NON-INTERVENTIONAL (NI) STUDY PROTOCOL

PASS 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	[redacted]
Protocol version identifier	V2.1
Date	10 January 2025
EU Post Authorization Study (PAS) register number	EUPAS1000000225
Active substance	B06AX03
Medicinal product	Oxbryta (voxelotor)
Product reference	N/A
Procedure Number	N/A
Marketing Authorization Holder	N/A
Joint PASS	No
Research question and objectives	The research question for this study will be: • Among individuals with sickle cell disease (SCD) and a history of red blood cell (RBC) transfusions, do individuals treated with voxelotor have a greater reduction in RBC transfusions per patient per year (PPPY) between baseline and follow-up (up to one year after index date) compared to matched individuals not treated with voxelotor? **Objectives:** Among individuals with SCD and a history of RBC transfusions, the following objectives will be compared between those treated with voxelotor and matched individuals who are not treated with voxelotor: • **Primary objective:** Compare the change in RBC transfusions (PPPY from baseline [one year prior to index date] to follow-up [up to one year after index date]). • **Exploratory objective 1:** **1.1** Compare the change in SCD-related complications including vaso-occlusive crises (VOCs), acute chest syndrome (ACS), and priapism (PPPY from baseline [one year prior to index date] to follow-up [up to one year after index date]); **1.2** compare the time-to-stroke during the follow-up period.

	• **Exploratory objective 2:** Compare the number of RBC transfusions PPPY during the follow-up period (including those with 0 RBC transfusions). The percentage of patients with more than a 30% reduction in RBC transfusions from baseline will also be compared. • **Exploratory objective 3:** Compare the time to RBC transfusion-associated complications (iron overload, iron chelation use, delayed hemolytic transfusion reactions, alloimmunization) during the follow-up period.
Country of study	USA
Author	[redacted]

1. TABLE OF CONTENTS

1. TABLE OF CONTENTS 3
2. LIST OF ABBREVIATIONS 5
3. RESPONSIBLE PARTIES 7
4. ABSTRACT 8
5. AMENDMENTS AND UPDATES 11
6. MILESTONES 14
7. RATIONALE AND BACKGROUND 14
8. RESEARCH QUESTION AND OBJECTIVES 15
9. RESEARCH METHODS 16
9.1. Study Design 16
9.2. Setting 17
9.2.1. Inclusion Criteria 20
9.2.2. Exclusion Criteria 21
9.3. Variables 22
9.4. Data Sources 28
9.5. Study Size 29
9.6. Data Management 29
9.7. Data Analysis 30
9.7.1. Descriptive Analysis 30
9.7.2. Baseline Balance 31
9.7.3. Outcome Analysis 33
9.8. Quality Control 37
9.8.1. Analytic Quality Control 37
9.8.2. Data Storage 37
9.9. Strengths and limitations of the research methods 38
9.10. Other Aspects 40
10. PROTECTION OF HUMAN PARTICIPANTS 40
10.1. Patient Information 40
10.2. Patient Consent 40
10.3. Institutional Review Board (IRB)/ Ethics Committee (EC) 40
10.4. Ethical Conduct of the Study 40

11. MANAGEMENT AND REPORTING OF ADVERSE EVENTS/ADVERSE REACTIONS 41
12. PLANS FOR DISSEMINATING AND COMMUNICATING STUDY RESULTS 41
13. REFERENCES 42
14. LIST OF TABLES 45
15. LIST OF FIGURES 45
ANNEX 1. LIST OF STANDALONE DOCUMENTS 45
ANNEX 2. ADDITIONAL INFORMATION 45

2. LIST OF ABBREVIATIONS

Abbreviation	Definition
ACS	Acute Chest Syndrome
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
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	Address
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]

Core Study Team and Additional Responsibility Parties

Name, Degree(s)	Job Title	Affiliation
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]

4. ABSTRACT

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

Version: 2.0

Date: 23 December 2024

Main author: [redacted]

Rationale and background: The current treatment landscape for 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 (VOC), and stroke. However, every RBC transfusion administered carries risks, such as iron overload (which can lead to end-organ damage), alloimmunization, infections, or delayed hemolytic transfusion reactions (DHTR) 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®) inhibits the polymerization of sickle hemoglobin (Hb) polymerization and has shown clinical benefit on Hb levels in individuals with SCD, which may limit the need for RBC transfusions in the management of SCD. Pfizer voluntarily withdrew voxelotor from the market on September 25, 2024, due to emerging preliminary data from clinical and registry-based real-world evidence (RWE) studies that showed the benefit of voxelotor no longer outweighs the risk; a comprehensive analysis of these studies is currently ongoing. In December 2022, voxelotor had been recommended by the ASH Guideline Monitoring Expert Working Group for consideration as a preventive maintenance therapy to improve the baseline Hb of those with SCD at high risk of DHTR if red cell transfusion is required.[1] This observational study using a retrospective cohort from a claims database aims to determine if treatment with voxelotor reduces RBC transfusion rate compared to a similar population of individuals with SCD not treated with voxelotor matched on age and calendar date. The results from the study are intended to add to the totality of evidence for voxelotor pertaining to internal and external decision making for marketing authorization.

Research question and objectives:

The research question for this study will be:

- Among individuals with SCD and a history of RBC transfusions, do those treated with voxelotor have a greater reduction in RBC transfusions per-patient-per year

[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

(PPPY) between baseline and follow-up (up to one year after index date) compared to matched individuals not treated with voxelotor?

Objectives: Among individuals with SCD and a history of RBC 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 RBC transfusions (PPPY from baseline [one year prior to index date] to follow-up [up to one year after index date]).
- **Exploratory objective 1:** **1.1** Compare the change in SCD-related complications including VOCs, acute chest syndrome (ACS), and priapism (PPPY from baseline [one year prior to index date] to follow-up [up to one year after index date]); **1.2** compare the time-to-stroke during the follow-up period.
- **Exploratory objective 2:** Compare the number of RBC transfusions PPPY during the follow-up period (including those with 0 RBC transfusions). The percentage of patients with more than a 30% reduction in RBC transfusions from baseline will also be compared.
- **Exploratory objective 3:** Compare the time to RBC transfusion-associated complications (iron overload, iron chelation use, DHTRs, alloimmunization) during the follow-up period.

Study design: This is a non-interventional, retrospective 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 in comparative analyses.

Population: The study population will include individuals with SCD between 12 and 85 years of age who are enrolled in a health insurance plan (commercial, Medicare, Medicaid) continuously during the 365 days prior to and including the index date (baseline) and 180 days post-index. The index selection period will be November 25, 2019 (date of FDA accelerated approval for voxelotor) through January 31, 2023. The exposed group will include patients with a new prescription claim for voxelotor (no 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 four non-voxelotor patients based on age and calendar date. Both the exposed and control groups may be treated with other 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 (e.g., age, sex, insurance type) as well as patient characteristics (e.g., specialty provider type, common SCD comorbidities, treatment history, Hb levels, number of prior RBC transfusions, and healthcare utilization). The primary outcome of interest is RBC transfusion. Exploratory outcomes include VOC, stroke, and RBC transfusion complications. Tables 4 and 5 define exposures, outcomes, and key covariates.

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 (defined as the date of a new voxelotor prescription claim for treated individuals, and the matched date for untreated individuals) from November 25, 2019, until January 31, 2023. 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 linear regression model with an adjustment factor of 1.65 which takes into account variance inflation in the analysis when propensity score adjustment will be applied. With 185 patients treated with voxelotor and 370 matched patients not treated with voxelotor (1:2 matching ratio), the study will have an 80% power to detect at least a 1-unit (1 RBC transfusion) 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 gaps), death, discontinuation of voxelotor (with a 45-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 RBC transfusions PPPY will be analyzed using a generalized linear model with an identity link to estimate the difference in RBC transfusion change rates between the two groups (primary objective). If any pre-index covariates that are found to be unbalanced after the PS method, they may be included in the regression as a covariate. Depending on the measurement, absolute rate differences or median time to event will also be reported for each of the transfusion-associated secondary complications and secondary effectiveness endpoints (exploratory objectives 1 and 3). Logistic regression will be used to compare the percentage of patients with zero RBC transfusions in follow-up (exploratory objective 2, endpoint 3) and the percentage of patients with a more than a 30% reduction in RBC transfusions from baseline (exploratory objective 2, endpoint 2). 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. 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

Version Identifier	Date	Amendment Type (substantial or administrative)	Protocol Section(s) Changed	Summary of Amendment(s)	Reason
v1.0	2024 August 20				
v1.1	2024 October 02	Administrative	Section 9.1	Removal of matching caliper as the default option for coarsened-exact matching	Increase precision of the age match in the coarsened-exact match process
			Section 9.1	Addition of exclusion criteria (iron overload, iron chelation use, DHTR, alloimmunization) for Exploratory Objective 3	To enable assessment of incident events

			Section 9.2 - Table 4	Stroke and transient ischemic attack removed from footnote definition of SCD-related comorbidities. Clarification added to Charlson Comorbidity index footnote.	Stroke will not be measured in baseline as it is an exclusion criterion.
			Section 9.6	Removal of VOC outcomes in Exploratory Analysis 1	VOC outcomes will be assessed in other work
			Section 9.6	Assessment of follow-up time and application of pairwise censoring as a sensitivity analysis (instead of a primary analysis)	To adjust for differences in follow-up time between voxelotor patients and their matched controls
			Section 9.6.3	Exploratory Objective 3 (transfusion-associated complications) will be assessed in a subpopulation of patients without baseline events using cox proportional hazards models for each outcome.	To enable assessment of incident transfusion-associated complications.

2.0	23 December 2024	Substantial	Section 8 Section 9.2 Section 9.6.3 Section 9.8	Adding VOC outcomes in Exploratory Analysis 1. Exploratory objective 1.1 updated to match the analysis and measurement of the primary objective.	VOC outcome is now of interest to study team and reflected throughout the study document. Exploratory objective 1.1 updated to match the analysis and measurement of the primary objective.
		Administrative	All sections	Ensured all sections were updated to be identical to CT-24 templates.	Updated all sections to ensure they followed the exact guidance of the CT-24 templates.
2.1	10 January 2025	Administrative	Section 4 Section 9.2	Update the coarsened exact match to include up to four non-voxelotor patients.	This will allow for propensity score matching of up to two referent patients.
2.2	27 January 2025	Substantial	Abstract, Figure 1 Study Schematic, Section 9.2.1, and Section , and 9.7.3	Voxelotor allowable treatment gap updated from 30-days to 45 days and required continuous enrollment of insurance and continuous voxelotor exposure 180-days after index.	Exploratory analysis prior to endpoint assessment revealed ~25-30% of voxelotor patients did not receive more than 30 days of treatment, which can bias the follow up times between the study groups. Providers often titrate patients when starting voxelotor, so a larger treatment gap allows us to keep more of these patients.

6. MILESTONES

Milestone	Planned Date
Registration in the HMA-EMA Catalogues of RWD studies	22 June 2024
Protocol v1.0 Approval Date	20 August 2024
Start of data collection	06 January 2025
End of data collection	06 January 2025
Interim study report	05 March 2025
Final study report	04 April 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 (VOC), 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 not treat the underlying cause of the disease, can have a high resource burden (e.g., 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 transfusions, 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, ACS, 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 (DHTR) and mortality, and transfusion transmitted infections can cause organ damage and mortality.[3,8–11] 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.[12] A significant unmet need exists for new therapies and interventions that address the clinical complications of SCD while mitigating the patient burden.

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.[13] 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. While 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.[14,15]

Pfizer voluntarily withdrew voxelotor from the market on September 25, 2024, due to emerging preliminary data from clinical and registry-based real-world evidence (RWE) studies that showed the benefit of voxelotor no longer outweighs the risk; a comprehensive analysis of these studies is currently ongoing. In December 2022, voxelotor was recommended by the ASH Guideline Monitoring Expert Working Group for consideration as a preventive maintenance therapy to improve the baseline Hb of those with SCD at high risk of DHTR if RBC transfusion is required. 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 $\geq$ 12 years of age in the US, who have a history of RBC transfusions using real-world data (RWD). The results from this study are intended to add to the totality of evidence for voxelotor pertaining to internal and external decision making for marketing authorization.

[redacted]

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. The research question for this study will be:

- Among individuals with SCD and a history of RBC transfusions, do individuals treated with voxelotor have a greater reduction in RBC transfusions per-patient-per year (PPPY) between baseline and follow-up (up to one year after index date) compared to matched SCD individuals not treated with voxelotor?

Objectives: Among individuals with SCD and a history of RBC transfusions, the following objectives will be compared between those treated with voxelotor and matched individuals with SCD who are not treated with voxelotor:

- **Primary objective:** Compare the change in RBC transfusions (PPPY from baseline [one year prior to index date] to follow-up [up to one year after index date]).
- **Exploratory objective 1:** **1.1** Compare the change in SCD-related complications including VOCs, ACS, and priapism (PPPY from baseline [one year prior to index date] to follow-up [up to one year after index date]); **1.2** compare the time-to-stroke during the follow-up period.
- **Exploratory objective 2:** Compare the number of RBC transfusions PPPY during the follow-up period (including those with 0 RBC transfusions). The percentage of patients with more than a 30% reduction in RBC transfusions from baseline will also be compared.
- **Exploratory objective 3:** Compare the time to RBC transfusion-associated complications (iron overload, iron chelation use, DHTRs, alloimmunization) during the follow-up period.

9. RESEARCH METHODS

9.1. Study Design

This will be a non-interventional, retrospective study of patients with SCD using US RWD 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 RBC 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. A robust study design and analytical methods will be employed, following best practices to minimize or eliminate biases where possible. This study will employ methods such as coarsened-exact-matching (CEM), propensity score (PS) adjustment, pair-wise censoring, and multiple sensitivity analyses to improve the study's internal validity and ensure confidence in the results. PS methods (e.g., matching or weighting) will be utilized to adjust for measured confounding including patient demographics, comorbidities and variations in treatment (e.g., concurrent medications).

A robust and reliable RWE approach overcomes challenges associated with conducting an RCT, eliminates the burden of trial participation for patients, and can appropriately address the research question. Key considerations to support the clinical meaningfulness of a reduction in RBC transfusions include defining the target SCD population for transfusions and ensuring there is no worsening of the clinical benefit of transfusions. These considerations were incorporated in this RWD study of a retrospective claims-based study in a US SCD population with a history of RBC transfusions and receiving marketed voxelotor or standard-of-care (SOC).

In the US, RBC transfusion documentation from RWD sources is unbiased, reliable, and readily available in medical claims. Pfizer worked with an independent study vendor to conduct a retrospective, observational study to evaluate the accuracy of RBC transfusions by using algorithms based on diagnosis codes and procedure codes from inpatient and outpatient records using Optum claims and EHR data (details provided in Section 9.4)[28].

9.2. Setting

The study population will include US 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) and 180 days post-index. The index selection period is November 25, 2019 (date of Food and Drug Administration [FDA] accelerated approval for voxelotor) through January 31, 2023. 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 have a history of stroke will be excluded from this analysis. Both the exposed and control groups may be treated with other SOC treatments such as HU and have 2 or more documented RBC transfusions prior to study index.

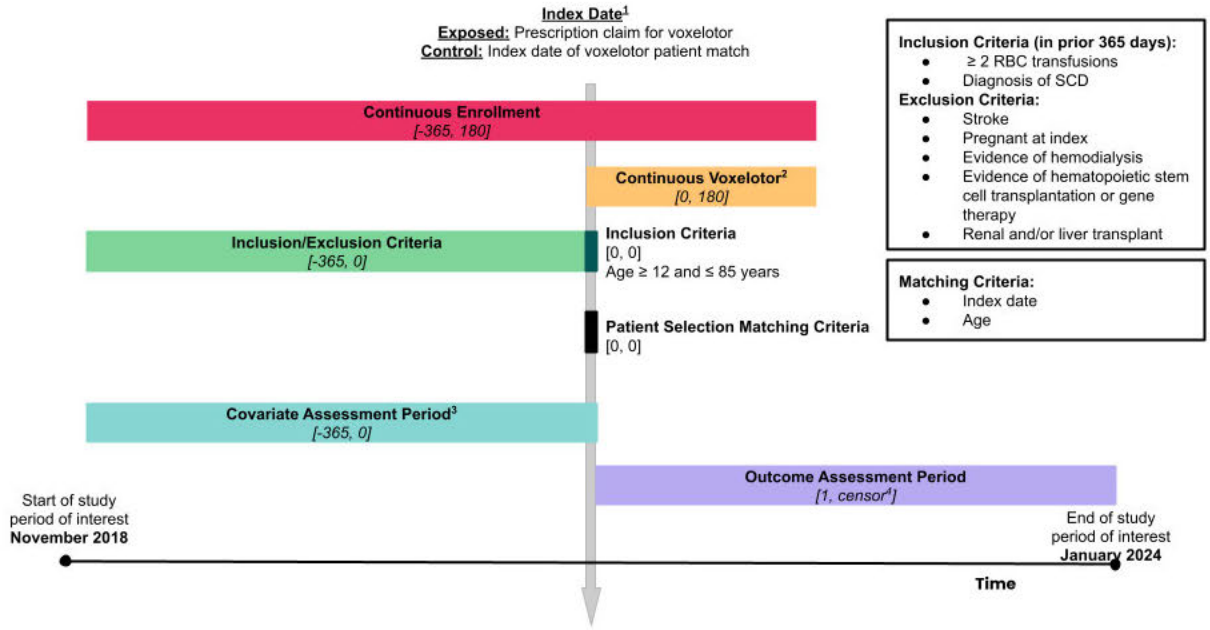
Depending on the source of information, approximately 50% of individuals with SCD in the US have received a RBC transfusion, and approximately 20% have received a transfusion in the past year.[16] In the feasibility assessment of Komodo Healthcare Map (KHM) dataset, race/ethnicity was documented for 90% (n=2230/2486) of patients with 87% (n=1929/2230) identifying as Black or African American. The Centers for Disease Control and Prevention (CDC) estimates 90% of individuals with SCD in the US are non-Hispanic Black or African American.[17] Socioeconomic status also plays an important factor in impacting patient behavior and access to care. While the KHM dataset does not have direct measures of patients' socioeconomic status, important indicators will be leveraged in the analysis for patients including geography by region, health insurance type, and whether a hematologist is involved in the patient's care. KHM is nationally representative of individuals with SCD. In the Komodo data, approximately 68% of patients have Medicaid as their primary insurance type.

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 CEM with up to four non-voxelotor patients based on exact age (year) and calendar date to achieve balance between groups.[18] 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 individuals with SCD have better access to care before adulthood and accumulate comorbidities through the natural course of the disease.[19] The matched controls will be selected from the pool of patients who meet all the eligibility criteria below and are the same age (year) on the same day as the

voxelotor claim. Depending on available matches on the same date as the voxelotor claim, an age caliper may be considered to preserve sample size.

Voxelotor individuals will be matched in chronological order where up to four referent non-voxelotor controls will be identified for each voxelotor individuals before additional referent non-voxelotor matches are identified for the remaining exposure group individuals. Individuals 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 individuals are matched to a voxelotor patient on the same 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 (e.g., COVID-19 pandemic). Available matches depend on the number of eligible individuals available in the data source.

Figure 1. Study Design Schematic



Abbreviations: RBC: Red blood cell; SCD: Sickle cell disease

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 4 non-voxelotor patients based on age and calendar date.
2. Continuous voxelotor use defined using the voxelotor ship date and days supply with an allowable gap of at most 45 days (voxelotor arm only)
3. The full list of potential covariates including definitions can be found in Table 3.
4. Follow-up will end at the earliest of the following events: 365 days after index, death, disenrollment, the occurrence of any exclusion criteria events (e.g., pregnancy, stroke, hemodialysis, evidence of HSCT or gene therapy, renal transplant, or liver transplant), 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 (January 31, 2024; the matched controls will index on the same date of their matched

Voxelotor

[redacted]

V2.0 23 December 2024

[redacted]

[redacted]

[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]]

[redacted]

9.2.1. Inclusion Criteria

Table 2. Summary of Inclusion 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)	Administrative claims study[20]
Age between 12 and 85 years	[Index]	Age in years since year of birth	Age range covers the initial accelerated approval indication.
≥ 2 RBC transfusions	[Baseline]	At least 2 medical claims with an 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).	The subset of SCD patients who would most benefit from a reduction in RBC transfusion and necessary to measure a change from baseline to follow-up.
≥ 365 days of continuous enrollment during baseline and at least 180 days in follow up	[Baseline] [Follow Up	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.
≥ 180 days of continuous voxelotor use post-index (with 45-day gaps allowed) – ***voxelotor arm only***	[Follow up]	At least 180 days of continuous voxelotor use post-index (with 45-day gaps allowed) defined using the voxelotor ship date and days supply reported on the claim– ***voxelotor arm only***	Adequate exposure to voxelotor is necessary to evaluate the impact on RBC transfusions and other outcomes.

9.2.2. Exclusion Criteria

Table 3. Summary of Exclusion Criteria

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. Individuals with SCD may receive RBC transfusions to ensure adequate oxygenation of the fetus during pregnancy. Voxelotor is not indicated during pregnancy.
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 or gene therapy, 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 management and will therefore be excluded.
Iron overload - Exploratory Objective 3 Subgroup Only	[Baseline]	At least one medical claim with an ICD-10-CM diagnosis code with evidence of iron overload	To assess incident events in follow-up for Exploratory Objective 3
Iron chelation use - Exploratory Objective 3 Subgroup Only	[Baseline]	At least one pharmacy claim with an NDC or medical claim with an HCPCS/CPT for iron chelation	To assess incident events in follow-up for Exploratory Objective 3
DHTR - Exploratory Objective 3 Subgroup Only	[Baseline]	At least one medical claim with an ICD-10-CM diagnosis code with evidence of DHTR	To assess incident events in follow-up for Exploratory Objective 3
Alloimmunization - Exploratory Objective 3 Subgroup Only	[Baseline]	At least one medical claim with an ICD-10-CM diagnosis code with evidence of alloimmunization	To assess incident events in follow-up for Exploratory Objective 3

9.3. Variables

The exposure of interest (i.e., voxelotor) will be defined using Claritas prescription data, lab assessments (i.e., Hb labs) will be defined using Quest Diagnostics data, and all other study variables will be defined using KHM database closed claims. See Section 9.4 for full

description of the data that will be used for this study. A list of variables for the study is outlined in Table 4 and the endpoints are described in Table 5. The operational definition for each variable listed in Tables 4 and 5 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 accurately because healthcare entities must bill to receive reimbursement.[21] 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.[22–26]

Table 4. 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/ethnicity	Baseline Characteristic; Covariate	[Index]	Frequency and proportion (N, %) Categorical: White, Black or African American, Asian or Pacific Islander, Hispanic or Latino, Other, Unknown
Insurance Type	Baseline Characteristic; Covariate	[Index]	Frequency and proportion (N, %) Categorical: Medicare, Medicaid,

			Commercial, Multiple, Other
Region	Baseline Characteristic; Covariate	[Index]	Frequency and proportion (N, %) Categorical: Northeast, Midwest, South, and West Regions (based on Census geographic regions), Missing
Top 10 Diagnosing Provider Specialty	Baseline Characteristic; Covariate	[Baseline]	Frequency and proportion (N, %) [Provider specialty categories will be included in statistical analysis plan (SAP)]
SCD Hematology Diagnosing Provider	Baseline Characteristic; Covariate	[Baseline]	Frequency and proportion (N, %) Binary: Yes, No (including non-hematology and null values)
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	[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]; [Follow-up]	Continuous: Mean (SD); Median (IQR); Minimum; Maximum.
Baseline Hb lab test	Baseline Characteristic	[Baseline]	Frequency and proportion (N, %) Categorical: Komodo lab test only, Quest lab test only, Komodo lab test and Quest lab test, No value
Baseline Hb value[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.
Healthcare Interactions - Inpatient/Outpatient Visit	Baseline Characteristic; Covariate	[Baseline]	Continuous: Mean (SD); Median (IQR); Minimum; Maximum.
Follow-up time	Baseline Characteristic; Covariate	[Follow-up]	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, %)
Death	Baseline Characteristic; Covariate	[Follow-up]	Frequency and proportion (N, %)
Voxelotor discontinuation	Baseline Characteristic; Covariate	[Follow-up]	Frequency and proportion (N, %)

Abbreviations: ESA: Erythropoietin Stimulating Agent; RBC: Red blood cell; SCD: Sickle cell disease.

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, transfusion-associated reaction, venous thromboembolism, avascular necrosis, cholelithiasis, chronic kidney disease, chronic pain, diabetes, heart failure, iron overload, leg ulcer, pulmonary hypertension, VOC, 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 (mild to moderate), renal disease (severe), any malignancy without metastasis including leukemia and lymphoma, moderate or severe liver disease, metastatic solid tumor, human immunodeficiency virus, 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 5. 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 change in the number of events (maximum 1 per day) PPPY from baseline to follow-up. VOC events need to occur more than 5 days apart to be considered separate events.	Exploratory Objective 1.1
ACS	[Baseline]; [Follow-up]	Mean change in the number of events (maximum 1 per day) PPPY from baseline to follow-up	Exploratory Objective 1.1
ACS Including Pneumonia	[Baseline]; [Follow-up]	Mean change in the number of events (maximum 1 per day) PPPY from baseline to follow-up	Exploratory Objective 1.1
Priapism	[Baseline]; [Follow-up]	Mean change in the number of events (maximum 1 per day) PPPY from baseline to follow-up	Exploratory Objective 1.1

Stroke (including transient ischemic attack)	[Follow-up]	Time to event during follow-up	Exploratory Objective 1.2
RBC transfusions in follow-up	[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	[Follow-up]	Proportion of patients with 0 RBC transfusions (maximum 1 per day) PPPY during follow-up	Exploratory Objective 2
Iron overload	[Follow-up]	Time to event during follow-up	Exploratory Objective 3
Iron chelation use	[Follow-up]	Time to event during follow-up	Exploratory Objective 3
Delayed hemolytic transfusion reactions	[Follow-up]	Time to event during follow-up	Exploratory Objective 3
Alloimmunization	[Follow-up]	Time to event during follow-up	Exploratory Objective 3
Composite outcome (iron overload, iron chelation use, delayed hemolytic transfusion reactions, and/or alloimmunization)	[Follow-up]	Time to 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

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.[27] The study will use Komodo's Healthcare Map (KHM) database linked to Claritas prescription data (proprietary voxelotor claims) and Quest Diagnostics lab data. While Komodo data is de-identified, Komodo requires that all patients have tokens from the original personally identifiable information; Datavant is used for the tokenization process. These tokens are used to link the information from unique patients together from the KHM database, Claritas prescription data, 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. 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.

Pfizer has conducted a validation study[28] 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.

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 January 2024. 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. The KHM data dictionary is included as a standalone document to this protocol.

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 (July 31, 2024). 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 November 25, 2019, until January 31, 2023. 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 RBC transfusions PPPY during follow-up).[14,15]

A conservative sample size estimate was calculated using a linear regression model with an adjustment factor of 1.65 which takes into account the variability that is introduced by applying propensity score adjustment methods.[29] With 185 patients treated with voxelotor and 370 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) 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

After assigning the tokens to each patient across the three unique data sources (described above in Section 9.4), Komodo Health delivers the data in one delivery. The data that is delivered is a subset of the KHM data and includes patients with at least one diagnosis for SCD at any time. 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 (i.e., non-missing counts, unique counts)
- Variable distribution (e.g., 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.[30] 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 6. 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 individuals are matched to unexposed patients using variable-ratio (1:2) greedy nearest neighbor matching with a caliper, will be the primary approach.[31] 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.[32] 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 (i.e., 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 6 below.

Table 6. 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 (i.e., 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.[33] Collinear covariates will be defined as any covariates with a Pearson correlation coefficient greater than or equal to 0.7.[34] 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.[35]	
3	Confirm baseline confounder balance	Balance of all potential confounders (i.e., covariates) will be confirmed for voxelotor vs referent patients. Covariate balance will be defined as an ASD ≤ 0.10.[36] Balance of these variables will still be confirmed after PS matching. Should any variables with small residual imbalance remain, 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 across the exposure and referent groups that are associated with the outcome (e.g., 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.[37] Any coefficient with extreme values that strongly predicts presence or absence of the exposure of interest (i.e., potential instrumental variables) or another covariate (i.e., 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 45-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 (e.g., 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. 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. To adjust for differences in follow-up time between voxelotor patients and their matched controls, matched or pairwise censoring will be applied to censor unexposed on the date of voxelotor discontinuation for their matched pair as a sensitivity analysis.[38] A summary of sensitivity analyses is included in Table 7 to evaluate how the magnitude of this bias will be assessed.

Primary Objective – Change in Number of RBC Transfusions PPPY

For each patient, the difference in the number of RBC transfusions PPPY from baseline to follow-up period will be reported as the change and will be calculated as follows:

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

*All patients have 365 days of baseline data

The change in the number of RBC transfusions PPPY will be analyzed using generalized linear model (GLM) with an identity link to estimate the difference in RBC 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 rate change and its 95% confidence intervals (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% CI will be reported for the difference between the groups. Full details will be provided in the SAP.

Exploratory Objective 1 – SCD-related Complications

Exploratory Objective 1.1 – Change in SCD-related Complications PPPY

VOCs, ACS, ACS + pneumonia, and priapism will be assessed as the change in the rate of events PPPY from baseline to follow-up. For each individual, the rate of each clinical outcome will be calculated as the number of events divided by the number of person-days. The change will be calculated as:

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

*All patients have 365 days of baseline data

The change in number of events PPPY will be analyzed using GLM with an identity link to estimate the difference in clinical outcome 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 rate change and its 95% CI will be reported.

Exploratory Objective 1.2 – Time to Stroke

Stroke will be assessed as a time-to-event outcome. Cox proportional hazards (PH) models will be used to calculate the hazard ratio (HR); 95% CIs will be calculated using a robust variance estimator.

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

The number of RBC transfusions that occur during the follow-up period will be assessed PPPY during available follow-up within each group. The mean number of RBC transfusions PPPY will be reported for each group. The mean difference between the groups will be assessed using a GLM similar to the primary analysis. The 95% CI will be reported for the difference between the groups.

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

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

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

*All patients have 365 days of baseline data

The mean percentage change in the number of RBC transfusions PPPY within each group will be reported. The proportion of patients with a 30% reduction in RBC transfusions will also be reported for each group. A 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. 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 and an odds ratio will be estimated. The 95% CIs will also be reported.

Exploratory Objective 3 – Time to Transfusion-Associated Complications

Each RBC-associated complication (including the composite outcome) will be assessed as a time-to-event outcome. Cox PH models will be used to calculate the HR; 95% CIs will be calculated using a robust variance estimator. The Cox PH regression model may incorporate confounders that remain unbalanced after PS adjustment.

Sensitivity and Subgroup Analyses

Table 7 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 7. 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	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.	To address the impact of informative censoring and differential follow-up
Matched censoring	If meaningful differences in follow-up time are observed, matched censoring will be performed to assess the impact of differential follow-up between the exposed and referent group for the primary objective. Bias due to differential follow-up can arise when the reasons for participants dropping out or being lost to follow-up are related to both the treatment and the outcome. Matched censoring mitigates this bias by censoring all patients within a matched set (i.e., patients that have been matched by PS matching) at the earliest time in which a patient in the set was censored. In the context of 1:2 PS matching, where each exposed patient is matched to two referent patients, the matched set will consist of at most three individuals. Thus, matched censoring in this scenario involves censoring data for all three individuals in a matched set at the earliest observed censoring time among them.	To address the impact of informative censoring and differential follow-up
Stratification by number of RBC transfusions in baseline	The number of RBC 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 at 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.

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 change in the rate of VOCs and its separate components from baseline through 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.[39–41]

The AEP maintains a date / time-stamped record of all analytic cohorts and corresponding analyses.[42]

9.8.2. 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.

9.9. Strengths and limitations of the research methods

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 (e.g., Hb levels, SCD genotype, comorbidity scores, etc.).[43] Additionally, in an RCT environment, healthcare interactions are on a protocolized schedule for enrolled patients. In RWD, differences exist in how often a patient sees 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. RWD allows for less restrictive inclusion/exclusion criteria and will allow for patients with more severe SCD to be evaluated; for example, all individuals with SCD and recent history of RBC transfusion are excluded from voxelotor clinical trials as well as other medications in development for SCD.[13]

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 (i.e., 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).[44] 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.[44] Additionally, as determined through exploratory analyses, 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 such patients will not be identifiable. However, this is estimated to be a small portion of patients within the study (less than 5% of pediatric patients with SCD receive chronic transfusions).[45]

When possible, validated claims-based algorithms will be utilized to define inclusion/exclusion criteria and outcomes. Pfizer conducted a validation study to assess the accuracy of medical claims billed for RBC transfusions. VOCs are more likely to be misclassified and underestimated in claims data, 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 (3 external hematologists) were 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.[25,26,46] 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 (e.g., 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 (e.g., 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 on 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 (i.e., 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. For most subgroup analyses, the PS obtained from the overall study cohort PS model will be used for PS adjustment. However, for the subgroup of patients with non-missing Hb values in baseline, the PS model will be re-fit and will include an additional covariate (mean Hb value). 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, matched censoring will be used to censor matched controls on the date of their voxelotor match's discontinuation date as a sensitivity analysis.

A summary of the potential limitations of the RWD in this study and the mitigation strategies that will be applied is provided in Annex Table 3.

9.10. Other Aspects

Missing data will be handled depending on the amount (e.g., more than x% of patients missing a value for a variable) and mechanism (e.g., missing completely at random, missing at random, or missing not at random) of missingness. Missingness will be documented 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 (e.g., diagnoses). 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 study was determined to be exempt under 45 CFR § 46.104(d)(4) by WCG IRB's IRB Affairs Department that reviewed this study under the Common Rule and applicable guidance on August 23, 2024.

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, individual patient data are not retrieved or validated, and it is not possible to link (i.e., identify a potential association between) a particular product and medical event for any individual. Thus, the minimum criteria for reporting an adverse event (AE) (i.e., 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 an interim study report inclusive of the primary objective and exploratory objectives 1 through 3. The interim study report and final study are planned to be provided to the FDA and EMA. Both the interim and final study reports will be disseminated to all the relevant members internally. The study is registered in the HMA EMA Catalogue.

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 (e.g., 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 GJ, Piel FB, Reid CD, et al. Sickle cell disease. *Nat Rev Dis Primers*. 2018;4(1):18010. doi:10.1038/nrdp.2018.10
2. Weatherall DJ. Current trends in the diagnosis and management of haemoglobinopathies. *Scandinavian Journal of Clinical and Laboratory Investigation*. 2007;67(1):1-2. doi:10.1080/00365510601051532
3. 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):327-355. doi:10.1182/bloodadvances.2019001143
4. Thomson AM, McHugh TA, Oron AP, et al. 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. *The Lancet Haematology*. 2023;10(8):e585-e599. doi:10.1016/S2352-3026(23)00118-7
5. Charache S, Terrin ML, Moore RD, et al. Effect of Hydroxyurea on the Frequency of Painful Crises in Sickle Cell Anemia. *N Engl J Med*. 1995;332(20):1317-1322. doi:10.1056/NEJM199505183322001
6. Niihara Y, Miller ST, Kanter J, et al. A Phase 3 Trial of L -Glutamine in Sickle Cell Disease. *N Engl J Med*. 2018;379(3):226-235. doi:10.1056/NEJMoa1715971
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. doi:10.1056/NEJMoa1611770
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 Cystic Fibrosis and Genetic Disorders Group, ed. *Cochrane Database of Systematic Reviews*. 2018;2018(8). doi:10.1002/14651858.CD012082.pub2
9. Porter J, Garbowski M. Consequences and management of iron overload in sickle cell disease. *Hematology*. 2013;2013(1):447-456. doi:10.1182/asheducation-2013.1.447
10. Meloni A, Puliyel M, Pepe A, Berdoukas V, Coates TD, Wood JC. Cardiac iron overload in sickle-cell disease. *American J Hematol*. 2014;89(7):678-683. doi:10.1002/ajh.23721
11. Thein SL, Pirenne F, Fasano RM, et al. Hemolytic transfusion reactions in sickle cell disease: underappreciated and potentially fatal. *Haematologica*. 2020;105(3):539-544. doi:10.3324/haematol.2019.224709
12. *The Voice of the Patient. A Series of Reports from the US FDA's Patient Focused Drug Development Initiative.* Food and Drug Administration; 2014. Accessed October 17, 2023. https://www.fda.gov/media/89898/download.
13. 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
14. 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 Review of Hematology*. 2022;15(2):167-173. doi:10.1080/17474086.2022.2031967
15. Blyden G, Bridges KR, Bronte L. Case series of patients with severe sickle cell disease treated with voxelotor (GBT440) by compassionate access. *American J Hematol*. 2018;93(8). doi:10.1002/ajh.25139
16. Osunkwo I, James J, El-Rassi F, et al. Burden of disease, treatment utilization, and the impact on education and employment in patients with sickle cell disease: A comparative analysis of high- and low- to middle-income countries for the international Sickle Cell

World Assessment Survey. *American J Hematol*. 2022;97(8):1055-1064. doi:10.1002/ajh.26576
17. *Data and Statistics on Sickle Cell Disease*. Centers for Disease Control and Prevention; 2024. https://www.cdc.gov/sickle-cell/data/index.html
18. Iacus SM, King G, Porro G. Causal Inference without Balance Checking: Coarsened Exact Matching. *Polit anal*. 2012;20(1):1-24. doi:10.1093/pan/mpr013
19. 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. *Pediatric Blood & Cancer*. 2019;66(1):e27463. doi:10.1002/pbc.27463
20. Reeves S, Garcia E, Kleyn M, et al. Identifying Sickle Cell Disease Cases Using Administrative Claims. *Academic Pediatrics*. 2014;14(5):S61-S67. doi:10.1016/j.acap.2014.02.008
21. *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
22. 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
23. 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. *JHEOR*. 2019;6(3):174-184. doi:10.36469/10512
24. 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
25. Friedman AM, Oberhardt M, Sheen JJ, et al. Measurement of hemorrhage-related severe maternal morbidity with billing versus electronic medical record data. *The Journal of Maternal-Fetal & Neonatal Medicine*. 2022;35(12):2234-2240. doi:10.1080/14767058.2020.1783229
26. 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
27. Gatto NM, Campbell UB, Rubinstein E, et al. The Structured Process to Identify Fit-For-Purpose Data: A Data Feasibility Assessment Framework. *Clin Pharma and Therapeutics*. 2022;111(1):122-134. doi:10.1002/cpt.2466
28. Gu Y, Purdie DM, Shambhu S, et al. A Validation Study of Red Blood Cell Transfusions in Patients with Sickle Cell Disease Using Optum Electronic Health Records and Administrative Claims Data. *Blood*. 2024;144(Supplement 1):7665-7665. doi:10.1182/blood-2024-205880
29. Austin PC. Informing power and sample size calculations when using inverse probability of treatment weighting using the propensity score. *Statistics in Medicine*. 2021;40(27):6150-6163. doi:10.1002/sim.9176
30. 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. *Statistics in Medicine*. 2015;34(28):3661-3679. doi:10.1002/sim.6607

31. Austin PC. A comparison of 12 algorithms for matching on the propensity score. *Statist Med*. 2014;33(6):1057-1069. doi:10.1002/sim.6004
32. Austin PC. Variance estimation when using inverse probability of treatment weighting (IPTW) with survival analysis. *Stat Med*. 2016;35(30):5642-5655. doi:10.1002/sim.7084
33. Schuster T, Lowe WK, Platt RW. Propensity score model overfitting led to inflated variance of estimated odds ratios. *Journal of Clinical Epidemiology*. 2016;80:97-106. doi:10.1016/j.jclinepi.2016.05.017
34. Mackinnon MJ, Puterman ML. Collinearity in generalized linear models. *Communications in Statistics - Theory and Methods*. 1989;18(9):3463-3472. doi:10.1080/03610928908830102
35. 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.
36. Cole SR, Hernan MA. Constructing Inverse Probability Weights for Marginal Structural Models. *American Journal of Epidemiology*. 2008;168(6):656-664. doi:10.1093/aje/kwn164
37. Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in Medicine*. 2009;28(25):3083-3107. doi:10.1002/sim.3697
38. 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. doi:10.1186/s12874-022-01623-8
39. FDA U. Guidelines for Good Epidemiologic Practice (GEP). Published online 2005. https://www.fda.gov/files/drugs/published/Good-Pharmacovigilance-Practices-and-Pharmacoepidemiologic-Assessment-March-2005.pdf
40. FDA. Best Practices for Conducting and Reporting Pharmacoepidemiologic Safety Studies Using Electronic Healthcare Data. Published online 2013. https://www.fda.gov/media/79922/download
41. International Society of Pharmacoepidemiology (ISPE). Good Pharmacoepidemiology Practices (GPP). Published online 2015. https://www.pharmacoepi.org/resources/policies/guidelines-08027/
42. Aetion Evidence Platform®. Software for real-world data analysis, which is validated for a range of studies. Published online 2021.
43. Blonde L, Khunti K, Harris SB, Meizinger C, Skolnik NS. Interpretation and Impact of Real-World Clinical Data for the Practicing Clinician. *Adv Ther*. 2018;35(11):1763-1774. doi:10.1007/s12325-018-0805-y
44. Funk, Michele Jonsson M, Landi SN. Misclassification in Administrative Claims Data: Quantifying the Impact on Treatment Effect Estimates. *Curr Epidemiol Rep*. 2014;1(4):175-185. doi:10.1007/s40471-014-0027-z
45. Tang AY, Zhou M, Maillis AN, Lai KW, Lane PA, Snyder AB. Trends in blood transfusion, hydroxyurea use, and iron overload among children with sickle cell disease enrolled in Medicaid, 2004–2019. *Pediatric Blood & Cancer*. 2023;70(3):e30152. doi:10.1002/pbc.30152
46. Patterson JA, Bowen JR, Francis S, Ford JB. Comparison of neonatal red cell transfusion reporting in neonatal intensive care units with blood product issue data: a validation study. *BMC Pediatr*. 2018;18(1):86. doi:10.1186/s12887-018-1005-2

14. LIST OF TABLES

Table 1. Summary of Study Time Periods

Table 2. Summary of Inclusion Criteria

Table 3. Summary of Exclusion Criteria

Table 4. List of Variables, Roles, and Assessment Period

Table 5. Description of Endpoints and Alignment to Study Objectives

Table 6. Diagnostic Checklist

Table 7. Planned Sensitivity and Subgroup Analyses

15. LIST OF FIGURES

Figure 1. Study Diagram

ANNEX 1. LIST OF STANDALONE DOCUMENTS

Number	Document Reference Number	Date	Title
001		23 December 2024	Pfizer SCD KRD+ Data Dictionary

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 • Asian or Pacific Islander • Hispanic or Latino • Other • Missing
Insurance Type	Insurance (categorical) • Medicare • Medicaid • Commercial • Multiple • Other
Region	Region, based on Census geographic regions (categorical) • Northeast • Midwest • South • West • Missing
Top 10 Diagnosing Provider Specialty	Top 10 provider specialties present on the same claim as a SCD diagnosis • *To be determined, pending data explorations*
SCD Hematology Diagnosing Provider	Categorical record of a medical claim with a hematologist provider, assessed as the value present on the same claims as an SCD diagnosis, • Yes • No (includes non-hematology and null values)
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,

[redacted]

	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
Transfusion-Associated Reaction	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: o Urgent care visit o Emergency visit o 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 lab test	Presence of a LOINC or CPT/HCPCS code, OR no presence of a LOINC or CPT/HCPCS code
Hb value	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 (Mild to Moderate)	Presence of an ICD-10-CM diagnosis code on any claim
Renal Disease (Severe)	Presence of an ICD-10-CM diagnosis code on any claim
Any Malignancy Without Metastasis Including Leukemia and 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
Human immunodeficiency virus	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 5 for output definition. See Annex Table 2 for RBC transfusion codes.
Stroke (including transient ischemic attack)	For each patient, stroke will be evaluated as the time to stroke in days. See covariate definition above for operational definition and Table 5 for output definition.

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 5 for output definition.
ACS	For each patient, the number of events will be evaluated PPPY in baseline and follow-up. See covariate definition above for operational definition and Table 5 for output definition.
ACS Including Pneumonia	For each patient, the number of events will be evaluated PPPY in baseline and follow-up. See covariate definition above for operational definition and Table 5 for output definition.
Priapism	For each patient, the number of events will be evaluated PPPY in baseline and follow-up. See covariate definition above for operational definition and Table 5 for output definition.
Iron overload	For each patient, iron overload will be evaluated as the time to event in days. See covariate definition above for operational definition and Table 5 for output definition.
Iron chelation use	For each patient, iron chelation will be evaluated as the time to event in days. See covariate definition above for operational definition and Table 5 for output definition.
Delayed hemolytic transfusion reactions Alloimmunization	For each patient, DHTR will be evaluated as the time to event in days. See covariate definition above for operational definition and Table 5 for output definition.
Composite outcome (iron overload, iron chelation use, Delayed hemolytic transfusion reactions, and/or alloimmunization)	For each patient, the composite outcome will be evaluated as the time to event (the first of any) in days. See covariate definition above for operational definition and Table 5 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

3023 0P1	Transfusion of Nonautologous Frozen Red Cells into Peripheral Vein, Open Approach	ICD10 PROC
3023 3N1	Transfusion of Nonautologous Red Blood Cells into Peripheral Vein, Percutaneous Approach	ICD10 PROC
3023 3P1	Transfusion of Nonautologous Frozen Red Cells into Peripheral Vein, Percutaneous Approach	ICD10 PROC
3024 0N1	Transfusion of Nonautologous Red Blood Cells into Central Vein, Open Approach	ICD10 PROC
3024 0P1	Transfusion of Nonautologous Frozen Red Cells into Central Vein, Open Approach	ICD10 PROC
3024 3N1	Transfusion of Nonautologous Red Blood Cells into Central Vein, Percutaneous Approach	ICD10 PROC
3024 3P1	Transfusion of Nonautologous Frozen Red Cells into Central Vein, Percutaneous Approach	ICD10 PROC
3025 0N1	Transfusion of Nonautologous Red Blood Cells into Peripheral Artery, Open Approach	ICD10 PROC
3025 0P1	Transfusion of Nonautologous Frozen Red Cells into Peripheral Artery, Open Approach	ICD10 PROC
3025 3N1	Transfusion of Nonautologous Red Blood Cells into Peripheral Artery, Percutaneous Approach	ICD10 PROC
3025 3P1	Transfusion of Nonautologous Frozen Red Cells into Peripheral Artery, Percutaneous Approach	ICD10 PROC
3026 0N1	Transfusion of Nonautologous Red Blood Cells into Central Artery, Open Approach	ICD10 PROC
3026 0P1	Transfusion of Nonautologous Frozen Red Cells into Central Artery, Open Approach	ICD10 PROC
3026 3N1	Transfusion of Nonautologous Red Blood Cells into Central Artery, Percutaneous Approach	ICD10 PROC
3026 3P1	Transfusion of Nonautologous Frozen Red Cells into Central Artery, Percutaneous Approach	ICD10 PROC
6A55 0Z0	Pheresis of Erythrocytes, Single	ICD10 PROC
6A55 1Z0	Pheresis of Erythrocytes, Multiple	ICD10 PROC

Annex Table 3. Addressing Potential Real-World Data Limitations and Concerns

Variable	Potential concern	Supportive Data and/or Rationale	Study Design Mitigation
RBC transfusion outcome	Misclassification bias	Gu et al. found that 42,106 out of 46,381 (91%) RBC transfusion medical records in patients with SCD from Optum documented having a medical claim. Random sample of 200 RBC transfusion episodes were reviewed with associated free text notes. Of the 200 RBC transfusions, 181 (90.5%) had documentation of an RBC transfusion within 3 days of the procedure claim.[1]	Medical claims data will be used to measure the RBC transfusion outcome in this study. Prior studies support that medical claims are an accurate source for RBC transfusion, and no evidence exists of misclassification bias. If missingness occurred, no theoretical basis for differential misclassification between exposed/unexposed groups exists.
Hydroxyurea exposure	Known confounder	Hydroxyurea increases hemoglobin F and has demonstrated in clinical studies to decrease the proportion of individuals with SCD to receive at least 1 RBC transfusion.	Hydroxyurea exposure will be captured from pharmacy prescription claims. Patients are required to have continuous enrollment in their pharmacy and medical benefits, so limited missing data should exist for hydroxyurea claims. Hydroxyurea will be included in the propensity score model as a covariate and both groups will have similar prior exposure to hydroxyurea. Similarly, as a sensitivity analysis will be performed to report the results stratified by prior treatment exposure to hydroxyurea, crizanlizumab, L-glutamine, and/or ESA.
Covariates	Misclassification bias	There is a possibility of missing information for patients is expected to be low but possible if patients seek healthcare without the use of medical or pharmacy insurance.	Since all patients regardless of exposure status will be required to have continuous enrollment over the period that covariates will be assessed if missingness occurred, no theoretical basis for differential misclassification between exposed/unexposed groups exists.
Hb labs	Missingness	Based on preliminary data explorations, approximately 17% of patients have at least one Hb result during the study timeframe. All lab values have an associated date and time.	Due to missingness in Hb lab results, Hb cannot be adjusted for in the primary analysis; however, a sensitivity analysis is planned that will stratify by baseline Hb and a separate propensity score model will be used to adjust for Hb results.

SCD genotype	Missingness	It is more likely severe genotypes (e.g., HbSS) should be requiring more than 2 RBC transfusions in the past year. Based on preliminary data explorations, approximately 93% have patients have genotype documented in our data set.	RBC transfusion by SCD genotype will be assessed in sensitivity analysis.
Comorbidities	Missingness	There is a possibility of missing information for patients is expected to be low but possible if patients seek healthcare without the use of medical or pharmacy insurance.	Continuous enrollment in a health plan is required as part of the analysis. Thus, all healthcare interactions through insurance will be captured and including documented comorbidities recorded in medical billing records. If missingness occurred, no theoretical basis for differential documentation between exposed/unexposed groups exists. Propensity score adjustment aggregates the probability of exposure to voxelotor. It is expected that more severe diseases that would have a larger impact on outcomes would be better captured in the data[2] and that limited underreporting of non-severe comorbidities should not have a substantial impact on the overall results.
Treatment adherence	Missingness	These data are not missing. Komodo data has near 100% of patients' pharmacy claims when requiring patients to have continuous enrollment in the pharmacy benefit. For exposed patients, we have 100% of the voxelotor claims through the use of Claritas data.	We have two planned analysis including as-treated and intent-to-treat analysis to ensure consistent results that best represent a RCT.
Race/ethnicity	Unmeasured confounder	Based on preliminary data explorations, approximately 90% of SCD patients have documented race/ethnicity in our data set. CDC estimates more than 90% are non-Hispanic Black or African American, and an estimated 3%–9% are Hispanic or Latino.	Race/ethnicity will be included in the propensity score model and is not missing from our data set.
Genetic factors	Unmeasured confounder	SCD genotype is captured in our data set. We are not aware of any other genetic factors that are important in the exposure-outcome relationship of interest.	None.
SCD Severity	Unmeasured confounder	SCD severity is not currently defined in the literature. We will have patients SCD genotype, comorbidities, Charleson	SCD genotype, comorbidities, Charleson Comorbidity Index, and past healthcare utilization

Voxelotor

[redacted]

V2.0 23 December 2024

[redacted]

Document Approval Record

Document Name:	[redacted]
Document Title:	[redacted]

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