



## The 65th ASH Annual Meeting Abstracts

## ONLINE PUBLICATION ONLY

**114. SICKLE CELL DISEASE, SICKLE CELL TRAIT AND OTHER HEMOGLOBINOPATHIES, EXCLUDING THALASSEMIA: CLINICAL AND EPIDEMIOLOGICAL****Characterization of Sickle Cell Disease Mortality in a National Electronic Health Record (EHR) System**

Gerard Hills, MD<sup>1</sup>, Kristin Wuichet<sup>2</sup>, Harlan Nelson, MS<sup>3</sup>, Robert Cronin, MDMS<sup>4</sup>, Shaun Grannis, MDMS<sup>5,6</sup>, Michael R. DeBaun, MD MPH<sup>7</sup>

<sup>1</sup>Department of Pediatrics / Division Pediatric Hematology/Oncology, Indiana University School of Medicine / Regenstrief Institute, Indianapolis, IN

<sup>2</sup>Vanderbilt University Medical Center, Nashville, TN

<sup>3</sup>Indiana University Health, Indianapolis, IN

<sup>4</sup>Department of Internal Medicine, The Ohio State University, Columbus, OH

<sup>5</sup>Department of Family Medicine, Indiana University School of Medicine, Indianapolis, IN

<sup>6</sup>Center of Biomedical Informatics, Regenstrief Institute, Indianapolis, IN

<sup>7</sup>Vanderbilt University School of Medicine, Nashville, TN

**Introduction**

Sickle cell disease (SCD) is a chronic condition with significant morbidity in children and early adult mortality. Identifying high-risk SCD populations has been challenging due to limited population cohorts of children and adults. Understanding the clinical course of SCD in both children and adults is crucial for informed treatment decisions around disease-modifying therapy vs. curative and transformative therapies. Recent use of Medicaid, Medicare, and private insurer claims datasets has helped define high-risk populations and estimate mortality, but they have limitations in clinical record depth. In contrast, electronic health record (EHR)-based datasets offer a better option for assessing the impact of disease-modifying therapy and allow comprehensive mortality risk prediction analyses, using high-dimensional laboratory data.

**Objective**

Previously, using the data warehouse at a single institution (Vanderbilt University School of Medicine) with data collected over 18 years, members of our team developed a contemporaneous cohort of children and adults with SCD. To do this our team developed a dual-approach algorithm that combined patients categorically selected by diagnosis using ICD-10 (International Classification of Diseases -10) codes and by laboratory data including hemoglobin fractionation values. The team estimated the median survival in adults with SCD and predicted mortality risk based on disease-modifying therapy and insurance status. We expanded this work by testing the hypothesis that we can first apply the laboratory portion of the Vanderbilt algorithm to identify children and adults with SCD living in the United States from the national Cerner Learning Health Network (Cerner LHN) EHR dataset.

**Methods**

The Cerner LHN is a nationwide dataset comprised of 88 million patients from over 74 hospital systems across 37 states (as of September 17, 2021). We used data from the Cerner LHN collected between 8/26/2002 and 12/13/2022 for our analysis. We extracted laboratory data elements required for the Vanderbilt algorithm including hemoglobin fractionation data with HbS values as well as total hemoglobin timepoints and blood transfusion time frames. The results of the laboratory portion of the Vanderbilt algorithm predict genotype of sickle cell disease in the following categories: SCA (HbSS or HbSB<sup>0</sup>), SC (Hb SC), HbSBeta<sup>+</sup> and HbSE. There were 5,827 patients identified with these laboratory components necessary for the algorithm. We applied an updated python script that detailed the logic from the Vanderbilt algorithm on the 5,827 patients and we identified 3,248 children and adults with SCD. Patients were then evaluated for demographics and hospital systems coverage within the Cerner LHN.

**Results**

Black/African Americans made up 2947 patients (90.7%) of the SCD cohort. There were 1325 (54%) male and 1150 (46.5%) female patients with SCA within the phenotyped cohort. Patients with SCA were collected from 44.6% of the Cerner LHN health systems.

**Conclusions**

The population was 90.7% Black/African American as expected in patients carrying a diagnosis of SCD. This contemporaneous national cohort represents 44.6% of the health systems within the Cerner LHN which denoted wide-reaching patient sampling. These findings show that the laboratory portion of the Vanderbilt SCD algorithm can, in an automated fashion, establish a phenotyped SCD cohort of children and adults in a national EHR system. We plan to power the total number by adding the ICD-10 portion of the algorithm and estimating mortality in the final cohort. Further, we will validate combined algorithm by chart review.

**Disclosures DeBaun:** *FORMA*: Consultancy; *Novartis Pharmaceuticals Corporation*: Other: Study steering committee member; *Global Blood Therapeutics*: Membership on an entity's Board of Directors or advisory committees, Research Funding; *Vertex/CRISPR*: Membership on an entity's Board of Directors or advisory committees.

<b>Table 1: Characteristics of SCD Lab Phenotyped Cohort from National EHR Population Dataset</b>				
Characteristic	SCA (HbSS or HbSB0) N=2475	SC (HbSC) N=751	Hb SBeta+ N=20	Hb SE N=2
Female	1150 (46.5%)	377 (50%)	10 (50%)	0 (0%)
Male	1325 (54%)	373 (50%)	10 (50)	2 (100%)
Age Range (years)	0.01 - 70	0.01 - 77.8	2 - 81.5	2.8 – 31.5
Black / African American	2080 (84%)	674 (90%)	17 (85%)	1 (50%)
White	90 (3.6%)	12 (1.6%)	2 (10%)	0 (0%)
Asian	12 (0.5%)	1 (0.1%)	0 (0%)	0 (0%)
Indigenous (North American)	2 (<0.1%)	1 (0.1%)	0 (0%)	0 (0%)
Middle Eastern	1 (<0.1%)	0 (0%)	0 (0%)	0 (0%)
Unknown	290 (11.7%)	63 (8.4%)	1 (5%)	1 (50%)
Hispanic	202 (8.2%)	35 (4.7%)	1 (5.0%)	0.0%
Non-Hispanic	2273 (91.8%)	716 (95%)	19 (95%)	2 (100%)
Number of Hospital Systems Represented out of 74 systems	33 (44.6%)	31 (41.9%)	2 (2.7%)	2 (2.7%)

**Figure 1**

<https://doi.org/10.1182/blood-2023-190257>