Predictive Modeling for 30-Day Emergency Room Visits among Sickle Cell Patients and Identifying Actionable Patient-Level Characteristics Using a Large Language Model

Sonali Tamhankar PhD, Marie C Vrablik MD MCR, Kleber Yotsumoto Fertrin, MD, PhD, M. Kennedy Hall MD MHS

Sickle Cell Disease (SCD) poses significant challenges for patients and healthcare providers, often leading to frequent emergency department (ED) visits. Seattle and Fred Hutchinson Cancer Center has created a patient-centered home for patients with SCD with a goal of addressing pain crises in the outpatient setting. Yet, many patients require ED management for breakthrough pain. We hypothesize that quality assurance (QA) can be enhanced using a novel QA methodology incorporating machine learning and LLMs to identify patients at high risk for 30-day ED visits and gain actionable insights into their individual risk factors.

The project was completed in two steps: first we employed a random forest algorithm, to develop a risk model using traditional structured covariates such as comorbidities, lab values, medications, number of ED visits among a cohort of 136 adult SCD patients from October 2022 until October 2023, 120 of whom were of Black ethnicity. Shapley values were utilized to understand the individual risk factors associated with ED visits. In a second step, we leveraged unstructured data through retrieval augmented generation technique with a vector store constructed from patients' clinical notes, utilizing the open-source large language model llama-2 by Meta to examine novel clinical characteristics associated with ED use. These clinical characteristics were presented to clinicians for judgement as a pilot proof of concept. All data and processing remained in a closed and HIPPA secure environment and transparency was enhanced by providing source excerpts from clinical notes.

Our random forest model performed well despite a small sample size, with an area under the receiver operator characteristic curve of 0.85. The patient specific risk factors generated by the structured data and the ability to interact with and extract novel information from clinical notes were highly valued in an initial ED Director and SSD provider review.

We demonstrate a proof-of-concept, in-house end-to-end QA tool using a 2-step approach can identify novel characteristics among a vulnerable SCD population. The incorporation of patient-specific risk factors and interpretability features the ability for healthcare provider to provide targeted and novel supportive interventions in the outpatient environment.

A screenshot of a computer

Description automatically generated

A screenshot of a medical survey

Description automatically generated