have shaped who I am today. I am forever grateful for the foundation you've provided and the enduring inspiration you continue to offer.

To my loving husband, your unwavering support, boundless encouragement, and quiet strength have been my constant source of inspiration throughout this journey. You have celebrated my triumphs, offered comfort during challenges, and provided the steadfast belief that propelled me forward. Your love and patience have been the guiding light in my darkest hours, and your faith in me has made every achievement possible. Beyond all this, you are also my goofy best friend, making me laugh and smile nonstop, even in the toughest times. Thank you for being my partner, my confidant, and the joy that enriches my life.

To my dear maternal grandparents (Nani and Nana), who left this world before I had the opportunity to know you: your strength, resilience, and love transcend time, guiding me from above. I have always felt your presence, support, and unwavering strength in my heart. Your blessings live on through Mom, and your legacy lives on in every step I take.

To my beloved grandfather (Dada), whose wish was for one of his grandbabies to earn a Ph.D.—this accomplishment is for you. Although you are not here to witness this milestone, I feel your love, encouragement, and pride from the heavens above. Your unwavering belief in me has carried me through this journey. I miss you profoundly, and your memory will forever be a cherished beacon in my life. To my grandmother (Baa), whom I never had the blessing to meet, your strength, grace, and spirit live on through the stories and values you passed down to our family. I feel your quiet guidance and loving support from the heavens above.

Acknowledgments

I am deeply appreciative of the steadfast mentorship of my dissertation committee chair, Dr. William Brown III, whose unwavering support, thoughtful feedback, and high standards challenged me to grow as a scholar and person. From the very beginning of my doctoral journey, Dr. Brown has been a guiding light—challenging me to think critically and always aim higher. His unwavering support, intellectual generosity, and ability to balance high expectations with deep compassion made an indelible impact on both my academic and personal growth. He consistently created space for reflection, encouraged me to prioritize my well-being, and reminded me of the power of purpose-driven research. I am endlessly grateful for his belief in me, even when I struggled to believe in myself. I am equally grateful to my committee members, Dr. Matthew Pantell and Dr. Yulin Hswen, for their insightful guidance and encouragement throughout this journey. Your mentorship helped shape the foundation of my dissertation. Thank you both for your deep clinical insight, thoughtful feedback, and genuine curiosity pushed me to think more holistically about the intersection of social determinants and health systems.

I extend sincere thanks to Dr. Urmimala Sarkar, my F31 co-sponsor, for her advocacy and invaluable mentorship; to Dr. Catie Oldenburg, whose thoughtful questioning during my qualifying exam helped sharpen my ideas and approaches; and to Eva Wong-Moy, whose support during my doctoral training was a steady source of encouragement and care.

Finally, to the Tenacious Ten—thank you for walking alongside me with unwavering belief, humor, and heart. Your steadfast encouragement carried me through every high and low. Erin and James, your comedic relief and constant presence brought joy and perspective when I needed it most. This journey was made lighter, funnier, and more meaningful because of you.

Artificial Intelligence Approaches to Identify Social Determinants of Health in Pediatric and Opioid-Related Care: Analyses of Transformer-Based Models, Structured Data, and Unstructured Clinical Notes

Shivani Mehta

Abstract

Social determinants of health (SDOH)—including housing instability, food insecurity, and transportation barriers—significantly impact health outcomes, especially among vulnerable pediatric and opioid-affected populations. Despite their importance, SDOH remain under-documented in structured fields of electronic health records (EHRs), limiting effective clinical and policy responses. This dissertation explores the utility of Artificial Intelligence (AI) for natural language processing (NLP), particularly transformer-based models, for improving the identification and analysis of SDOH in unstructured clinical notes across three distinct studies.

In Chapter 1, a comparative analysis evaluates the performance of RoBERTa, ClinicalBERT, and BioBERT in detecting SDOH mentions from pediatric inpatient notes. RoBERTa outperformed other models, achieving an F1-score of 0.83 and recall of 0.99, highlighting its potential for high-sensitivity detection of social risks in pediatric settings. ClinicalBERT and BioBERT underperformed, particularly in handling implicit or nuanced mentions, underscoring the challenges of applying domain-specific models to complex real-world data.

Chapter 2 examines the association between housing and transportation-related SDOH and pediatric hospital readmissions. This study compares structured ICD-10-CM Z-code data with unstructured clinical text analyzed via a validated NLP pipeline. Incorporating unstructured data

increased the identified prevalence of SDOH from 0.8% to 31.7%, significantly strengthening associations between social risk exposure and both readmission and emergency department utilization. These findings demonstrate the added value of NLP in revealing latent risk factors and improving predictive modeling in pediatric care.

Chapter 3 applies AI and prompt engineering, specifically zero-shot inference using GPT-4 and LLaMA-3.2 models to classify SDOH in patients with opioid-related disorders (ORDs). While both models excelled at identifying explicit social needs such as housing and transportation, they struggled with mental health categories like anxiety and depression, often misclassifying overlapping symptoms. Prompt engineering and model refinement improved classification performance but revealed persistent challenges in discerning complex psychosocial conditions from clinical narratives.

Together, these studies underscore the promise and limitations of transformer-based NLP models in healthcare. While models like RoBERTa and GPT-4 effectively detect explicit SDOH indicators, enhancing specificity and contextual inference remains a critical frontier. By integrating structured and unstructured data, this dissertation highlights how advanced NLP tools can inform risk stratification, guide interventions, and support health equity efforts across diverse patient populations.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AI: Artificial Intelligence

API: Application Programming Interface

AUC: Area Under the Curve

BCH-SF: Benioff Children's Hospital San Francisco

BRAT: Brat Rapid Annotation Tool

CCI: Charlson Comorbidity Index

CDW: Clinical Data Warehouse

CI: Confidence Interval

DEID: De-identified

EHR: Electronic Health Record

ED: Emergency Department

F1-score: Harmonic Mean of Precision and Recall

GPT-4o: Generative Pre-trained Transformer version 4 omni

H&P: History and Physical

HR: Hazard Ratio

ICD: International Classification of Diseases

LLaMA: Large Language Model Meta AI

NLP: Natural Language Processing

ORD: Opioid-Related Disorder

Optuna: Optimization Framework for Hyperparameter Tuning

ROC: Receiver Operating Characteristic

SD: Standard Deviation

SDOH: Social Determinants of Health

TPE: Tree-structured Parzen Estimator

UCSF: University of California, San Francisco

Chapter 1: Leveraging Transformer-Based Models for Identifying Social Determinants of Health in Pediatric Clinical Notes: A Comparative Analysis of RoBERTa, ClinicalBERT, and BioBERT

INTRODUCTION

Social determinants of health (SDOH), including food insecurity, housing instability, and transportation barriers, significantly influence children's health and outcomes.[1–4] Children in food-insecure households experience worse general health, limited access to care, and increased emergency department utilization compared to food-secure children.[5,6] Housing insecurity correlates with higher hospitalization rates, while transportation barriers hinder access to necessary medical care and preventive services, leading to missed healthcare appointments.[4,7] Many children face adverse SDOH, with approximately 1.3 million experiencing homelessness annually, 4% missing medical appointments due to transportation barriers, and 14% living in food-insecure households.[1,4,8,9] The cumulative impact of multiple unmet social needs exacerbates these challenges, worsening health outcomes due to the compounding effects of SDOH.[4]

In 2019, over 5 million children were hospitalized in the United States, placing them at high risk for adverse SDOH.[10–12] Professional organizations, such as the American Academy of Pediatrics and the Children's Hospital Association, recommend screening for social risk factors, including food insecurity, housing instability, and transportation barriers.[10–13] Despite this, identifying SDOH remains a challenge due to the lack of standardized documentation in electronic health records (EHRs).[14] The absence of systematic SDOH screening tools within EHRs limits both clinical decision-making and research.[14]

Three major approaches currently exist for capturing SDOH in EHRs: extracting structured data using ICD codes, integrating external data sources such as surveys, and manually reviewing unstructured clinical notes.[14] These methods have substantial limitations. Structured data fields often lack standardization, leading to missing information and biases.[14] ICD-10-CM Z-codes remain underutilized in inpatient settings, contributing to gaps in structured data.[14] External data sources, while informative, frequently suffer from small sample sizes and lack representativeness. Community-level SDOH indices also fail to strongly correlate with individual-level SDOH.[15,16] Unstructured clinical notes provide rich qualitative information about social risk factors, but manual review is time-consuming and labor-intensive.[14]

Natural language processing (NLP) offers an efficient solution for extracting SDOH-related information from unstructured clinical notes.[17] Studies in various healthcare contexts demonstrate that NLP can uncover substantial amounts of information beyond what is available in structured data.[17,18] Advanced transformer-based models, such as Bidirectional Encoder Representations from Transformers (BERT), have significantly improved text classification. BioClinical-BERT achieved high performance in detecting SDOH, with precision, recall, and F1-scores of 0.87, 0.89, and 0.88, respectively.[19] Comparative studies suggest that RoBERTa, another transformer-based model, outperforms other BERT variants in detecting SDOH within clinical text.[20] However, these models remain untested in the pediatric setting, where proxy-reported social risk factors and documentation patterns may differ from adult care.

This study aimed to evaluate the performance of RoBERTa, ClinicalBERT, and BioBERT in detecting SDOH mentions within pediatric inpatient clinical notes. Manually annotated clinical notes were used to train and validate these models, employing advanced techniques to address class imbalance and optimize performance. By leveraging automated hyperparameter tuning and threshold optimization, we assessed the models' ability to distinguish SDOH presence and absence in pediatric clinical notes using precision, recall, F1-score, and area under (AUC) the ROC curve. This study provides insight into the applicability of NLP models for automating SDOH detection in pediatric settings.

METHODS

Study Population

This study utilized a cohort of de-identified pediatric patients hospitalized and subsequently discharged from inpatient wards at the University of California, San Francisco Benioff Children's Hospital (BCH-SF), a tertiary children's hospital serving a diverse pediatric population in Northern California. Patients were included if they had a first-time hospitalization discharge from a pediatric inpatient ward between January 2018 and January 2023 and were aged 2 to 17 years at the time of admission. Patients had to be cared for by the Pediatric Hospital Medicine service, which manages general pediatric patients.

Exclusion criteria encompassed patients who were transferred to another care facility at discharge, those who died during their index admission, those under two years of age, and patients with prior hospitalizations at BCH-SF. Excluding patients with previous hospitalizations ensured that SDOH classification reflected first-time hospitalization experiences rather than cumulative social risk factors documented across multiple admissions. Although SDOH exposures vary over time, this study focused on a single time point—patients' first hospitalization—to assess the documentation of social risk factors at initial inpatient encounters. Neonatal and infant populations were excluded due to their distinct health characteristics and care requirements.

For each eligible patient, baseline structured data variables and clinical notes were extracted at the index discharge date. The dataset comprised 101,665 clinical notes corresponding to 8,536 unique patients. Stratified random sampling ensured a representative dataset, resulting in a selection of 2,500 notes distributed as follows: 50% interdisciplinary notes, 16.7% history and physical (H&P) notes, 16.7% progress notes, and 16.7% discharge summaries. Pilot analyses revealed that interdisciplinary notes contained the highest density of SDOH documentation. A subset of 400 notes was randomly selected and manually annotated by two trained annotators to identify SDOH categories, including housing insecurity, food insecurity, and transportation needs. We received ethical approval from the UCSF Institutional Review Board (study # 23-38740).

Annotation and data preproccesing

Annotations were performed using the Brat Rapid Annotation Tool (BRAT), a web-based platform for text annotation. Each note was independently reviewed, and Cohen's kappa score was used to assess inter-rater agreement. The annotated dataset underwent preprocessing to align labels with tokenized text, with clinical notes segmented into individual sentences using Python version 3.11.8 NLTK tokenizer. Tokenization was conducted using model-specific tokenizers to ensure compatibility, and entities were mapped to their corresponding start and end indices within the text.

NLP Model Training and Optimization

The annotated dataset exhibited a highly skewed class distribution, with most tokens labeled as 'O' (outside any entity). Token-level sequence labeling approaches, such as bio-tagging, struggled with this imbalance. To address this, the classification task was reframed from sequence labeling to document-level classification, categorizing entire notes as indicating SDOH presence or absence. To further address imbalance within the dataset, oversampling techniques were employed, incorporating additional samples for minority classes from the DEID CDW SDOH note concepts data table, a data table that leveraged an a priori developed UCSF NLP SDOH pipeline.[21] Weighted loss functions were implemented to penalize misclassification of minority-class instances, improving class balance. Additionally, focal loss was used to dynamically scaled loss contributions, reducing the impact of well-classified examples while emphasizing underrepresented classes.

RoBERTa, ClinicalBERT, and BioBERT were evaluated for their ability to detect SDOH mentions at the document level. RoBERTa served as the baseline, while ClinicalBERT and BioBERT assessed domain-specific adaptations. Each model was fine-tuned using its designated tokenizer, with tokenized inputs constrained to a maximum of 512 tokens. Preprocessing steps included token normalization and selective removal of stop words.

Hyperparameter and Threshold Optimization

Hyperparameter tuning was conducted using Optuna, a state-of-the-art optimization framework designed to efficiently explore hyperparameter spaces. The flexibility of Optuna allowed us to automate the search for optimal hyperparameters, leveraging its ability to adaptively narrow down the search space based on intermediate results. The objective was to maximize the F1-score for SDOH presence detection, ensuring the models achieved a balance between precision and recall.

Hyperparameter tuning was conducted through automated grid search with cross-validation. The search explored a range of learning rates (1e-5 to 5e-5), batch sizes (16 and 32), and epoch counts (3 and 5). Five-fold cross-validation ensured robust performance evaluation, with metrics such as F1-score guiding parameter selection. Optuna employs a Tree-structured Parzen Estimator (TPE) sampler, which iteratively models the relationship between hyperparameters and the objective metric to prioritize promising configurations. Each trial consisted of training a model on the training set and evaluating it on the validation set. The F1-score on the validation set was used as the guiding metric for optimization. Optuna's visualization tools were employed to analyze the sensitivity of the objective function to each hyperparameter. This analysis revealed that the learning rate had the most significant impact on model performance, with batch size and epochs contributing to a lesser extent. These findings underscore the importance of fine-tuning learning rates, particularly for transformer-based models, where subtle changes can lead to significant improvements in performance. By leveraging Optuna for hyperparameter tuning, we ensured a systematic and efficient exploration of the hyperparameter space, enabling each model to achieve its best possible performance. This automated approach eliminated the need for manual tuning and significantly reduced computational overhead compared to exhaustive grid or random search techniques.

To ensure robustness, we used a stratified k-fold cross-validation approach within Optuna. This ensured that the validation metric was not overly dependent on a particular data split, reducing the risk of overfitting to the validation data. The algorithm adaptively balanced exploration of new hyperparameter combinations with exploitation of previously observed high-performing configurations.

Following model training, threshold tuning was performed to optimize classification thresholds, balancing precision and recall. Thresholds ranging from 0.1 to 0.9 were tested to identify the decision boundary that maximized the F1-score.

The dataset was stratified into training, validation, and test subsets to maintain class balance across all phases. The validation set consisted of 1,537 samples labeled as SDOH presence (Class 1) and 609 samples labeled as SDOH absence (Class 0). Evaluation metrics included precision, recall, F1-score, accuracy, and area under the ROC curve (AUC). Confusion matrices highlighted patterns of false positives and false negatives. Error analysis explored common misclassification patterns and areas for improvement.

The best-performing hyperparameters for each model were: RoBERTa: learning rate of 1.34e-5, batch size of 16, 3 epochs. ClinicalBERT: learning rate of 1.76e-5, batch size of 16, 3 epochs. BioBERT: learning rate of 1.21e-5, batch size of 16, 3 epochs. Final models used these parameters for training on balanced datasets, and performance was assessed on validation and test datasets.

RESULTS

Population

We identified a cohort of pediatric patients admitted to the Pediatric Hospital Medicine service between January 1, 2018, and December 31, 2023. 8,518 unique patients met inclusion criteria. Table 1 provides detailed demographic data. The average age of the cohort at the time of first discharge was 9.45 years (SD = 5.08). Majority of the patients were females (50.25%). Among the race and ethnicity groups, the largest proportion identified as Latinx (33.47%), followed by non-Latinx White (17.72%), non-Latinx Black or African American (10.11%), and non-Latinx Asian (9.08%). Insurance coverage included 42.85% on public insurance, 34.39% on private insurance, and 22.76% self-pay.

Inter-rater agreement was high, with a Cohen's kappa score of 0.63, indicating substantial agreement between annotators during the manual validation. Threshold tuning was performed by systematically varying the decision threshold from 0.1 to 0.9 to optimize the trade-off between precision and recall. On the validation set, performance metrics were calculated at each threshold to identify the point that maximized the F1-score.

RoBERTa

Following threshold tuning, RoBERTa achieved an F1-score of 0.83 with a precision of 0.72 and a recall of 0.99 for detecting SDOH presence. The confusion matrix (Figure 2) indicates that

1,523 SDOH-positive cases were correctly identified (true positive) and only 14 were missed (false negative), while among the negatives, merely 17 were correctly classified (true negative) and 592 were incorrectly predicted as positive (false positive). Although this high recall (0.99) underscores RoBERTa's strong sensitivity and ability to capture nearly all true SDOH cases, the substantial number of false positives reflects a trade-off in specificity. With an AUC of 0.87 (Figure 5), the model demonstrates robust discriminative power overall; however, these findings suggest that additional optimization is warranted to reduce the false positive rate and achieve a more balanced performance profile.

ClinicalBERT

ClinicalBERT achieved a precision of 0.87 and a recall of 0.10, yielding an F1-score of 0.17 and an AUC of 0.48 (Figure 5). The confusion matrix (Figure 3) reveals that the model correctly classified 587 true negatives and 147 true positives, while misclassifying 1390 cases as false negatives and 22 as false positives. These results indicate that although ClinicalBERT demonstrates high precision in detecting SDOH presence, its low recall reflects a significant inability to capture the full spectrum of true positive cases. The large number of false negatives underscores the model's challenges in generalizing to the diverse linguistic patterns associated with SDOH, leading to a pronounced bias toward predicting the absence of SDOH.

BioBERT

BioBERT exhibited suboptimal performance compared to both RoBERTa and ClinicalBERT, achieving an F1-score of 0.02 with a precision of 0.80 and a recall of 0.01. Its AUC was 0.57 (Figure 5), reflecting only marginally better discriminative ability than ClinicalBERT yet still far below acceptable performance. The updated confusion matrix (Figure 4) shows that BioBERT correctly classified 605 true negatives and only 15 true positives, while misclassifying 1522 cases as false negatives and 4 as false positives. These results indicate a pronounced bias toward predicting the majority class, resulting in an overwhelming number of false negatives. This poor sensitivity to SDOH presence likely stems from BioBERT's pretraining on biomedical literature, which may not capture the contextual nuances and linguistic diversity inherent to pediatric clinical notes. Consequently, the model struggles to generalize effectively in this domain, limiting its utility for accurately identifying SDOH mentions.

Error analysis

An error analysis of RoBERTa’s predictions revealed that the majority of false positives were linked to ambiguous phrases. These phrases often included references to financial or transportation contexts, such as mentions of "delayed payments" or "transportation issues," which did not explicitly indicate insecurity or unmet needs. These misclassifications suggest that the model occasionally overgeneralized contextual clues as indicative of SDOH presence. Additionally, some false negatives occurred when SDOH indicators were embedded in longer, complex sentences, making it difficult for the model to associate them with the target class. ClinicalBERT and BioBERT, on the other hand, exhibited challenges with both explicit and implicit mentions of SDOH. ClinicalBERT frequently failed to capture nuanced expressions requiring inference beyond the surface text, such as subtle implications of housing instability without explicit terms like “eviction” or “homelessness.” BioBERT showed a strong bias toward the majority class, resulting in significant under-detection of SDOH mentions. For instance, phrases with indirect references to unmet needs were often misclassified as negatives, indicating limited contextual understanding by the model.

The ROC curves further illustrated the performance disparities among the models. RoBERTa displayed a pronounced curve with a clear separation from the diagonal line of random guessing, achieving an AUC of 0.87 (Figure 5). This indicated its strong ability to distinguish between SDOH presence and absence across a variety of clinical notes. ClinicalBERT and BioBERT, in contrast, demonstrated flatter ROC curves, with AUCs of 0.48 and 0.57, respectively, reflecting their weaker discriminative power. These results reinforced the superior generalizability of RoBERTa, which maintained consistent performance across diverse note types, including interdisciplinary notes, history and physicals, progress notes, and discharge summaries. RoBERTa robustness in detecting SDOH mentions, while underscoring the limitations of ClinicalBERT and BioBERT in handling nuanced and implicit contexts. This disparity in performance emphasizes the importance of model selection and fine-tuning to optimize NLP tools for clinical applications.

DISCUSSION

In this study, we evaluated the performance of transformer-based NLP models, including RoBERTa, ClinicalBERT, and BioBERT, to identify SDOH within pediatric clinical notes. Leveraging a dataset of annotated clinical notes, our results demonstrated the strengths and limitations of these models for detecting SDOH-related information, offering valuable insights into the application of advanced NLP methodologies in a pediatric clinical context.

RoBERTa emerged as the most effective model, achieving an F1-score of 0.83, a recall of 0.99, and an AUC of 0.87, underscoring its ability to generalize across diverse note types. Despite this strong performance, error analysis revealed that RoBERTa struggled with ambiguous phrases and complex sentences, often misclassifying them as SDOH-related. This highlights the importance of further refining the model to improve precision while maintaining its high recall. ClinicalBERT, while achieving a high precision of 0.87, struggled with recall of 0.10, leading to an F1-score of 0.17 and an AUC of 0.48. These results suggest that ClinicalBERT may be better suited for detecting explicit SDOH mentions but lacks the contextual depth required for nuanced inferences. BioBERT demonstrated the weakest performance, with an F1-score of 0.02 and an AUC of 0.57, likely due to its pretraining on biomedical literature that does not fully capture the social and linguistic nuances of pediatric clinical notes.

Threshold tuning was a critical component of this study, particularly for RoBERTa, where a threshold of 0.7 optimized the balance between precision and recall, increasing its F1-score to 0.84. In contrast, threshold adjustments for ClinicalBERT and BioBERT offered limited improvements, reflecting inherent challenges in these models' ability to generalize across diverse linguistic patterns and contexts. The results highlight the importance of model-specific threshold optimization in enhancing NLP performance for diverse clinical populations and applications.

Error analysis further revealed important insights into model performance. False positives for RoBERTa were often linked to phrases referencing financial or transportation contexts without explicit indicators of insecurity. This suggests that while RoBERTa effectively identifies overt mentions of SDOH, its reliance on contextual clues can sometimes lead to overgeneralization.

ClinicalBERT and BioBERT struggled more broadly, misclassifying both explicit and implicit SDOH mentions. This was particularly evident in BioBERT's strong bias toward the majority class, resulting in significant under-detection of SDOH presence. These findings underscore the challenges of accurately capturing the complex and nuanced nature of SDOH documentation within clinical notes.

The ROC curves provided additional validation of these findings. RoBERTa's curve demonstrated clear separation from the diagonal line of random guessing, reflecting its superior discriminative ability, while ClinicalBERT and BioBERT exhibited flatter curves indicative of weaker performance. These results affirm RoBERTa's robustness in distinguishing between SDOH presence and absence, particularly in a dataset comprising diverse note types such as interdisciplinary notes, history and physicals, progress notes, and discharge summaries.

Our study has several strengths. First, the use of a diverse and well-annotated dataset ensured that the models were exposed to a wide range of SDOH expressions, enhancing their generalizability. Second, the rigorous evaluation framework, including hyperparameter tuning, threshold optimization, and stratified data splits, ensured robust and clinically relevant results. Third, error analysis provided actionable insights into model limitations, informing potential directions for future improvement.

However, this study is not without limitations. The annotated dataset was derived from a single urban academic medical center, which may limit the generalizability of the findings to other healthcare settings, such as safety-net hospitals or rural clinics. Additionally, some SDOH domains, such as food insecurity and finances, were underrepresented in the dataset, potentially constraining the models' ability to learn and generalize from these categories. Furthermore, the reliance on binary classification for SDOH detection oversimplifies the multifaceted nature of social risk factors, which often interact in complex ways. While SDOH are inherently time-varying exposures, this study focused on a single time point—patients' first hospitalization—to assess initial documentation of SDOH mentions. The exclusion of patients with prior hospitalizations ensured that the models were trained on initial SDOH documentation rather than cumulative mentions across multiple admissions. However, this approach may not capture changes in social risk factors over time, limiting insight into longitudinal patterns of SDOH documentation. Finally, while RoBERTa demonstrated strong performance, its susceptibility to false positives indicates a need for further refinement, such as incorporating domain-specific pretraining or ensemble learning approaches.

This study highlights the potential of transformer-based models, particularly RoBERTa, for identifying SDOH mentions in clinical notes. By leveraging advanced preprocessing, hyperparameter optimization, and error analysis, we have demonstrated the feasibility of using NLP tools to automate the extraction of SDOH information from unstructured clinical text. Future work should focus on enhancing model specificity, expanding the annotated dataset to include a broader range of SDOH domains, and validating these findings in diverse clinical settings. These efforts will be critical to realizing the full potential of NLP technologies in supporting social risk factor screening and addressing health disparities in pediatric populations.

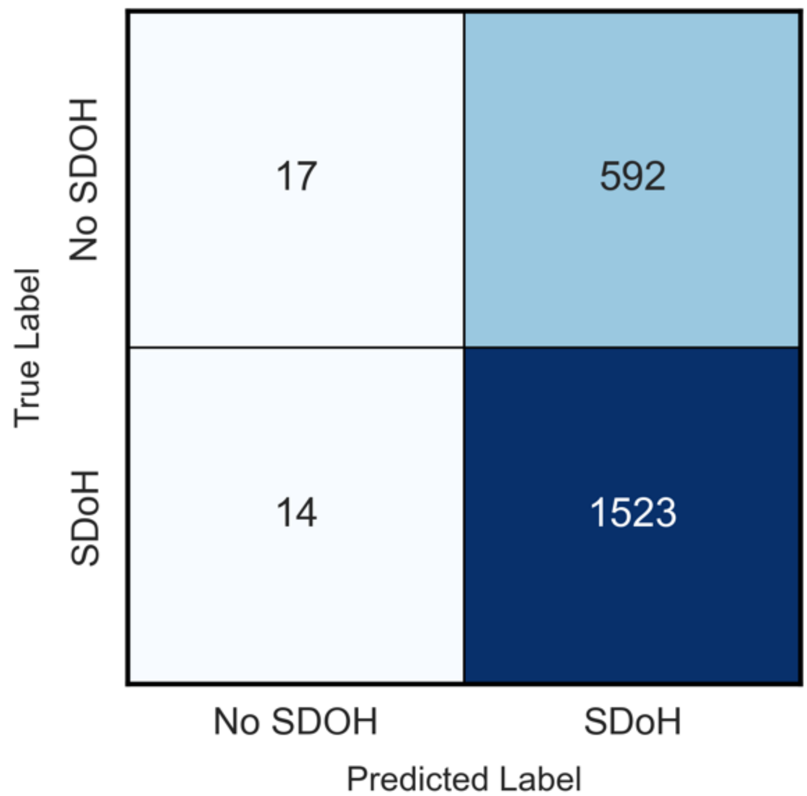

Figure 1.1 Confusion matrix for RoBERTa.

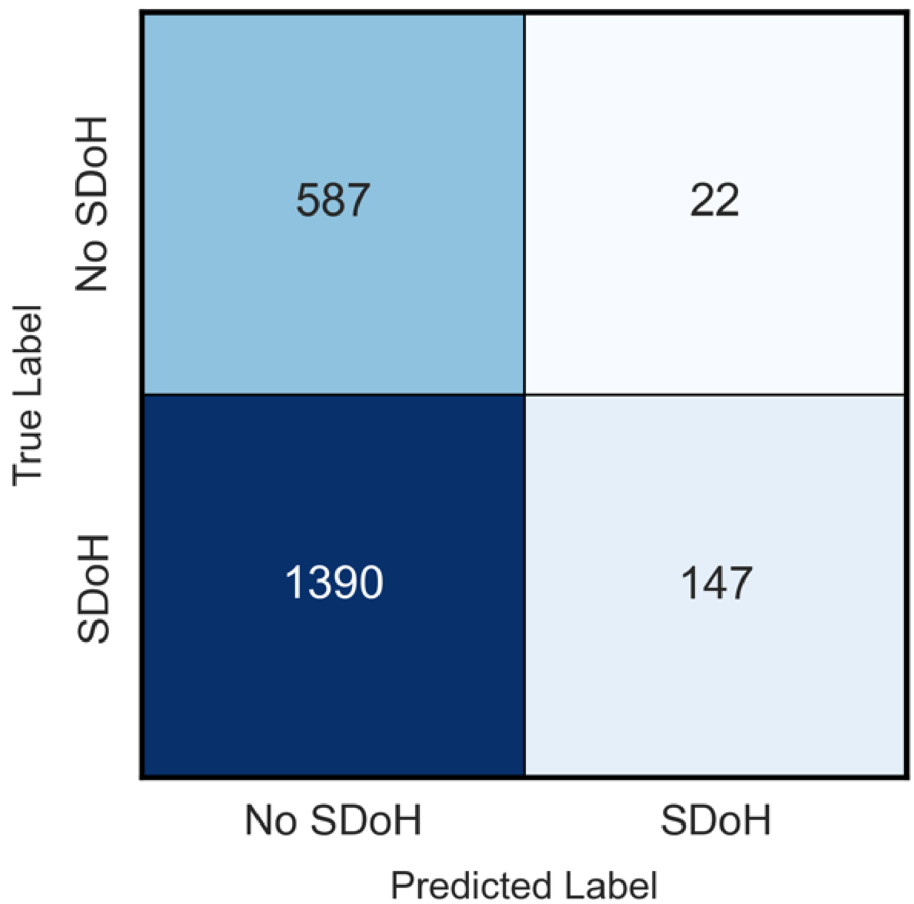

Figure 1.2 Confusion matrix for ClinicalBERT.

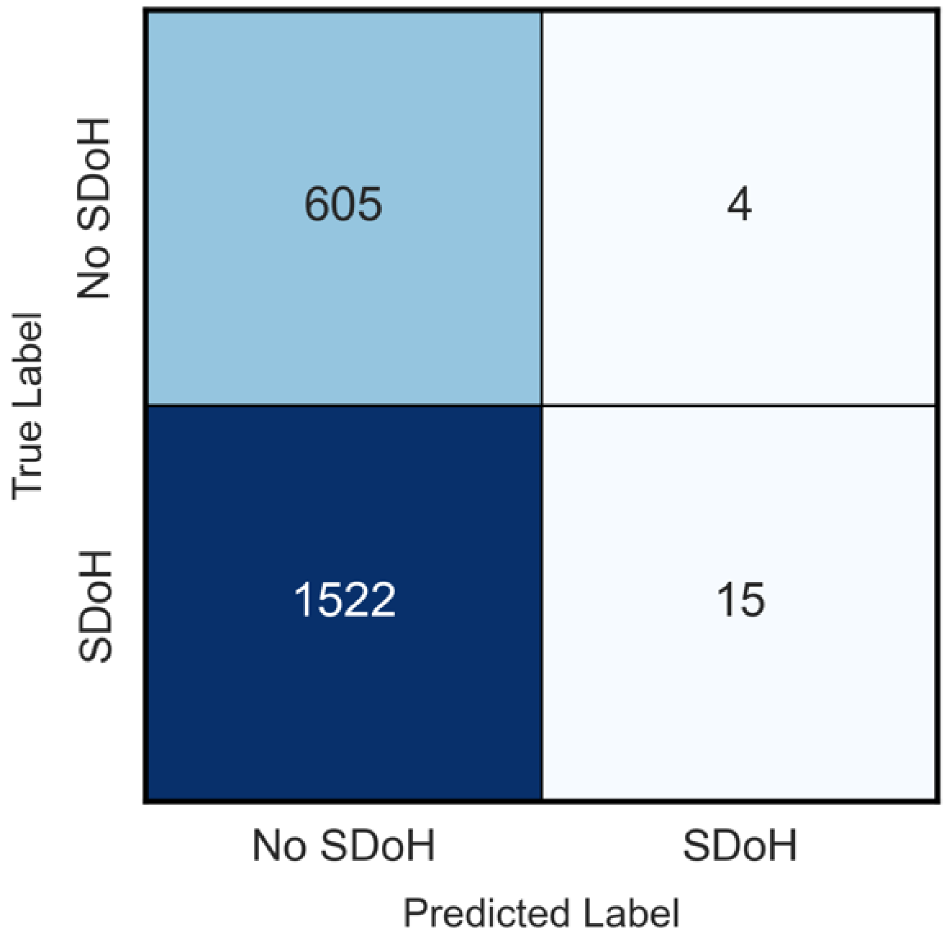

Figure 1.3 Confusion matric for BioBERT.

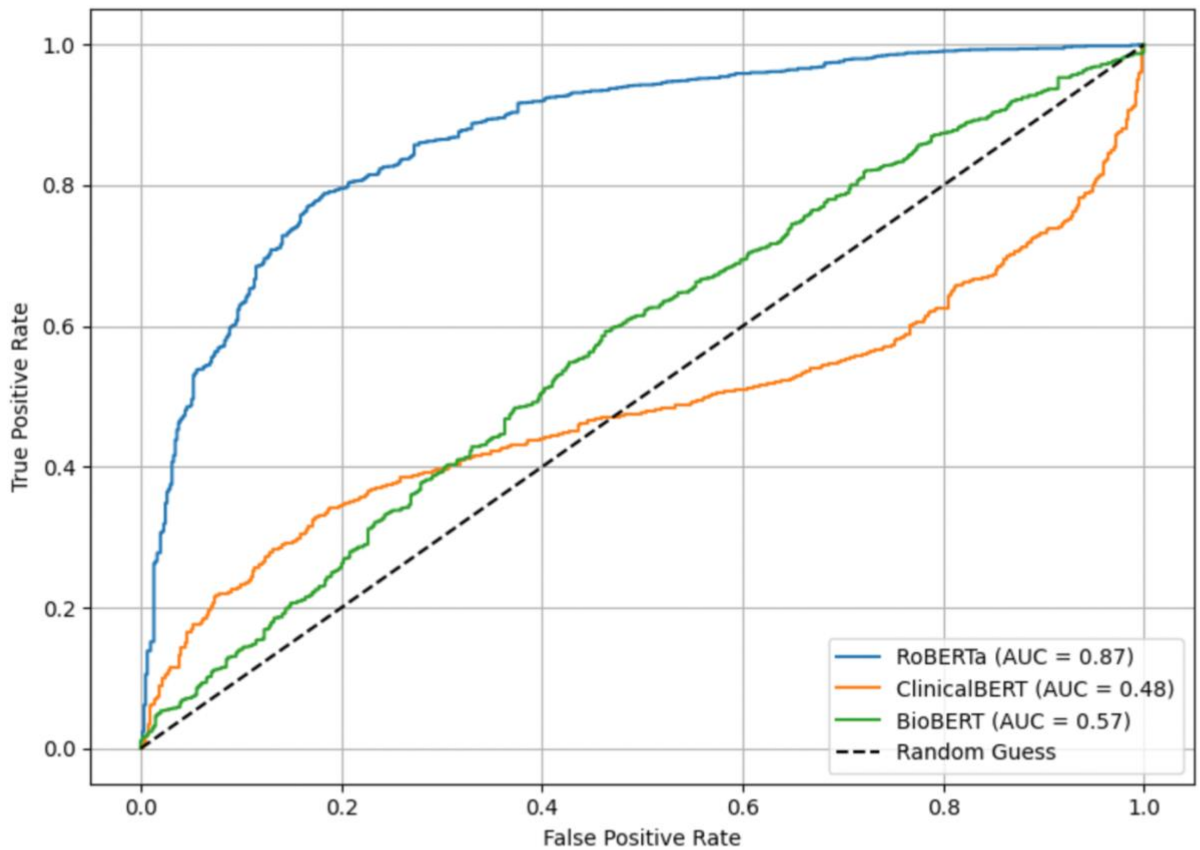

Figure 1.4 ROC curve.

Table 1.1 Patient demographics.

	Cohort (N=8,928)
Age, mean (SD)	9.45 (5.1)
Length of stay in days, mean (SD)	3.98 (10.0)
Sex	
Male	4,429 (49.6)
Female	4,495 (50.4)
Other	4 (0.04)
Race and ethnicity	
Non-Latinx Asian	811 (9.1)
Non-Latinx Black or African American	903 (10.1)
Latinx	2,988 (33.5)
Non-Latinx White	1,582 (17.7)
Other	1,136 (12.7)
Unknown	1,508 (16.9)
Insurance	
Private	3,070 (34.4)
Public	3,826 (42.9)
Self-pay	2,032 (22.8)
Primary Language	
English	6,964 (78.0)
Spanish	1,619 (18.1)
Other	345 (3.9)
Charlson Comorbidity Index	
0	5,446 (61.0)

	Cohort (N=8,928)
1	1,205 (13.5)
2+	2,277 (25.5)

Chapter 2: Examining Housing Insecurity and Transportation Barriers in Pediatric Hospital Readmissions: Insights from Structured and Unstructured Data

INTRODUCTION

Hospital readmissions among pediatric patients remain a significant challenge in healthcare, highlighting disruptions in continuity of care and contributing to increased healthcare expenditures.[22] Identifying predictors of these readmissions is critical to designing targeted interventions that improve pediatric health outcomes and reduce healthcare system burdens.[23] Social determinants of health (SDOH), hereinafter interchangeably referred to as social risk factors, such as housing insecurity and transportation barriers, have been increasingly recognized as key drivers of health outcomes. While prior research has examined the role of SDOH in adult readmissions, the specific impact of housing and transportation-related social risk factors on pediatric populations remains underexplored.[24]

Children, particularly those with chronic or complex health needs, are uniquely vulnerable to social risk factors like housing instability and transportation barriers. These factors contribute to health inequities, limit access to essential medical care, and result in poor health outcomes.[25] Pediatric patients from socioeconomically disadvantaged families face greater risks of fragmented healthcare, which increases their likelihood of readmission.[26] Targeting housing and transportation-related challenges offers an opportunity to optimize interventions and allocate resources efficiently to support at-risk children and families.[27]

Traditional methods for identifying social risk factors rely on structured data, such as ICD-10-CM Z-codes, which often fail to fully capture the scope of these factors due to inconsistent coding practices.[14] Clinical notes, an unstructured data source, frequently contain nuanced information about patients' social contexts that remain unaccounted for in structured fields. Natural language processing (NLP) enables the extraction of this valuable information, providing a more comprehensive understanding of patients' social risks.[28] Integrating structured Z-codes with insights derived from unstructured clinical notes allows for a broader and more accurate assessment of SDOH, but has not been well-explored among inpatient pediatric populations.[29]

This study investigated the association between housing and transportation-related social risk factors and hospital readmissions in pediatric inpatients. We compared two approaches: (1) structured data using ICD-10-CM Z-codes for housing and transportation problems and (2) a combined analysis integrating structured Z-codes with NLP-extracted information from clinical notes.

METHODS

Study Population

The study population consisted of de-identified pediatric patients hospitalized and subsequently discharged from inpatient pediatric wards at the University of California, San Francisco Benioff Children's Hospital (BCH-SF). BCH-SF is a tertiary children's hospital serving Northern California that caters to a diverse population of children. Patients were eligible for inclusion if they had a first-time hospitalization discharge from a pediatric inpatient ward between January 2018 and January 2023 and were aged 2 to 17 years. We excluded patients under 2 years old, as neonatal and infant patients require separate analysis due to the heterogeneous nature of their health characteristics. We also excluded patients who were transferred to another care facility at discharge, patients who died during the index admission, and patients with a prior hospital admission to an inpatient pediatric ward at BCH-SF A complete list of inclusion and exclusion criteria is provided in Table 2.1. We received approval from the UCSF Institutional Review Board (study #23-38740).

Data collection and exposure definition

We extracted structured data from the University of California, San Francisco (UCSF) de-identified Clinical Data Warehouse (CDW), including demographic information, clinical characteristics, and hospitalization details. The UCSF CDW contains de-identified data on patients in the UCSF Health system, including the hospital used for this study. For each patient included in our sample, we also extracted their clinical notes at the index date of discharge.

We determined exposure status at each patient's initial discharge in two ways. First, structured data only (Analysis 1): We defined exposure for the structured data-only analysis as the presence of at least one ICD-10-CM Z-code related to housing insecurity and/or transportation problems during the initial hospitalization. These codes included those for housing-related problems (Z59.0, Z59.1, Z59.10, Z59.18X) and transportation problems (Z59.811, Z59.812, Z59.819). We classified patients without any such codes documented in their structured data as unexposed. Second, structured and unstructured data (analysis 2): We additionally defined exposure using both the ICD-10-CM Z-codes and information extracted from unstructured clinical notes using NLP techniques. The NLP pipeline was designed to identify mentions of social risk factors corresponding to the same SDOH categories as the Z-codes. We classified patients as exposed if either the structured data contained relevant Z-codes or the unstructured data revealed social risk factors.

Natural language processing pipeline

We utilized a previously developed NLP pipeline created by UCSF to process unstructured clinical notes and extract information on social risk factors.[30] This pipeline was designed to identify and categorize mentions of social risk factors corresponding to the same SDOH categories as the ICD-10-CM Z-codes used in the structured data analysis. The key features of the NLP pipeline included:

- Data preprocessing: The clinical notes were prepared for analysis in three main ways.de-identification, tokenization, and normalization techniques.
- Entity recognition and classification: The pipeline employed advanced NLP techniques, including machine learning algorithms and rule-based methods, to recognize entities related to social risk factors within the text and classify them into predefined SDOH categories.
- Integration with structured data: The extracted information from the unstructured notes was mapped to the corresponding SDOH categories used in the structured data analysis, allowing for a combined assessment of exposure.
- Validation and quality assurance: The NLP pipeline had been previously validated by UCSF researchers to ensure accuracy and reliability in extracting social risk factors from clinical notes.[30]

Outcome

The primary outcome of interest was hospital readmission within 365 days post-discharge. This was identified from hospital records and defined as any unplanned readmission to the hospital within 365 days of the initial discharge date. Unplanned readmissions were defined as readmissions that were not scheduled as part of the patient's established care plan, encompassing any hospitalizations that were not pre-arranged for ongoing treatments or procedures. The secondary outcome of interest was time to all-cause emergency department (ED) visit within 365 days post-discharge. This outcome captures any return visit to the ED, regardless of whether it resulted in hospital admission or discharge home. By including all ED encounters, we aimed to assess the broader impact of SDOH on urgent healthcare utilization, beyond inpatient readmissions alone. Patients who did not return to the ED were censored at 365 days post-discharge or at the time of death if it occurred within the follow-up period.

Statistical analysis

We conducted separate analyses for the two exposure definitions to evaluate the impact of incorporating unstructured data on effect estimates. Baseline characteristics of the study population were summarized for both exposure groups, with continuous variables reported as means with standard deviations (SD) or medians with interquartile ranges (IQR), and categorical variables presented as frequencies and percentages.

For each exposure definition, we constructed Kaplan-Meier survival curves to visualize differences in time to hospital readmission and compared survival distributions using log-rank tests.

To quantify the association between social risk factor exposure and time to inpatient hospital readmission and ED utilization, we applied multivariable Cox proportional hazards models, fitting separate models for each outcome. For hospital readmission, we analyzed structured data alone, using ICD-10-CM Z-codes, and a combined approach incorporating both structured Z-codes and unstructured data extracted through NLP from clinical notes. Similarly, for time to ED visit, we examined structured data alone and the combined structured and unstructured data model.

Each model adjusted for age, sex, race/ethnicity, preferred language, insurance type, length of stay, and Charlson Comorbidity Index (CCI) to account for potential confounders. The proportional hazards assumption was tested using Schoenfeld residuals to ensure model validity.

For both hospital readmission and ED utilization, we compared hazard ratios and confidence intervals across the structured-only and combined-data models to assess the impact of incorporating unstructured data on effect estimates. These comparisons allowed us to evaluate differences in the magnitude, direction, and statistical significance of the associations between social risk factors and healthcare utilization, highlighting the added value of NLP-extracted data in identifying at-risk pediatric patients.

RESULTS

Sample

This study included 8,928 pediatric patients discharged from UCSF Benioff Children's Hospital San Francisco (BCH-SF) between January 2018 and January 2023. Patients were stratified into two cohorts based on exposures of interest: one using structured data alone and another combining structured and unstructured data from clinical notes. In the structured data cohort, only 71 patients (0.8%) were classified as exposed to SDOH, while 8,857 patients (99.2%) were unexposed. However, in the cohort that incorporated unstructured data, 2,827 patients (31.7%) were identified as exposed, with 6,101 patients (68.3%) categorized as unexposed. There is a substantially increased prevalence of identified social risk factors when leveraging unstructured clinical notes, suggesting that relying solely on structured data may significantly underestimate the burden of SDOH-related risks in the patient population.

Table 2.2 summarizes the demographic and clinical characteristics of the study population. The distribution of race and ethnicity differed between exposed and unexposed groups, with a higher proportion of Black and Latinx patients in the exposed groups across both the structured and combined data cohorts. Exposed patients had higher rates of public insurance, and a greater proportion reported a primary language other than English. Differences in comorbidity burden were also observed, with a lower proportion of exposed patients in the structured data cohort having a CCI score of 2 or more, whereas in the combined data cohort, the proportion of exposed patients with a CCI score of 2 or more was higher.

Time to unplanned hospital readmission

A Kaplan-Meier survival analysis (Figure 2.1) was conducted to compare time to hospital readmission between patients with and without documented social determinants of health (SDOH) concerns identified via structured and unstructured data. The log-rank test showed a significant difference in readmission rates between the groups ($\chi^2 = 381$, df = 1, p<0.0001).

Patients in the exposed group (with documented SDOH concerns) had a higher observed readmission rate while the unexposed group had fewer readmissions than expected These findings indicate that patients with documented SDOH concerns have significantly shorter times to readmission compared to those without.

When using structured data alone, patients with social risk factors demonstrated a hazard ratio (HR) of 1.99 (95% CI: 1.27–3.13) for unplanned readmission compared to patients without social risk factors, signifying a nearly two-fold increased risk (Figure 2.2). In contrast, the addition of unstructured data strengthened this association, with the HR for patients with social risk factors increasing to 2.64 (95% CI: 2.34–2.98). This amplification reflects the broader detection of social risk exposures via unstructured data and its significant association with adverse outcomes.

Other covariates were also associated with readmission risk. In the structured-only analysis, public insurance was linked to a 39% higher probability of readmission compared to private insurance (HR: 1.39; 95% CI: 1.20–1.61). Patients with a CCI of 2 or more had a 47% increased probability of readmission compared to those with a CCI of 0 (HR: 1.47; 95% CI: 1.29–1.67). Age had a statistically significant but modest effect size (HR: 1.01; 95% CI: 1.00–1.02). However, in the combined analysis, the impact of public insurance was attenuated (HR: 1.18; 95% CI: 1.02–1.37), and the hazard associated with a CCI of 2 or more was slightly reduced to 1.34 (95% CI: 1.18–1.53).

Time to emergency department (ED) hospital visit

In the structured data-only analysis, exposed patients demonstrated a significantly higher hazard of ED readmission among patients with social risk factors compared to unexposed patients, with a hazard ratio (HR) of 2.26 (95% CI: 1.65–3.10) (Figure 2.3).

Several other covariates were also associated with ED readmission. Age was inversely associated with readmission risk, with each additional year of age corresponding to a 2.8% reduction in the hazard (HR: 0.97; 95% CI: 0.96–0.98). Male sex was protective, with a hazard ratio of 0.92 (95% CI: 0.85–1.00), compared to female patients. Race and ethnicity showed notable inequities. Black patients had the highest risk, with an HR of 2.05 (95% CI: 1.77–2.38), followed by Latinx patients (HR: 1.46; 95% CI: 1.28–1.67). Patients with public insurance and those who were self-pay also had significantly higher hazards of ED readmission compared to patients with private insurance, with HRs of 1.31 (95% CI: 1.18–1.45) and 1.30 (95% CI: 1.15–1.46), respectively. Patients with a CCI of 1 or ≥2 had 39% (HR: 1.39; 95% CI: 1.24–1.55) and 33% (HR: 1.33; 95% CI: 1.22–1.45) higher hazards, respectively, compared to patients with a CCI of 0.

In the combined analysis incorporating both structured and unstructured data, the association between exposure and ED readmission was slightly attenuated, with an HR of 1.49 (95% CI: 1.37–1.62). Similar to the structured-only analysis, age remained protective, with an HR of 0.97 (95% CI: 0.96–0.98), and male sex was associated with a reduced hazard of readmission (HR: 0.93; 95% CI: 0.86–1.00). Race and ethnicity disparities persisted. Black patients had an HR of 2.02 (95% CI: 1.74–2.35), while Latinx patients had an HR of 1.46 (95% CI: 1.28–1.67). Public insurance and self-pay were again significant predictors, although the HRs were slightly attenuated at 1.25 (95% CI: 1.13–1.38) and 1.26 (95% CI: 1.12–1.42), respectively. Patients with a CCI of 1 or ≥2 demonstrated elevated hazards of ED readmission, with HRs of 1.34 (95% CI: 1.20–1.49) and 1.27 (95% CI: 1.16–1.39), respectively.

A Kaplan-Meier survival analysis (Figure 2.4) was conducted to compare time to ED readmission between patients with and without documented SDOH, identified via structured and unstructured data. The log-rank test showed a significant difference in ED readmission rates between the groups ($\chi^2 = 135$, df = 1, $p < 0.001$). These findings indicate that patients with documented SDOH concerns experience significantly shorter times to ED readmission compared to those without.

DISCUSSION

This study highlights the importance of incorporating both structured and unstructured data to comprehensively assess the impact of social risk factors on pediatric unplanned hospital readmissions. Using structured data alone, the hazard of readmission for exposed pediatric patients was nearly double that of unexposed patients, emphasizing the significance of social determinants of health (SDOH) documented in ICD-10-CM Z-codes. However, the addition of unstructured data from clinical notes via an NLP pipeline revealed an even stronger association between social risk factors and readmissions. The hazard ratio increased substantially when unstructured data was incorporated, underscoring the potential underestimation of social risk factors when relying solely on structured data.

Our findings are consistent with prior research demonstrating the relationship between social risk factors and hospital utilization, but this study advances the literature by demonstrating the additional value of leveraging unstructured data to capture nuanced and often undocumented social factors. Importantly, the Kaplan-Meier curves showed a marked differentiation between exposed and unexposed groups, with the combined analysis revealing greater inequities in survival probabilities. These findings suggest that unstructured data not only identifies more patients with social risks but also captures those at the highest risk for adverse outcomes. Using unstructured data in addition to structured data gets a more accurate prediction of the prevalence of social risk factors, which can help health systems understand the burden of risk factors among their patient populations and invest in resources to address common social risk factors

This study importantly contributes to the relatively sparse body of research focusing on pediatric patients and the influence of social determinants of health on their readmission rates. While existing literature extensively examines adult populations, pediatric patients, especially those facing socio-economic adversities, remain significantly under-examined. Our findings not only underscore the heightened vulnerability of this group but also illuminate the critical gaps in current healthcare practices that could benefit from targeted research and tailored interventions.

Other significant predictors of readmission included public insurance, which remained a consistent risk factor in both analyses, and a higher CCI, reflecting the cumulative burden of chronic conditions. This highlights the importance of understanding how social and clinical vulnerabilities interact to drive disparities in health outcomes.

Despite the strengths of this study, there are limitations to consider. First, while the use of unstructured data improved exposure ascertainment, the accuracy of NLP pipelines can vary depending on the quality and standardization of clinical notes. Second, this study was conducted in a single academic medical center, which may limit generalizability to other pediatric populations or healthcare settings. Third, while we adjusted for several confounders, residual confounding cannot be ruled out, particularly in the context of social risk factors that are difficult to quantify comprehensively.

This study demonstrates that leveraging unstructured data through NLP pipelines in addition to structured data significantly enhances the identification of social risk factors and their association with unplanned hospital readmissions. Structured data alone may underestimate the burden of social determinants of health, highlighting the need for a hybrid approach in future research and clinical interventions. Policymakers and healthcare providers should consider integrating unstructured data into routine analytics to improve risk stratification and target interventions for vulnerable pediatric populations as addressing social determinants of health remains a critical pathway for reducing hospital readmissions and achieving equitable health outcomes in pediatric care.

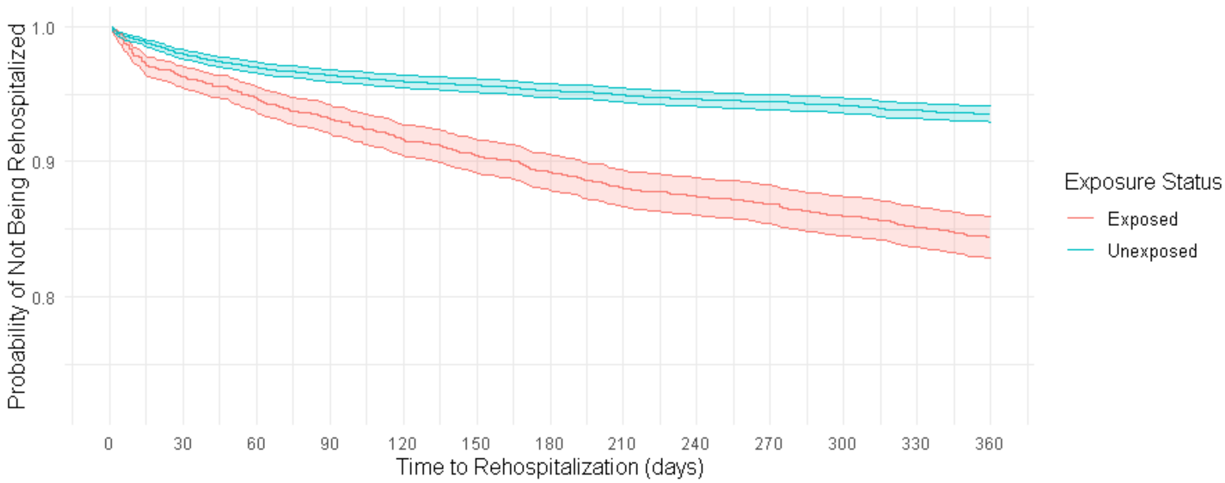

Figure 2.1 Kaplan-Meier Survival Curve for Time to Unplanned Hospital Readmission Stratified by Exposure Group.

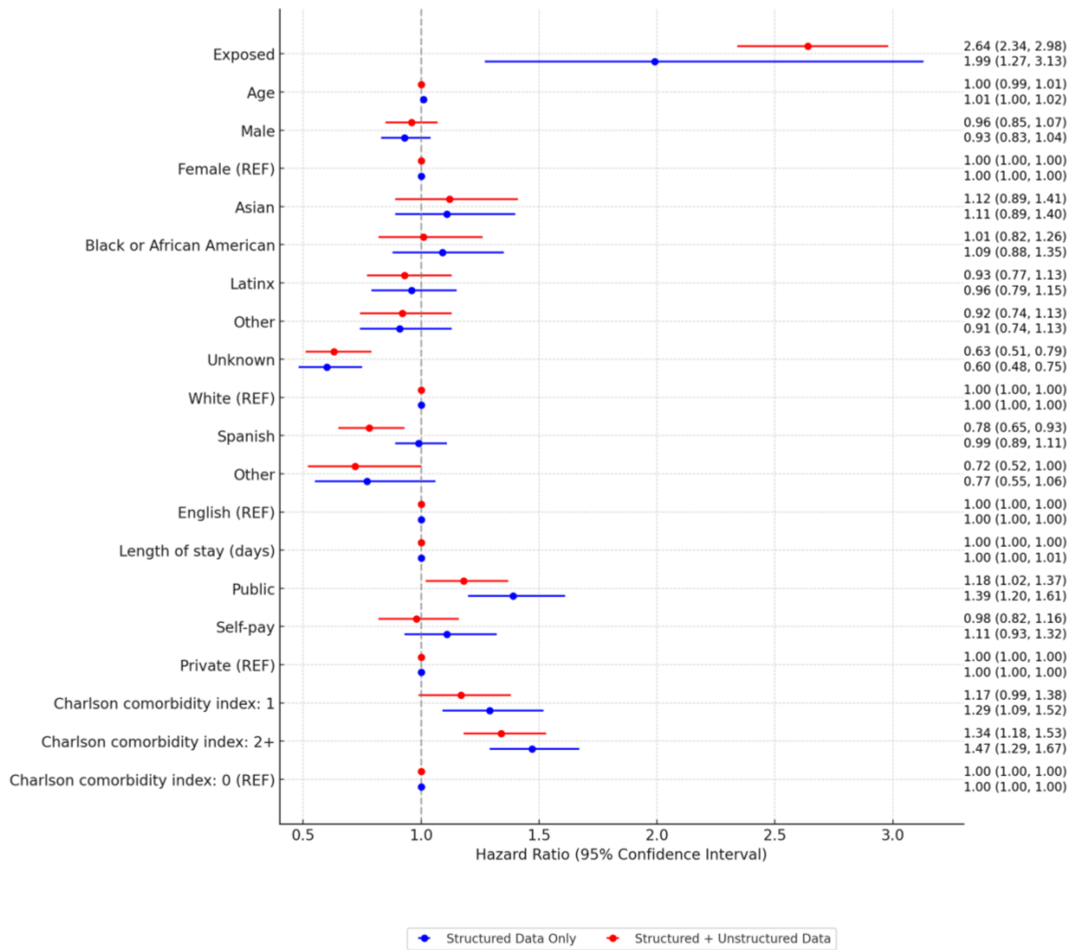

Figure 2.2 Hazard Ratios for Time to Unplanned Hospital Readmission by Exposure and Covariates Using Structured Data and Combined Structured and Unstructured Data**.** Abbreviations: REF, reference.

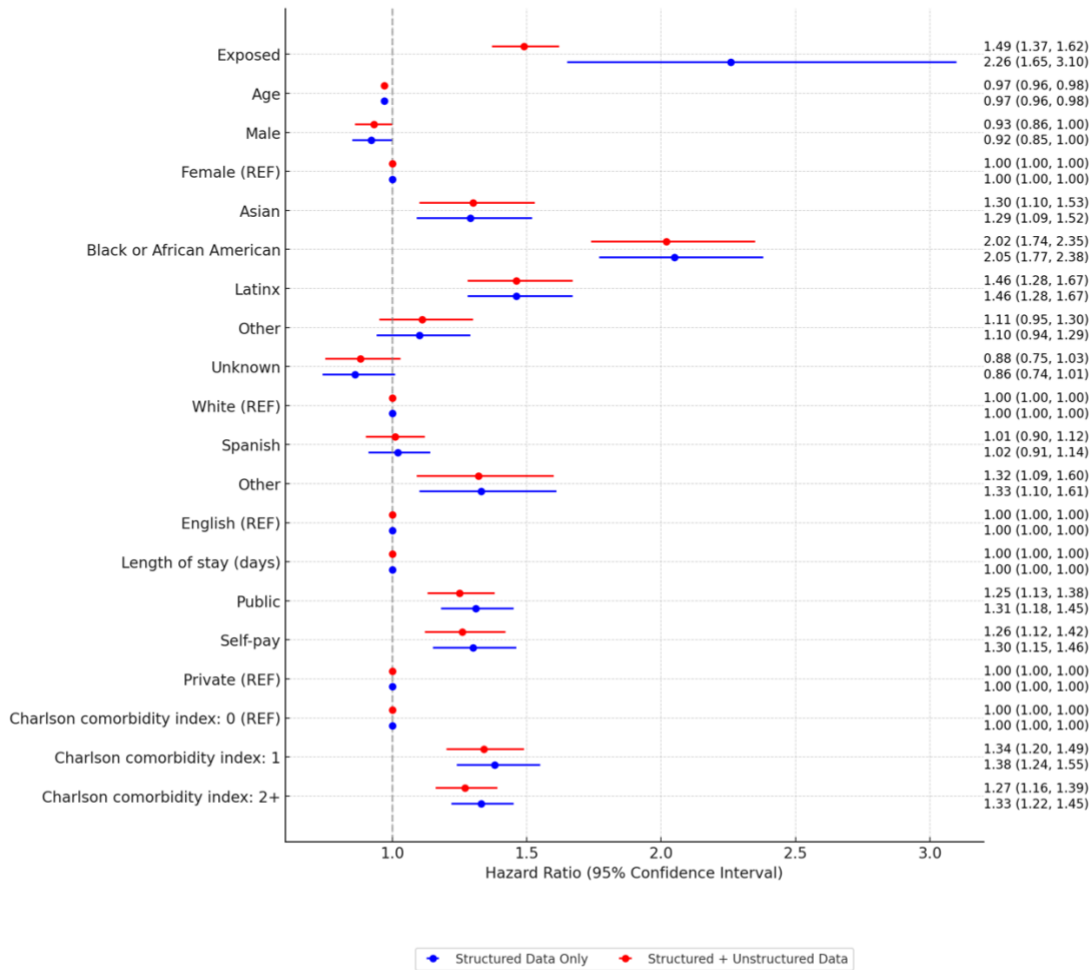

Figure 2.3 Hazard Ratios for Time to ED Readmission by Exposure and Covariates Using Structured Data and Combined Structured and Unstructured Data
Abbreviations: REF, reference.

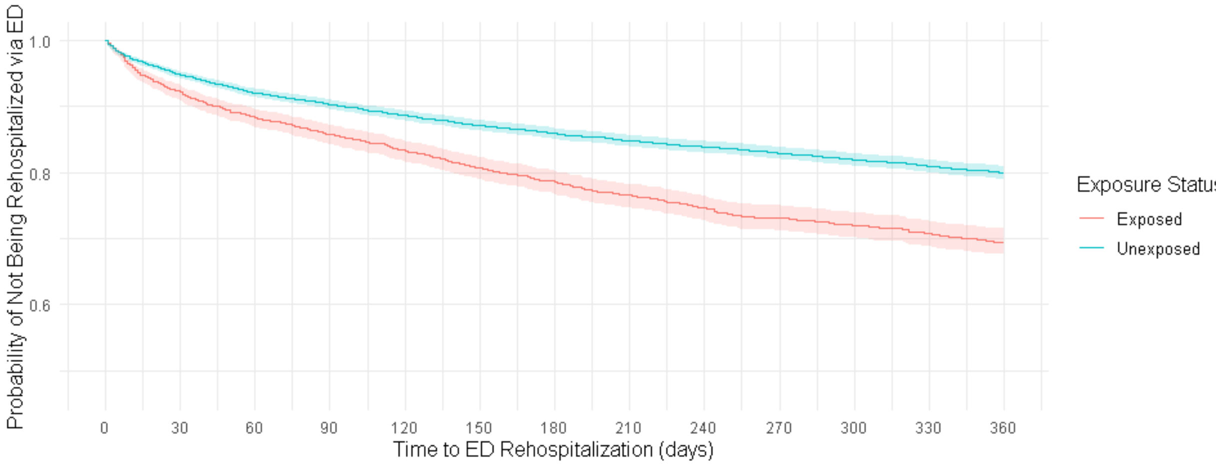

Figure 2.4 Kaplan-Meier Survival Curve for Time to ED Readmission Stratified by Exposure Group.

Table 2.1 Inclusion and exclusion criteria for cohort selection**.**

Inclusion criteria	Exclusion criteria
• Aged 2-17 years • First-time hospitalization discharge from inpatient pediatric wards at University of California, San Francisco Benioff Children's Hospital from January 2018 to January 2023	• Patients who were transferred to another care facility at discharge • Patient discharges who died on index admission (before pediatric discharge) • Patients with a prior hospital admission to an inpatient pediatric ward at BCH-SF

Table 2.2 Patient demographics.

	Structured data		**Unstructured and structured data**	
	Exposed (N=71)	Unexposed (N=8,857)	Exposed (N=2,827)	Unexposed (N=6,101)
Age, mean (SD)	10.50 (5.43)	9.44 (5.08)	10.46 (5.01)	8.98 (5.04)
Length of stay in days, mean (SD)	8.69 (37.49)	3.94 (9.41)	5.81 (16.20)	3.13 (4.60)
Sex				
Male	28 (39.44)	4,401 (49.69)	1,323 (46.8)	3,106 (50.9)
Female	43 (60.56)	4,452 (50.27)	1,502 (53.1)	2,993 (49.1)
Other	--	4 (0.05)	2 (0.07)	2 (0.03)
Race and ethnicity				
Asian	1 (1.41)	810 (9.15)	215 (7.6)	596 (9.8)
Black or African American	17 (23.94)	886 (10.00)	390 (13.8)	513 (8.4)
Latinx	28 (39.44)	2,960 (33.42)	1,090 (38.6)	1,898 (31.1)
White	3 (4.23)	1,579 (17.83)	409 (14.5)	1,173 (19.2)
Other	12 (16.90)	1,124 (12.69)	322 (11.4)	814 (13.3)
Unknown	10 (14.08)	1,498 (16.91)	401 (14.2)	1,107 (18.1)
Insurance				
Private	3 (4.23)	3,067 (34.63)	631 (22.3)	2,439 (40.0)
Public	46 (64.79)	3,780 (42.68)	1,501 (53.1)	2,325 (38.1)
Self-pay	22 (30.99)	2,010 (22.69)	695 (24.6)	1,337 (21.9)
Primary Language				
English	51 (71.8)	6,913 (78.1)	2,085 (73.8)	4,879 (80.0)
Spanish	16 (22.5)	1,603 (18.1)	622 (22.0)	997 (16.3)
Other	4 (5.6)	341 (3.9)	120 (4.2)	225 (3.7)

	Structured data		Unstructured and structured data	
	Exposed (N=71)	Unexposed (N=8,857)	Exposed (N=2,827)	Unexposed (N=6,101)
CCI categorical:				
0	39 (54.9)	5,407 (61.0)	1,489 (52.7)	3,956 (64.8)
1	15 (21.1)	1,190 (13.4)	472 (16.7)	734 (12.0)
2+	17 (23.9)	2,260 (25.5)	866 (30.6)	1411 (23.1)

Chapter 3: Artificial Intelligence for Natural Language Processing: Evaluating Transformer-Based Models and Prompt-Engineering Strategies for Social Determinants of Health Identification in Clinical Text of Patients with Opioid-Related Disorders

INTRODUCTION

Social determinants of health (SDOH) are pivotal in shaping both health outcomes and disparities, particularly in vulnerable populations such as those with opioid-related disorders (ORD).[31–34] These determinants encompass a wide range of socioeconomic and environmental factors, including housing stability, education level, socioeconomic status, access to essential services like transportation and healthcare, and more.[35] Given the strong link between mental health and social circumstances, this study also examined categories including generalized anxiety disorder, history of anxiety state, signs and symptoms of anxiety, major depressive disorder, and symptoms of depression as SDOH.[36–39] Recognized by the World Health Organization as primary contributors to health inequities, SDOH play a critical role in comprehensive health care planning and intervention, extending beyond medical treatment to address the underlying causes of health issues.[35,40]

For individuals affected by the opioid epidemic—a public health crisis that disproportionately impacts marginalized communities—the integration of SDOH into healthcare strategies is especially crucial.[41,42] These individuals often encounter numerous social risk factors that exacerbate their health conditions, making the identification and effective management of SDOH essential for improving health outcomes and reducing disparities.[43]

The challenge, however, lies in the extraction of SDOH data, predominantly contained within the unstructured text of clinical notes.[14] These notes are diverse in format and narrative style, filled with nuanced language that requires an in-depth understanding of medical context and terminology.[44] Traditional methods of SDOH data extraction, such as manual annotation, are labor-intensive and fraught with inconsistencies, and they do not scale well with large datasets.[14,44] Moreover, conventional rule-based or machine learning approaches often fall short in handling the ambiguity and contextual complexities presented by clinical narratives.[45] Hospitals and health systems are increasingly incorporating social determinants of health (SDOH) into their EHRs to enhance patient care.[46] Efforts to integrate SDOH data have been characterized by various approaches, including the development and use of specific frameworks and tools aimed at addressing individual and system-level determinants of health.[46] For example, some studies have identified structured initiatives where hospitals followed specific steps to integrate SDOH data into hospital care practices with EHRs.[46] Despite these efforts to document SDOH more efficiently, challenges in adherence remain significant.

The advent of advanced natural language processing (NLP) technologies, particularly transformer-based models like GPT-4 and LLaMA-3.2, offers new avenues for overcoming these challenges.[47] These models leverage deep learning architectures that process words in relation to all other words in a text, thereby excelling in generating and understanding complex, context-rich narratives. GPT-4, trained on a vast array of internet-based text, brings extensive domain knowledge and robust text generation capabilities to a variety of applications.[48] LLaMA-3.2, although newer, has shown potential to match or even exceed GPT-4's performance in tasks requiring specific domain expertise.[49]

This study employs the capabilities of GPT-4 and LLaMA-3.2 to automate the classification of SDOH from clinical notes within a cohort of patients diagnosed with opioid-related disorders. By harnessing these cutting-edge models, the research aims to enhance the accuracy, consistency, and scalability of SDOH data processing, significantly reducing the reliance on manual extraction and facilitating more effective integration of crucial social determinants into patient care and health policy development.

METHODS

Study Population

The study cohort consisted of patients diagnosed with ORDs, identified through EHR at two UCSF hospitals, part of a larger health system. Inclusion criteria were centered around specific ICD-10 diagnosis codes indicative of opioid-related disorders, notably those codes starting with 'F11'. The cohort was restricted to inpatient admissions, focusing solely on individuals who had been hospitalized due to ORDs between January 1, 2018, and December 31, 2019.

Social risk factors and mental health outcomes

This study utilized de-identified clinical note excerpts containing unstructured text relevant to SDOH. The study focused on several predefined categories of SDOH that are critical in assessing the social and environmental factors affecting individuals with ORDs. Additionally, recognizing the significant interplay between mental health and social factors, this study also examined specific categories, including generalized anxiety disorder, history of anxiety state, signs and symptoms of anxiety, major depressive disorder, and symptoms of depression. These categories were:

1. Anxiety: generalized anxiety disorder
2. Anxiety: history of anxiety state
3. Anxiety: signs and symptoms of anxiety
4. Depression: major depressive disorder
5. Depression: symptoms of depression
6. Housing: homeless
7. Marital or partnership status: divorced
8. Marital or partnership status: separated
9. Marital or partnership status: widowed
10. Social isolation: lives alone
11. Transportation: transportation problems

These categories were selected to encompass a broad spectrum of social conditions that are often intertwined with the health trajectories of individuals suffering from ORDs.

Prompt-engineering strategies

In our study, the prompt engineering process was vital to optimizing the performance of both GPT-4 and LLaMA-3.2 for classifying social determinants of health (SDOH) from clinical text. The process was iterative, beginning with basic prompt structures and gradually incorporating more sophisticated strategies to refine the models' responses.

Initially, our prompts were straightforward, asking the model to classify text into predefined SDOH categories. However, early results revealed that the models sometimes struggled with ambiguity in the text and frequently missed subtle cues critical for accurate classification. To address these issues, we revised the prompts to include more explicit instructions that defined the task context and the expected response format clearly. For instance, an early version of a prompt might simply have asked the model to "Identify the SDOH category for the following snippet." This was later expanded to include specific instructions such as "Classify the following medical snippet into one SDOH category from the list provided. Ensure your response is concise and directly correlates with evidence from the text."

We also incorporated elements that defined the models' roles more clearly, such as stating "You are a medical expert assistant tasked with analyzing clinical text to identify relevant social determinants of health." This approach helped to set a clear goal and appropriate tone for the interaction, making it more aligned with clinical reasoning.

Further refinements included the use of structured outputs and the introduction of 'chain of thought' reasoning, where the model was prompted to first explain its reasoning before providing a classification. This not only improved the transparency of the model's decision-making process but also enhanced the accuracy by ensuring that the classifications were well-founded.

Examples of prompt evolution include:

- Initial prompt: "Classify this clinical note."

- Revised prompt: "As a medical expert, analyze the following clinical snippet and determine the most relevant SDOH category. Provide a brief justification for your choice based on the content."

Each iteration of prompt refinement was tested, and improvements were noted in the models' ability to handle complex clinical narratives. For example, after introducing prompts that required justification, there was a noticeable improvement in precision and recall, particularly in distinguishing between closely related categories such as 'Anxiety' and 'Depression.'
These modifications, alongside our use of a structured prompt format that delineated sections for system role, user query, and expected model response, significantly enhanced the performance of the NLP models in our tests. This detailed approach to prompt engineering, emphasizing clarity, context specificity, and reasoning, was critical in leveraging the full capabilities of advanced transformer-based models for clinical text analysis.

Inference large language models (LLMs)

We utilized two advanced language models, GPT-4o and LLaMA-3.2, to classify social determinants of health (SDOH) from clinical notes using a zero-shot inference approach. This method leverages the inherent capabilities of the models without requiring additional training on domain-specific data

Prompt engineering was critical in guiding the models to perform the classification task accurately. For both GPT-4 and LLaMA-3.2, prompts were meticulously designed to include a clear description of the task—classifying snippets into specific SDOH categories defined a priori. Each prompt framed a system role that described the clinical snippet, followed by direct instructions to classify the snippet into one of the predefined categories. This structure helped maintain the models' focus within the defined boundaries, ensuring that the outputs were consistently relevant to the provided categories. Moreover, the prompts were designed to solicit not just a category label but also a rationale for the choice, which encouraged the models to engage in a deeper level of reasoning.

The aim was to elicit precise, contextually appropriate responses from the models. The classification process for each snippet was conducted through a single API call to the respective model. The response from each model was formatted to return the most relevant SDOH category based on the content of the snippet. This setup allowed us to evaluate and compare the effectiveness of GPT-4 and LLaMA-3.2 in identifying and categorizing key social determinants directly from text data, reflecting the models' potential to support healthcare professionals in recognizing and addressing the broader social factors impacting patient health. Results from this zero-shot classification were then parsed and analyzed to assess the accuracy, precision, recall, and F1 score of each model.

Model validation

In our study, the SDOH concepts extracted by the UCSF NLP pipeline were used as the ground truth. This validated pipeline is specifically designed to process unstructured clinical notes and reliably extract information on social risk factors.[21]

RESULTS

Sample characteristics

In this study, we analyzed clinical notes from a cohort comprising 925 unique patients. Each patient had at least one inpatient admission attributed to opioid related disorders (ORDs) within the specified period from January 1, 2018, to December 31, 2019. The average age at the first encounter for patients within the cohort was approximately 49.92 years, with a standard deviation of 15.77 years, indicating a wide age range from 18 to 85 years. Among the cohort showed a slight male predominance, with 502 male patients compared to 422 female patients (Table 3.1).

GPT-4o classification results analysis

In the study, the GPT-4o model demonstrated substantial proficiency in identifying and classifying snippets related to Housing and Transportation, with F1 scores of 0.94 and 0.96 respectively (Table 2). These results suggest that GPT-4o was highly effective at accurately determining categories that involve explicit SDOH factors, which are typically described in clear and concrete terms within clinical notes. In contrast, the model exhibited moderate difficulties with the Depression category, achieving a lower F1 score of 0.68 despite a relatively high recall of 0.73 (Table 3.2). This indicated that while GPT-4o could reliably identify relevant cases of Depression, it was less precise in categorizing them accurately, potentially due to overlapping symptoms with other mental health conditions.

The detailed-level evaluation revealed significant variances within the Anxiety sub-categories. The model adeptly recognized "Anxiety: Signs and symptoms of anxiety" with a notable recall of

0.80 but struggled with precision, resulting in an F1 score of 0.33 (Table 3.3). Conversely, GPT-4o showed considerable difficulty with "Anxiety: Generalized Anxiety Disorder," achieving a notably low F1 score of 0.13, highlighting challenges in differentiating between general anxiety, history of anxiety state, and signs and symptoms of anxiety (Table 3.3).

LLaMA-3.2 classification results analysis

LLaMA-3.2 mirrored GPT-4o in its exceptional accuracy in specific categories such as Housing and Marital or partnership status, with F1 scores of 0.97 and 0.95, respectively (Table 3.4). However, its performance varied significantly across broader mental health categories. For instance, in anxiety and depression, LLaMA-3.2 exhibited notably poorer performance, with F1 scores of 0.56 and 0.43, respectively. The model's high recall in depression, in particular, suggested a propensity to over-classify this condition, possibly due to the model's sensitivity to commonly used descriptors related to mood disorders.

The detailed analysis for LLaMA-3.2 exposed a stark incapacity to effectively categorize specific Anxiety conditions, with the model failing to recognize any valid cases in "Anxiety: history of anxiety state" and only identifying a minuscule fraction in "Anxiety: generalized anxiety disorder." This inefficacy pointed to a critical limitation in the model’s ability to parse and interpret the nuanced expressions of anxiety disorders (Table 3.5).

Error analysis and model performance

Both GPT-4o and LLaMA-3.2 demonstrated high accuracy in identifying and classifying SDOH related to Housing and Transportation, likely due to the explicit and clear language commonly used to describe these conditions in clinical notes. These categories achieved F1 scores nearing perfection, indicating strong model performance where the contextual cues are clear and less ambiguous.

However, both models displayed notable deficiencies in accurately distinguishing between mental health conditions, particularly between Anxiety and Depression. This issue was evident as both models sometimes incorrectly identified Depression as Anxiety and vice versa, reflecting a fundamental challenge in differentiating between these closely related mental health categories. The nuanced language of clinical notes, which often subtly conveys symptoms common to both conditions, likely contributed to these classification errors. For instance, symptoms like "feeling overwhelmed" or "lack of interest" can be indicative of either condition depending on additional contextual factors that the models occasionally misinterpreted.

DISCUSSION

In this investigation, we assessed the efficacy of transformer-based NLP models, specifically GPT-4o and LLaMA-3.2, to discern SDOH from clinical text. Utilizing a dataset of clinical notes from a cohort of patients with opioid-related disorders, this study showcased the models' capabilities and limitations in identifying SDOH-related information. This endeavor is particularly significant given the complex nature of SDOH and their profound impact on health outcomes, especially within populations struggling with opioid use disorders.

The study demonstrated that both models, GPT-4o and LLaMA-3.2, effectively identified explicit SDOH mentions, such as those related to housing and transportation, confirming their utility in extracting clear-cut SDOH data. For instance, the precision and recall for housing in GPT-4o were remarkably high, indicating a strong ability to recognize and classify clear SDOH references accurately. However, the performance varied significantly when dealing with more nuanced SDOH categories like anxiety and depression. Both models often confused these categories with one another, suggesting a challenge in distinguishing between related but distinct mental health conditions. This difficulty underscores a limitation in the models' ability to parse more complex and subtly expressed SDOH, where the language used may not explicitly outline the category, requiring a deeper contextual understanding.

Furthermore, error analysis indicated that while both models excelled in identifying straightforward SDOH mentions, they struggled with the ambiguity and context-dependent nature of language in clinical notes. For example, phrases that ambiguously suggested financial or transportation issues without explicit distress were sometimes incorrectly classified. This suggests that while the models are adept at recognizing overt SDOH mentions, their ability to infer based on broader context or less explicit cues remains an area for improvement.

For patients with ORDs, understanding SDOH is crucial because these factors heavily influence treatment outcomes and overall health. SDOH can affect access to healthcare services, adherence to treatment, and the risk of relapse. For instance, unstable housing or lack of transportation can hinder consistent access to treatment programs, which is critical for recovery from OUD. Similarly, mental health issues like anxiety and depression, which are prevalent among individuals with OUD, complicate the treatment landscape and are essential for comprehensive care planning. Recognizing and addressing these SDOH through advanced NLP models can significantly enhance personalized care strategies and improve health outcomes in this vulnerable population.

Our study has several strengths, first this study utilized a zero-shot inference approach, highlighting the models' ability to adapt to new tasks without specific prior training. This demonstrates the robustness and flexibility of transformer-based models in handling diverse and complex datasets. The detailed categorization of SDOH enabled a nuanced understanding of model performance across various domains, providing valuable insights into both model strengths and areas needing improvement. By leveraging the advanced capabilities of GPT-4 and LLaMA-3.2, the study addressed significant challenges in SDOH data extraction, enhancing the process's accuracy and scalability.

However, this study is not without limitations. First the models struggled to distinguish between closely related SDOH categories, particularly in mental health contexts, indicating a need for improved model training or finer categorization. The performance drop in contexts with ambiguous or implicit SDOH mentions suggests that further advancements in model training to handle such nuances are necessary. The study's findings are based on a dataset from a specific patient demographic, which may limit the generalizability of the results to other populations or settings.

Our research highlights the potential of NLP models, particularly GPT-4o and LLaMA-3.2, for identifying SDOH mentions in clinical notes. By leveraging advanced preprocessing, hyperparameter optimization, and error analysis, we have demonstrated the feasibility of using

NLP tools to automate the extraction of SDOH information from unstructured clinical text. Future work should focus on enhancing model specificity, expanding the annotated dataset to include a broader range of SDOH domains, and validating these findings in diverse clinical settings. These efforts will be critical to realizing the full potential of NLP technologies in supporting social risk factor screening and addressing health disparities in diverse populations. This study highlights the utility of transformer-based NLP models, particularly GPT-4o and LLaMA-3.2, in identifying and classifying SDOH among patients with ORDs. By implementing zero-shot learning techniques, we demonstrated these models' proficiency in accurately identifying explicit SDOH categories such as housing and transportation, which are pivotal in crafting targeted interventions for ORD patients. However, the models exhibited challenges in consistently recognizing nuanced mental health categories, which are frequently relevant to the ORD population. Such difficulties underscore the necessity for ongoing enhancements in model training to improve their sensitivity to the subtle linguistic cues associated with mental health. Integrating these advanced NLP tools into clinical practice could revolutionize the management of ORD by enabling more comprehensive assessments of patient needs, thereby informing more effective healthcare strategies and interventions. This research not only confirms the transformative potential of NLP in healthcare but also emphasizes the critical need for further advancements to ensure these tools can adeptly handle the complexities of clinical narratives, particularly in populations struggling with ORD and associated health disparities.

Table 3.1 Description of sample.

	Cohort (N=925)
Age, mean (SD)	49.92 (15.77)
Sex	
Male	502 (54.27)
Female	422 (45.62)
Other	1 (0.11)
Race and ethnicity	
Asian	27 (2.92)
Black or African American	117 (12.65)
Latinx	90 (9.73)
White	501 (54.16)
Other	90 (9.73)
Unknown	100 (10.81)

Table 3.2 High level results, GPT-4

Category	Precision	Recall	F1 Score
Anxiety	0.94	0.86	0.90
Depression	0.64	0.73	0.68
Housing	0.89	1.00	0.94
Marital or partnership status	0.82	0.97	0.89
Social isolation	0.55	0.97	0.70
Transportation	0.95	0.97	0.96

Table 3.3 Detailed-level results, GPT-4o.

Category	Precision	Recall	F1 Score
Anxiety: generalized anxiety disorder	0.54	0.07	0.13
Anxiety: signs and symptoms of anxiety	0.21	0.80	0.33
Anxiety: history of anxiety state	0.61	0.31	0.41
Depression: symptoms of depression	0.58	0.27	0.37
Housing: homeless	0.89	1.00	0.94
Marital or partnership status: divorced	0.98	0.96	0.97
Marital or partnership status: separated	0.80	0.94	0.86
Marital or partnership status: widowed	0.64	1.00	0.78
Social isolation: lives alone	0.46	0.66	0.54
Transportation: transportation problems	0.95	0.97	0.96

Table 3.4 High level results, LLaMA-3.2.

Category	Precision	Recall	F1 Score
Anxiety	0.98	0.39	0.56
Depression	0.28	0.95	0.43
Housing	0.97	0.96	0.97
Marital or partnership status	0.99	0.91	0.95
Social isolation	0.80	0.98	0.88
Transportation	0.79	1.00	0.88

Table 3.5 Detailed level results, LLaMA-3.2.

Category	Precision	Recall	F1 Score
Anxiety: generalized anxiety disorder	0.50	0.00	0.00
Anxiety: signs and symptoms of anxiety	0.20	0.53	0.29
Anxiety: history of anxiety state	0	0	0
Depression: symptoms of depression	0.30	0.94	0.45
Housing: homeless	0.98	0.96	0.97
Marital or partnership status: divorced	1.00	0.91	0.95
Marital or partnership status: separated	0.95	0.94	0.94
Marital or partnership status: widowed	1.00	0.89	0.94
Social isolation: lives alone	0.80	0.98	0.88
Transportation: transportation problems	0.79	1.00	0.88

References

1. Gultekin LE, Brush BL, Ginier E, Cordom A, Dowdell EB. Health risks and outcomes of homelessness in school-age children and youth: A scoping review of the literature. *J Sch Nurs*. 2020;36(1):10-18.

2. Cook JT, Frank DA, Berkowitz C, et al. Food Insecurity Is Associated with Adverse Health Outcomes among Human Infants and Toddlers. *J Nutr*. 2004;134(6):1432-1438. doi:10.1093/jn/134.6.1432

3. Sandoval VS, Jackson A, Saleeby E, Smith L, Schickedanz A. Associations Between Prenatal Food Insecurity and Prematurity, Pediatric Health Care Utilization, and Postnatal Social Needs. *Acad Pediatr*. 2021;21(3):455-461. doi:10.1016/j.acap.2020.11.020

4. Grant R, Gracy D, Goldsmith G, Sobelson M, Johnson D. Transportation Barriers to Child Health Care Access Remain After Health Reform. *JAMA Pediatr*. 2014;168(4):385-386. doi:10.1001/jamapediatrics.2013.4653

5. Gonzalez JV, Hartford EA, Moore J, Brown JC. Food Insecurity in a Pediatric Emergency Department and the Feasibility of Universal Screening. *West J Emerg Med*. 2021;22(6):1295-1300. doi:10.5811/westjem.2021.7.52519

6. Mendoza JA, Haaland W, D'Agostino RB, et al. Food Insecurity is Associated with High Risk Glycemic Control and Higher Health Care Utilization among Youth and Young Adults with Type 1 Diabetes. *Diabetes Res Clin Pract*. 2018;138:128-137. doi:10.1016/j.diabres.2018.01.035

7. Chang L, Stewart AM, Kester K, Routhier G, Michelson KA. Association of Homelessness With Emergency Department Use Among Children in New York. *JAMA Pediatr*. 2023;177(6):637-640. doi:10.1001/jamapediatrics.2023.0478

8. Syed ST, Gerber BS, Sharp LK. Traveling Towards Disease: Transportation Barriers to Health Care Access. *J Community Health*. 2013;38(5):976-993. doi:10.1007/s10900-013-9681-1

9. Coleman-Jensen A. Household Food Security in the United States in 2020. Published online 2020.

10. Lopez MA, Yu X, Hetrick R, et al. Social Needs Screening in Hospitalized Pediatric Patients: A Randomized Controlled Trial. *Hosp Pediatr*. 2023;13(2):95-114. doi:10.1542/hpeds.2022-006815

11. Fritz CQ, Thomas J, Gambino J, Torok M, Brittan MS. Prevalence of Social Risks on Inpatient Screening and Their Impact on Pediatric Care Use. *Hosp Pediatr*. 2020;10(10):859-866. doi:10.1542/hpeds.2020-0094

12. Vaz LE, Wagner DV, Ramsey KL, et al. Identification of Caregiver-Reported Social Risk Factors in Hospitalized Children. *Hosp Pediatr*. 2020;10(1):20-28. doi:10.1542/hpeds.2019-0206

13. Coker TR, Gottschlich EA, Burr WH, Lipkin PH. Early Childhood Screening Practices and Barriers: A National Survey of Primary Care Pediatricians. *Pediatrics*. 2024;154(2):e2023065552. doi:10.1542/peds.2023-065552

14. Gianfrancesco MA, Goldstein ND. A narrative review on the validity of electronic health record-based research in epidemiology. *BMC Med Res Methodol*. 2021;21(1):234. doi:10.1186/s12874-021-01416-5

15. Cottrell EK, Hendricks M, Dambrun K, et al. Comparison of Community-Level and Patient-Level Social Risk Data in a Network of Community Health Centers. *JAMA Netw Open*. 2020;3(10):e2016852. doi:10.1001/jamanetworkopen.2020.16852

16. Brown EM, Franklin SM, Ryan JL, et al. Assessing Area-Level Deprivation as a Proxy for Individual-Level Social Risks. *Am J Prev Med*. 2023;65(6):1163-1171. doi:10.1016/j.amepre.2023.06.006

17. Hatef E, Rouhizadeh M, Tia I, 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. doi:10.2196/13802

18. Han B, Volkow ND, Compton WM, McCance-Katz EF. Reported Heroin Use, Use Disorder, and Injection Among Adults in the United States, 2002-2018. *JAMA*. 2020;323(6):568-571. doi:10.1001/jama.2019.20844

19. Richie R, Ruiz VM, Han S, Shi L, Tsui F (Rich). Extracting social determinants of health events with transformer-based multitask, multilabel named entity recognition. *J Am Med Inform Assoc*. 2023;30(8):1379-1388. doi:10.1093/jamia/ocad046

20. 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 JAMIA*. 2021;28(12):2608-2616. doi:10.1093/jamia/ocab194

21. Lituiev DS, Lacar B, Pak S, Abramowitsch PL, De Marchis EH, Peterson TA. Automatic extraction of social determinants of health from medical notes of chronic lower back pain patients. *J Am Med Inform Assoc JAMIA*. 2023;30(8):1438-1447. doi:10.1093/jamia/ocad054

22. Berry JG, Toomey SL, Zaslavsky AM, et al. Pediatric readmission prevalence and variability across hospitals. *JAMA*. 2013;309(4):372-380. doi:10.1001/jama.2012.188351

23. Hain PD, Gay JC, Berutti TW, Whitney GM, Wang W, Saville BR. Preventability of Early Readmissions at a Children's Hospital. *Pediatrics*. 2013;131(1):e171-e181. doi:10.1542/peds.2012-0820

24. Lax Y, Martinez M, Brown NM. Social Determinants of Health and Hospital Readmission. *Pediatrics*. 2017;140(5):e20171427. doi:10.1542/peds.2017-1427

25. Fieldston ES, Zaniletti I, Hall M, et al. Community Household Income and Resource Utilization for Common Inpatient Pediatric Conditions. *Pediatrics*. 2013;132(6):e1592-e1601. doi:10.1542/peds.2013-0619

26. McKay S, Parente V. Health Disparities in the Hospitalized Child. *Hosp Pediatr*. 2019;9(5):317-325. doi:10.1542/hpeds.2018-0223

27. COUNCIL ON COMMUNITY PEDIATRICS, Briggs MA, Granado-Villar DC, et al. Providing Care for Children and Adolescents Facing Homelessness and Housing Insecurity. *Pediatrics*. 2013;131(6):1206-1210. doi:10.1542/peds.2013-0645

28. Patra BG, Sharma MM, Vekaria V, 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):2716-2727. doi:10.1093/jamia/ocab170

29. Navathe AS, Zhong F, Lei VJ, et al. Hospital Readmission and Social Risk Factors Identified from Physician Notes. *Health Serv Res*. 2018;53(2):1110-1136. doi:10.1111/1475-6773.12670

30. Lituiev DS, Lacar B, Pak S, Abramowitsch PL, De Marchis EH, Peterson TA. Automatic extraction of social determinants of health from medical notes of chronic lower back pain patients. *J Am Med Inform Assoc*. 2023;30(8):1438-1447. doi:10.1093/jamia/ocad054

31. Sulley S, Ndanga M. Inpatient Opioid Use Disorder and Social Determinants of Health: A Nationwide Analysis of the National Inpatient Sample (2012-2014 and 2016-2017). *Cureus*. 12(11):e11311. doi:10.7759/cureus.11311

32. Nicholson Jr. HL. Social determinants of opioid use behaviors among adults in the United States. *J Subst Use*. 2023;28(3):371-378. doi:10.1080/14659891.2022.2050317

33. Radic M, Parlier-Ahmad AB, Wills B, Martin CE. Social determinants of health and emergency department utilization among adults receiving buprenorphine for opioid use disorder. *Drug Alcohol Depend Rep*. 2022;3:100062. doi:10.1016/j.dadr.2022.100062

34. Do VM, Simpson S, Fisch KM, Gabriel RA. Associations Between Social Determinants of Health and Opioid-Use Disorder Among Chronic Pain Patients From a Multi-Institutional Dataset. *Anesth Analg*. Published online March 18, 2022:10.1213/ANE.0000000000007247. doi:10.1213/ANE.0000000000007247

35. Marmot M, Wilkinson R. *Social Determinants of Health*. OUP Oxford; 2005.

36. Kirkbride JB, Anglin DM, Colman I, et al. The social determinants of mental health and disorder: evidence, prevention and recommendations. *World Psychiatry*. 2024;23(1):58-90. doi:10.1002/wps.21160

37. Bukten A, Skjærvø I, Stavseth MR. Exploring mental health comorbidities and opioid agonist treatment coverage among people in prison: A national cohort study 2010–2019. *Drug Alcohol Depend*. 2023;250:110896. doi:10.1016/j.drugalcdep.2023.110896

38. Santo T, Campbell G, Gisev N, et al. Prevalence of mental disorders among people with opioid use disorder: A systematic review and meta-analysis. *Drug Alcohol Depend*. 2022;238:109551. doi:10.1016/j.drugalcdep.2022.109551

39. Davis MA, Lin LA, Liu H, Sites BD. Prescription Opioid Use among Adults with Mental Health Disorders in the United States. *J Am Board Fam Med*. 2017;30(4):407-417. doi:10.3122/jabfm.2017.04.170112

40. Health WC on SD of, Organization WH. *Closing the Gap in a Generation: Health Equity Through Action on the Social Determinants of Health : Commission on Social Determinants of Health Final Report*. World Health Organization; 2008.

41. Park JN, Rouhani S, Beletsky L, Vincent L, Saloner B, Sherman SG. Situating the Continuum of Overdose Risk in the Social Determinants of Health: A New Conceptual Framework. *Milbank Q*. 2020;98(3):700-746. doi:10.1111/1468-0009.12470

42. Roberts DL. Reframing the response to the opioid crisis: The critical role of resilience in public health. *Open Health*. 2024;5(1). doi:10.1515/ohe-2023-0006

43. Altekruse SF, Cosgrove CM, Altekruse WC, Jenkins RA, Blanco C. Socioeconomic risk factors for fatal opioid overdoses in the United States: Findings from the Mortality Disparities in American Communities Study (MDAC). *PLoS ONE*. 2020;15(1):e0227966. doi:10.1371/journal.pone.0227966

44. Ghiasvand O, Kate RJ. Biomedical Named Entity Recognition with less Supervision. In: *2015 International Conference on Healthcare Informatics*. ; 2015:495-495. doi:10.1109/ICHI.2015.85

45. van Es B, Reteig LC, Tan SC, et al. Negation detection in Dutch clinical texts: an evaluation of rule-based and machine learning methods. *BMC Bioinformatics*. 2023;24(1):10. doi:10.1186/s12859-022-05130-x

46. Rangachari P, Thapa A, Sherpa DL, et al. Characteristics of hospital and health system initiatives to address social determinants of health in the United States: a scoping review of the peer-reviewed literature. *Front Public Health*. 2024;12:1413205. doi:10.3389/fpubh.2024.1413205

47. Zarfati M, Nadkarni GN, Glicksberg BS, et al. Exploring the Role of Large Language Models in Melanoma: A Systematic Review. *J Clin Med*. 2024;13(23):7480. doi:10.3390/jcm13237480

48. OpenAI, Achiam J, Adler S, et al. GPT-4 Technical Report. Published online March 4, 2024. doi:10.48550/arXiv.2303.08774

49. Touvron H, Lavril T, Izacard G, et al. LLaMA: Open and Efficient Foundation Language Models. Published online February 27, 2023. doi:10.48550/arXiv.2302.13971

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:
Shivani Mehta
9A7D4F6098E94C9...

Author Signature

5/14/2025

Date