NON-INTERVENTIONAL (NI) STUDY PROTOCOL

Study Information

Title	Effectiveness of voxelotor in individuals with sickle cell disease and a history of red blood cell transfusions: A non-interventional, retrospective cohort study using real-world data in the United States
Protocol number	C5341060
Protocol version identifier	V1.0.
Date	01 August 2024
EU Post Authorization Study (PAS) register number	EUPAS1000000225
Active substance	B06AX03 (voxelotor)
Medicinal product	Oxbryta
Research question and objectives	Among patients with sickle cell disease (SCD) and a history of transfusions, the following objectives will be compared between those treated with voxelotor and matched patients who are not treated with voxelotor: • **Primary objective:** Compare the change in red blood cell (RBC) transfusions per patient per year (PPPY) from baseline (one year prior to index date) to follow-up (up to one year after index date). • **Exploratory objective 1:** Compare the rate of SCD-related outcomes including vaso-occlusive crises (VOCs), stroke, acute chest syndrome [ACS], and priapism from baseline (one year prior to index date) to follow-up (up to one year after index date). • **Exploratory objective 2:** Compare the number of RBC transfusions PPPY during follow-up (including those with 0 transfusions). Patients with more than a 30% reduction in transfusions from baseline will also be reported. • **Exploratory objective 3:** Compare the rate of RBC transfusion-associated complications (iron overload, iron chelation use, delayed hemolytic transfusion reactions, alloimmunization) during the follow-up period. All objectives will be reported overall and stratified by the number

	of transfusions received during the baseline period**.**
Country of study	USA
Author	[redacted]

1. TABLE OF CONTENTS

2. LIST OF ABBREVIATIONS

Abbreviation	Definition
ACS	Acute Chest Syndrome
ANCOVA	Analysis of Covariance
ATT	Average Treatment Effect of the Treated
CEM	Coarsened-Exact-Matching
CI	Confidence Interval
CCI	Charlson Comorbidity Index
COVID-19	Coronavirus Disease 2019
DBC	Database Connection
DHTR	Delayed Hemolytic Transfusion Reaction
EC	Ethics Committee
EHR	Electronic Health Record
ESA	Erythropoietin Stimulating Agent
FDA	Food and Drug Administration
GPP	Good Pharmacoepidemiology Practices
Hb	Hemoglobin
HbS	Sickle Hemoglobin
HCPCS	Healthcare Common Procedure Coding System
HSCT	Hematopoietic stem cell transplantation
HU	Hydroxyurea
ICD	International Classification of Diseases
IPTW	Inverse Probability Treatment Weighting
IRB	Institutional Review Board

Abbreviation	Definition
ISPE	International Society for Pharmacoepidemiology
ITT	Intention-to-Treat
KHM	Komodo's Healthcare Map
MSM	Marginal Structural Model
NDC	National Drug Code
PI	Principal Investigator
PPPY	Per-Patient-Per-Year
PPV	Positive Predictive Value
PS	Propensity Score
RBC	Red Blood Cell
RWD	Real-World Data
RWE	Real-World Evidence
SAP	Statistical Analysis Plan
SOC	Standard of Care
SPIFD	Structured Process to Identify Fit-for-Purpose Data
US	United States
VIF	Variance Inflation Factor
VOC	Vaso-occlusive Crises

3. RESPONSIBLE PARTIES

Principal Investigator(s) of the Protocol

Name, Degree(s)	Job Title	Affiliation
[redacted]	Rare Hematology Integration Lead—Rare Disease Pfizer	Pfizer
[redacted]	Hematology Lead, RWE Science and Epidemiology	Pfizer
[redacted]	SCD Global Medical Director	Pfizer
[redacted]	Hematology RWE Scientist	Pfizer
[redacted]	Global Medical Lead-Voxelotor	Pfizer
[redacted]	Senior Director, US Medical Affairs	Pfizer
[redacted]	Biostatistician	Pfizer
[redacted]	Senior Director, HTA, Value & Evidence	Pfizer
[redacted]	Director, Science	Aetion
[redacted]	Senior Scientist, Science	Aetion
[redacted]	Associate Director, Science	Aetion

4. ABSTRACT

Title: Effectiveness of voxelotor in individuals with sickle cell disease and a history of red blood cell (RBC) transfusions: A non-interventional, retrospective cohort study using real-world data in the United States

Rationale and background: The current treatment landscape for sickle cell disease (SCD) is limited and some options utilized for the management of symptoms have associated risks and can be burdensome for patients. One such treatment is RBC transfusions, which are indicated to treat or prevent complications like severe anemia, acute multi-organ failure, vaso-occlusive crises (VOCs), and stroke. However, there are risks associated with each transfusion procedure, such as iron overload, which can lead to end-organ damage, or alloimmunization, which can lead to delayed hemolytic transfusion reactions (DHTR) or infections that can be life-threatening for individuals with SCD. Furthermore, transfusion administration requires frequent venous access, which can result in excessive burden to patients.

Voxelotor (Oxbryta®) is an SCD treatment that inhibits the polymerization of sickle hemoglobin (Hb) polymerization and has shown clinical benefit on Hb levels, which may limit the need for RBC transfusions in the management of SCD. Voxelotor was granted accelerated approval in November 2019 from the US Food and Drug Administration (FDA). Voxelotor has meanwhile also been recommended by the ASH Guideline Monitoring Expert Working Group for consideration as a preventive maintenance therapy to improve the baseline Hb of patients at high risk of DHTR if red cell transfusion is required.[1] Due to challenges in conducting a randomized trial in this setting, this study aims to conduct a robust real-world evidence (RWE) study to determine if treatment with voxelotor reduces RBC transfusion rate compared to a similar population of SCD patients not treated with voxelotor matched on age and calendar date.

Research question and objectives: For individuals with SCD and a history of transfusions (ie, at least two within one year of index), does voxelotor reduce the number of RBC transfusions over the following year compared to a control group of matched individuals with SCD who are not treated with voxelotor?

The study will include the following specific objectives:

- **Primary objective:** Compare the change in RBC transfusions per patient per year (PPPY) from baseline (one year prior to index date) to follow-up (up to one year after index date).

- **Exploratory objective 1:** Compare the rate of SCD-related outcomes including VOCs, stroke, acute chest syndrome [ACS], and priapism from baseline (one year prior to index date) to follow-up (up to one year after index date)

[1] ASH Chou ST, Hendrickson JE, Fasano RM. Transfusion therapy for sickle cell disease: what's new? *Blood Adv*. 2023;7(11):2551-2553. doi:10.1182/bloodadvances.2022009283.

- **Exploratory objective 2:** Compare the number of RBC transfusions PPPY during follow-up (including those with 0 transfusions). Patients with more than a 30% reduction in transfusions from baseline will also be reported.

- **Exploratory objective 3:** Compare the rate of RBC transfusion-associated complications (iron overload, iron chelation use, delayed hemolytic transfusion reactions, alloimmunization) during the follow-up period.

All objectives will be reported overall and stratified by the number of transfusions received in baseline**.**

Study design: This is a non-interventional, retrospective cohort study of individuals with SCD using a US real-world dataset composed of linked closed claims, proprietary voxelotor prescription claims, and laboratory data. Propensity score (PS) methods will be utilized to adjust for measured confounding covariates in comparative analyses.

Population: The study population will include patients with SCD between the ages of 12 and 85 years of age who are enrolled in a health insurance plan (commercial, Medicare, Medicaid) continuously during the baseline period. The index selection period will be 25 November 2019 (date of FDA accelerated approval for voxelotor) through one year prior to the end of the study period (TBD). The exposed group will include patients with a new prescription claim for voxelotor (no prior use 365 days prior), while the control group will include matched patients who do not have a prescription claim for voxelotor. The exposed group will index on the voxelotor prescription claim date and will be sequentially coarsened-exact-matched with up to two non-voxelotor patients based on age and calendar date. Both the exposed and control groups may be treated with other background standard of care (SOC) treatments, such as hydroxyurea, during the study period.

Variables: The exposure of interest will be a voxelotor prescription claim. Baseline characteristics and covariates for PS modeling will include demographic information (eg, age, sex, insurance type) as well as patient characteristics (eg, specialty provider type, common SCD comorbidities and treatment history, Hb levels, number of RBC transfusions, and healthcare utilization).

Data sources: The study will use the Komodo Healthcare Map (KHM) database linked to Claritas prescription data (proprietary voxelotor claims) and Quest Diagnostics lab data. Data sources are anchored on a closed system of administrative claims, including patient enrollment, physician, facility, and pharmacy claims.

Study size: The study will include all the eligible patients with an index date from 25 November 2019, until the end of the study period. An initial sample size assessment conducted with Komodo data through September 2023 identified ~94,000 patients with SCD, and among those patients, ~7,700 were treated with voxelotor. A conservative sample size estimate was calculated using a Welch's t-test with an adjustment factor of 1.65 which takes into account the lost-to-follow-up, variance inflation in the analysis.[26] With 177 patients treated with voxelotor and 354 matched patients not treated with voxelotor (1:2 matching ratio), the study will ensure an 80% power to detect at least a 1-unit (1 transfusion per day)

difference between the two groups (standard deviation = 3.0) in the change in RBC transfusions PPPY from baseline to follow-up.

Data analysis: Baseline covariates will be evaluated in the 365 days prior to index date using descriptive statistics (continuous variables: mean, standard deviation, median, interquartile range; categorical variables: number of patients, percent). A PS model will be fit using logistic regression after baseline balance of covariates is assessed using absolute standardized differences and clinical input. A stepwise approach will be utilized to select the appropriate PS method based on diagnostic criteria to estimate the average treatment effect of the treated (ATT). For the primary analysis, an as-treated approach will be utilized and censoring criteria will include the following: disenrollment from health insurance (greater than 45-day allowable gaps), death, discontinuation of voxelotor (with a 30-day allowable gap), or the occurrence of exclusion criteria.

After ensuring balance of baseline covariates using PS-matching or PS-weighting, change in number of transfusions PPPY will be analysed using generalized linear model with an identity link to estimate the difference in transfusion change rates between the two groups (primary objective). If any pre-index covariates that are found to be unbalanced after the PS method, it will be included in the regression as a covariate. The incidence rate change and its 95% CI will be reported. Absolute rate differences between groups will be reported to calculate the proportion of patients with any reduction in VOCs (exploratory objective 1) and >30% reduction in RBC transfusions (exploratory objective 2, endpoint 2); logistic regression will be used to assess the difference between groups for both objectives. Absolute rate differences will also be reported for each of the transfusion-associated secondary complications and secondary efficacy endpoints (exploratory objectives 1 and 3). Logistic regression will be used to compare the number of patients with zero RBC transfusions in follow-up among patients with at least 365 days of follow-up (exploratory objective 2, endpoint 3). Confidence intervals (95%) and/or standard deviations will be reported for analyses where applicable. Prior to assessing outcomes, the follow-up time in each arm will be described. For sensitivity analysis, transformations of the outcome may also be considered as appropriate. 95% confidence interval (CI) will be reported for the difference between the groups. Full details on the diagnostic checks and selection of a regression model will be provided in the statistical analysis plan (SAP).

5. AMENDMENTS AND UPDATES

None.

6. MILESTONES

Milestone	Planned Date
Estimated Protocol Approval Date	12 August 2024
Start of data collection	01 September 2024
End of data collection	31 January 2025
Registration in the HMA-EMA Catalogues of RWD studies	22 June 2024
Final study report	TBD; Q1 2025

7. RATIONALE AND BACKGROUND

Sickle cell disease (SCD) is a group of inherited progressive disorders caused by a mutation in the β-chain of hemoglobin (Hb), resulting in malformed, sickle-shaped red blood cells (RBCs) that are rigid, sticky, and die early.[1] Patients with SCD have acute and chronic complications from repeated RBC sickling, including hemolytic anemia and vaso-occlusion, which result in chronic pain, vaso-occlusive crises (VOCs), acute chest syndrome (ACS), and cumulative multi-organ damage, among other manifestations.[1] The most common genotype of SCD is referred to as sickle cell anemia (HbSS or HbS β0 thalassemia) where an individual produces mainly sickle hemoglobin (HbS), either from inheriting two genes that code for hemoglobin "S." (HbSS) or one gene that codes for HbS and one that does not produce any beta globin (HbS β0 thalassemia). Other subtypes include HbSC, a milder disease with fewer complications, and people with HbSβ+ thalassemia have some, albeit decreased healthy beta-globin resulting in HbA production and is less severe.[2] SCD is associated with severe disability, lower quality of life, and reduced life-expectancy by 25-30 years.[3] The most recent Global Burden of Disease estimates suggest the prevalence of SCD to be approximately 100,000 people in the United States (US) and more than 7.5 million people worldwide.[4]

Most current treatment options for SCD are aimed at management of symptoms to improve quality of life and survival but can be burdensome and associated with risk. SCD specific pharmacologic interventions, which include hydroxyurea (HU), L-glutamine, and crizanlizumab, are primarily indicated and used for prevention of pain crises. However, these treatments do often not address hemolytic anemia, can have a high resource burden (eg, requiring frequent monitoring or intravenous administration), are often not well-tolerated, and can be limited by poor adherence.[5-7]

An important non-pharmacologic intervention for SCD is RBC transfusion.[3] RBC transfusions are indicated for severe symptomatic anemia, acute complications like stroke or acute multi-organ failure, and/or secondary prevention of stroke or recurrent pain crises in individuals with SCD.[3] The two primary types of RBC transfusion, simple (addition of RBCs without removal of the patient's blood) and exchange (removal of the patient's blood and replacement with donor packed RBCs), have both acute and chronic indications. Treatment guidelines recommend acute transfusions for symptomatic anemia, acute chest syndrome, multiorgan failure, and preoperatively; chronic transfusions, usually administered monthly, are recommended for stroke prevention.[3] While blood transfusions can improve SCD complications, they are also associated with important risks that can occur at each transfusion: resultant iron overload can lead to end organ damage, alloimmunization can lead to delayed hemolytic transfusion reactions and mortality, and transfusion transmitted infections can cause organ damage and mortality.[3,8-12] Additionally, transfusion administration can be burdensome to patients due to the need for frequent venous access, which can cause pain, psychological distress, and interruptions to daily life.[13] There is an unmet need for new therapies and interventions that address the clinical complications of SCD while mitigating the patient burden.

Oxbryta® (voxelotor) is one of only four Food and Drug Administration (FDA) approved medicines for SCD and is approved for treatment of SCD in adults and pediatric patients ≥4 years of age in the United States (US). Once-daily oral administration of voxelotor has been shown to reduce red-cell sickling, blood viscosity and improve red-cell deformability in vitro and to extend red-cell half-life and reduce anemia and hemolysis in vivo.[14] Voxelotor is a first-in-class inhibitor of sickle Hb polymerization that can be self-administered as an oral treatment and does not require laboratory monitoring. Evidence from clinical studies has demonstrated that voxelotor has a beneficial effect on Hb levels, a critical factor in the decision for an RBC transfusion. However, no clinical trials have demonstrated the impact of voxelotor on rates of RBC transfusion. A randomized clinical trial (RCT) to evaluate RBC transfusion rates is likely to be limited by a small voxelotor naïve patient pool and lack of clinical equipoise to suggest an uncertain treatment effect. While some preliminary real-world evidence (RWE) supports a reduction in RBC transfusions associated with voxelotor use, these studies are limited by small sample sizes and lack of comparator populations.[15,16] Due to the challenges in conducting a RCT in this setting and limitations in previous RWE analyses, this study aims to determine if voxelotor reduces transfusions compared to the current care in the general population of individuals with SCD but not treated with voxelotor using a robust comparative RWE study design.

The goal of this non-interventional, retrospective cohort study is to evaluate the effect of voxelotor on the reduction of RBC transfusions in individuals with SCD ≥12 years of age in the United States (US), who have a history of RBC transfusions using real-world data. Results from this study will help inform management of those with SCD and intended to support regulatory decision making and labeling.

This non-interventional study is designated as a PASS and is conducted voluntarily by Pfizer.

8. RESEARCH QUESTION AND OBJECTIVES

The aim of this study is to evaluate if individuals in the US with SCD treated with voxelotor receive fewer RBC transfusions compared to matched controls who are not treated with voxelotor.

Among patients with SCD and a history of transfusions, the following specific objectives will be included to compare RBC transfusion and other clinical outcomes between those treated with voxelotor and matched patients with SCD who are not treated with voxelotor:

- **Primary objective:** Compare the change in RBC transfusions per patient per year (PPPY) from baseline (one year prior to index date) to follow-up (up to one year after index date).

- **Exploratory objective 1:** Compare the rate of SCD-related outcomes including VOCs, stroke, acute chest syndrome [ACS], and priapism from baseline (one year prior to index date) to follow-up (up to one year after index date).

- **Exploratory objective 2:** Compare the number of RBC transfusions PPPY during follow-up (including those with 0 transfusions. Patients with a 30% reduction in transfusions from baseline will also be reported.

- **Exploratory objective 3:** Compare the rate of RBC transfusion-associated complications (iron overload, iron chelation use, delayed hemolytic transfusion reactions, alloimmunization) during the one-year follow-up period.

All objectives will be reported overall and stratified by the number of transfusions received in baseline**.**

9. RESEARCH METHODS

This will be a non-interventional, retrospective cohort study of patients with SCD using a US real-world dataset composed of linked closed commercial claims, proprietary voxelotor prescription claims, and laboratory data. The goal of this study is to evaluate the effectiveness of voxelotor in reducing the number of RBC transfusions (defined assuming at most one transfusion per day) compared to matched control patients with SCD who do not use voxelotor. Other SCD-related clinical outcomes and complications associated with transfusions will be explored. Propensity score (PS) weighting will be utilized to adjust for measured confounding covariates including patient demographics, comorbidities and variations in treatment (eg, concurrent medications).

9.1. Study Design

This is a non-interventional, retrospective cohort study of individuals with SCD using a US real-world dataset composed of linked closed claims, proprietary voxelotor prescription claims, and laboratory data. Propensity score (PS) methods will be utilized to adjust for measured confounding covariates in comparative analyses.

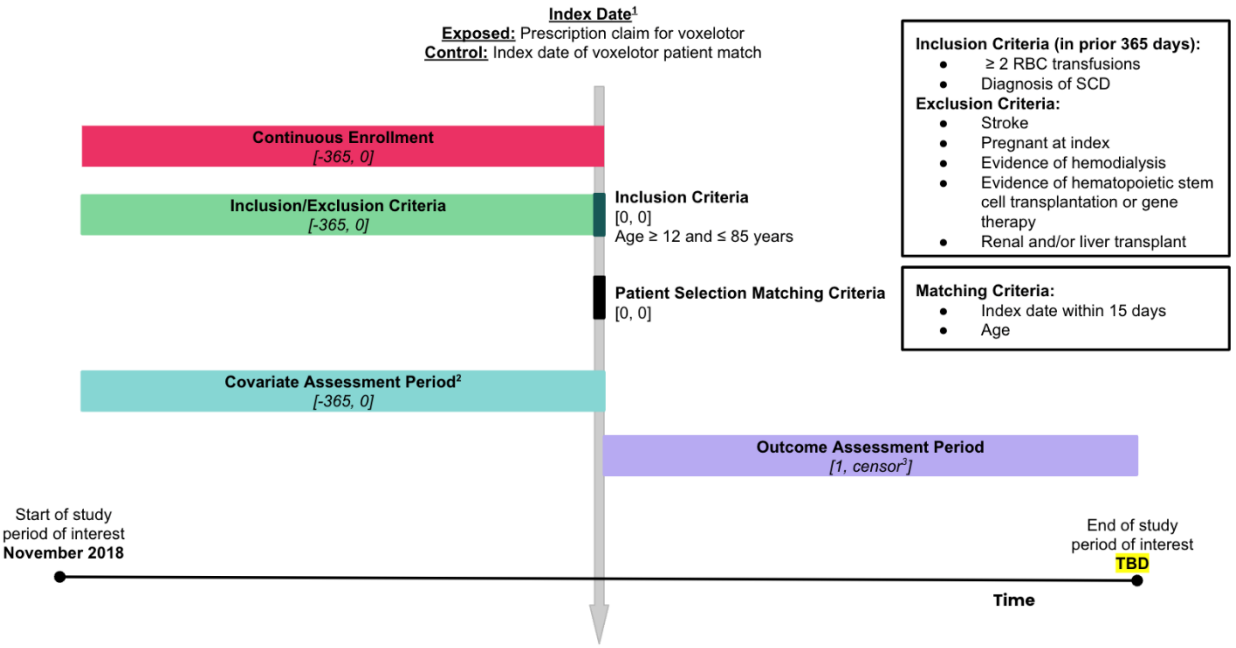
9.2. Setting

The study population will include patients with SCD between the ages of 12 and 85 years who are enrolled in a health insurance plan (commercial, Medicare, Medicaid) continuously during the 365 days prior to and including the index date (baseline). The index selection period is 25 November 2019 (date of FDA accelerated approval for voxelotor) through one year prior to the end of the study period. The exposed group will include patients with a new prescription claim for voxelotor (no prior use 365 days prior to index date), while the control group will include matched patients who do not have a prescription claim for voxelotor. Non-voxelotor controls will include those receiving no SCD modifying therapies as well as those receiving non-voxelotor SCD treatments to be reflective of clinical practice. Patients who are receiving chronic transfusion therapy for stroke prevention will be excluded from this analysis. Both the exposed and control groups may be treated with other SOC treatments such as HU.

After applying the inclusion/ exclusion criteria, to determine patient index date, the exposed group will index on the first voxelotor prescription claim date and will be sequentially matched using coarsened-exact-matching (CEM) with up to two non-voxelotor patients based on age and calendar date to achieve balance between groups.[17] All exposed patients will be sampled before selecting control patients so that there will be no potential for voxelotor patients to be included in the control group. Age is a critical matching criterion, as patients with SCD have better access to care before adulthood and accumulate comorbidities through

the natural course of the disease.[18] The matched controls will be selected from the pool of patients who meet all the eligibility criteria below and are the same age within 15 days of the voxelotor claim; the patients who meet the criteria closest to the matched voxelotor prescription claim will be prioritized. Depending on available matches within 15 days of the voxelotor claim, this criterion may be widened to preserve sample size (eg, within 30 or 45 days). Voxelotor patients will be matched in chronological order where up to two referent non-voxelotor controls will be identified for each voxelotor patient before additional referent non-voxelotor matches are identified for the remaining exposure group patients. Patients will be sampled once to either the matched voxelotor or non-voxelotor cohort and contribute to one group only. This matching approach ensures that all referent patients are matched to a voxelotor patient within 15 days of index date, mitigating potential selection bias by controlling for calendar time and thus any secular trends in available therapies, clinical guidelines, and other confounders that may occur during the study period (eg, COVID-19 pandemic). Available matches depend on the number of eligible patients available in the data source.

Figure 1. Study Design Schematic



Abbreviations: RBC: Red blood cell; SCD: Sickle cell disease; TBD: To be decided

1. Index date for the exposed group will be the date of voxelotor prescription claim date and will be sequentially coarsened-exact-matched with up to 2 non-voxelotor patients based on age and calendar date
2. The full list of potential covariates including definitions can be found in Table 3
3. Follow-up will end at the earliest of the following events: 365 days after index, death, disenrollment, pregnancy, end of data, or discontinuation of voxelotor (voxelotor arm).

This study will use administrative closed claims data between November 2018 (one year prior to voxelotor FDA accelerated approval in the US) and the end of the study period; the matched controls will index within 15 days (before or after) of their matched voxelotor counterpart. For example, if a voxelotor patient has an index date of 20 June 2022, a control

patient meeting all eligibility criteria and the same age as the voxelotor patient can have an index date between June 5th and July 5th. If there are not enough available matches within 15 days of a voxelotor index date, a 30-day or 45-day window may be utilized instead. The baseline period will include the 365 days prior to and including the index date and follow-up will start one day after the index date until censoring.

Table 1. Summary of Study Time Periods

Period	Description	Time frame
Study period	The full range of available data for the study assessment periods	01 November 2018 - *TBD*
Cohort identification period	The period of the study during which individuals can be selected into the cohort after meeting the definition for exposed or referent groups	01 November 2019 - *TBD*
Index date	The date on which an individual first meets the eligibility criteria for exposed or referent groups and enters the cohort	[Day 0]
Baseline period	The 1-year period prior to and including the index date over which baseline characteristics and study eligibility will be assessed	[Day -365 to Day 0]
Follow-up period	The up to 1-year period over which the outcomes will be assessed	[Day 1 to Day 365 or outcome or censoring]

9.2.1. Inclusion and Exclusion Criteria

A summary of inclusion and exclusion criteria for the study population is included in Table 2 below.

Table 2. Summary of Inclusion/Exclusion Criteria

Inclusion Criteria	Assessment Period	Operational Definition	Rationale/Validated Source
Confirmed diagnosis of SCD	[Baseline]	1 inpatient medical claim with an International Classification of Diseases (ICD), Tenth Revision, Clinical Modification (ICD-10-CM) diagnosis code for SCD or 2 outpatient claims with an ICD-10-CM diagnosis code for SCD at least 30 days apart (including all genotypes as well as unknown type)	Reeves S, Garcia E, Kleyn M, et al. Identifying sickle cell disease cases using administrative claims. Acad Pediatr. 2014;14(5 suppl):S61–S67. doi:10.1016/j.acap.2014.02.008
Age between 12 and 85 years	[Index]	Age in years	Age range covers the initial accelerated approval indication.
≥ 2 RBC transfusions	[Baseline]	At least 2 medical claims with an International Classification of	The subset of SCD patients who would most benefit from a

Inclusion Criteria	Assessment Period	Operational Definition	Rationale/Validated Source
		Diseases (ICD), Tenth Revision, Procedure Coding System (ICD-10-PCS) procedure code for an RBC transfusion at least 1 day apart (maximum 1 transfusion counted per day).	reduction in transfusion and necessary to measure a change from baseline to follow-up.
≥ 365 days of continuous enrollment	[Baseline]	Both medical and pharmacy enrollment is required for observability (with a 45-day allowable gap between claims).	Observability is required to assess whether outcomes occurred and were not assumed to be missing.
Exclusion Criteria	**Assessment Period**	**Operational Definition**	**Rationale/Validated Source**
Stroke or history of stroke	[Baseline]	At least one medical claim with an ICD-10-CM diagnosis code with evidence of stroke.	This study is intended to be generalizable to SCD patients. Patients with a history of stroke may rely on RBC transfusions for disease management and will therefore be excluded.
Pregnancy	[Baseline]	At least one medical claim with an ICD-10-CM diagnosis code with evidence of pregnancy	This study is intended to be generalizable to SCD patients and this is a common exclusion for RCTs.
Hemodialysis or history of hemodialysis	[Baseline]	At least one medical claim with an ICD-10-PCS procedure code with evidence of hemodialysis.	Patients with a history of hemodialysis may rely on RBC transfusions for disease management and will therefore be excluded.
Hematopoietic stem cell transplantation (HSCT) or gene therapy or history of HSCT or gene therapy	[Baseline]	At least one medical claim with an ICD-10-PCS procedure code with evidence of HSCT or gene therapy.	Patients with a history of HSCT may rely on RBC transfusions for disease management and will therefore be excluded. After HSCT, their disease may be markedly improved.
Renal transplant or history of renal transplant	[Baseline]	At least one medical claim with an ICD-10-PCS procedure code with evidence of renal transplant.	Patients with a history of renal transplants may rely on RBC transfusions for disease management and will therefore be excluded.
Liver transplant or history of liver transplant	[Baseline]	At least one medical claim with an ICD-10-PCS procedure code with evidence of liver transplant.	This study is intended to be generalizable to SCD patients. Patients with a history of liver transplants may rely on RBC transfusions for disease

Inclusion Criteria	Assessment Period	Operational Definition	Rationale/Validated Source
			management and will therefore be excluded.

9.3. Variables

A list of variables for the study is outlined in Table 3 and the endpoints are described in Table 4. The operational definition for each variable listed in Table 3 and Table 4 is provided in Annex Table 1. All variables on medical history, covariates, and endpoints of interest will be identified with
ICD-10-CM diagnosis/procedure codes and Healthcare Common Procedure Coding System (HCPCS) codes in inpatient and or outpatient medical claims, and with National Drug Codes (NDC) in prescription claims. All codes and/or code algorithms will be reviewed by all protocol authors as well as the study Scientific Review Committee.

When possible, validated algorithms in administrative claims data will be used to define the study variables. For the primary and exploratory outcomes, RBC transfusions will be defined using procedure codes from medical claims and evaluated as a maximum of 1 per day. RBC transfusions are assumed to be documented correctly because healthcare entities must bill to receive reimbursement.[19] A literature review was conducted on the validity of RBC transfusions in RWD such as claims and electronic health records (EHR) prior to drafting the protocol and results were consistently positive.[20-24]

Table 3. List of Variables, Roles, and Assessment Period

Variables	Role	Assessment period	Output
Age	Baseline Characteristic; Covariate	[Index]	Continuous: Mean (SD); Median (IQR); Minimum; Maximum. Frequency and proportion (N, %) of patients with missing values will also be reported.
Sex	Baseline Characteristic; Covariate	[Index]	Frequency and proportion (N, %) Categorical: Male, Female, Unknown, Missing
Race	Baseline Characteristic; Covariate	[Index]	Frequency and proportion (N, %) Categorical: White, Black or African American, American Indian or Alaska Native, Asian, Native Hawaiian or Other Pacific Islander, Missing

Variables	Role	Assessment period	Output
Insurance Type	Baseline Characteristic; Covariate	[Index]	Frequency and proportion (N, %) Categorical: Medicare, Medicaid, Commercial, Other
Region	Baseline Characteristic; Covariate	[Index]	Frequency and proportion (N, %) Categorical: Northeast, Midwest, South, and West Regions (based on Census geographic regions)
Diagnosing Provider Specialty	Baseline Characteristic; Covariate	[Baseline]	Frequency and proportion (N, %) [Provider specialty categories will be included in statistical analysis plan (SAP)]
SCD Genotype[1]	Baseline Characteristic	[Baseline]	Frequency and proportion (N, %) Categorical: Homozygous SS, HbSC, HbSB0 thalassemia, HbSB+ thalassemia, Unknown, Multiple genotypes reported
Comorbidities – SCD specific morbidities[2]	Baseline Characteristic; Covariate	[Baseline]	Frequency and proportion (N, %)
Hydroxyurea	Baseline Characteristic; Covariate; Time-varying Confounder	[Baseline]; [Follow-up]	Frequency and proportion (N, %)
Crizanlizumab	Baseline Characteristic; Covariate	[Baseline]; [Follow-up]	Frequency and proportion (N, %)
L-glutamine	Baseline Characteristic; Covariate	[Baseline]; [Follow-up]	Frequency and proportion (N, %)
Erythropoietin Stimulating Agent (ESA)	Baseline Characteristic; Covariate	[Baseline]; [Follow-up]	Frequency and proportion (N, %)
RBC transfusion	Baseline Characteristic; Covariate	[Baseline];	Continuous: Mean (SD); Median (IQR); Minimum; Maximum.
Baseline Hb[3]	Baseline Characteristic	[Baseline]	Continuous: Mean (SD); Median (IQR); Minimum; Maximum.
Charlson Comorbidity Index (CCI)[4]	Baseline Characteristic; Covariate	[Baseline]	Continuous: Mean (SD); Median (IQR); Minimum; Maximum. Frequency and proportion (N, %) For individual categories see SAP.

Variables	Role	Assessment period	Output
Healthcare Interactions - Inpatient/Outpatient Visit	Baseline Characteristic; Covariate	[Baseline]	Continuous: Mean (SD); Median (IQR); Minimum; Maximum.
Voxelotor	Exposure	[Index]; [Follow-up]	Frequency and proportion (N, %)
Iron overload	Baseline Characteristic; Covariate	[Baseline]	Frequency and proportion (N, %)
Iron chelation use	Baseline Characteristic; Covariate	[Baseline]	Frequency and proportion (N, %)
Delayed hemolytic transfusion reactions	Baseline Characteristic; Covariate	[Baseline]	Frequency and proportion (N, %)
Alloimmunization	Baseline Characteristic; Covariate	[Baseline]	Frequency and proportion (N, %)

Abbreviations: ESA: Erythropoietin Stimulating Agent; RBC: Red blood cell; SCD: Sickle cell disease; VOC: Vaso-occlusive crisis.

1. Baseline genotype will not be included in the PS model because patients commonly have multiple values due to coding/billing errors and this value is more likely to be misclassified. The most frequently occurring value (not including unknown/other) will be selected.

2. SCD-specific comorbidities include the following: acute chest syndrome, acute coronary syndrome, acute osteomyelitis, acute renal failure, aplastic crisis, intracerebral and subarachnoid hemorrhage, myocardial infarction, priapism, splenic sequestration crisis, stroke, transfusion-associated reaction, transient ischemic attack, vaso-occlusive crisis, venous thromboembolism, avascular necrosis, cholelithiasis, chronic kidney disease, chronic pain, diabetes, heart failure, iron overload, leg ulcer, pulmonary hypertension, and retinopathy.

3. Given feasibility analyses have indicated a low proportion of patients (<25%) have lab values, Hb will be excluded from the PS model. Hb will be evaluated excluding those measurements within 2 days of an acute event or within 30-60 days of a transfusion event. Full algorithm will be provided in a separate statistical analysis plan.

4. CCI categories include: myocardial infarction, congestive heart failure, peripheral vascular disease, cerebrovascular disease, dementia, chronic pulmonary disease, rheumatic disease, peptic ulcer disease, mild liver disease, diabetes without chronic complication, diabetes with chronic complication, hemiplegia or paraplegia, renal disease, any malignancy without metastasis, leukemia, lymphoma, moderate or severe liver disease, metastatic solid tumor, acquired immunodeficiency syndrome

The table above includes a summary of baseline characteristics and possible covariates for PS modeling. Covariates will be selected using the diagnostic criteria outlined in the data analysis plan (Section 9.7).

Table 4. Description of Endpoints and Alignment to Study Objectives

Endpoint	Assessment Period	Output	Objective Alignment
Change in RBC transfusions	[Baseline]; [Follow-up]	Mean change in the number of RBC transfusions (maximum 1 per day) PPPY from baseline to follow-up	Primary Objective
VOCs	[Baseline]; [Follow-up]	Mean number of VOCs PPPY from baseline to follow-up	Exploratory Objective 1
Stroke	[Baseline]; [Follow-up]	Rate of event during follow-up	Exploratory Objective 1

Endpoint	Assessment Period	Output	Objective Alignment
ACS	[Baseline]; [Follow-up]	Mean number of events PPPY from baseline to follow-up	Exploratory Objective 1
ACS + Pneumonia	[Baseline]; [Follow-up]	Mean number of events PPPY from baseline to follow-up	Exploratory Objective 1
Priapism	[Baseline]; [Follow-up]	Rate of event among males during follow-up	Exploratory Objective 1
RBC transfusions in follow-up	[Baseline]; [Follow-up]	Mean number of RBC transfusions (maximum 1 per day) PPPY in follow-up	Exploratory Objective 2
Percent reduction in RBC transfusion > 30%	[Baseline]; [Follow-up]	Proportion of patients with a percentage decrease in the number of RBC transfusions (maximum 1 per day) PPPY from baseline to follow-up that is greater than 30%	Exploratory Objective 2
Zero RBC transfusions in follow-up	[Baseline]; [Follow-up]	Proportion of patients with 0 RBC transfusions (maximum 1 per day) PPPY during follow-up	Exploratory Objective 2
Iron overload	[Baseline]; [Follow-up]	Rate of event during follow-up	Exploratory Objective 3
Iron chelation use	[Baseline]; [Follow-up]	Rate of event during follow-up	Exploratory Objective 3
Delayed hemolytic transfusion reactions	[Baseline]; [Follow-up]	Rate of event during follow-up	Exploratory Objective 3
Alloimmunization	[Baseline]; [Follow-up]	Rate of event during follow-up	Exploratory Objective 3

Abbreviations: ACS: Acute Chest Syndrome; PPPY: per patient per year; RBC: Red blood cell; VOC: Vaso-occlusive crisis.

The table above includes a summary of the outcomes of interest for this analysis. Operational definitions can be found in the data analysis plan (Section 9.7).

9.4. Data Sources

Pfizer has conducted a validation study of RBC transfusions using the Optum Electronic Health Records and Claims Data among patients with SCD to support the use of procedure codes to identify these events in claims data, with the following objectives:

- To assess the concordance of RBC transfusion documentation based on medical claims data (ICD-10 procedure codes) compared to structured EHR data (and vice versa).

- To confirm RBC transfusions identified in structured EHR/claims data via procedures codes through clinical notes review.

- To identify and record the details of RBC transfusion administrations (specifically the reason for, type of transfusion and volume) and secondary complications of transfusions.

The results from this validation study inform the accuracy of claims data based on procedure codes in measuring RBC transfusions as an endpoint. The validation study included 13,327 patients with a documented diagnosis of SCD and at least one RBC transfusion documented in claims or EHR within the Optum Market Clarity database between January 2016 and March 2023. Among these patients, there were 46,381 RBC transfusions performed over the study period. The vast majority (91%) of these transfusions were documented in claims (42,106) with some (22%) overlap in EHR (10,394). A minority (9%) of RBC transfusions were only documented in EHR (4,275). Given these results, we believe that claims are the more reliable data source and will be used primarily for this study.

The study will use Komodo's Healthcare Map (KHM) database linked to Claritas prescription data (proprietary voxelotor claims) and Quest Diagnostics lab data. Data sources are anchored on a closed system of administrative claims, including patient enrollment, physician, facility, and pharmacy claims. Administrative claims submitted for payment by providers and pharmacies are verified, adjudicated, and de-identified before inclusion. A data feasibility assessment was conducted using the Structured Process to Identify Fit-for-Purpose Data (SPIFD) framework to determine fit-for-purpose RWD sources for this study.25 An initial data completeness check demonstrated full coverage for encounters based on ICD-10-CM codes and NDC on all continuously enrolled patients.

Overall, claims data are expected to be a nearly complete capture of the patient's healthcare interactions, as claims restrict data inclusion to the time when patients are enrolled in both medical and pharmacy coverage. We expect minimal missing data on the treatment of interest (voxelotor) by linking the data from Komodo with Claritas Rx and high completeness of key data elements. However, guidance from Komodo recommends study periods end 9 months prior to the end of data to ensure complete capture of closed claims. Each data source is described in further detail below.

Komodo

KHM is a real-world dataset that integrates disparate sources of patient-level data to map longitudinal patient journeys. Komodo pulls de-identified, patient-level claims data from clearinghouses, payers (150+ payers including Medicare and Medicaid), and provider data sources to follow patients as they move through the healthcare system. The patient-centric database includes both open and closed claims data and is enriched with data from EHR and other sources. Komodo's closed dataset (payer complete dataset) will be used for this analysis. Data will be included from November 25, 2018, to the most recently available dataset. The database covers all the US census regions. The all-payer claims data is fully adjudicated. Komodo is a patient-centered claims dataset that complies with the Health Insurance Portability and Accountability Act.

Claritas

Claritas Rx captures patient-level data from channel partners such as specialty pharmacies, patient services hubs (including patient assistance programs), specialty distributors, claims data providers, labs and other patient services providers to enable insights into the market and patient journey. This data source contains proprietary voxelotor claims and is linked to KHM.

Quest
Quest laboratory database contains over 56 billion laboratory test results nationwide. For this study, Quest data will be linked to the KHM from November 25, 2018, through the end of data. Important baseline variables like Hb will be captured from lab data.

9.5. Study Size

The study will include all the eligible patients with an index date from 25 November 2019, until the end of the study period. An initial sample size assessment conducted with Komodo data through September 2023 identified ~94,000 patients with SCD, and among those patients, ~7,700 were treated with voxelotor. Published information on the annual rate of transfusions for patients with SCD was the basis for estimating an expected range for the primary outcome (number of transfusions PPPY during follow-up).[15,16]

A conservative sample size estimate was calculated using a Welch's t-test with an adjustment factor of 1.65 which takes into account the lost-to-follow-up, variance inflation in the analysis.[26] With 177 patients treated with voxelotor and 354 matched patients not treated with voxelotor (1:2 matching ratio), the study will ensure an 80% power to detect at least a 1-unit (1 transfusion per day) difference between the two groups (standard deviation = 3.0) in the change in RBC transfusions PPPY from baseline to follow-up.

9.6. Data Management

Komodo Health runs data quality checks across data ingestion/mastering/normalization and standardization process by developing tests of the data pipeline with data quality assurance tools.
At Aetion, as part of the data ingestion process, raw data review is routinely conducted to understand contents of the database and scientific integrity checks are performed to ensure the contents of the data are consistent with the expected data as laid out in the applicable data usage agreement. Some of the key characteristics explored in this process include:

- Table structure (number of rows, columns, column names, etc.)
- Summary counts per table (ie, non-missing counts, unique counts)
- Variable distribution (eg, min, mean, median, max for numeric variables; top frequencies for categorical variables)
- Date range (min, max, and distribution over a time period)
- Missingness percentage of attributes

Raw data files as well as transformed data files are retained on the platform; rows are dropped if they do not contain dates or have start dates that are earlier than end dates. Aetion performs an additional deidentification step to map the raw data unique patient identifier to an Aetion patient identifier.

Following receipt and review of the raw data, each data cut is connected to the Aetion Evidence Platform (AEP). A data connector specification is drafted by a data scientist, which provides a map for transformation of raw data to the Aetion longitudinal patient timeline. Validation of the database connection (DBC), via double programming, is completed to ensure that the implementation of DBC logic leads to transformed data output that connects to and behaves within AEP exactly as intended. Following validation, the specification files are used to create an Aetion data dictionary for the dataset. Prior to deployment on the AEP, a manual test is conducted to check the baseline details of the dataset, the generation of measures, cohorts, and analyses, any resulting outputs, and any relevant coding systems. The test further ensures that platform features and dataset values are visible, testable, and working as expected on the front-end.

Analyses will be done within the AEP [Version 4.99.6] (Aetion Evidence Platform® 2024) or using R version 4.2.3. For analytic datasets created off platform, original (or raw) files will never be modified.

9.7. Data Analysis

Detailed methodology for summary and statistical analyses of data collected in this study will be documented in the SAP, which will be dated, filed, and maintained by Pfizer. The SAP may modify the plans outlined in the protocol; any major modifications of primary endpoint definitions or their analyses would be reflected in a protocol amendment.

9.7.1. Descriptive Analysis

After identifying the populations of interest using CEM on age and calendar time, the distribution of baseline covariates will be evaluated in the 365 days prior to index. Continuous variables will be described using mean values with standard deviations and median values with interquartile range; categorical variables will be described as the number of patients and percent. For variables derived from diagnoses, procedures, and prescription codes, patients are assumed to have experienced the event of interest if the relevant code(s) are found among their claim records. Otherwise, it is assumed that the patient did not experience the event, thus resulting in no missing data for these variables.

9.7.2. Baseline Balance

PS methods will be used to account for baseline balance of covariates between the treated group and matched controls. PS adjustment allows for the inclusion of a pre-defined comprehensive list of confounding covariates to ensure comparable exposure and referent groups, avoiding non-convergent issues of standard outcome models. PS models estimating the ATT will be considered for this analysis. An iterative approach to estimating the PS will be used.

PS Estimation

Potential covariates that affect treatment decisions or outcomes and are eligible for inclusion in the PS model are listed in Table 5.[27] Potential confounding covariates will be assessed for balance at baseline using absolute standard differences (ASD). Covariates with an ASD > 0.1 will be considered imbalanced and will be included in a PS model fit using logistic

regression. With this approach, the PS will be modeled as the conditional probability of receiving voxelotor given the selected covariates for each individual. The PS modeling will be performed blinded of any outcome to ensure endpoints will not be assessed before a PS method is selected. Diagnostic criteria used to assess PS model estimate are described below and in Table 5. Multiple iterations of the PS model estimation may be performed depending on diagnostic evaluation results.

Use of PS in Baseline Confounder Balance

Multiple PS methods will be considered and a step-wise approach will be used to select the most appropriate method. PS matching, in which voxelotor-exposed patients are matched to unexposed patients using variable-ratio (1:2) greedy nearest neighbor matching with a caliper, will be the primary approach.[28] If the sample size falls below the minimum required sample size for the primary and secondary endpoint to be 80% powered during the matching process, weighting methods, such standardized mortality/morbidity ratio (SMR) weights, will be evaluated for their suitability and to preserve sample size of the exposed group. With the SMR approach, weights for the voxelotor group are equal to 1; weights for the referent arm are defined as PS/(1-PS). The distribution of SMR weights will be reviewed; truncation and trimming will be considered to account for extreme weights. If neither PS model type is deemed appropriate, a non-ATT model (overlap weights) will be considered. With the overlap approach, overlap weights are defined as 1-PS in the voxelotor arm; weights for individuals in the referent arm are defined as the PS. The distribution of overlap weights will be reviewed to evaluate the appropriate trimming threshold. PS-matching and weighting methods will be further described in the SAP.

Diagnostic Evaluation of PS Methods

An *a priori* list of diagnostic criteria will be set to ensure analytic assumptions are met before a PS method is selected. For example, ASD will be used to assess the comparability of the groups before and after the application of PS methods.[29] Observations for any covariate leading to zero cells in either group may lead to an inadequate overlap between groups in their PS distributions. Therefore, collapsing similar categories within variables if there are categories with frequencies <10% to minimize concerns for non-positivity will be considered. Further, variables with substantial non-positivity may be removed from the model (ie, covariables with very low or zero values in either group). Patients with missing information for covariables may be excluded or imputed from the analysis set depending on the degree of missingness and the variable. When applicable, a robust variance estimator may be utilized in the outcome model(s) to account for the weighted design.[28] Demographics and baseline clinical characteristics of patients within each study group will be described before and after the applications of PS adjustment.

The diagnostic results will be presented to the study team who will provide a final confirmation that all the necessary criteria have been satisfied. The inferential analysis phase will only begin after sign-off of the final protocol by the study team, certified by the Principal investigator (PI). More information on the methods used to balance baseline covariates can be found in SAP and in Table 5 below.

Table 5. Diagnostic Checklist

#	Diagnostic	Description	✓
1	Confirm positivity of variables	PS distribution will be visually inspected and overlap in all areas of the PS distribution will be confirmed. Additionally, any patients with an observation for covariates not meeting the positivity assumption and may be considered outliers. Further detail on methodology is available in SAP.	
2	Confirm PS model fit	Any covariate with no variability (ie, a constant value across all observations) will be removed from the PS model. The PS model will be determined as not overfit if each covariate contains ≥ 12 exposed subjects.[30] Collinear covariates will be defined as any covariates with a Pearson correlation coefficient greater than or equal to 0.7.[31] Collinear variables may be removed from the PS model if large standard errors are observed because including them may lead to biased coefficient estimates. If removed, the model will be refitted.[32]	
3	Confirm baseline confounder balance	Balance of all potential confounders (ie, covariates) will be confirmed for voxelotor vs referent patients. Covariate balance will be defined as an ASD ≤ 0.10.[33] Balance of these variables will still be confirmed after PS matching. Should any variables with small residual imbalance remain (0.10 ≤ ASD ≤ 0.20), balance may be deemed acceptable if the variable is not a predictor of the outcome among the referent group (defined as an ASD < 0.10 when comparing the risk of the outcome to those with the variable versus those without it). This determination will be made by clinical experts or key opinion leaders on the study team. Evaluation of imbalance in outcome prediction will only be conducted once at the end of the diagnostic sub-phase if all other diagnostic criteria are met. Variables with residual imbalance (0.10 ≤ ASD≤ 0.20) across the exposure and referent groups that are associated with the outcome (eg, baseline Hb) will be included as an independent variable in the subsequent outcome model.	
4	Assess PS model coefficients for extreme values	Evaluate PS model coefficients to identify extreme coefficients which may indicate collinearity or positivity violations.[34] Any coefficient with extreme values that strongly predicts presence or absence of the exposure of interest (ie, potential instrumental variables) or another covariate (ie, multicollinearity) may be dropped, combined, or re-categorized within the PS model.	

9.7.3. Outcome Analysis

Operational definitions for each outcome are described below. Each outcome will be assessed using the primary assessment method for the follow-up period where follow-up will end on the earliest of the following occurrences:

- Disenrollment from health insurance (with a 45-day allowable gap)
- Death
- Discontinuation of voxelotor (with a 30-day allowable treatment gap, voxelotor arm only)
- End of follow-up: up to 1 year from index date (365 days)
- The occurrence of any exclusion criteria events (eg, pregnancy, stroke, hemodialysis, evidence of HSCT or gene therapy, renal transplant, or liver transplant)

The primary analyses will employ an as-treated follow-up approach that ends follow-up upon treatment changes. Patients will be followed from the index date until treatment change (discontinuation of voxelotor for voxelotor group), or the end of follow-up, whichever is earlier. To assess potential differences in follow-up time between voxelotor patients and their matched controls, the median and mean follow up times will be assessed between the cohorts and further assessed with pairwise censoring in the sensitivity analysis.[35] This as-treated method of analysis allows for an estimate of the effect while the patient is taking the drug of interest and addresses intercurrent events. However, it may introduce bias into the estimates. A summary of sensitivity analyses is included in Table 6 to evaluate how the magnitude of this bias will be assessed.

Primary Objective – Change in Number of Transfusions PPPY
For each patient, the difference in the number of transfusions PPPY from baseline to follow-up period will be reported as the change and will be calculated as follows:

$$Change = \frac{Number\ of\ transfusions\ in\ follow-up}{Available\ days\ of\ follow-up} - \frac{Number\ of\ transfusions\ in\ baseline}{365\ days\ *} \times 365$$

*All patients have 365 days of baseline data

The change in number of transfusions PPPY will be analyzed using generalized linear model with an identity link to estimate the difference in transfusion change rates between the two groups. If any pre-index covariates that are found to be unbalanced after the PS method, it will be included in the regression as a covariate. The incidence rate change and its 95% CI will be reported.

For sensitivity analysis, transformations of the outcome may also be considered as appropriate. The outcome data will only be viewed overall prior to selecting a method of regression. 95% confidence interval (CI) will be reported for the difference between the groups. Full details will be provided in the SAP.

Exploratory Objectives 1 & 3 - Rate of Clinical Outcomes and Transfusion-Associated Complications

The rate of each clinical outcome (eg, VOCs) and transfusion-associated complication (listed in Table 5) will be described separately and defined as the number of patients with the outcome divided by the total person-time in the follow-up period. Absolute rate differences between the two groups will be reported for each of the transfusion-associated secondary complications and secondary efficacy endpoints. The number of persons, time-at-risk, and number of outcomes in each group will also be reported.

Exploratory Objective 2, Endpoint 1 – Number of RBC Transfusions PPPY in Follow-Up

The number of transfusions that occur during the follow-up period will be assessed PPPY during available follow-up within each group. The mean number of transfusions PPPY will be reported for each group. The mean difference between the groups will be assessed using a regression model similar to the primary analysis. A weighted analysis of covariance (ANCOVA) may also be utilized for confounders that remain unbalanced after the PS adjustment. The 95% CI will be reported for the difference between the groups.

Exploratory Objective 2, Endpoint 2 – Proportion of Patients with > 30% Reduction in Transfusions

This outcome will be defined as the proportion of patients with a percent reduction in transfusions between baseline and follow-up that is greater than 30%. For each patient, the percent change will be calculated as follows:

$$Percent\ Change = \frac{\left(\frac{Number\ of\ transfusions\ in\ follow-up}{Available\ days\ of\ follow-up} - \frac{Number\ of\ transfusions\ in\ baseline}{365\ days*}\right)}{\frac{Number\ of\ transfusions\ in\ baseline}{365\ days*}} x\ 100$$

*All patients have 365 days of baseline data

The mean percentage change in the number of transfusions PPPY within each group will be reported. The proportion of patients with a 30% reduction in transfusions will also be reported for each group. A weighted logistic regression model will be used to compare the difference of the proportion of patients with the reduction of more than 30% between the groups with an estimated odds ratio. The 95% CIs will also be reported.

Exploratory Objective 2 Endpoint 3 – Proportion of Patients with Zero RBC Transfusions PPPY in Follow-Up

This outcome will be defined as the proportion of patients with zero RBC transfusions PPPY during follow-up. The absolute risk difference between the two groups will be reported. A logistic regression model will be used to compare the percentage of patients without a transfusion in the follow-up time frame between the groups. The 95% CIs will also be reported.

Sensitivity and Subgroup Analyses

Table 6 below includes a summary of the sensitivity analyses and the rationale for performing each analysis. The following sensitivity analyses will be performed for the primary endpoint unless otherwise specified:

Table 6. Planned Sensitivity and Subgroup Analyses

Analysis	Description	Rationale
Intention-to-Treat (ITT)	The follow-up period will be changed from as-treated to follow an ITT approach with the exposed group requiring at least one voxelotor claim. All patients will be followed from the index date until disenrollment, death, or end of study period. Follow-up will not end on treatment discontinuation or switches.	To estimate the effect of the "assigned" treatment.
Required 1-year minimum follow-up and pairwise censoring subgroup analysis	Two subgroup analyses have been proposed to evaluate bias introduced by differential follow-up. First, a subgroup analysis of patients with a minimum of 1-year of follow-up will be performed to evaluate the impact of informative censoring. Second, an analysis pairwise censoring will be performed to censor unexposed on the date of voxelotor discontinuation for their matched pair.	To address the impact of informative censoring and differential follow-up
Stratification by number of transfusions in baseline	The number of transfusions PPPY in baseline will be categorized and the primary objective will be reported separately for those categories. This analysis will be performed for all objectives.	To understand the effect of the treatment on high-risk subgroups.
Subgroup analysis by Hb level	The primary analysis will also be performed on a subgroup of patients with Hb values in baseline. A separate PS model will be utilized for this subgroup.	To understand the effect of the treatment on high-risk subgroups and patients with available baseline data.
Subgroup analysis by prior medications	A subgroup analysis will be performed on patients who received hydroxyurea, crizanlizumab, L-glutamine, and/or ESA in baseline the exposed and referent groups.	To understand the effect of other SCD treatments.
Subgroup analysis by concomitant medications	A subgroup analysis will be performed on patients who received hydroxyurea, crizanlizumab, L-glutamine, and/or ESA in baseline the exposed and referent groups. This subgroup analysis will be performed for the Primary Objective and the Exploratory Objective 1 (VOCs).	To understand the effect of other SCD treatments.

Abbreviations: ITT: Intention-to-treat; PPPY: per patient per year; SCD: Sickle cell disease.

All subgroup analyses and stratifications must meet a minimum sample size criteria in each group to be performed (see SAP for more detail). Exploratory objective 1 which evaluates the number of VOCs and its separate components over follow-up will also include a sensitivity analysis utilizing alternative definitions for VOCs, given a higher likelihood of misclassification in claims using ICD-10 codes alone; these alternative VOC definitions will be detailed in code lists provided in the SAP. Other sensitivity analyses such as e-values may be considered if other possible confounders or sources of bias are identified, depending on effect size in the primary outcome or differences in baseline characteristics and/or follow-up time between the exposed and matched controls. Details on potential sensitivity analyses will be included in the SAP.

9.8. Quality Control

9.8.1. Analytic Quality Control

Aetion will build measures for cohort identification, outcomes, and other variables of interest based on codes and algorithms described in this protocol. All measures created will undergo quality control review by at least one additional analyst or scientist under the supervision of the Study Lead.

This protocol will be strictly followed when conducting the analysis of this study. All cohorts developed, statistical analyses implemented, and tables completed will undergo quality control review by at least one additional analyst or scientist under the supervision of the Study Lead.

The Study Lead will review all results tables and other final deliverables to confirm accuracy, logical flow, and appropriate format.

This secondary data collection study follows the Guidelines for Good Epidemiologic Practice laid out in 2005 FDA Good Pharmacoepidemiology Practices (GPP), Best Practices for Conducting and Reporting Pharmacoepidemiologic Safety Studies Using Electronic Healthcare Data Sets and the 2015 International Society of Pharmacoepidemiology GPP.[36-38]

The AEP maintains a date / time-stamped record of all analytic cohorts and corresponding analyses.[39]

9.9. Data Storage

Aetion may retain raw and transformed data for as long as it is permitted by the data vendor. Once a contract expires that is no longer planned for renewal, Aetion will remove and destroy all raw and transformed data from Amazing Web Services (AWS) S3 as well as the Substantiate platform. Results can be downloaded from the platform and saved at any time. Results can remain on platform as long as the client has an active Substantiate license.

9.10. Strengths and Limitations of the Research Methods

This study is intended to support regulatory decision making and labeling enhancements for voxelotor because a RCT was deemed unfeasible. There are some distinct differences between RWE and RCT settings. RWD allows for more generalizable patient populations, but there is less standardization and higher potential for missing values for important patient characteristics (eg, Hb levels, SCD genotype, comorbidity scores, etc.). Additionally, in a RCT environment, healthcare interactions are typically on a protocolized schedule for enrolled patients. In RWD, there will likely be differences in how often a patient is seeing a provider and the type of provider both across groups and within groups, which is why both of these elements will be important for PS modeling. However, RWE allows for less restrictive inclusion/exclusion criteria (ie, a more generalizable population) and will allow for patients with more severe SCD to be evaluated. Although there are many benefits to utilizing RWD, there are some inherent limitations.

Claims data sources are subject to potential misclassification of the outcome, exposure, or covariates in observational studies (ie, the presence of a diagnosis code on a medical claim or a generic medication on a prescription claim does not indicate the presence of disease or compliance to treatment given rule out diagnosis or coding errors). The extent of miscoding is believed to be small, as study variables are drawn directly from medical claims that are used for clinical documentation and insurance reimbursement. Given there are no validated algorithms to identify SCD genotype currently available and there is a high possibility for misclassification through genotype-specific ICD-10 codes, genotype will not be a covariate included in the PS model. However, because two transfusions are required in baseline to be included in the study, it is most likely that patients will have SCD genotype Hb SS or $S\beta^0$ thalassemia. Additionally, a small proportion of patients have available lab data after linking across all three data sources; Hb will also be excluded from the PS model but will have a separate subgroup analysis because it may be an important confounder in this study. Patients will be excluded from the analysis if they have evidence of stroke in baseline so that patients receiving chronic transfusions for secondary stroke prevention are not included in the analysis. Although patients may also receive transfusions for primary stroke prevention, claims data will not have information on transcranial doppler velocity, and thus will not be identifiable. However, this is estimated to be a small portion of patients within the study.

When possible, validated claims-based algorithms will be utilized to define inclusion/exclusion criteria and outcomes. Pfizer is currently conducting a validation study to assess the accuracy of medical claims billed for RBC transfusions. VOCs are more likely to be misclassified in claims, and underestimated as only more severe VOC will be captured through claims, and there are a variety of claims-based algorithms in similar research studies. Clinical advisors will be consulted on how to best capture VOCs using medical claims, and sensitivity analyses will be conducted by using both a specific and broader algorithm informed by clinical feedback. Furthermore, a literature review was conducted to evaluate the validity of RBC transfusion procedures in both claims and EHR databases. A high positive predictive value (PPV) was found in multiple studies including both US and non-US data validating RBC transfusion procedures in claims or EHR against blood bank data.[17-21] Additionally, this analysis defines a RBC transfusion as at most one transfusion per day because the data does not allow us to distinguish between the claims for transfusions billed as

multiple units or multiple procedures. As a result, we may be undercounting the true number of transfusions patients receive. However, units of blood are not well captured in claims or EHR and will be a limitation of this analysis. Furthermore, if patients receive a transfusion outside of the KHM data contributors, it will not be captured but is assumed to impact both the exposed and control groups equally.

Unmeasured confounding leading to an unobservable imbalance between the voxelotor and referent groups may be present given not all demographic or clinical characteristics are available in the KHM database (eg, race or ethnicity, provider characteristics, genotypes, etc.). The study population will consist of voxelotor patients who are able to be linked to proprietary claims and with standardized fields for demographic variables that may not be as generalizable (eg, specific race or ethnic populations like 'Middle Eastern' are not reflected in claims). The COVID-19 pandemic will overlap with the study period, however, the impact on healthcare utilization is expected to be equal among the exposed and control groups. Adherence to voxelotor will be assumed with each prescription fill per the days supply. If patients are not adherent to daily doses of voxelotor, the treatment effect may be underestimated in the exposed population. However, these limitations are inherent to the data source, and thus cannot be addressed in this study.

Because this study does not include an active comparator as a control group, CEM will be used to select the index date of matched controls within 15 days of the initiation of voxelotor for exposed patients. All exposed patients will be selected prior to controls to preserve sample size. Sampling bias may be introduced by prioritizing exposures, but the impact is limited due to a larger pool of untreated control patients than potential exposures. There may be differences in patient characteristics dependent on when patients enter the study population (ie, initiating voxelotor versus periods without treatment for matched controls). However, PS methods will be utilized to control for observed baseline confounding between the exposed and referent groups. Linked Quest data will be used to evaluate Hb levels in baseline (an important diagnostic criterion for RBC transfusions). Feasibility analyses will be conducted prior to implementation to assess the number of patients in this study with available Hb data. Assuming a significant number of patients will not have Hb data in baseline, a sensitivity analysis has been proposed to assess the primary outcome among patients with this lab data. However, there may be differences in the frequency of Hb testing between the exposed and control group, relative to when patients enter the study cohort. A separate PS model will be fit for this subgroup analysis to address these differences. Depending on the utilization of other SOC treatments in both the exposed and control groups, other subgroup analyses may be conducted. Additionally, the primary analysis will be as-treated, which includes discontinuation of voxelotor as a censoring criterion in the exposed group. This creates potential for confounding via differential censoring between the exposed and referent groups. However, to account for these potential differences in follow-up time, pairwise censoring will be used to censor matched controls on the date of their voxelotor match's discontinuation date in a sensitivity analysis.

9.11. Other Aspects

Missing data will be handled depending on the amount (eg, more than x% of patients missing a value for a variable) and mechanism (eg, missing completely at random, missing at

random, or missing not at random) of missingness. Missingness will be considered only for variables for which a value would be expected for all observations, such as age or sex, and not for variables where the absence of a value can be reasonably interpreted as the absence of a condition (eg, data are missing by design). Furthermore, whenever applicable, a category for missing values will be reported for categorical variables. For more details on how missing data is imputed see companion SAP.

10. PROTECTION OF HUMAN PARTICIPANTS

10.1. Patient Information

This study involves data that exist in deidentified/anonymized structured format and contain no patient personal information.

10.2. Patient Consent

As this study involves deidentified/anonymized structured data, which according to applicable legal requirements do not contain data subject to privacy laws, obtaining informed consent from patients by Pfizer is not required.

10.3. Institutional Review Board (IRB)/ Ethics Committee (EC)

This protocol will be submitted to IRB for an approval waiver. No analysis of data will begin until this waiver is obtained.

10.4. Ethical Conduct of the Study

The study will be conducted in accordance with legal and regulatory requirements, as well as with scientific purpose, value, and rigor and follow generally accepted research practices described in Guidelines for Good Epidemiologic Practice and Best Practices for Conducting, Reporting Pharmacoepidemiologic Safety Studies Using Electronic Healthcare Data Sets, and the IPSE GPP.

11. MANAGEMENT AND REPORTING OF ADVERSE EVENTS/ADVERSE REACTIONS

This study involves data that exist as structured data by the time of study start. In these data sources, it is not possible to link (ie, identify a potential association between) a particular product and medical event for any individual. Thus, the minimum criteria for reporting an adverse event (AE) (ie, identifiable patient, identifiable reporter, a suspect product, and event) cannot be met.

12. PLANS FOR DISSEMINATING AND COMMUNICATING STUDY RESULTS

Study results will be summarized in a study report and will be disseminated to all the relevant members internally.

Study results, separately by dataset, will be communicated, presented, and/or published in scientific journals or other scholarly media.

Authorship of study manuscripts and presentations at scientific conferences will follow the guidelines established by the International Committee of Medical Journal Editors (https://www.icmje.org/).

In the event of any prohibition or restriction imposed (eg, clinical hold) by an applicable competent authority in any area of the world, or if the party responsible for collecting data from the participant is aware of any new information which might influence the evaluation of the benefits and risks of a Pfizer product, Pfizer should be informed immediately.

13. REFERENCES

1. Kato, G., Piel, F., Reid, C. et al. Sickle cell disease. Nat Rev Dis Primers 4, 18010 (2018). https://doi.org/10.1038/nrdp.2018.10.

2. Weatherall DJ. Current trends in the diagnosis and management of haemoglobinopathies. Scand J Clin Lab Invest. 2007;67(1):1–2. doi: 10.1080/00365510601051532.

3. Chou ST, Alsawas M, Fasano RM, Field JJ, Hendrickson JE, Howard J, Kameka M, Kwiatkowski JL, Pirenne F, Shi PA, Stowell SR, Thein SL, Westhoff CM, Wong TE, Akl EA. American Society of Hematology 2020 guidelines for sickle cell disease: transfusion support. Blood Adv. 2020 Jan 28;4(2):327-355. doi: 10.1182/bloodadvances.2019001143. PMID: 31985807; PMCID: PMC6988392.

4. GBD 2021 Sickle Cell Disease Collaborators. Global, regional, and national prevalence and mortality burden of sickle cell disease, 2000-2021: a systematic analysis from the Global Burden of Disease Study 2021. Lancet Haematol. 2023 Aug;10(8):e585-e599. doi: 10.1016/S2352-3026(23)00118-7.

5. Charache S, Terrin ML, Moore RD, et al; Investigators of the Multicenter Study of Hydroxyurea in Sickle Cell Anemia. Effect of hydroxyurea on the frequency of painful crises in sickle cell anemia. N Engl J Med. 1995;332(20):1317-1322.

6. Niihara Y, Miller ST, Kanter J, et al; Investigators of the Phase 3 Trial of l-Glutamine in Sickle Cell Disease. A phase 3 trial of l-glutamine in sickle cell disease. N Engl J Med. 2018;379(3):226-235.

7. Ataga KI, Kutlar A, Kanter J, et al. Crizanlizumab for the prevention of pain crises in sickle cell disease. N Engl J Med. 2017;376(5):429-439.

8. Fortin PM, Hopewell S, Estcourt LJ. Red blood cell transfusion to treat or prevent complications in sickle cell disease: an overview of Cochrane reviews. Cochrane Database of Systematic Reviews. 2018(8).

9. Chou ST, Alsawas M, Fasano RM, et al. American Society of Hematology 2020 guidelines for sickle cell disease: transfusion support. Blood Advances. 2020;4(2):327-55.

10. Porter J, Garbowski M. Consequences and management of iron overload in sickle cell disease. Hematology 2013, the American Society of Hematology Education Program Book. 2013;2013(1):447-56.

11. Meloni A, Puliyel M, Pepe A, et al. Cardiac iron overload in sickle-cell disease. American Journal of Hematology. 2014;89(7):678-83.

12. Thein SL, Pirenne F, Fasano RM, et al. Hemolytic transfusion reactions in sickle cell disease: underappreciated and potentially fatal. Haematologica. 2020;105(3):539-44.

13. FDA. The Voice of the Patient. A Series of Reports from the US FDA's Patient Focused Drug Development Initiative. Food and Drug Administration; 2014. https://www.fda.gov/media/89898/download.

14. Vichinsky E, Hoppe CC, Ataga KI, et al. A Phase 3 Randomized Trial of Voxelotor in Sickle Cell Disease. N Engl J Med. 2019;381(6):509-519. doi:10.1056/NEJMoa1903212.

15. Shah N, Lipato T, Alvarez O, et al. Real-world effectiveness of voxelotor for treating sickle cell disease in the US: a large claims data analysis. Expert Rev Hematol. 2022;15(2):167-173. doi:10.1080/17474086.2022.2031967.

16. Blyden G, Bridges KR, Bronte L. Case series of patients with severe sickle cell disease treated with voxelotor (GBT440) by compassionate access. Am J Hematol. 2018;93(8). doi:10.1002/ajh.25139.

17. Iacus SM, King G, Porro G. Causal Inference without Balance Checking: Coarsened Exact Matching. Political Analysis. 2012;20(1):1-24. doi:10.1093/pan/mpr013.

18. Kayle M, Docherty SL, Sloane R, et al. Transition to adult care in sickle cell disease: A longitudinal study of clinical characteristics and disease severity. Pediatr Blood Cancer. 2019;66(1):e27463. doi:10.1002/pbc.27463.

19. AABB Billing Guide for Blood Products and Related Services. The Association for the Advancement of Blood and Biotherapeutics; 2023. Available at https://www.aabb.org/docs/default-source/default-document-library/resources/aabb-billing-guide-for-blood-products-and-related-services.pdf.

20. Elenga N, Loko G, Etienne-Julan M, Al-Okka R, Adel AM, Yassin MA. Real-World data on efficacy of L-glutamine in preventing sickle cell disease-related complications in pediatric and adult patients. Front Med. 2022;9:931925. doi:10.3389/fmed.2022.931925.

21. Kanter J, Bhor M, Li X, Li Y, Paulose J. High Healthcare Utilization in Adolescents with Sickle Cell Disease Prior to Transition to Adult Care: A Retrospective Study. J Health Econ Outcomes Res. 2019;6(3):174-184. doi:10.36469/10512.

22. Desai RJ, Mahesri M, Globe D, et al. Clinical outcomes and healthcare utilization in patients with sickle cell disease: a nationwide cohort study of Medicaid beneficiaries. Ann Hematol. 2020;99(11):2497-2505. doi:10.1007/s00277-020-04233-w.

23. Friedman AM, Oberhardt M, Sheen JJ, et al. Measurement of hemorrhage-related severe maternal morbidity with billing versus electronic medical record data. J Matern Fetal Neonatal Med. 2022;35(12):2234-2240. doi:10.1080/14767058.2020.1783229.

24. Patterson JA, Francis S, Ford JB. Assessing the Accuracy of Reporting of Maternal Red Blood Cell Transfusion at Birth Reported in Routinely Collected Hospital Data. Matern Child Health J. 2016;20(9):1878-1885. doi:10.1007/s10995-016-1992-2.

25. Gatto NM, Campbell UB, Rubinstein E, et al. The Structured Process to Identify Fit-For-Purpose Data: A Data Feasibility Assessment Framework. Clin Pharmacol Ther. 2022;111(1):122-134. doi:10.1002/cpt.2466.

26. Austin PC. Informing power and sample size calculations when using inverse probability of treatment weighting using the propensity score. Stat Med. 2021;40(27):6150-6163. doi:10.1002/sim.9176.

27. Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. Stat Med. 2015;34(28):3661-3679. doi:10.1002/sim.6607.

28. Austin PC. A comparison of 12 algorithms for matching on the propensity score. Stat Med. 2014;33(6):1057-1069. doi:10.1002/sim.6004.

29. Austin PC. Variance estimation when using inverse probability of treatment weighting (IPTW) with survival analysis. Stat Med. 2016 Dec 30;35(30):5642-5655. doi: 10.1002/sim.7084. Epub 2016 Aug 22. PMID: 27549016; PMCID: PMC5157758.

30. Schuster T, Lowe WK, Platt RW. Propensity score model overfitting led to inflated variance of estimated odds ratios. J Clin Epidemiol. 2016;80:97-106. doi:10.1016/j.jclinepi.2016.05.017.

31. Mackinnon MJ, Puterman ML. Collinearity in generalized linear models. Commun Stat - Theory Methods. 1989;18(9):3463-3472. doi:10.1080/03610928908830102.

32. Yoo W, Mayberry R, Bae S, Singh K, (Peter) He Q, Lillard JW. A Study of Effects of MultiCollinearity in the Multivariable Analysis. Int J Appl Sci Technol. 2014;4(5):9-19.

33. Cole SR, Hernán MA. Constructing inverse probability weights for marginal structural models. American journal of epidemiology. 2008 Sep 15;168(6):656-64.

34. Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. Stat Med. 2009;28(25):3083-3107. doi:10.1002/sim.3697.

35. Lefort M, Sharmin S, Andersen JB, et al. Impact of methodological choices in comparative effectiveness studies: application in natalizumab versus fingolimod comparison among patients with multiple sclerosis. BMC Med Res Methodol. 2022;22(1):155. Published 2022 May 30. doi:10.1186/s12874-022-01623-8.

36. FDA. Guidelines for Good Epidemiologic Practice (GEP). 2005. https://www.fda.gov/files/drugs/published/Good-Pharmacovigilance-Practices-and-Pharmacoepidemiologic-Assessment-March-2005.pdf .

37. FDA. Best Practices for Conducting and Reporting Pharmacoepidemiologic Safety Studies Using Electronic Healthcare Data Sets. 2013. https://www.fda.gov/media/79922/download.

38. International Society of Pharmacoepidemiology (ISPE). Good Pharmacoepidemiology Practices (GPP). 2015. https://www.pharmacoepi.org/resources/policies/guidelines-08027/.

39. Aetion Evidence Platform® (2021). Software for real-world data analysis, which is validated for a range of studies.* Aetion, Inc. https://www.aetion.com.

1. LIST OF TABLES

2. LIST OF FIGURES

ANNEX 1. LIST OF STANDALONE DOCUMENTS

Number	Document Reference Number	Date	Title
			None

ANNEX 2. ADDITIONAL INFORMATION

Annex Table 1. Covariate and Outcome Operational Definitions

Code lists will be included in SAP document.

Variables	Operational Definition[1]
Demographic Variables	
Age	Age in years (continuous)
Sex	Sex (categorical) • Male • Female • Unknown • Missing
Race	Race/Ethnicity (categorical) • White • Black or African American • American Indian or Alaska Native • Asian, Native Hawaiian or Other Pacific Islander • Missing
Insurance Type	Insurance (categorical) • Medicare • Medicaid • Commercial • Other
Region	Region, based on Census geographic regions (categorical) • Northeast • Midwest • South • West
Diagnosing Provider Specialty	Provider specialty • *To be determined, pending data source selection*
SCD Genotype[1]	Genotype (categorical) • Homozygous SS • HSC • HbSB0 thalassemia • HbSB+ thalassemia • Unknown/Other
SCD Comorbidities	

Acute Chest Syndrome	Presence of an ICD-10-Clinical Modification (CM) diagnosis code on any claim
Acute Chest Syndrome + Pneumonia	Presence of an ICD-10-CM diagnosis code on any claim for the following conditions: Acute chest syndrome, AND Pneumonia
Acute Coronary Syndrome	Presence of an ICD-10-CM diagnosis code on any claim
Acute Osteomyelitis	Presence of an ICD-10-CM diagnosis code on any claim
Acute Renal Failure	Presence of an ICD-10-CM diagnosis code on any claim
Aplastic Crisis	Presence of an ICD-10-CM diagnosis code on any claim
Intracerebral And Subarachnoid Hemorrhage	Presence of an ICD-10-CM diagnosis code on any claim
Myocardial Infarction	Presence of an ICD-10-CM diagnosis code on any claim
Priapism	Presence of an ICD-10-CM diagnosis code on any claim
Splenic Sequestration Crisis	Presence of an ICD-10-CM diagnosis code on any claim
Stroke	Presence of an ICD-10-CM diagnosis code on any claim
Transfusion-Associated Reaction	Presence of an ICD-10-CM diagnosis code on any claim
Transient Ischemic Attack	Presence of an ICD-10-CM diagnosis code on any claim
Vaso-Occlusive Crisis (VOC)	Definition 1 (primary analysis, strict definition): • Presence of ICD-10 diagnoses code for SCD with VOC (D57.09, D57.21, D57.41, D57.43, D57.45, D57.81) or dactylitis (D57.04, D57.214, D57.414, D57.434, D57.454, D57.814) on a claim from one of the following healthcare settings:

	○ Urgent care visit ○ Emergency visit ○ Inpatient visit Definition 2 (sensitivity analysis, broad definition): • Presence of ICD-10 diagnoses code for VOCs (definition 1) and/or VOC component (e.g., ACS, priapism, stroke, splenic sequestration, or crisis unspecified) on a claim from one of the following healthcare settings: ○ Telemedicine visit ○ Office clinic visit: Outpatient visit ○ Urgent care visit ○ Emergency visit ○ Inpatient visit Final definition to be included in SAP.
Venous Thromboembolism	Presence of an ICD-10-CM diagnosis code on any claim
Avascular Necrosis	Presence of an ICD-10-CM diagnosis code on any claim
Cholelithiasis	Presence of an ICD-10-CM diagnosis code on any claim
Chronic Kidney Disease	Presence of an ICD-10-CM diagnosis code on any claim
Chronic Pain	Presence of an ICD-10-CM diagnosis code on any claim
Diabetes	Presence of an ICD-10-CM diagnosis code on any claim
Heart Failure	Presence of an ICD-10-CM diagnosis code on any claim
Iron Overload	Presence of an ICD-10-CM diagnosis code on any claim
Leg Ulcer	Presence of an ICD-10-CM diagnosis code on any claim
Pulmonary Hypertension	Presence of an ICD-10-CM diagnosis code on any claim
Retinopathy	Presence of an ICD-10-CM diagnosis code on any claim

Treatments	
Voxelotor	Presence of a national drug code (NDC) or generic name on a pharmacy event claim
Hydroxyurea	Presence of a national drug code (NDC) or generic name on a pharmacy event claim
Crizanlizumab	Presence of a national drug code (NDC) or generic name on a pharmacy event claim OR Presence of an ICD-10-CM procedure, Current Procedural Terminology (CPT), or Healthcare Common Procedure Coding System (HCPCS) code on any claim
L-glutamine	Presence of a national drug code (NDC) or generic name on a pharmacy event claim
Erythropoietin Stimulating Agent (ESA)	Presence of a national drug code (NDC) or generic name on a pharmacy event claim OR Presence of an ICD-10-CM procedure, Current Procedural Terminology (CPT), or Healthcare Common Procedure Coding System (HCPCS) code on any claim
Other Covariates	
RBC transfusion	Presence of an ICD-10-CM procedure, Current Procedural Terminology (CPT), or Healthcare Common Procedure Coding System (HCPCS) code on any claim **Multiple codes or claims occurring on the same day will be assessed as a single RBC transfusion event.**
Hb	Presence of a LOINC code, AND A non-missing, numeric lab result value
Healthcare Interactions - Inpatient/Outpatient Visit	Presence of any claim in the following tables: • Inpatient • Non-inpatient
Iron Overload	Presence of an ICD-10-CM diagnosis code on any claim

Iron Chelation Use	Presence of a national drug code (NDC) or generic name on a pharmacy event claim OR Presence of an ICD-10-CM procedure, Current Procedural Terminology (CPT), or Healthcare Common Procedure Coding System (HCPCS) code on any claim
Delayed Hemolytic Transfusion Reactions	Presence of an ICD-10-CM diagnosis code on any claim
Alloimmunization	Presence of an ICD-10-CM diagnosis code on any claim
Charlson Comorbidity Index (CCI)[42]	CCI (continuous)
Charlson Comorbidities	
Myocardial Infarction	See SCD comorbidities definition above
Congestive Heart Failure	Presence of an ICD-10-CM diagnosis code on any claim
Peripheral Vascular Disease	Presence of an ICD-10-CM diagnosis code on any claim
Cerebrovascular Disease	Presence of an ICD-10-CM diagnosis code on any claim
Dementia	Presence of an ICD-10-CM diagnosis code on any claim
Chronic Pulmonary Disease	Presence of an ICD-10-CM diagnosis code on any claim
Rheumatic Disease	Presence of an ICD-10-CM diagnosis code on any claim
Peptic Ulcer Disease	Presence of an ICD-10-CM diagnosis code on any claim
Mild Liver Disease	Presence of an ICD-10-CM diagnosis code on any claim
Diabetes Without Chronic Complication	Presence of an ICD-10-CM diagnosis code on any claim

Diabetes With Chronic Complication	Presence of an ICD-10-CM diagnosis code on any claim
Hemiplegia or Paraplegia	Presence of an ICD-10-CM diagnosis code on any claim
Renal Disease	Presence of an ICD-10-CM diagnosis code on any claim
Any Malignancy Without Metastasis	Presence of an ICD-10-CM diagnosis code on any claim
Leukemia	Presence of an ICD-10-CM diagnosis code on any claim
Lymphoma	Presence of an ICD-10-CM diagnosis code on any claim
Moderate or Severe Liver Disease	Presence of an ICD-10-CM diagnosis code on any claim
Metastatic Solid Tumor	Presence of an ICD-10-CM diagnosis code on any claim
Acquired Immunodeficiency Syndrome	Presence of an ICD-10-CM diagnosis code on any claim
Patient-Level Outcome Definitions	
RBC transfusions	For each patient, the number of RBC transfusions will be evaluated PPPY in baseline and follow-up. See covariate definition above for operational definition and Table 4 for output definition. See Annex Table 2 for RBC transfusion codes.
VOCs	For each patient, the number of VOCs will be evaluated PPPY in baseline and follow-up. See covariate definition above for operational definition and Table 4 for output definition.
Stroke	For each patient, the number of events will be divided by person-time in follow-up. See covariate definition above for operational definition and Table 4 for output definition.
ACS	For each patient, the number of events will be divided by person-time in follow-up. See covariate definition above for operational definition and Table 4 for output definition.
ACS + Pneumonia	For each patient, the number of events will be divided by person-time in follow-up. See covariate definition above for operational definition and Table 4 for output definition.

Priapism	For each patient, the number of events will be divided by person-time in follow-up. See covariate definition above for operational definition and Table 4 for output definition.
Iron overload	For each patient, the number of events will be divided by person-time in follow-up. See covariate definition above for operational definition and Table 4 for output definition.
Iron chelation use	For each patient, the number of events will be divided by person-time in follow-up. See covariate definition above for operational definition and Table 4 for output definition.
Delayed hemolytic transfusion reactions Alloimmunization	For each patient, the number of events will be divided by person-time in follow-up. See covariate definition above for operational definition and Table 4 for output definition.

[1] Diagnosis (ICD), procedure (ICD, CPT, HCPCS), and drug codes (NDC) or generic names will be documented in the code list accompanying the statistical analysis plan.

Annex Table 2. RBC Transfusion Codes

Code	Description	Code Type
36430	Transfusion, blood or blood components	CPT
36440	Push transfusion, blood, 2 years or under	CPT
36450	Exchange transfusion, blood, newborn	CPT
36455	Exchange transfusion, blood; other than newborn	CPT
36455	EXCHNG TRANSFUSION BLOOD OTHER/THAN NEW BORN	CPT
36516	Therapeutic apheresis; with extracorporeal immunoadsorption, selective adsorption, or selective filtration and plasma reinfusion	CPT
P9010	BLOOD (WHOLE), FOR TRANSFUSION, PER UNIT	HCPCS
P9011	BLOOD, SPLIT UNIT	HCPCS
P9016	Red blood cells, leukocytes reduced, each unit	HCPCS
P9021	Red blood cells, each unit	HCPCS
P9022	Red blood cells, washed, each unit	HCPCS
P9038	Red blood cells, irradiated, each unit	HCPCS
P9039	Red blood cells, deglycerolized, each unit	HCPCS
P9040	Red blood cells, leukocytes reduced, irradiated, each unit	HCPCS
P9051	Whole blood or red blood cells, leukocytes reduced, cmv-negative, each unit	HCPCS
P9054	Whole blood or red blood cells, leukocytes reduced, frozen, deglycerol, washed, each unit	HCPCS
P9057	Red blood cells, frozen/deglycerolized/washed, leukocytes reduced, irradiated, each unit	HCPCS
P9058	Red blood cells, leukocytes reduced, cmv-negative, irradiated, each unit	HCPCS
30230N1	Transfusion of Nonautologous Red Blood Cells into Peripheral Vein, Open Approach	ICD10 PROC
30230P1	Transfusion of Nonautologous Frozen Red Cells into Peripheral Vein, Open Approach	ICD10 PROC
30233N1	Transfusion of Nonautologous Red Blood Cells into Peripheral Vein, Percutaneous Approach	ICD10 PROC
30233P1	Transfusion of Nonautologous Frozen Red Cells into Peripheral Vein, Percutaneous Approach	ICD10 PROC
30240N1	Transfusion of Nonautologous Red Blood Cells into Central Vein, Open Approach	ICD10 PROC
30240P1	Transfusion of Nonautologous Frozen Red Cells into Central Vein, Open Approach	ICD10 PROC
30243N1	Transfusion of Nonautologous Red Blood Cells into Central Vein, Percutaneous Approach	ICD10 PROC
30243P1	Transfusion of Nonautologous Frozen Red Cells into Central Vein, Percutaneous Approach	ICD10 PROC

30250N1	Transfusion of Nonautologous Red Blood Cells into Peripheral Artery, Open Approach	ICD10 PROC
30250P1	Transfusion of Nonautologous Frozen Red Cells into Peripheral Artery, Open Approach	ICD10 PROC
30253N1	Transfusion of Nonautologous Red Blood Cells into Peripheral Artery, Percutaneous Approach	ICD10 PROC
30253P1	Transfusion of Nonautologous Frozen Red Cells into Peripheral Artery, Percutaneous Approach	ICD10 PROC
30260N1	Transfusion of Nonautologous Red Blood Cells into Central Artery, Open Approach	ICD10 PROC
30260P1	Transfusion of Nonautologous Frozen Red Cells into Central Artery, Open Approach	ICD10 PROC
30263N1	Transfusion of Nonautologous Red Blood Cells into Central Artery, Percutaneous Approach	ICD10 PROC
30263P1	Transfusion of Nonautologous Frozen Red Cells into Central Artery, Percutaneous Approach	ICD10 PROC
6A550Z0	Pheresis of Erythrocytes, Single	ICD10 PROC
6A551Z0	Pheresis of Erythrocytes, Multiple	ICD10 PROC

Document Approval Record

Document Name:	C5341060 Non Interventional Study Protocol V01_01 August 2024
Document Title:	C5341060 Non Interventional Study Protocol V01_01 August 2024

Signed By:	Date(GMT)	Signing Capacity
[redacted]	[redacted]	[redacted]