ASH RESEARCH COLLABORATIVE®

Accelerating Progress in Hematology

Sickle Cell Disease

Data Hub Report 2025

ABOUT THE ASH RESEARCH COLLABORATIVE

The ASH Research Collaborative® (ASH RC) is a non-profit organization established by the American Society of Hematology (ASH) to improve the lives of people affected by blood diseases by fostering collaborative partnerships that accelerate progress in hematology. The ASH RC comprises disease-based research networks that conduct collaborative research and are powered by a real-world data (RWD) program called the Data Hub. The Data Hub captures Health Insurance Portability and Accountability Act (HIPAA)-compliant data on individuals with sickle cell disease (SCD) and is a state-of-the-art data-sharing platform. The Data Hub SCD data are aggregated to enable researchers and pharmaceutical companies to quickly and efficiently answer research questions, develop new treatments for individuals suffering from hematologic conditions, and maximize the value of those treatments post-approval.

ABOUT THE SCD DATA HUB PROGRAM

The SCD Data Hub is a comprehensive, real-world data resource designed to advance research and enhance care for individuals with SCD. By aggregating high quality clinical data from diverse sources, the Data Hub supports the generation/testing of research hypotheses, assessment of best practices, identification of meaningful and quantifiable clinical trial endpoints, and characterization of the patient population specifically within the Data Hub. Through strategic collaboration, it serves as a vital tool for pursuing better treatments for those living with SCD.

SUGGESTED CITATION

ASH Research Collaborative SCD Data Hub
2025 Data Hub Report
Washington, DC

SPECIAL ACKNOWLEDGEMENTS

This report was made possible through the collaborative efforts of numerous partners across the ASH RC Research Network and the SCD community. We would like to gratefully acknowledge the following people:

- Individuals living with SCD, as well as their families, for providing advice during the creation of the SCD Data Hub
- Reviewers for this report: Biree Andemariam, MD, Betty S. Pace, MD, Alexis A. Thompson, MD, MPH, William A. Wood, MD, MPH, Charles S. Abrams, MD
- Programmatic and statistical support: Travis Cleaves, BS, ADN, RN, CCM, Michael W. Craige, PhD, MBA, Aliza Fink, DSc, MS, Nicole Mortier, MHS, PA-C, Susan Rogers, MSN, RN, Emily Semmel, MS, Sam Walters, MBA, Kathleen Torres, DNP

We are especially thankful to the SCD **Data Hub sites that contributed data for these analyses:**

Baylor College of Medicine/Texas Children's Hospital
Boston Medical Center
Children's Medical Center, Dallas
Children's Hospital of Philadelphia
Cincinnati Children's Hospital Medical Center
Emory University School of Medicine
Johns Hopkins University School of Medicine
Medical College of Wisconsin Pediatric/Adult
Montefiore Medical Center
Parkland Health & Hospital System
UCSF Benioff Children's Hospital of Oakland
University of Alabama at Birmingham
University of Illinois at Chicago
University of Texas Southwestern Medical Center

Dear colleagues and friends,

We are pleased to share the **2025 Sickle Cell Disease Data Hub Report**, the first of its kind from the ASH Research Collaborative's Sickle Cell Disease (SCD) Data Hub Program.

In this inaugural report, we provide insights drawn from over 10,000 individuals' data submitted to the SCD Data Hub with a **physician-attested** diagnosis of SCD. These data represent more than 45 million clinical data points collected from 2015 to 2024. This resource captures real-world experiences across diverse care settings, providing an initial understanding of the disease burden, treatment patterns, and health outcomes of individuals with SCD treated across 17 sites who submitted data to the Data Hub.

While this report does not fully represent the clinical practices and patient outcomes for *all* individuals living in the United States today, I think you will find it helpful to see the breadth of the data captured by the Data Hub program and consider how they will inform and inspire impactful research as this important program continues to grow.

Further, these data are just the start; they serve as a prominent foundation for new discoveries, reflect treatment variations, and illustrate that continued work is needed to ensure that electronic health record data are of high quality. We are deeply grateful to the individuals living with SCD, their families, care teams, and the staff at participating Data Hub sites, who have made this first report possible. Their contributions shape a future where real-world data can provide meaningful insights for research, policy, and patient-centered care.

We invite you to review the full report, reflect on the insights, and join us in asking, "What questions can we now answer together?"

Sincerely,

Charles Abrams

Charles Abrams, MD
Chair, Sickle Cell Disease Research Network
ASH Research Collaborative

Table of Contents

List of Tables

List of Figures

Introduction

Overview of the SCD Data Hub Program

The Sickle Cell Disease (SCD) Data Hub is a real-world data platform designed to enhance research, therapeutic development, and care delivery for people with SCD. Through partnerships with clinical sites, researchers, patients, and other stakeholders, the Data Hub aggregates and standardizes electronic health record (EHR) data with the following goals:

- To enable the longitudinal analysis of individuals living with SCD to better understand the natural history of the disease
- To support comparative effectiveness and outcomes research
- To facilitate the development of computable phenotypes and real-world endpoints
- To provide a national resource for evaluating the associated demographics, disease severity, comorbidities, prescribing patterns, guideline adherence, disease natural history, healthcare system utilization, and opportunities for structural improvements to enhance care availability and quality

Background on Sickle Cell Disease

SCD is a genetic blood disorder characterized by chronic hemolytic anemia, recurrent painful episodes, and progressive organ damage. It affects approximately 100,000 individuals in the United States, and despite advances in treatment, SCD remains associated with substantial morbidity, early mortality, and high levels of healthcare utilization. (Centers for Disease Control and Prevention, 2023)

Key Clinical and Research Focus Areas

The clinical and research focus areas of the Data Hub include the following:

- Aligning with the SCD Research Priorities 2024 update published by the American Society of Hematology (ASH)
- Tracking the use and real-world effectiveness of disease-modifying therapy
- Understanding the progression of SCD-related comorbidities (e.g., neurological complications, acute chest syndrome, renal disease, etc.)
- Evaluating treatment patterns and site-specific variations in care
- Capturing longitudinal outcomes to support evidence-based care guidelines
- Supporting the identification of objective, quantifiable, and clinically meaningful therapeutic research trial endpoints

Data Hub Methodology

Overview

The Data Hub is a multi-center data repository that aggregates and harmonizes real-world data of individuals suffering from hematologic diseases, including SCD. All data in the Data Hub represent electronic health record (EHR) data of patients receiving care at participating sites.

The participating sites are academic medical centers where experienced hematologists or internists provide outpatient longitudinal care. They submit data quarterly, which includes demographic information, encounter records, laboratory and clinical results, and medication prescriptions.

Submissions undergo validation checks for completeness and consistency before being integrated into the central Data Hub.

Data Hub Sites: Roles and Responsibilities

Participating Data Hub sites have the following responsibilities:

- Submitting quarterly EHR data to the Data Hub for patients with an EHR-documented SCD diagnosis (see Appendix A)
    - The data are submitted through one of the following methods, depending on the site's local infrastructure:
        - Observational Medical Outcomes Partnership (OMOP): Sites extract and transform data into the OMOP Clinical Data Model (CDM) format. The files are then securely transferred via a secure file transfer protocol (SFTP) to the Data Hub's staging environment, where they undergo schema and content validation before being incorporated into the Data Hub.
        - Health Level Seven Fast Healthcare Interoperability Resources (HL7 FHIR): Sites may also elect to submit data by conforming to HL7 FHIR R4. FHIR data are securely accessed by the Data Hub via facility-approved application programming interfaces (APIs). These APIs are typically designed to comply with United States Federal regulations for health data interoperability (see 45 CFR 170, 171), which aim to enable direct access to electronic health information for patients and healthcare providers.
- Attesting SCD diagnoses (including genotype or hemoglobin phenotype) to improve diagnostic accuracy
- Participating in data validation activities, such as attesting SCD diagnoses and data audits, to ensure data accuracy

To date, 17 clinical sites have contributed data, and 14 sites have submitted physician-attested diagnoses for inclusion in this report.

Human Subjects Protections, Institutional Review Board Framework, and Confidentiality

The ASH RC Data Hub operates under a central-umbrella protocol that guides ethical data collection, management, and sharing. Key elements include those listed below:

- **Institutional Review Board (IRB) Oversight**: Participating institutions must comply with both local and central IRB requirements.
- **Informed Consent**: The patient (or their guardian) may provide informed consent for the use of data that can directly identify them. Until consent is obtained, sites submit HIPAA-limited datasets that include only the patient's dates and zip codes. A limited data set does not include direct identifiers such as names or medical record numbers.
- **Withdrawal of Consent**: Participants may withdraw their consent at any time. No further data are transmitted post-withdrawal, and data can be deleted upon request.
- **Data Privacy**: All data are encrypted (AES-256) and transmitted via secure protocols (SFTP, HTTPS, and SSL). IQVIA, ASH RC's technology partner, processes and stores all data using strict access control and logging procedures.

Patient Inclusion Criteria, Exclusion Criteria, and Physician-Attested SCD Diagnosis

Patients are eligible for inclusion if they fulfill the following criteria:

- They have a documented diagnosis code of SCD in their EHR, as identified by the site's EHR system using a validated list of diagnostic codes (see Appendix A).
- They are receiving or have received care at a participating site.

Once the data are submitted, physicians from the participating sites further attest to the SCD diagnosis of each patient identified via an EHR diagnostic code (see Appendix A).

About the Data in this Report

The Data Hub is a valuable resource for advancing our understanding of SCD, evaluating and standardizing care, and supporting clinical trials. To ensure valid research, stakeholders must understand the data it contains and its limitations. This section provides an overview of the Data Hub's contents and key considerations for analysis.

Data Submission, Patient Identification, and Physician-Attested SCD Diagnosis

All the data in this report are from sites that submit to the Data Hub using the OMOP submission method. The sites were provided with a list of International Classification of Diseases (ICD) and Systematized Nomenclature of Medicine (SNOMED) codes to run through their electronic health systems to identify individuals who may have SCD (see Appendix A). This list is designed to be highly sensitive, enabling the algorithm to identify all potential cases of SCD; however, it may consequently have lower specificity.

There are known challenges associated with reliably identifying SCD diagnoses and the specific SCD type (e.g., hemoglobin SS) utilizing the EHR system data alone. **To improve the validity and accuracy of the SCD diagnosis type of the individuals included in the Data Hub, a specific process has been implemented for each site to submit physician-attested SCD diagnosis types. Most of the data presented in this report focuses on ONLY patients who have a physician-attested SCD diagnosis.**

Work is ongoing to refine the Data Hub's case-finding algorithm, and the validation of computable phenotypes based on combinations of the ICD, SNOMED, LOINC and other coding categories with high specificity for identifying SCD cases. These activities will help streamline and improve efficiency in the attestation process.

Identifying Unique Individuals

The data in the Data Hub represents a HIPAA-limited data set that includes dates and zip codes. Each individual included in the Data Hub is assigned a random, anonymized identification number. Steps are in place to prevent duplicate entries as part of the individual site submission process. Therefore, within each site, all individuals included are unique. At this time patient data is not linked across sites, in instances where a patient has received care at more than one site participating in the Data Hub. For example, if an individual in the Data Hub transitions from a pediatric to adult site or moves and receives care at a different site, they will be included in the Data Hub with a separate identification number from each of the sites where they have received care and may be represented in the platform more than once.

The total number of individuals included in the Data Hub will increase as new participating sites begin submitting data for patients with a physician-attested SCD diagnosis and as individuals begin receiving care at participating sites.

Timeframe

When new participating Data Hub sites are onboarded, they are asked to submit EHR data dating back to 2015. This report includes data submissions received up to December 31, 2024.

Records were included in this report if the event date (encounter, laboratory measurement, or prescription) **occurred between January 1, 2015, and December 31, 2024.** When data are presented for a single calendar year, they include all individuals that have at least one record with an occurrence date between January 1st and December 31st.

Due to processing time, there can be a delay of three to four months before the EHR data are available for submission to the Data Hub. **Therefore, the data for 2024 included in this report do not represent the whole year and are not considered final. When reporting data from the most recent year, 2023 was used, as it was the most current year with complete data.**

Once a data submission is received, extensive data processing and quality-control assessments are performed. Data submissions received at the end of the first quarter of 2025, including additional data from 2024, are currently undergoing validation and analysis. Therefore, these data were not included in this report.

Variable Definitions

Throughout the report, the term "Active" defines an individual's participation in the Data Hub during a given calendar year and overall. Further, "Ever Active" entails that the individual had at least one inpatient, emergency department, or outpatient encounter recorded in the Data Hub between 2015 and 2024. To be considered active in a specific calendar year, the individual must have one or more records of an inpatient, emergency department, or outpatient encounter during the calendar year. This definition of Active is used to define denominators for data reporting and research and may not reflect clinical definitions of the term "Active." Current limitations related to characterizing outpatient encounters prevent the specific tracking and utilization of outpatient hematology visits to define active hematology follow-ups.

The Data Hub contains a data set that lists all individuals with one record for each person. For each person, the file includes their date of birth, sex, race, ethnicity, Data Hub site, state of residence and zip code, and vital status. Every time a site submits quarterly data to the Data Hub, each individual's demographic data are overwritten to reflect the most recent information in the electronic health record. Currently, the Data Hub does not track changes in the recorded demographic information over time.

This report presents data showing the distribution of individuals in the Data Hub by age, sex, race, ethnicity, and de-identified Data Hub site.

Age: Throughout the report, some data are stratified by age, as described within the individual data tables. To compare pediatric and adult populations, those who were 18 years or older were considered adults. In some cases, alternative age group categories are presented; for example, when examining age distribution across the different Data Hub sites, age was broken down into three groups: less than 18 years, 18–26 years (to capture the period of time during which many individuals transition from pediatric to adult care), and 26 years and older. When examining the distribution of reported comorbidities and prescription records, two pediatric and two adult subcategories are used (<12 years and 12–17 years for pediatrics and 18–30 years and >30 years for adults).

Location: In this report, the distribution of individuals is reported based on the location of the site from which their data was submitted. Work is ongoing to review the completeness and accuracy of the state of residence and zip code reported in the EHR. Based on the findings, it will be determined whether it is appropriate to include this level of data in future reports.

Vital Status: If an individual's occurrence and date of death are recorded in their EHR, this information is included as a part of the Data Hub submission. Currently, the number of recorded deaths is substantially lower than expected for a population of people with SCD; therefore, vital status data are not included in this report. Additional approaches for obtaining complete information about individuals' vital status are being considered; once they are implemented, patient vital status and survival metrics will be included in reports.

SCD Diagnosis Type: As a part of the physician-attested SCD diagnosis process, sites are asked to provide an individual's type of SCD. The options are listed below:

- Hemoglobin SS (HbSS)
- Hemoglobin SC (HbSC)
- Hemoglobin S beta zero thalassemia (HbS/β^0)
- Hemoglobin S beta plus thalassemia (HbS/β^+)
- Hemoglobin S Lepore (HbS/Lepore)
- Hemoglobin S/O-Arab (HbS/O Arab)
- HbS/delta beta zero thalassemia (HbS/$\delta\beta^0$)
- Hemoglobin SD (HbSD)
- Hemoglobin SE (HbSE)
- Hemoglobin S Other, Other sickle cell disorder not captured above (HbS/Other)

While some participating Data Hub clinical sites have shifted to using DNA sequencing to confirm SCD genotype, most Data Hub sites continue to use a quantitative measurement of hemoglobin type (e.g., high-performance liquid chromatography [HPLC], capillary electrophoresis, etc.) to confirm SCD diagnosis and type. To avoid confusion, this report utilizes the more general term "Type of SCD" instead of "SCD Genotype."

Limitations

The Data Hub cohort in this report reflects data submitted by the participating sites. While we expect the number of contributing sites will continue to grow, the current cohort represents a subset of the ASH RC SCD Research Network sites and, by extension, a subset of individuals living with SCD in the US. The data presented in this report were submitted by academic medical centers, which may not fully reflect where all individuals with SCD receive care. Additional work is needed to verify the representativeness of the broader SCD population, and ASH RC remains committed to expanding the size and scope of the Data Hub to enhance generalizability.

The Data Hub is a real-world dataset that currently utilizes information obtained from EHRs, which offers a wealth of clinical data and many advantages but also has known limitations. For example, variability between sites in terms of coding, recording, and submission of EHR data can create internal validity challenges. While the ASH RC has undertaken efforts to improve data accuracy, completeness, and understanding, this work remains ongoing. Exploration is also underway to obtain EHR data from unstructured notes, which would also enhance data accuracy.

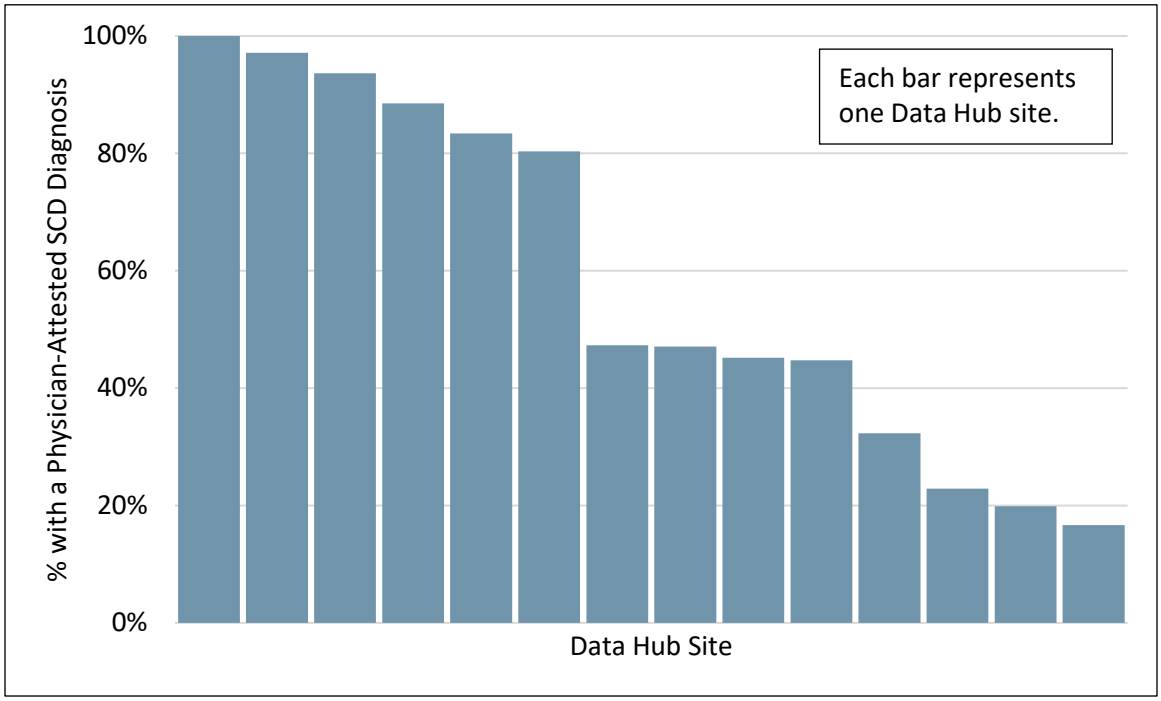
Throughout this report, each table and figure is accompanied by a corresponding paragraph that provides the reader with context for the data, including a discussion of their known and potential limitations, to facilitate appropriate interpretation.

Description and Characteristics of the Data

Individuals Included in the Data Hub and the Report

In total, 17 clinical sites have submitted data for patients with recorded encounters from January 1, 2015 – December 31, 2024 (as mentioned, complete data for the full year of 2024 was not available at the time this report was created). Of these sites, three have not yet begun physician attestation of the SCD diagnoses of their patients.
Figure 1 displays the percentage of individuals within each of the 14 sites for whom the physician attestation process for SCD diagnosis has been started. Each bar represents one of these 14 Data Hub sites.

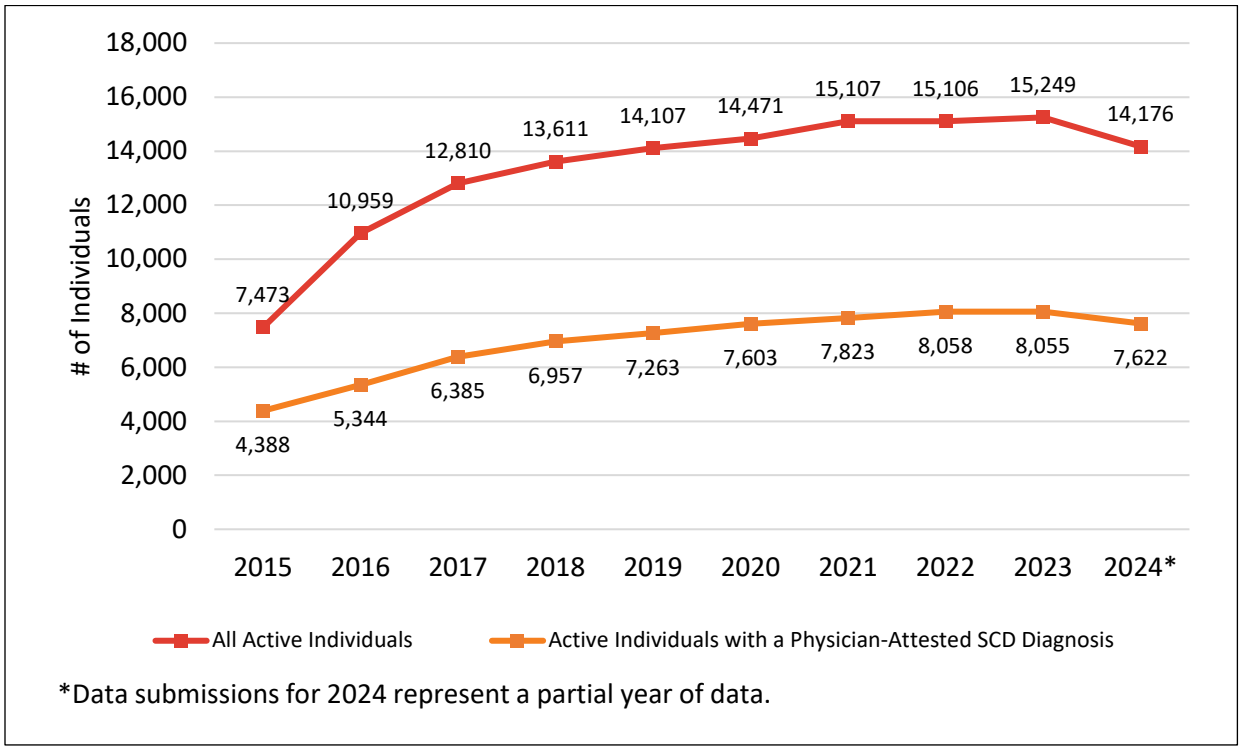
Figure 1: Percentage of Individuals with a Physician-Attested SCD Diagnosis by Data Hub Sites



There are 23,580 total active individuals in the Data Hub, 10,412 of whom have a physician-attested SCD diagnosis.

Figure 2 displays the total number of active individuals in the Data Hub and the number of active individuals with a physician-attested SCD diagnosis for each year from 2015 to 2024.

Figure 2: Number of Active Individuals in the Data Hub Each Year from 2015 to 2024



*Data submissions for 2024 represent a partial year of data.

The Data Hub provides data on a dynamic cohort of individuals with SCD who have received care at any one of the institutions that are a part of the Data Hub. Individuals with a physician-attested SCD diagnosis are a subset of this cohort. The extent to which these two cohorts are generalizable to the demographics and disease characteristics of the full population of individuals living with SCD in the United States has not yet been determined.

Table 1 displays the characteristics of individuals with and without a physician-attested SCD diagnosis. Sites contributing to the Data Hub are asked to self-identify whether they are a pediatric site (primarily providing care for patients under the age of 18), an adult site (primarily providing care for patients over the age of 18), or a lifespan site (providing care for patients of all ages). Among the sites contributing data on individuals with a physician-attested SCD diagnosis, six are pediatric sites; four are adult sites, and four are lifespan sites. Overall, the physician-attested population is younger than the non-attested population (42% younger than 18 years versus 24% younger than 18 years) and has a higher percentage of individuals with five or more active years recorded in the Data Hub (73% among the attested versus 49% among the non-attested). Pediatric sites contributed the majority (51%) of patients in the physician-attested SCD diagnosis group. In contrast, contributions were more evenly distributed by sites in the non-physician SCD diagnosis group (36% from pediatric sites, 29% from adult sites, and 35% from Lifespan sites).

The Data Hub provides data on a dynamic cohort of individuals with SCD who have received care at any one of the institutions that are a part of the Data Hub. Individuals with a physician-attested SCD diagnosis are a subset of this cohort. The extent to which these two cohorts are generalizable to the demographics and disease characteristics of the full population of individuals living with SCD in the United States has not yet been determined.

Therefore, the demographics of the physician-attested subset should be taken into consideration when comparing distributions from the Data Hub with other population-based data sources.

Table 1: Number and Percentage of Individuals by Demographic Variables, Stratified by Physician Attestation Status

	Individuals with a Physician-Attested SCD Diagnosis	Individuals without a Physician-Attested SCD Diagnosis
Total Population	10,412	13,168
	n (%)	n (%)
Age as of December 31, 2024		
<18	4,393 (42%)	3,186 (24%)
18–26	2,027 (19%)	2,150 (16%)
26+	3,992 (38%)	7,832 (59%)
Sex		
Female	5,418 (52%)	7,412 (56%)
Male	4,994 (48%)	5,756 (44%)
Year Added to the Data Hub		
2015–2017	6,860 (65%)	7,691 (58%)
2018–2020	2,185 (21%)	2,641 (20%)
2021–2024	1,367 (13%)	2,836 (22%)
Years Included in the Data Hub		
1	410 (4%)	2,344 (18%)
2–5	2,447 (24%)	4,398 (33%)
5+	7,555 (73%)	6,426 (49%)
Type of Sites		
Pediatric	5,285 (51%)	4,710 (36%)
Adult	2,703 (26%)	3,830 (29%)
Lifespan	2,424 (23%)	4,628 (35%)

The following data represent the cohort of individuals with a physician-attested SCD diagnosis.

Table 2 presents information about the number of patients and their ages from the data submitted by the pediatric, adult, and lifespan sites in 2023. The table provides the range of sample sizes at individual sites for consideration of site-specific analyses and clustering. In general, the age distribution matches the descriptive site type, and 98% of adult patients have transitioned from pediatric sites by age 26.

Table 2: Site Characteristics for the 8,055 Individuals included in the Data Hub in 2023

	Pediatric Sites	Adult Sites	Lifespan Sites
Number of Sites	6	4	4
Total Number of Individuals	4,288 (53%)	1,976 (25%)	1,791 (22%)
Patient Count Across Sites			
Minimum Patient Count	225	401	154
Maximum Patient Count	1,200	579	853
Age on December 31, 2023			
<18	3,230 (75%)	23 (1%)	478 (27%)
18–26	965 (22%)	306 (15%)	282 (16%)
26+	93 (2%)	1,647 (83%)	1,031 (58%)

Data Hub Records and Data Available and Used in this Report

Encounter Records

Each time an individual receives care or comes in contact with a healthcare system or institution participating Data Hub, a record is included in their Data Hub encounter file. This file includes the start and end date of the encounter and a code for the type of encounter. Code sets have been developed to classify these codes as inpatient, emergency department, and outpatient encounters.

At some Data Hub sites, visits that began in the emergency department and resulted in hospital admission were coded as one encounter. Within this report, these are considered inpatient encounters and are, therefore, not included in the counts for emergency department visits.

There are differences across Data Hub sites in how inpatient encounters are recorded, for example, some sites submit one record per stay and others a separate record for each inpatient day. As a result, differences in the number of inpatient records reported across sites may reflect differences in site data reporting approaches and not differences in the actual rates of inpatient stays. Work is being done to clarify these data in order to accurately reflect hospitalization frequency and utilization data.

There are also differences in how sites report outpatient encounters, including a lack of specificity in coding regarding what constitutes an outpatient encounter. For example, the outpatient encounter count in this report is not specific to a hematology outpatient clinic visit for routine or urgent care and may also include primary care or specialty clinic visits, outpatient therapy visits, infusion center visits, pharmacy consultations, etc. We are working to define these encounters more clearly to better understand care utilization in the outpatient setting.

Laboratory and Clinical Measurement Records

The laboratory and clinical measurements dataset includes clinical measurements (e.g., height and weight) and laboratory results. Each file includes the date and type of measurements, which are entered as a Logical Observations, Identifiers, Names, and Codes (LOINC), SNOMED, Current Procedural Terminology (CPT4), or Healthcare Common Procedural Coding System (HCPCS) code, along with the

numerical result and the unit of the measurements. Each record in the dataset corresponds to one measurement. Therefore, one complete blood count (CBC) measurement will have individual records for each laboratory element, including hemoglobin, hematocrit, erythrocyte indices, white blood cell counts, and differential. As a result, this dataset contains tens of millions of records. Continuous work is needed to arrange the codes from the different nomenclatures into code sets that can be used to assign the records to specific lab tests.

This report includes an initial list of laboratory measurements for which code sets have been developed, and validation testing has been completed. These measurements are hemoglobin, total bilirubin, aspartate aminotransferase (AST), lactate dehydrogenase (LDH), serum and urine creatinine, and urine albumin. The report focuses on laboratory measurements from 2023, which is the most recent year for which complete data are available. Individuals with no recorded measurements for the entire year were not included in the denominator for the analyses of the number, percentage, and mean number of tests recorded per individual. The Data Hub only includes laboratory measurements recorded in patients' EHRs and may not reflect testing ordered and performed at other laboratories.

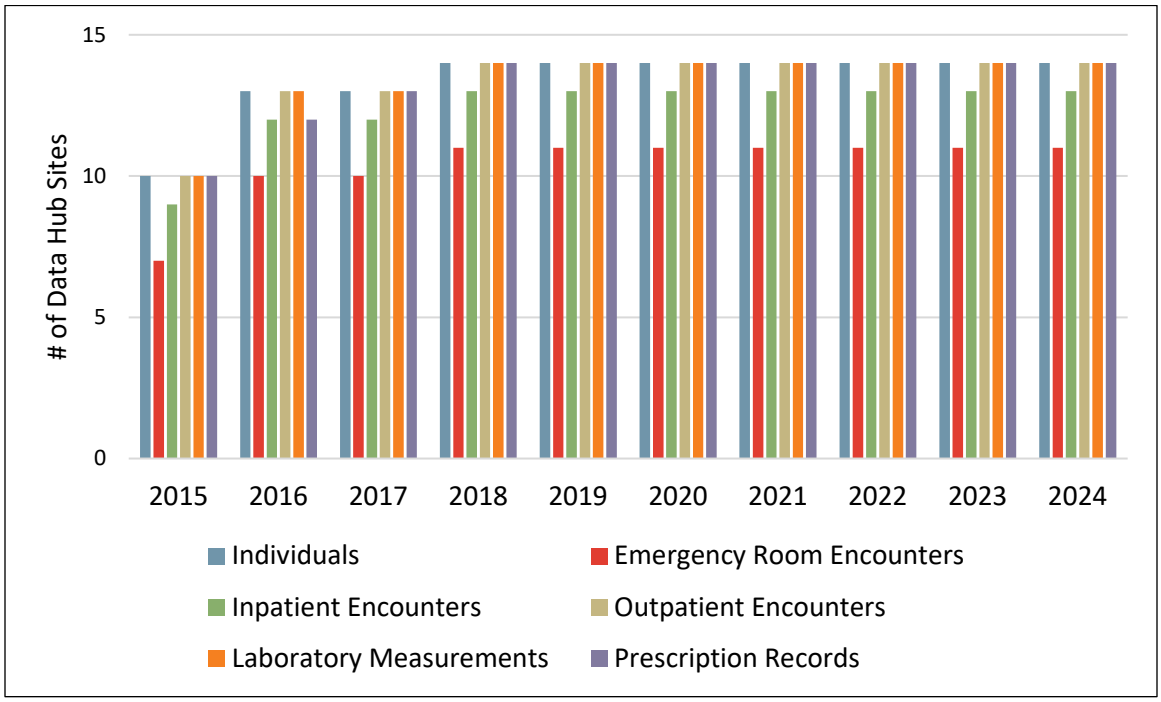
Prescription/Medication Records

The medications dataset contains the prescription data recorded in individuals' EHRs and includes the prescription date, an Rx Drug Nomenclature (RXNorm) or CVX code for the specific medication or vaccine, the route and dose prescribed, and the number of refills included in the prescription. Each record in the dataset reflects one prescription/medication record added to the EHR. While all prescriptions/medications recorded in the EHR are transferred to the Data Hub, this report focuses on the subset of SCD-related disease-modifying therapies. Specifically, it provides information on hydroxyurea, L-glutamine, crizanlizumab, and voxelotor. Across the Data Hub sites, there are differences in terms of the utilization of EHRs for recording prescription/medication orders and variability in how individuals fill prescriptions. This may explain some of the differences in prescription/medication records across Data Hub sites. The prescription/medication record counts included in this report may not reflect the full picture of the actual prescribing practices and medication utilization.

Data Availability and Variability

When interpreting the data presented in this report, it is necessary to consider the differences in the number of Data Hub sites submitting data and the types of data submitted by the participating sites each year. **Error! Reference source not found.** displays the number of sites that have submitted emergency department, inpatient, outpatient, laboratory/clinical measurements, and prescription records each year. One site has not yet contributed any inpatient records, and three sites have not yet contributed any emergency department records. This needs to be taken into consideration when analyzing these data and making comparisons across different encounter types.

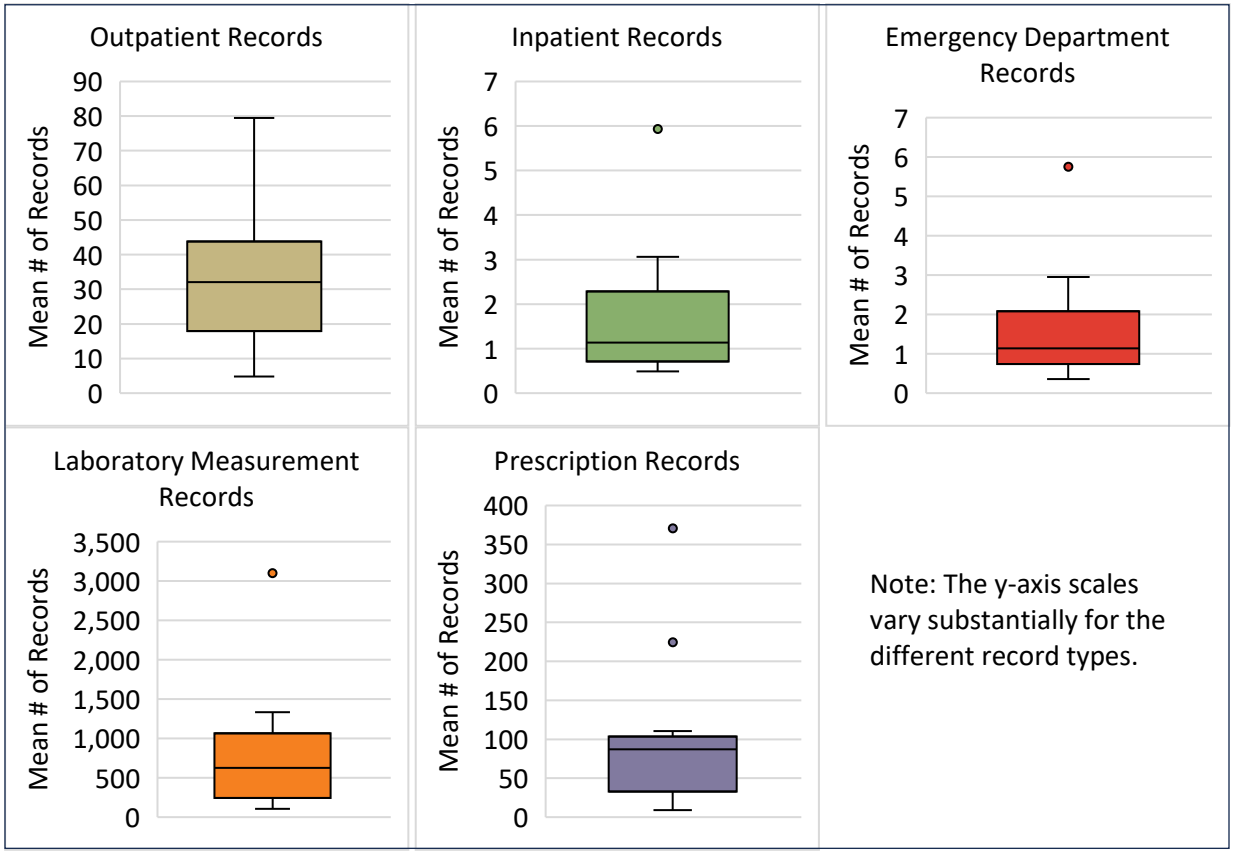
Figure 3: Number of Data Hub Sites Submitting Data by Data Type and Year



The advantage of having multiple institutions provide data to the Data Hub is that it enables the possibility of observing differences in the provision of care across sites and examining the impact of these differences on patient outcomes. As a first step, it is necessary to understand the variability across Data Hub sites, which is likely driven by numerous factors, including differences in patient population characteristics such as age, social determinants of health, and disease severity. Additionally, there are differences in site approaches to care and treatment as well as variability in their access to institutional resources and support. Before comparisons can be made, it is also necessary to understand the differences between Data Hub sites in terms of how they capture and submit records.

Figure 4 displays box and whisker charts showing the distribution of the mean number of records per individual reported by each Data Hub site during 2023. The box represents the interquartile range (25th to 75th percentile), while the line represents the median value. Substantial variability is observed across the 14 Data Hub sites in the mean number of records per individual for each record type. Many potential factors contribute to the inter-site variability seen in the submitted records, including institutional differences in how events are coded (e.g., one record per inpatient stay versus one record per day of an inpatient stay), capture of prescription/medication records within the EHR, and the proportion of laboratory and other services that occur within versus outside the institution. Some sites' EHR systems include outpatient services, but not all of them. Steps are being taken within the Data Hub to understand and characterize these differences, facilitating future analysis and research.

Figure 4: Distribution of the Mean Number of Records per Individual Reported by Each Data Hub Site in 2023 by Record Type



Throughout its first ten years, the Data Hub collected a substantial number of records.

Table 3 provides the number of total records collected each year. There are currently millions of encounters, laboratory/testing measurements, and prescription records in the Data Hub. As expected, the number of records for 2024 is lower; however, this is anticipated to increase with the processing of additional data submissions in 2025, which will include all patients who received care in 2024.

Table 3 is intended to provide information about the type and breadth of the data available within the Data Hub for research and analysis. As noted previously, the group of clinical sites contributing data for each given year varied, and differences exist across clinical sites in how events and records are coded and submitted. Therefore, conclusions or inferences should not be drawn about patient care within a given year or over time.

Table 3: Number of Encounter, Laboratory, and Prescription Records Counts by Year

Year	Inpatient Encounter Records	Emergency Room Encounter Records	Outpatient Encounter Records	Lab Values/Clinical Measurement Records	Prescription Records
2015	9,604	5,356	145,324	1,818,174	314,616
2016	10,768	6,259	163,141	2,164,445	407,310
2017	11,888	7,609	186,179	4,294,161	526,376
2018	12,409	7,872	181,190	4,627,432	575,087
2019	13,043	7,695	189,180	5,028,447	709,482
2020	10,801	6,513	207,066	4,937,045	602,368
2021	13,270	8,037	248,626	5,602,974	693,927
2022	14,206	8,724	261,366	5,778,055	782,581
2023	13,665	9,800	273,847	6,076,895	803,791
2024*	11,138	7,845	220,009	4,990,126	576,932
Total	120,792	75,710	2,075,928	45,317,754	5,992,470

*Data for 2024 represent partial data for that year.

Longitudinal Data

A strength of the Data Hub is its inclusion of longitudinal data, which enables the ability to follow individuals over time and investigate changes in the natural history of the disease. The Data Hub contains data starting from 2015, so it currently includes up to ten years of follow-up data for some individuals.

Figure 5 outlines the number of years of follow-up for all individuals. The red portion of the bar reflects the number of people whose follow-up spanned consecutive years. The gold portion represents individuals with non-consecutive years of treatment, meaning there were one or more years between their first and last active year, during which they did not meet the criteria for being active in the Data Hub. The Data Hub has ten years of follow-up data for 2,626 individuals and eight or more years of follow-up data for 4,812 individuals (2,626 + 1,185 + 1,001).

Figure 5: Number of Individuals by Years of Follow-up in the Data Hub (n=10,412)

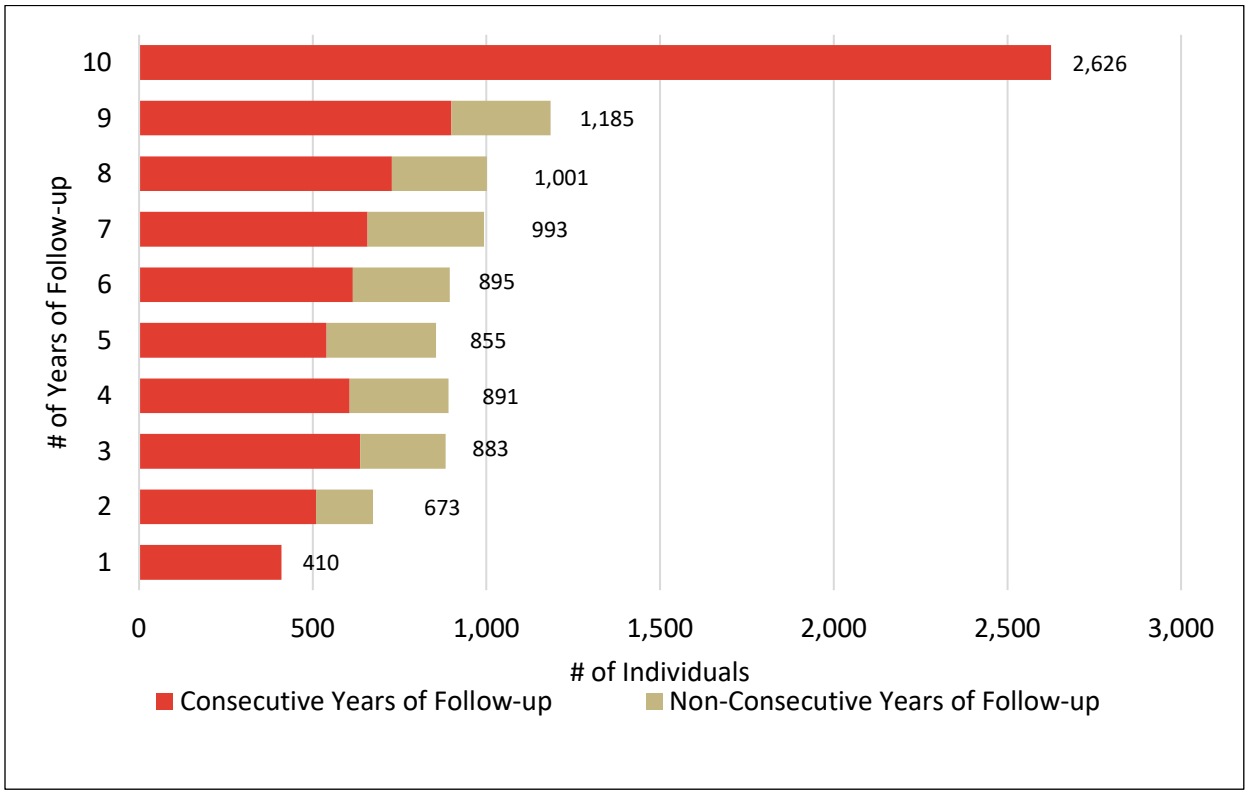


Figure 6 presents the age of each patient when they were first added to the Data Hub, which represents their age at the beginning of their follow-up period. Many of the patients were over the age of 18 at the time of addition to the Data Hub (41%). The Data Hub currently includes a cohort of 1,484 individuals who were less than one year old when they were added to the Data Hub. This cohort offers the opportunity to conduct longitudinal follow-ups starting in infancy and observe the natural history and progression of the disease.

The screening of newborns for SCD began in the 1970s and, as of 2006, had been adopted across the United States. It is expected that many individuals included in the Data Hub were diagnosed with SCD as infants and received care for their condition before being added to the Data Hub. The Data Hub only includes the events recorded for these individuals from the time they were added to the Data Hub until 2024. Researchers conducting longitudinal analyses will need to consider how to define the eligibility of the individuals to be included in their sample.

Figure 6: Number and Percentage of Individuals by Age at the Time of Addition to the Data Hub

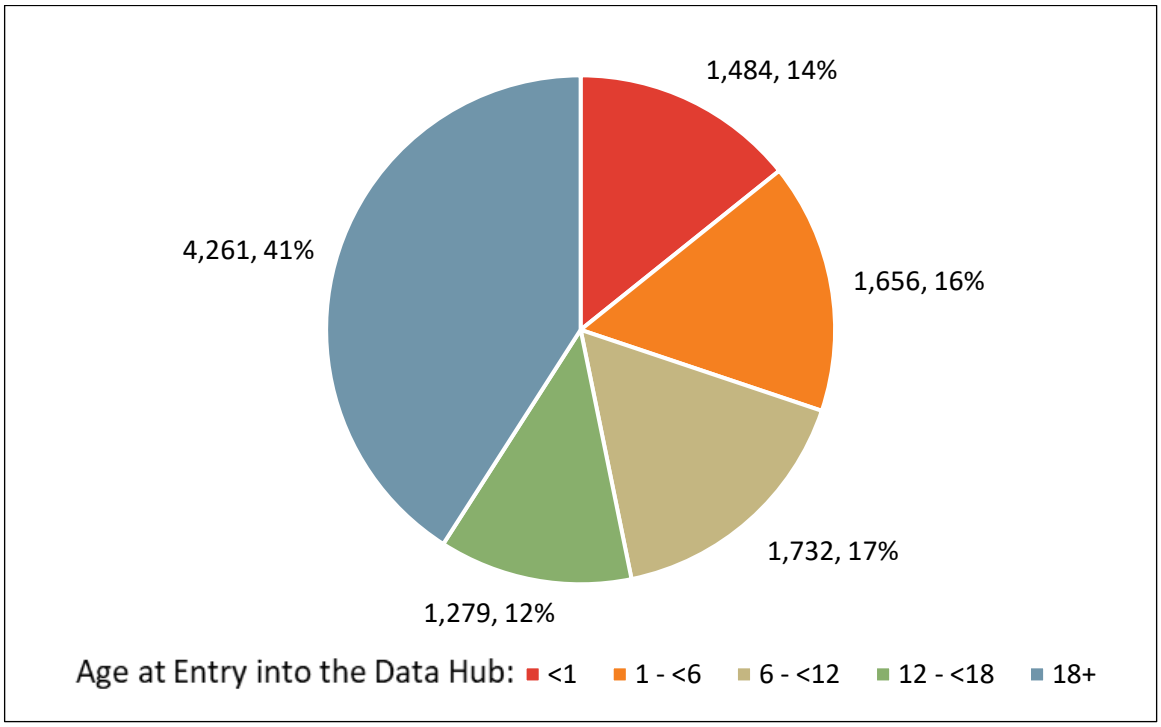


Table 4 displays the total population and three subsets of the population based on the available longitudinal data. The first is the subset of the population for which there are ten years of data. Compared to the total population, this subset has a higher proportion of individuals aged 12–17, a lower proportion who are younger than 12, and fewer individuals from adult sites. The second subset comprises individuals for whom there are consecutive years of data, regardless of the length of follow-up. This subset generally mirrors the total population in terms of age distribution, sex, site type, and type of SCD. Similarly, individuals who were still active in the Data Hub in 2023 or 2024 are comparable to the total population with respect to the distribution of the variables examined.

Table 4: Number and Percentage of Individuals by Demographic Characteristics Classified by Follow-Up Attributes

	Total Population	Individuals with 10 Years of Follow-Up	Individuals with Consecutive Follow-Up	Individuals who were Active in 2023 or 2024
Total Population	10,412	2,626	8,228	8,481
	n (%)	n (%)	n (%)	n (%)
Age in 2024				
<12	2,662 (26%)	323 (12%)	2,296 (28%)	2,400 (28%)
12–17	1,731 (17%)	711 (27%)	1.391 (17%)	1,520 (18%)
18–30	2,767 (27%)	748 (28%)	2,100 (25%)	2,150 (25%)
30+	3,252 (31%)	844 (32%)	2,441 (30%)	2,411 (28%)
Sex				
Female	5,418 (52%)	1,416 (54%)	4,289 (52%)	4,416 (52%)
Male	4,994 (48%)	1,210 (46%)	3,939 (48%)	4,065 (48%)
Type of Site				
Pediatric	4,446 (42%)	1,325 (50%)	3,611 (44%)	3,827 (45%)
Adult	2,703 (26%)	282 (11%)	2,033 (25%)	2,105 (25%)
Lifespan	3,263 (31%)	1,019 (39%)	2,551 (31%)	2,549 (30%)
Sickle Cell Type				
HbSS	6,782 (65%)	1,764 (67%)	5,473 (67%)	5,530 (65%)
HbS/β^0	251 (2%)	68 (3%)	203 (2%)	212 (2%)
HbSC	2,602 (25%)	625 (24%)	1,954 (24%)	2,105 (25%)
HbS/β^+	669 (6%)	146 (5%)	506 (6%)	550 (6%)
HbS/Other	108 (1%)	23 (1%)	92 (1%)	84 (1%)

The data presented in this section represent active individuals with a Physician-Attested SCD Diagnosis only.

Study Population Characteristics

Demographic and Disease Characteristics

Figure 7 displays the distribution of the type of SCD for all individuals included in the Data Hub. The most common type is HbSS, representing 65% of the population, while the next most common subtype is HbSC, which includes 25% of the population. Further, HbS/β^0 and HbS/β^+ are found in 2% and 7% of the population, respectively. The category of HbS/Other includes HbS/Lepore, HbS/O Arab, HbS/$\delta\beta^0$, HbSD, and HbSE, and an additional 59 individuals who are recorded as Other SCD in the Data Hub.

Figure 7: Percentage of Individuals by SCD Type

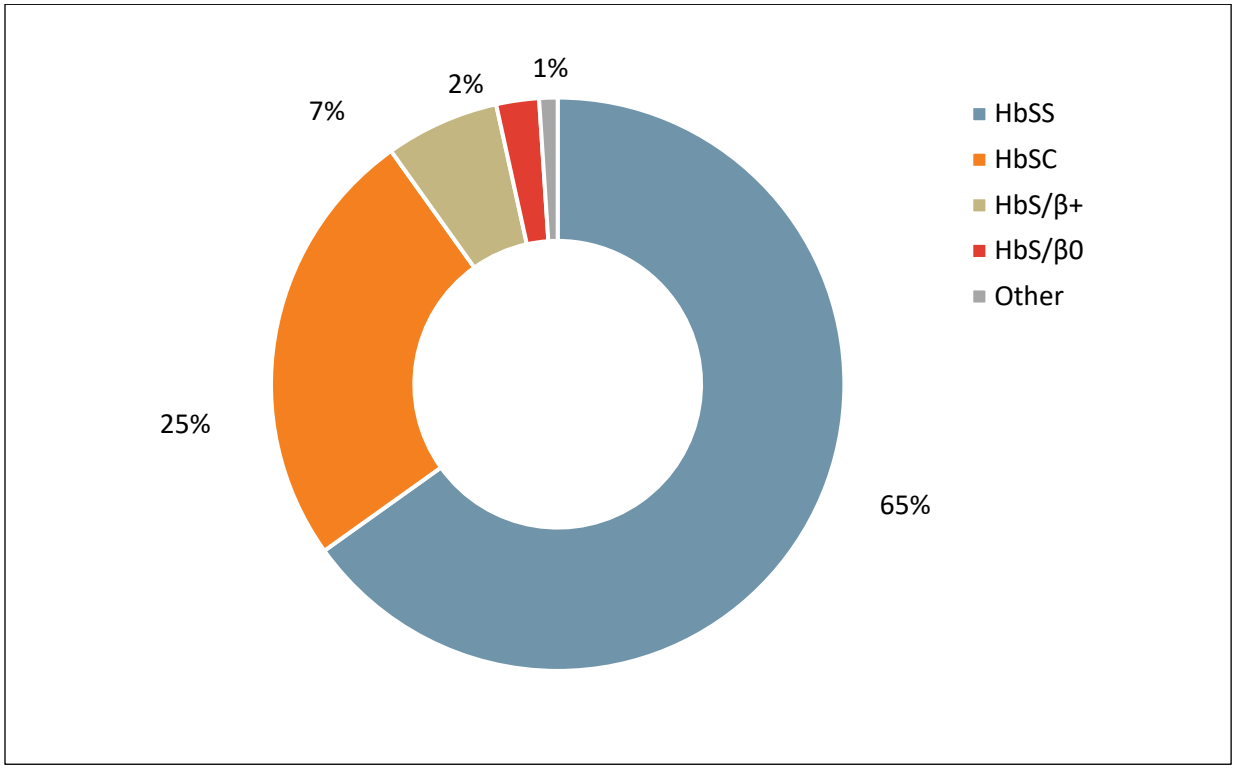


Table 5 displays the demographic characteristics of the Data Hub population. The first column provides data on the entire Data Hub population, while the subsequent columns show cross-sectional subsets of the population active in 2015, 2019, and 2023. The total population consists of 48% males and 52% females, 98% of the individuals are African American or Black, and 10% are Hispanic. Overall, the attributes of the population included in the Data Hub have remained relatively consistent over the years.

Table 5: Number and Percentage of Individuals by Demographic Characteristics in the Entire Data Hub Population and Those Active in 2015, 2019, and 2023

Demographic Characteristic	Active between 2015–2024	Active in 2015	Active in 2019	Active in 2023
	n (%)	n (%)	n (%)	n (%)
Total Population	10,412	4,388	7,263	8,055
Age as of December 31st of active year				
<5	734 (7%) *	799 (18%)	1,029 (14%)	878 (11%)
5–9	1,338 (13%)	816 (19%)	1,129 (16%)	1,165 (14%)
10–14	1,445 (14%)	721 (16%)	1,168 (16%)	1,198 (15%)
15–19	1,470 (14%)	474 (11%)	899 (12%)	1,244 (15%)
20–24	1,225 (12%)	424 (10%)	684 (9%)	801 (10%)
25–34	1,768 (17%)	542 (12%)	1,033 (14%)	1,163 (14%)
35–44	1,160 (11%)	271 (6%)	626 (9%)	777 (10%)
45–54	633 (6%)	193 (4%)	364 (5%)	440 (5%)
55–65	371 (4%)	119 (3%)	230 (3%)	245 (3%)
65+	268 (3%)	29 (1%)	101 (1%)	144 (2%)
Sex				
Female	5,418 (52%)	2,290 (52%)	3,798 (52%)	4,203 (52%)
Male	4,994 (48%)	2,098 (48%)	3,465 (45%)	3,852 (48%)
Race				
Data Provided				
Black, African American	9,419 (97%)	4,120 (98%)	6,683 (98%)	7,370 (97%)
White	211 (2%)	78 (2%)	140 (2%)	174 (2%)
Asian	42 (<1%)	10 (<1%)	22 (<1%)	28 (<1%)
American Indian/Alaskan Native	9 (<1%)	5 (<1%)	7 (<1%)	8 (<1%)
Native Hawaiian or Pacific Islander	2 (<1%)	2 (<15)	2 (<1%)	2 (<1%)
Missing	729 (7%)	173 (4%)	409 (6%)	473 (6%)
Ethnicity				
Data Provided				
Hispanic	953 (9%)	383 (9%)	640 (9%)	786 (10%)
Non-Hispanic	9,181 (91%)	3,943 (91%)	6,503 (91%)	7,117 (90%)
Missing	278 (3%)	62 (1%)	120 (2%)	152 (2%)
Type of Site				
Pediatric	5,285 (51%)	2,171 (49%)	3,631 (50%)	4,288 (53%)
Adult	2,703 (26%)	503 (11%)	1,761 (24%)	1,976 (25%)
Lifespan	2,424 (23%)	1,714 (39%)	1,871 (26%)	1,791 (22%)

*The age categories for "Ever Active Between 2015–2024" reflect the individuals' ages on December 31st, 2024.

Following standard practice within the SCD literature, the types of SCD included in this report were grouped based on hemoglobin phenotype. The demographic characteristics for HbSS, HbS/β^0, HbSC and HbSβ^+ were generally similar, as displayed in Table 6. In contrast, the individuals classified as HbS/Other tended to be younger, were less likely to be Black or African American, were more likely to be Hispanic, and were usually treated at pediatric sites.

Table 6: Number and Percentage of Individuals by Demographic Characteristics, Stratified by SCD Type

Demographic Characteristic	HbSS or HbS/β^0	HbSC	HbS/β^+	HbS/Other
	n (%)	n (%)	n (%)	n (%)
Total Population	7,033	2,602	669	108
Age as of December 31, 2024				
<5	451 (6%)	216 (8%)	51 (8%)	16 (15%)
5–9	897 (13%)	326 (13%)	94 (14%)	21 (19%)
10–14	967 (14%)	368 (14%)	90 (13%)	20 (19%)
15–19	989 (14%)	375 (14%)	89 (13%)	17 (16%)
20–24	865 (12%)	268 (10%)	81 (12%)	11 (10%)
25–34	1,305 (19%)	340 (13%)	114 (17%)	9 (8%)
35–44	832 (12%)	265 (10%)	55 (8%)	8 (7%)
45–54	415 (6%)	165 (6%)	48 (7%)	5 (5%)
55–65	195 (3%)	147 (6%)	28 (4%)	1 (1%)
65+	117 (2%)	132 (5%)	19 (3%)	0
Sex				
Female	3,632 (52%)	1,384 (53%)	348 (52%)	54 (50%)
Male	3,401 (48%)	1,218 (47%)	321 (48%)	54 (50%)
Race				
Data Provided				
Black, African American	6,363 (97%)	2,380 (99%)	601 (96%)	75 (80%)
White	157 (2%)	23 (1%)	19 (3%)	12 (13%)
Asian	24 (<1%)	8 (<1%)	3 (<1%)	7 (7%)
American Indian/Alaskan Native	6 (<1%)	3 (<1%)	0	0
Native Hawaiian or Pacific Islander	1(<1%)	1 (<1%)	0	0
Missing	482 (7%)	187 (7%)	46 (7%)	14 (13%)
Ethnicity				
Data Provided				
Hispanic	683 (10%)	199 (8%)	49 (8%)	22 (21%)
Non-Hispanic	6,168 (90%)	2,330 (92%)	600 (92%)	83 (79%)
Missing	182 (3%)	73 (3%)	20 (3%)	3 (3%)
Type of Site				
Pediatric	3,508 (50%)	1,352 (52%)	347 (52%)	78 (72%)
Adult	1,848 (26%)	664 (26%)	175 (26%)	16 (15%)
Lifespan	1,677 (24%)	586 (23%)	147 (22%)	14 (13%)

SCD Comorbidities

The ASH RC SCD Data Hub Program is working with clinical informaticists and hematology experts to inform and prioritize the creation of core clinical data elements and metrics of interest specific to SCD-related clinical care and research. Extensive work is underway, both within the Data Hub and globally, to develop a database of computable clinical phenotypes comprising curated sets of EHR codes with high specificity and validity for specific health metrics. Code sets for clinical concepts such as SCD-associated comorbidities are constructed using ICD-10 and SNOMED CT codes. While many comorbidities and sequelae of SCD are of clinical and research interest, in this report, we focus on the key comorbidities for which there are ASH RC clinician leader-defined code sets, namely acute chest syndrome, acute kidney injury, chronic kidney disease, cerebrovascular accidents, deep-vein thrombosis, hypertension, pulmonary hypertension, pulmonary embolism, priapism (among males), vaso-occlusive pain episodes, and venous thromboembolism.

In the remaining figures and tables presented in this section, individuals were included if they had a record of one of the previously listed comorbidities at any time during their active period in the Data Hub (up to 10 years) between 2015 and 2024. While there is some variation across age groups, the average number of years of follow-up data for individuals included in the Data Hub is seven. Some of the comorbidities included are chronic, while others are acute, and individuals may have experienced multiple episodes within the follow-up period. In this report, comorbidities are counted only once per person. Some counts may reflect comorbidities that occurred or started before the participant's addition to the Data Hub but were added to their EHR at some point during their follow-up period. Future work will aim to differentiate comorbidities prevalent at the time of addition to the Data Hub and define discrete episodes of acute comorbidities, enabling further investigations into the cross-sectional prevalence, frequency, and recurrence rates of these comorbidities.

The proportions of individuals who had at least one record of a comorbidity during their follow-up period are displayed in Figure 8. This figure includes all active individuals regardless of age. Subsequent tables and figures provide percentages by current age group. The most reported comorbidity was vaso-occlusive pain episodes, with 73% of individuals having at least one episode, followed by acute chest syndrome, with 38% of individuals having at least one episode during their follow-up period.

Figure 8: Percentage of All Active Individuals with at Least One Report of the Comorbidity During Data Hub Follow-up

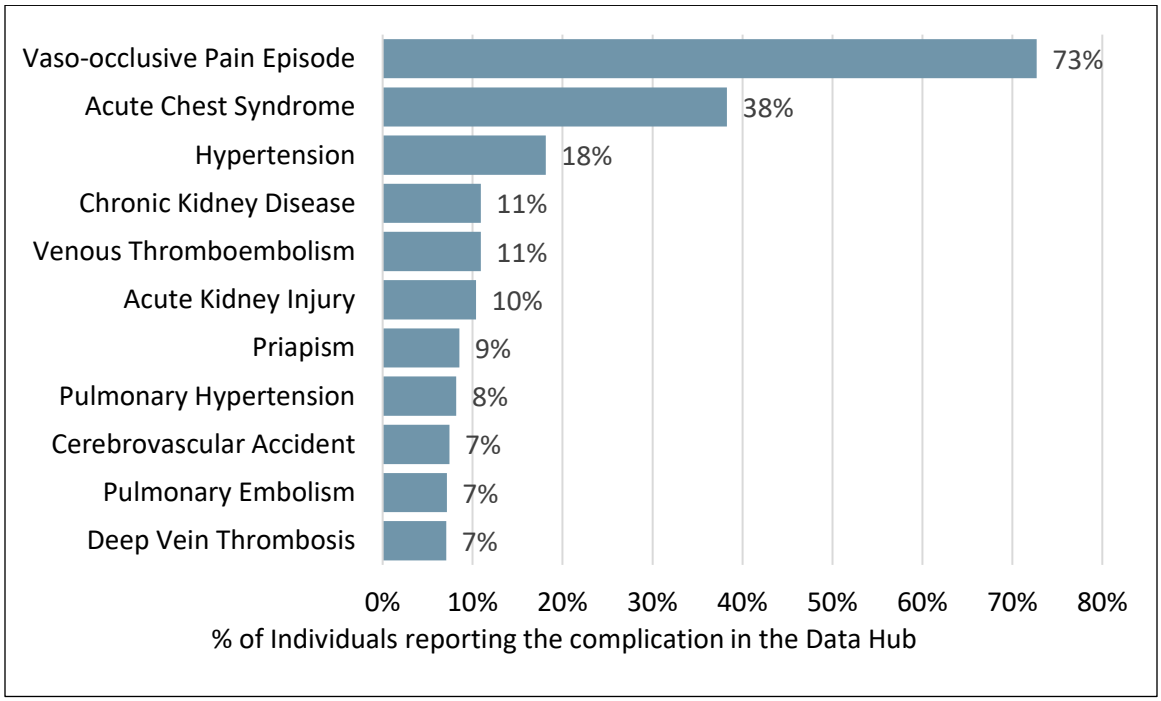


Table 7 displays the percentages of individuals with at least one record of a comorbidity from the time they were added to the Data Hub until December 31, 2024, stratified by age group. The age groups represent young children under the age of 12, adolescents aged 12 to 17, young adults aged 18 to 30, and adults aged 30 and older. It should be noted that, for this report, individuals were categorized based on their age as of December 31, 2024, and not their age at the time the comorbidity was recorded. In each row, the number and percentage represent a subset of the total population from that age group with at least one record of the comorbidity during the follow-up period (up to 10 years). The percentage of individuals affected by all comorbidities increases with age.

Table 7: Number and Percentage of Individuals with a Record for Each Comorbidity, Stratified by Age Group

	Age Group in 2024			
	<12	**12–17**	**18–30**	**30+**
Total Population	2,662	1,731	2,767	3,252
	n (%)	n (%)	n (%)	n (%)
Acute Chest Syndrome	789 (30%)	689 (40%)	1,049 (38%)	1,459 (45%)
Acute Kidney Injury	35 (1%)	51 (3%)	195 (7%)	799 (25%)
Cerebrovascular Accident	47 (2%)	90 (5%)	261 (9%)	377 (12%)
Chronic Kidney Disease	27 (1%)	37 (2%)	207 (7%)	865 (27%)
Deep-Vein Thrombosis	11 (<1%)	19 (1%)	175 (6%)	530 (16%)
Hypertension	72 (3%)	137 (8%)	427 (15%)	1,252 (38%)
Pulmonary Embolism	4 (<1%)	9 (1%)	157 (6%)	573 (18%)
Priapism (among males*)	19 (1%)	57 (6%)	151 (11%)	199 (15%)
Pulmonary Hypertension	25 (1%)	22 (1%)	136 (5%)	670 (21%)
Vaso-Occlusive Pain Episode	1,425 (53%)	1,179 (68%)	2,184 (79%)	2,783 (86%)
Venous Thromboembolism	14 (1%)	26 (2%)	262 (9%)	833 (26%)

Age was calculated as of December 31, 2024.
*Male Population: <12 = 1,399, 12–17 = 907, 18–30 = 1,334, and 30+ = 1,354

Table 8 presents the percentage of individuals with a reported comorbidity by type of SCD and Table 9 provides the percentages by type of SCD and age group (pediatric vs. adult). The numbers and percentages reflect the subset of the total population with that type of SCD and from that age group with at least one report of the comorbidity during the follow-up period. The prevalence of comorbidities is typically higher among individuals with HbSS or HbS/β^0 as compared to those with HbSC or HBS/β^+. However, a substantial percentage of individuals across all types of SCD have a record of these comorbidities and a subsequent burden of disease. The most significant differential between SCD types is observed in the reports of strokes, pulmonary hypertension, and priapism.

Table 8: Number and Percentage of Individuals with Reported Comorbidities, Stratified by SCD Type

	HbSS & HbS/β^0	HbSC	HbS/β^+	HbS/Other
Total Population	7,033	3,271	1,520	108
	n (%)	n (%)	n (%)	n (%)
Acute Chest Syndrome	2,967 (42%)	793 (30%)	201 (30%)	25 (23%)
Acute Kidney Injury	790 (11%)	232 (9%)	51 (8%)	7 (6%)
Cerebrovascular Accident	672 (9%)	81 (3%)	20 (3%)	2 (2%)
Chronic Kidney Disease	838 (12%)	243 (9%)	46 (7%)	9 (8%)
Deep-Vein Thrombosis	556 (8%)	135 (5%)	40 (6%)	4 (45%)
Hypertension	1,239 (18%)	515 (20%)	123 (18%)	11 (10%)
Pulmonary Embolism	505 (7%)	187 (7%)	46 (7%)	5 (5%)
Priapism (among males)	347 (10%)	66 (5%)	12 (4%)	1 (2%)
Pulmonary Hypertension	696 (10%)	121 (5%)	30 (4%)	5 (6%)
Vaso-occlusive Pain Episode	5,299 (75%)	1,801 (69%)	420 (63%)	51 (47%)
Venous Thromboembolism	824 (12%)	237 (9%)	67 (10%)	7 (6%)

Age was calculated as of December 31, 2024.
Male Population: HbSS and HbS/β0: 3,401; HbSC: 1,218; HbS/β+: 321; HbS/Other: 54

Table 9 provides a further breakdown of the number and percentage of individuals with a reported comorbidity record stratified by both SCD type and pediatric and adult age groups. Age represents the individual's age as of December 31, 2024, not the age at the time of the event.

Table 9: Number and Percentage of Individuals with Reported Comorbidities, Stratified by SCD Type and Age Group

	HbSS & HbS/β^0		HbSC		HbS/β^+		HbS/Other	
	<18	18+	<18	18+	<18	18+	<18	18+
Total Population	2,882	4,151	1,150	1,452	293	376	68	40
	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)
Acute Chest Syndrome	1,081 (38%)	1,886 (45%)	315 (27%)	478 (33%)	69 (24%)	132 (35%)	13 (19%)	12 (30%)
Acute Kidney Injury	64 (2%)	726 (17%)	17 (1%)	215 (15%)	4 (1%)	47 (12%)	1 (1%)	6 (15%)
Cerebrovascular Accident	124 (4%)	548 (13%)	9 (1%)	72 (5%)	3 (1%)	17 (5%)	1 (1%)	1 (2%)
Chronic Kidney Disease	33 (1%)	805 (19%)	23 (2%)	220 (15%)	7 (2%)	39 (10%)	1 (1%)	8 (20%)
Deep-Vein Thrombosis	25 (1%)	531 (13%)	4 (<1%)	131 (9%)	0	40 (11%)	1 (1%)	3 (8%)
Hypertension	159 (6%)	1,080 (26%)	40 (3%)	475 (33%)	8 (3%)	115 (31%)	2 (3%)	9 (22%)
Pulmonary Embolism	8 (<1%)	497 (12%)	4 (<1%)	183 (13%)	1 (<1%)	45 (12%)	0	5 (12%)
Priapism (among males)	52 (3%)	295 (15%)	16 (3%)	50 (8%)	7 (4%)	5 (3%)	1 (3%)	0
Pulmonary Hypertension	39 (1%)	657 (16%)	5 (<1%)	116 (8%)	3 (1%)	27 (7%)	0	6 (15%)
Vaso-occlusive Pain Episode	1,782 (62%)	3,517 (85%)	669 (58%)	1,132 (78%)	126 (43%)	294 (78%)	27 (40%)	24 (60%)
Venous Thromboembolism	31 (1%)	793 (19%)	7 (1%)	230 (16%)	1 (<1%)	66 (18%)	1 (1%)	6 (15%)

Age was calculated as of December 31, 2024.

Male Population: HbSS and HbS/β0: <18 = 1,489, 18+ = 1,912; HbSC: <18 = 611, 18+ = 607; HbS/β+: <18 169, 18+ = 152 ; HbS/Other: <18 = 37, 18+ = 17

Care and Treatment Characteristics

The Data Hub provides a unique opportunity to examine the care and treatments received by individuals living with SCD. This section of the report provides an overview of the records included in the Data Hub related to prescriptions for disease-modifying therapies for SCD, lab tests performed, and emergency department and inpatient encounter records. These findings provide a framework for researchers to understand the extent of the data available in the Data Hub regarding care to inform their plans and designs for future studies.

Medication Prescriptions

Between 2015 and 2024, four FDA-approved disease-modifying therapies for SCD were available. Hydroxyurea was approved for adults in 1998 and children in 2017, L-glutamine was approved in 2017 and crizanlizumab was approved in 2019. Voxelotor was approved in 2019 and was voluntarily withdrawn from the market in September 2024 due to safety concerns. Prescription records/medication orders for these four treatments are provided by type of SCD, year, age, and Data Hub site. Most clinical studies and treatment guidelines for these four treatments have focused on the HbSS and HbS/β^0 populations. Therefore, throughout this section, the data are displayed by type of SCD. When time trends and years of prescriptions are presented, data are provided only for the HbSS and HbS/β0 groups.

Across the Data Hub sites, there are differences in the utilization of EHRs for recording prescription data and medication orders, as well as variability in how individuals fill those prescriptions. This impacts the Data Hub's ability to capture prescription data accurately and may explain some of the differences in prescription rates among Data Hub sites. **The prescription statistics included in this report may not reflect the complete picture of the prescribing practices and medication utilization employed at the clinical sites.**

Additionally, a prescription record noted in an EHR does not necessarily mean that the individual received the prescription, chose to fill it, and, if filled, took the medication as prescribed.

Table 10 displays the number and percentage of individuals with at least one prescription for hydroxyurea, L-glutamine, crizanlizumab, or voxelotor recorded during their active Data Hub follow-up period (up to 10 years) between 2015 and 2024. These data are stratified by type of SCD and age group. It should be noted that, for this report, the age category was determined by the age of the individuals as of December 31, 2024. Hence, the 18+ category may include individuals whose prescription/medication order was recorded when they were younger than 18. The proportion of individuals with a prescription/medication record for L-glutamine, crizanlizumab, or voxelotor is less than 10%. The data suggest that individuals with HbSC, HbS/$\beta^{+,}$ and other types of SCD have an off-label prescription record for hydroxyurea in the Data Hub.

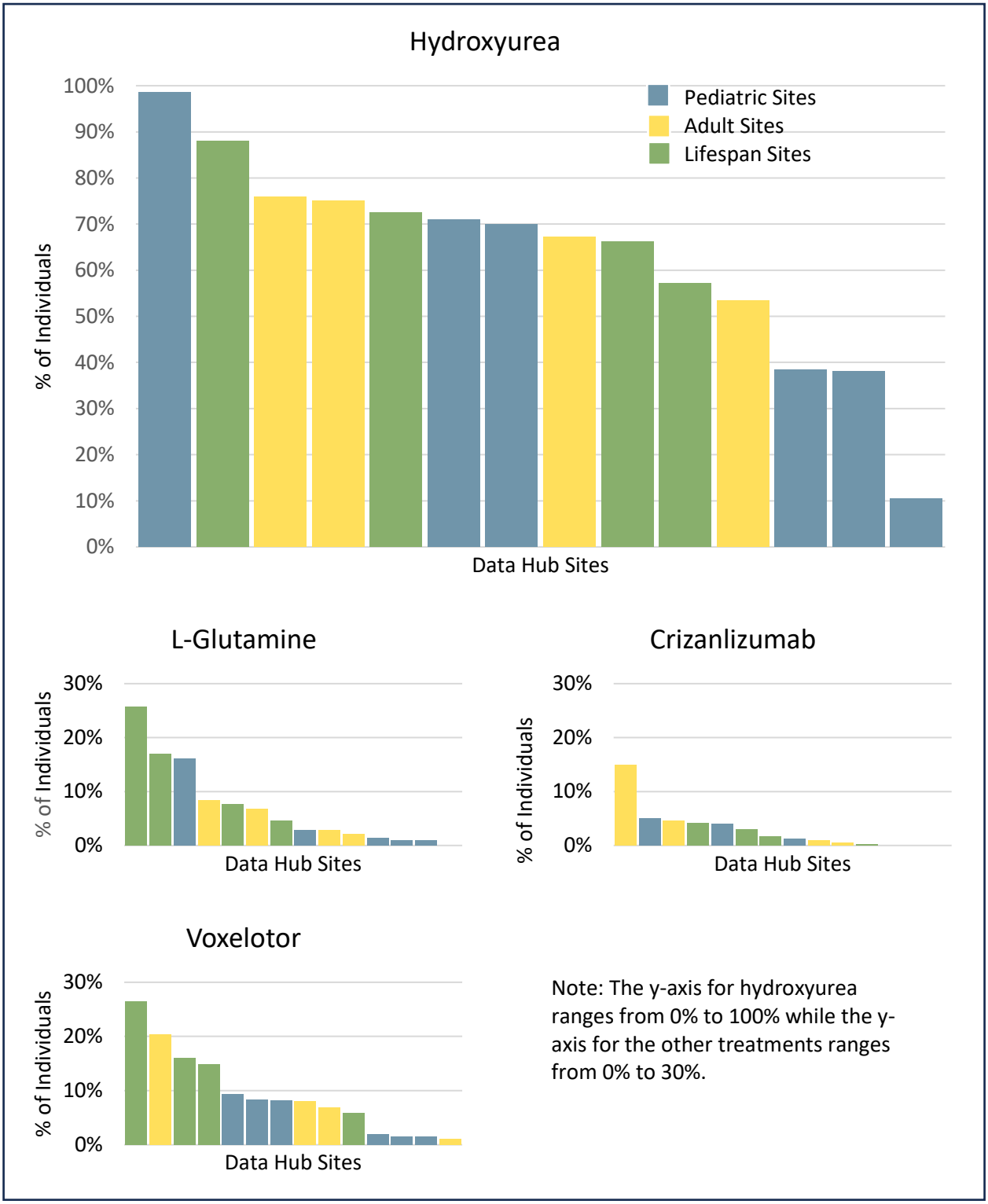
Table 10: Number and Percentage of Individuals with a Recorded Disease-Modifying Therapy Prescription, Stratified by Type of SCD and Age Group

Type of SCD	HbSS & HbS/β^0		HbSC		HbS/β^+		HbS/Other	
Age	<18	18+	<18	18+	<18	18+	<18	18
Population	2,882	4,151	1,150	1,452	293	376	68	40
	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)
Hydroxyurea	1,318 (46%)	2,790 (67%)	79 (7%)	374 (26%)	23 (8%)	143 (38%)	7 (10%)	17 (42%)
L-Glutamine	114 (4%)	317 (8%)	31 (3%)	55 (4%)	7 (2%)	20 (5%)	0	1 (2%)
Crizanlizumab	6 (<1%)	209 (5%)	1 (<1%)	51 (4%)	1 (<1%)	17 (5%)	0	1 (2%)
Voxelotor	186 (6%)	392 (9%)	2 (<1%)	6 (<1%)	0	6 (2%)	0	1 (2%)

Age was calculated as of December 31, 2024.

Figure 9 displays the proportions of individuals with HbSS and HbS/β^0 at each Data Hub site with at least one record of a prescription/medication order for one of the four disease-modifying therapies between 2015 and 2024. For the more recently approved medications, the records are concentrated among fewer Data Hub sites. There is substantial variability in the proportion of individuals with a prescription record for hydroxyurea across Data Hub sites, with no clear pattern between pediatric, adult, or lifespan sites. *The differences in how the sites record prescription and medication data in the EHRs are reflected in these distributions. Therefore, conclusions about prescribing practices and medication utilization should be avoided.*

Figure 9: Percentage of Individuals with a Prescription Record by Data Hub Site



In 2014, the National Heart Lung and Blood Institute (NIH) issued guidelines recommending that hydroxyurea be offered to all patients with HbSS and HbS/β^0 beginning from the age of nine months. The number and percentage of individuals with at least one prescription record for hydroxyurea in the Data Hub during each year is displayed in Figure 10 and the percentages with a prescription record by type of SCD and age group are displayed in Table 11. "Ever Prescribed" refers to the number and percentage of individuals with at least one hydroxyurea prescription record at any point during their follow-up period (up to 10 years). It should be noted that Table 11 was intended to provide information about the breadth of the records available within the Data Hub. The group of clinical sites contributing data during each year was not consistent, which means that any conclusions or inferences about changes in prescribing practices will be invalid.

Figure 10: Number of Individuals with a Report of a Hydroxyurea Prescription by Calendar Year

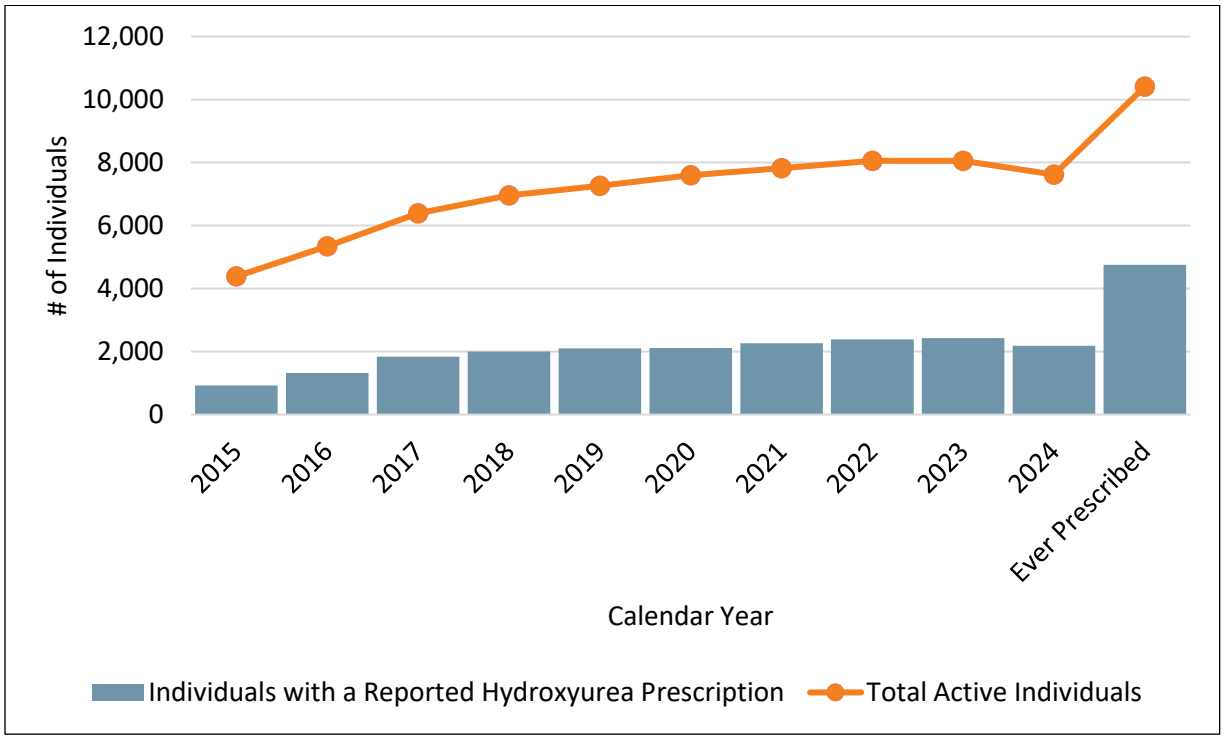


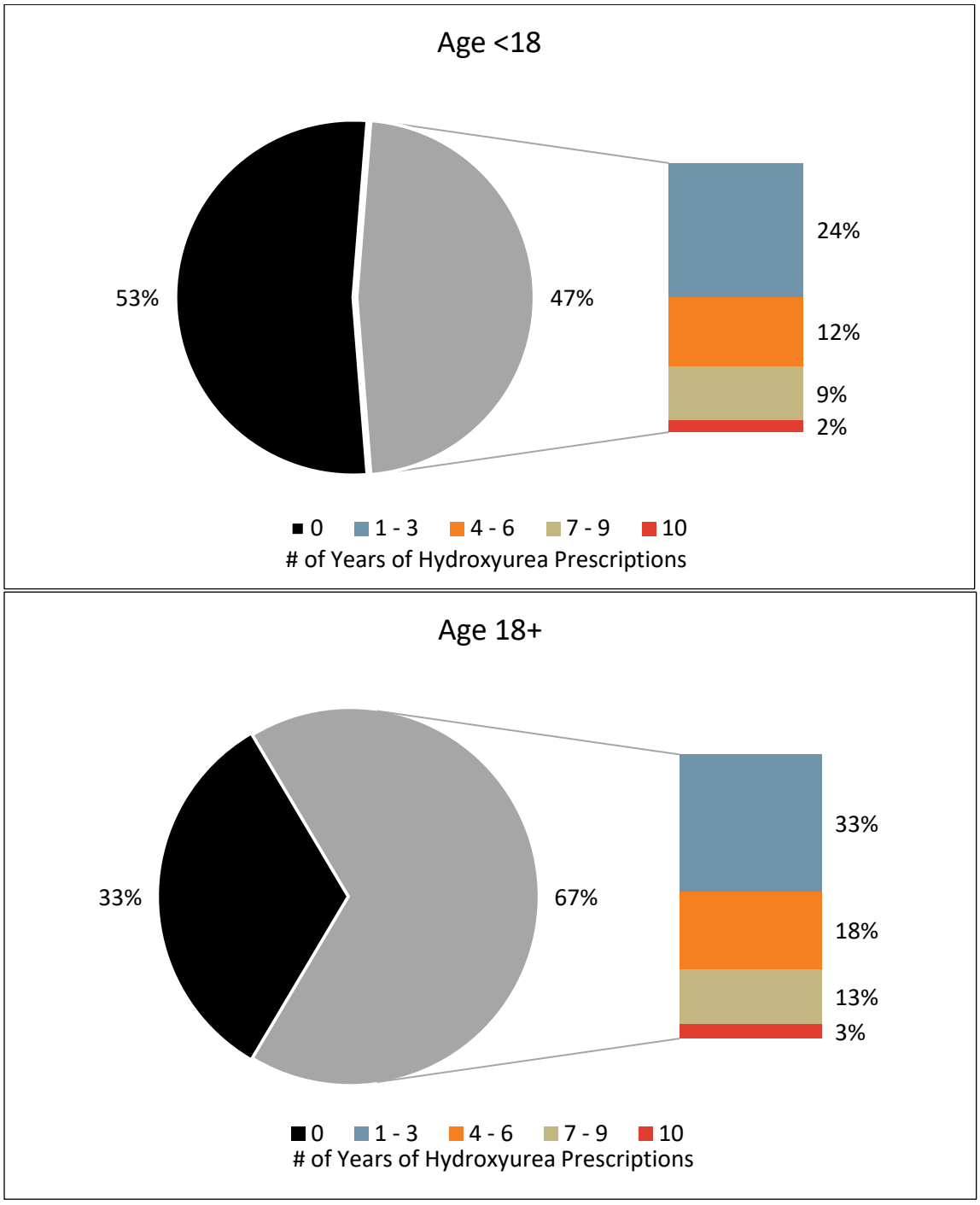
Table 11: Number and Percentage of Individuals with a Reported Prescription for Hydroxyurea by Year, Stratified by Type of SCD

	HbSS & HbS/β^0		HbSC		HbS/β^+		HbS/Other	
	Total	n (%)	Total	n (%)	Total	n (%)	Total	n (%)
2015	3,007	850 (28%)	1,091	52 (5%)	253	23 (9%)	37	4 (11%)
2016	3,645	1,203 (33%)	1,328	84 (6%)	322	28 (9%)	49	5 (10%)
2017	4,339	1,641 (38%)	1,580	133 (8%)	407	54 (13%)	59	8 (14%)
2018	4,716	1,784 (38%)	1,724	153 (9%)	450	54 (12%)	67	12 (18%)
2019	4,926	1,871 (38%)	1,806	162 (9%)	466	54 (12%)	65	12 (18%)
2020	5,176	1,910 (37%)	1,877	149 (8%)	482	46 (10%)	68	12 (18%)
2021	5,328	2,039 (38%)	1,930	154 (8%)	489	60 (12%)	76	12 (16%)
2022	5,492	2,121 (39%)	1,973	181 (9%)	511	76 (15%)	82	14 (17%)
2023	5,478	2,152 (39%)	1,978	189 (10%)	521	79 (15%)	78	11 (14%)
2024*	5,198	1,921 (37%)	1,866	189 (10%)	481	62 (13%)	77	9 (12%)
Ever Prescribed	7,033	4,108 (58%)	2,602	453 (17%)	669	166 (25%)	108	24 (22%)

*Data for 2024 represent partial data for that year.

Figure 11 displays the number of years during which individuals with HbSS and HbS/β^0 had at least one record for a hydroxyurea prescription. Among those under 18 years of age, 47% have at least one record of a hydroxyurea prescription during their follow-up period, while 23% have four or more years with a recorded prescription. Similarly, 67% of individuals aged 18 years and older have at least one hydroxyurea prescription, half (34%) of whom have a hydroxyurea prescription record for four or more years. This suggests that research projects involving a multi-year follow-up of individuals with a hydroxyurea prescription may be feasible using the Data Hub.

Figure 11: Percentage of Individuals with HbSS and HbS/β^0 by the Number of Years with Hydroxyurea Prescription Records, Stratified by Age Group



Laboratory Test Records

Laboratory testing is a vital component of routine and acute care for individuals with SCD. Real-world laboratory data can support and enhance research, help better understand disease progression and outcomes, and provide a metric for screening and preventive care implementation as well as therapeutic monitoring. While several clinically relevant laboratory tests are utilized for SCD clinical care and monitoring, this section includes information about a subset of tests for which code sets have been previously developed.

The information reported here includes Data Hub laboratory records from 2023. The following tables include data for the 6,896 individuals with one or more records of laboratory tests in 2023.

Table 12 displays the mean and median number of test records per individual and the number and percentage of individuals with no test records in 2023. On average, the hemoglobin and AST test were reported the most frequently, while urine creatine and albumin tests were reported the least frequently. The difference between the mean and median values suggests that a subset of individuals—likely those with multiple encounters with the healthcare system, either for acute visits or treatment monitoring—underwent many more tests, which led to an increase in the mean but not the median. For instance, the maximum number of individual hemoglobin test records seen in the dataset in 2023 was 294, which is substantially higher than the mean. Most individuals (93%) had at least one report of a hemoglobin test, while only 30% reported a urine albumin test.

Table 12: Mean and Median Number of SCD-Specific Laboratory Tests per Individual and Percentage of Individuals with Zero Tests in 2023

Laboratory Test	Mean Number of Tests/Individual	Median Number of Tests/Individual	N (%) Individuals with Zero Tests
Hemoglobin	10.8	5	491 (7%)
Total Bilirubin	8.4	3	1,543 (22%)
AST	11.3	4	1,536 (22%)
LDH	2.8	0	3,543 (51%)
Serum Creatinine	6.8	2	1,179 (17%)
Urine Creatinine	0.9	0	4,236 (61%)
Urine Albumin	0.8	0	4,795 (70%)

For all the tests considered, Figure 12 indicates that the mean number of laboratory test records increases with age. Among individuals under 18 years of age, hemoglobin has the highest mean number of records in 2023, while among individuals aged 18 and older, AST has the highest mean number of records in the Data Hub.

Figure 12: Mean Number of Laboratory Records in 2023 per Individual, Stratified by Age Group

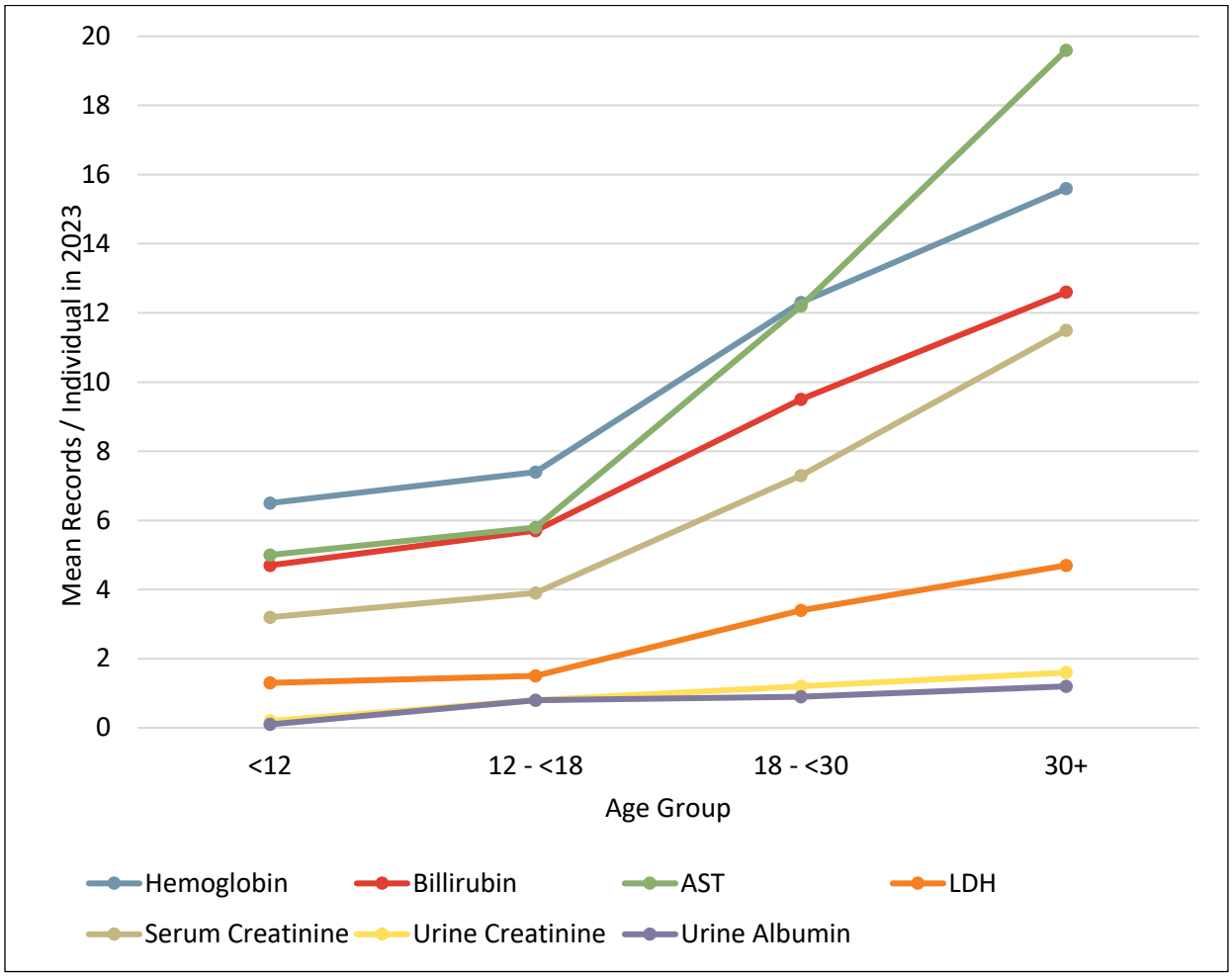


Table 13 presents the mean number of laboratory records for 2023 stratified by type of SCD and pediatric/adult age groups. For both age groups, a higher number of laboratory records are available for individuals with HbSS and HbS/β^0 than those with HbSC and HbS/β^+. This may reflect differences with respect to laboratory collection frequency related to age, disease severity, and monitoring associated with treatment utilization (e.g., hydroxyurea and chronic blood transfusion).

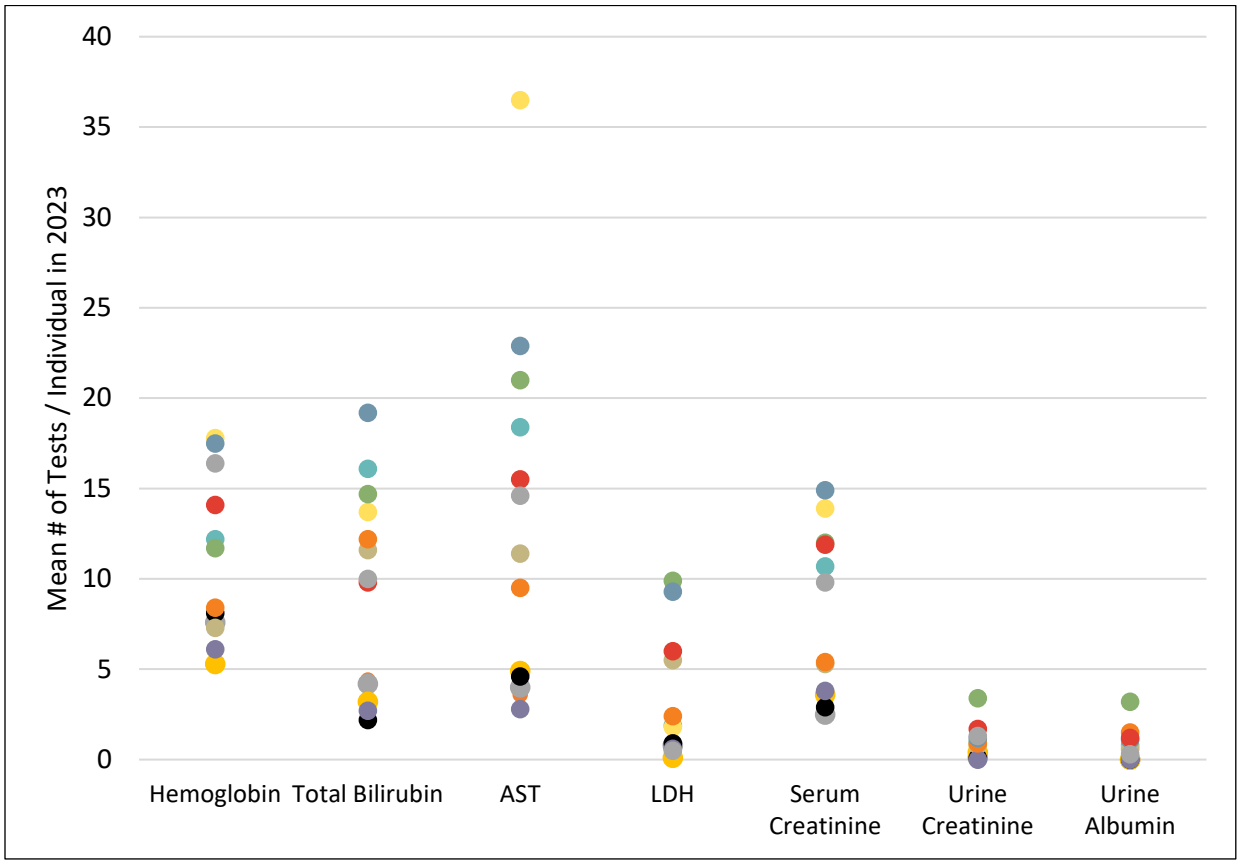
Table 13: Mean Number of SCD-Specific Records of Laboratory Tests in 2023, Stratified by Type of SCD and Age Group

	HbSS & HbS/β^0		HbSC		HbS/β^+		HbS/Other	
Age Group	<18	18+	<18	18+	<18	18+	<18	18+
Population	2,085	2,594	802	908	202	239	44	22
	Mean	Mean	Mean	Mean	Mean	Mean	Mean	Mean
Hemoglobin	8.6	16.5	3.6	8.3	2.8	9.9	3.6	17.0
Total Bilirubin	6.4	13.0	2.5	6.7	2.1	8.7	2.4	10.9
AST	6.6	18.6	2.8	10.3	2.5	14.7	2.8	17.9
LDH	1.7	4.7	0.7	2.7	0.7	2.8	0.8	1.3
Serum Creatinine	4.3	10.9	1.8	6.4	1.7	7.7	1.6	11.3
Urine Creatinine	0.5	1.5	0.3	1.2	0.2	1.2	0.4	0.7
Urine Albumin	0.4	1.2	0.3	0.9	0.3	0.8	0.5	0.2

Data calculated as per data submissions received by December 31, 2024.

Figure 13 illustrates the distribution of the mean number of laboratory records per person across the participating Data Hub sites in 2023. Each dot represents one site. AST shows the most variability in absolute mean value across the Data Hub sites, ranging from a value of 2.8 to 36.5, while urine albumin has the least variability, with means ranging from 0 to 3.2. Further investigation is needed to determine whether this variability reflects differences in care practices vs. differences in the inclusion of laboratory records to patients' EHRs across the Data Hub sites.

Figure 13: Mean Number of Laboratory Reports per Individual by Site in 2023



Healthcare Utilization

The data discussed in this section are restricted to the 13 sites that reported inpatient encounters and the 11 sites that reported emergency room encounters. Due to the variability across centers regarding the reporting of inpatient encounters, the mean numbers of records per person have not been presented. Further research is needed to understand the nature of the outpatient encounters recorded in the Data Hub before metrics, such as percentages and means, can be presented.

At some Data Hub sites, visits that began in the emergency department and resulted in hospital admission were coded as one encounter. Within this report, these are counted as inpatient encounters and, therefore, are not included in the counts for emergency department visits.

There are also differences across Data Hub sites with respect to how inpatient encounters are recorded and whether there is one record per stay or separate records for each inpatient day. Hence, differences in the number of inpatient records reported across sites may reflect differences in how the sites reported the data and not in the actual rates of inpatient stays.

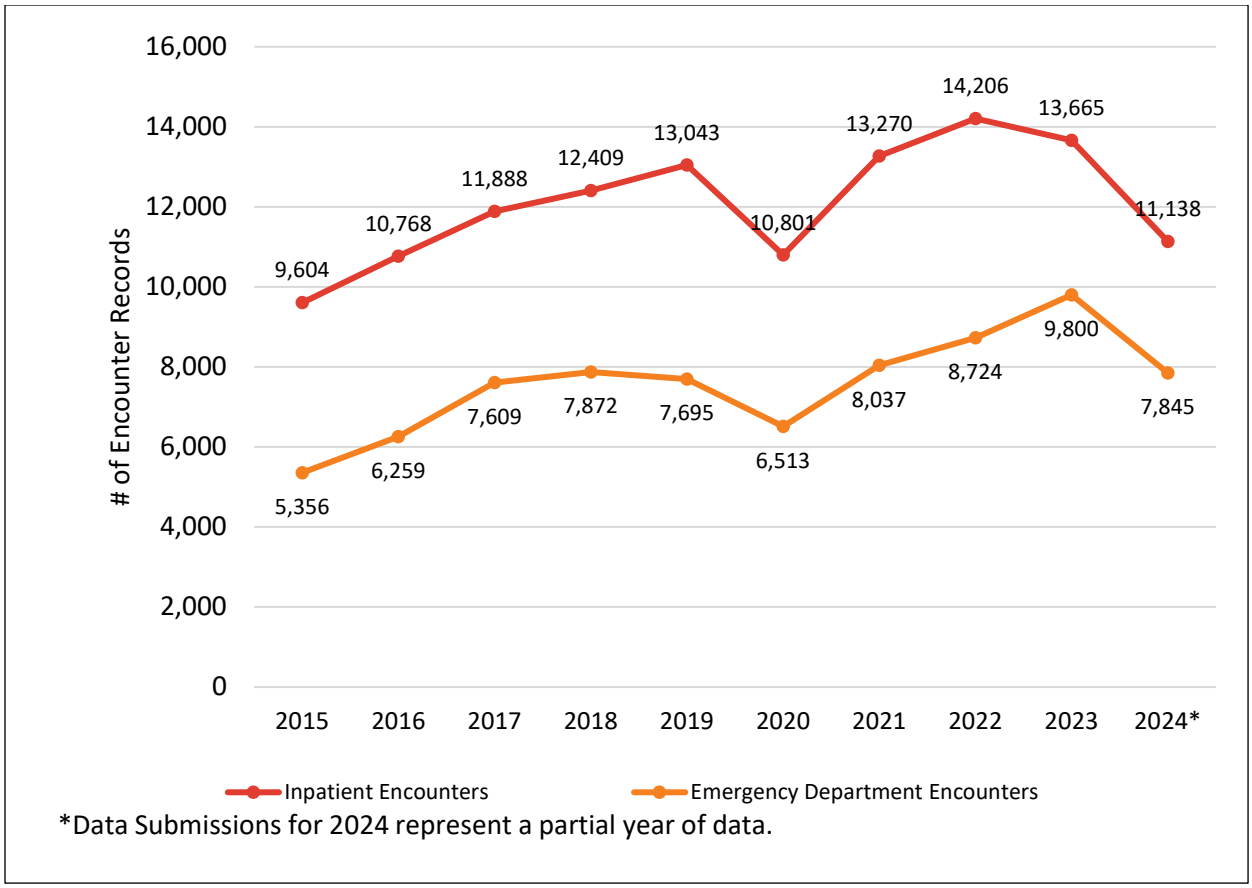
The data presented reflects the number of emergency department and inpatient encounter records added to the Data Hub during each year. Overall, there were 75,710 emergency department records and 120,792 inpatient records. The overall and annual numbers are impacted by differences across Data Hub sites in terms of how these encounters are recorded and coded.

Figure 14 displays the numbers of emergency department and inpatient encounters recorded in the Data Hub for each year.

It should be noted that the number of sites contributing data was not consistent across the years, particularly during the early years, so care should be taken when inferring changes in care utilization.

As expected, the number of records for 2024 is relatively lower but is anticipated to increase as data from late 2024 are processed and submitted. A decrease in inpatient and emergency department encounters is observed in 2020 compared to 2019 and 2021, with comparable numbers of sites contributing data. This may reflect the changes that occurred in care utilization during the early months of the COVID-19 pandemic.

Figure 14: Number of Emergency Department and Inpatient Encounters by Year



To provide a sense of the number of encounter records available for research,

Figure 15 provides a breakdown of the number of emergency department and inpatient encounters by age group. The age groups are defined based on the individual's age at the time of the encounter. The differences in the coding and recording of emergency department and inpatient encounters across Data Hub sites need to be considered when evaluating these data. Inferences about healthcare utilization should not be made.

Figure 15: Number of Emergency Department and Inpatient Encounters, Stratified by Age Group

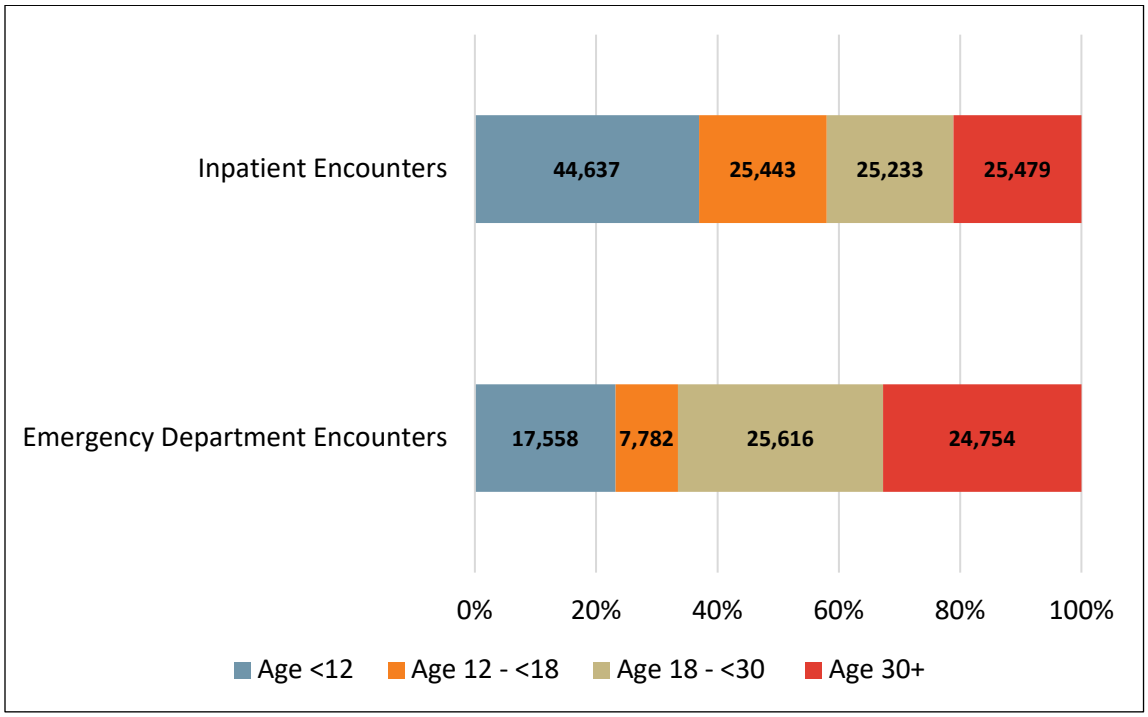


Table 14 displays the number of individuals with at least one emergency department or inpatient visit recorded at any point during their active Data Hub follow-up period (up to 10 years) between 2015 and 2024. The proportion of individuals who have one or more emergency department encounter records does not differ substantially between the HbSS and HbS/β^0 and the HbSC and HbS/β^+ SCD type groups. The proportion of individuals with inpatient encounter records is highest among individuals with HbSS and HbS/β^0 under the age of 18 years.

Table 14: Number and Percentage of Individuals with a Recorded Emergency Department or Inpatient Encounter, Stratified by Type of SCD and Age

Type of SCD	HbSS & HbS/β^0		HbSC		HbS/β^+		HbS/Other	
Age Group	<18	18+	<18	18+	<18	18+	<18	18+
	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)
Total Population	1,992	3,564	822	1,326	217	349	36	37
Emergency Department	1,580 (79%)	2,462 (69%)	643 (78%)	937 (71%)	154 (71%)	266 (76%)	19 (53%)	27 (73%)
Total Population	2,731	4,102	1,085	1,434	281	372	62	39
Inpatient	2,252 (82%)	3,100 (76%)	790 (73%)	999 (70%)	201 (72%)	273 (73%)	46 (74%)	29 (74%)

Age is calculated as of December 31, 2024.

Conclusion and Advancing Research Together

The Data Hub holds tremendous potential to advance our understanding of the natural history, clinical care, and outcomes of individuals living with SCD. This report is intended to provide researchers and clinicians with an overview of the vast extent of the data elements available in the Data Hub, insights into potential sample sizes, and an understanding of the number and types of measurements captured—all to better inform research proposals. It also highlights areas of uncertainty that exist when analyzing EHR data.

We hope that this report inspires you to formulate impactful research questions that the Data Hub can help address and that, together, we can continue to advance science and improve the lives of people living with SCD.

References

Centers for Disease Control and Prevention. (2023). *Data and statistics on sickle cell disease.* https://www.cdc.gov/sickle-cell/data/index.html.

Wood WA, P Marks, RM Plovnick, K Hewitt, DS Neuberg, S Walters, BK Dolan, EA Tucker, CS Abrams, AA Thompson, KC Anderson, P Kluetz, A Farrell, D Rivera, M Gertzog, G Pappas. (2021). *ASH Research Collaborative: A real-world data infrastructure to support real-world evidence development and learning healthcare systems in hematology.* Blood Advances: 5(23)5429–38.

Appendix

Appendix A: Electronic Health Record Code List for SCD Patient Identification

The code list provided below represents the inclusion criteria for the patients included in the SCD Data Hub. Sites may

- use ICD codes only,
- use SNOMED codes only,
- use both ICD and SNOMED codes, or
- use a local custom patient identification option (Note: Approval for sites that request the local patient identification option is considered case-by-case by the applicable committee.)

Code System	Code	Label	Athena Concept ID Map (for OMOP pathway only)
ICD10CM	D57.0	Sickle cell anemia with crisis	45571634
ICD10CM	D57.00	HbSS disease with crisis, unspecified	45595765
ICD10CM	D57.01	HbSS disease with acute chest syndrome	45542708
ICD10CM	D57.02	HbSS disease with splenic sequestration	45561923
ICD10CM	D57.03	HbSS disease with cerebral vascular involvement	725216
ICD10CM	D57.09	HbSS disease with crisis and other specified complication	725217
ICD10CM	D57.1	Sickle cell disease without crisis	35206724
ICD10CM	D57.2	Sickle cell/Hb-C disease	1567846
ICD10CM	D57.20	Sickle cell/Hb-C disease without crisis	45542710
ICD10CM	D57.21	Sickle cell/Hb-C disease with crisis	45542711
ICD10CM	D57.211	Sickle cell/Hb-C disease with acute chest syndrome	45576406
ICD10CM	D57.212	Sickle cell/Hb-C disease with splenic sequestration	45571635
ICD10CM	D57.213	Sickle cell/Hb-C disease with cerebral vascular involvement	725218
ICD10CM	D57.218	Sickle cell/Hb-C disease with crisis and other specified complication	725219
ICD10CM	D57.219	Sickle cell/Hb-C disease with crisis, unspecified	45561924
ICD10CM	D57.4	Sickle cell thalassemia	1567847
ICD10CM	D57.40	Sickle cell thalassemia without crisis	45561925
ICD10CM	D57.41	Sickle cell thalassemia with crisis	1567848
ICD10CM	D57.411	Sickle cell thalassemia with acute chest syndrome	45586115
ICD10CM	D57.412	Sickle cell thalassemia with splenic sequestration	45561926
ICD10CM	D57.413	Sickle cell thalassemia, unspecified, with cerebral vascular involvement	725220
ICD10CM	D57.418	Sickle cell thalassemia, unspecified, with crisis and other specified complication	725221
ICD10CM	D57.419	Sickle cell thalassemia with crisis, unspecified	45566702
ICD10CM	D57.42	Sickle cell thalassemia beta zero without crisis	725222
ICD10CM	D57.43	Sickle cell thalassemia beta zero with crisis	725223
ICD10CM	D57.431	Sickle cell thalassemia beta zero with acute chest syndrome	725224
ICD10CM	D57.432	Sickle cell thalassemia beta zero with splenic sequestration	725225

Code System	Code	Label	Athena Concept ID Map (for OMOP pathway only)
ICD10CM	D57.433	Sickle cell thalassemia beta zero with cerebral vascular involvement	725226
ICD10CM	D57.438	Sickle cell thalassemia beta zero with crisis and other specified complication	725227
ICD10CM	D57.439	Sickle cell thalassemia beta zero with crisis, unspecified	725228
ICD10CM	D57.44	Sickle cell thalassemia beta plus without crisis	725229
ICD10CM	D57.45	Sickle cell thalassemia beta plus with crisis	725230
ICD10CM	D57.451	Sickle cell thalassemia beta plus with acute chest syndrome	725231
ICD10CM	D57.452	Sickle cell thalassemia beta plus with splenic sequestration	725232
ICD10CM	D57.453	Sickle cell thalassemia beta plus with cerebral vascular involvement	725233
ICD10CM	D57.458	Sickle cell thalassemia beta plus with crisis and other specified complication	725234
ICD10CM	D57.459	Sickle cell thalassemia beta plus with crisis, unspecified	725235
ICD10CM	D57.8	Other sickle cell disorders	1567849
ICD10CM	D57.80	Other sickle cell disorders without crisis	45552357
ICD10CM	D57.81	Other sickle cell disorders with crisis	1567850
ICD10CM	D57.811	Other sickle cell disorders with acute chest syndrome	45595766
ICD10CM	D57.812	Other sickle cell disorders with splenic sequestration	45600606
ICD10CM	D57.813	Other sickle cell disorders with cerebral vascular involvement	725236
ICD10CM	D57.818	Other sickle cell disorders with crisis and other specified complication	725237
ICD10CM	D57.819	Other sickle cell disorders with crisis, unspecified	45581324
ICD9CM	282.41	Sickle cell thalassemia without crisis	44829908
ICD9CM	282.42	Sickle cell thalassemia with crisis	44824096
ICD9CM	282.6	Sickle cell disease	44820699
ICD9CM	282.60	Sickle cell disease, unspecified	44831065
ICD9CM	282.61	HbSS disease without crisis	44827638
ICD9CM	282.62	HbSS disease with crisis	44825276
ICD9CM	282.63	Sickle cell/Hb-C disease without crisis	44835766
ICD9CM	282.64	Sickle cell/Hb-C disease with crisis	44819529
ICD9CM	282.68	Other sickle cell disease without crisis	44836942
ICD9CM	282.69	Other sickle cell disease with crisis	44825277
SNOMED CT	23269001	Double heterozygous sickling disorder (disorder)	4049095
SNOMED CT	25472008	Sickle cell-hemoglobin D disease (disorder)	4103763
SNOMED CT	35434009	Sickle cell-hemoglobin C disease (disorder)	24006
SNOMED CT	36472007	Sickle cell-thalassemia disease (disorder)	315523
SNOMED CT	47024008	Sickle cell-hemoglobin E disease (disorder)	4264046
SNOMED CT	49938009	Sickle cell	4177967
SNOMED CT	61777009	Thalassemia-hemoglobin C disease (disorder)	4264909
SNOMED CT	127040003	Sickle cell-hemoglobin SS disease (disorder)	22281
SNOMED CT	127041004	Sickle cell-beta-thalassemia (disorder)	4130067
SNOMED CT	127042006	Sickle cell-beta^+^-thalassemia (disorder)	4130678
SNOMED CT	127043001	Sickle cell-beta^0^-thalassemia (disorder)	4130068
SNOMED CT	127044007	Sickle cell-delta beta^0^-thalassemia (disorder)	4131913

Code System	Code	Label	Athena Concept ID Map (for OMOP pathway only)
SNOMED CT	127045008	Sickle cell anemia with coexistent alpha-thalassemia (disorder)	4131914
SNOMED CT	127047000	Sickle cell-hemoglobin Lepore disease (disorder)	4131126
SNOMED CT	127048005	Sickle cell-hemoglobin O Arab disease (disorder)	4130190
SNOMED CT	416180004	Hemoglobin SS disease without crisis (disorder)	30683
SNOMED CT	416214006	Sickle cell-hemoglobin D disease without crisis (disorder)	4166099
SNOMED CT	416290001	Hemoglobin S sickling disorder without crisis (disorder)	4166115
SNOMED CT	416484003	Sickle cell-hemoglobin E disease with crisis (disorder)	4219568
SNOMED CT	416638004	Sickle cell-hemoglobin E disease without crisis (disorder)	4211932
SNOMED CT	416826005	Sickle cell-thalassemia disease with crisis (disorder)	443738
SNOMED CT	417048006	Sickle cell-thalassemia disease without crisis (disorder)	321263
SNOMED CT	417425009	Hemoglobin SS disease with crisis (disorder)	26942
SNOMED CT	417517009	Sickle cell-hemoglobin C disease with crisis (disorder)	443721
SNOMED CT	417683006	Sickle cell-hemoglobin C disease without crisis (disorder)	443726
SNOMED CT	417748003	Sickle cell-hemoglobin D disease with crisis (disorder)	4167364
SNOMED CT	440206000	Hemoglobin SS disease with vaso-occlusive crisis (disorder)	4209085
SNOMED CT	444108000	Acute sickle cell splenic sequestration crisis (disorder)	40485018
SNOMED CT	769167005	Vaso-occlusive pain co-occurrent and due to sickle cell disease	35626024
SNOMED CT	372146004	Acute chest syndrome (disorder)	254062
SNOMED CT	417357006	Sickling disorder due to hemoglobin S (disorder)	4213628
SNOMED CT	417279003	Hemoglobin S sickling disorder with crisis (disorder)	4216915