STATISTICAL ANALYSIS PLAN

Study Title:	**A Phase 3, Randomized, Double-Blind, Placebo-Controlled Study of Voxelotor (GBT440) in Pediatric Participants with Sickle Cell Disease (HOPE Kids 2)**
Phase:	3
Protocol No.:	GBT440-032 (C5341021), Amendment 4
Protocol Date:	24 April 2023
Investigational Product:	Voxelotor (GBT440 / PF-06759497)
Study Sponsor:	Global Blood Therapeutics, Inc., a wholly owned subsidiary of Pfizer Inc. 181 Oyster Point Blvd. South San Francisco, CA 94080 United States of America
SAP Version:	Version 3.0
SAP Date:	27 January 2025
Prepared By:	PPD, MS

CONFIDENTIAL AND PROPRIETARY INFORMATION

STATISTICAL ANALYSIS PLAN REVIEW AND APPROVAL

Prepared by:

PPD

27 Jan 2025 17:23:054-0500

PPD , MS
PPD
Global Blood Therapeutics, Inc., a wholly owned subsidiary of Pfizer Inc.

Date

Approved by:

PPD

27 Jan 2025 17:24:045-0500

PPD , PhD
PPD
Global Blood Therapeutics, Inc., a wholly owned subsidiary of Pfizer Inc.

Date

PPD

27 Jan 2025 17:23:019-0500

PPD , PhD
PPD
Pfizer Inc.

Date

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY OF SAP CHANGES

Version	Version Date	Summary of Changes
1.0	20 May 2023	N/A (original document)
2.0	11 Jan 2024	- Section 6.3 (Analysis Timing): Updated timing of the primary study analysis to be performed after all subjects have completed the study through Week 48 - Section 6.4.2.1 (TCD Change from Baseline): For subjects that discontinued early due to abnormal TCD, changed imputation method from last observation carried forward to multiple imputation - Section 6.4.2.3 (TCD Responder): Updated missing data rule such that all subjects with a missing TCD flow velocity at the assessment timepoint will be classified as non-responders. - Section 6.4.2.4 (Hemoglobin and Hemolysis Change from Baseline): Added a sensitivity analysis for change from baseline in hemoglobin. - Sections 6.4.3.2 (Stop Dates for Concomitant Medications) and 6.4.3.4 (Stop Dates for Adverse Events): Removed sections as not applicable - Section 8.2.1 (Primary Endpoint): Added multiple imputation - Section 8.2.2.2 (Time to Event Endpoints): Added censoring rule for two or more consecutively missed TCD assessment to Table 2 - Section 8.2.2.3 (TCD Response Endpoint): Added clarification on CI method; Updated missing data rule such that all subjects with a missing TCD flow velocity at the assessment timepoint will be classified as non-responders; Changed Week 24 assessment based on average of Week 24 and Week 28 to be a sensitivity analysis - Section 8.2.6 (Adjustment for Multiple Comparisons): Updated the order of testing for secondary endpoints - Section 8.2.8 (Sensitivity Analysis): Added worst observation carried forward and tipping point analyses as sensitivity - Minor edits for clarity and consistency

Version	Version Date	Summary of Changes
3.0	27 Jan 2025	- Section 2 (Introduction): Added description of dose pause and early discontinuation of the study - Section 6.3 (Analysis Timing): Clarified summaries of Week 96 and time to event efficacy endpoints to be exploratory due to the early termination of the study - Section 8.2.4 (Annualized Incidence Rate of VOCs): Added analysis of VOCs in the first 28 days after last dose for subjects ongoing in the study at time of dose pause CCI [redacted] - Section 8.2.6 (Adjustment for Multiple Comparisons): Removed change from baseline in TCD flow velocity at Week 96 and time to conversion to abnormal TCD flow velocity from the formal testing sequence due to the early termination of the study - Section 8.2.7 (Subgroup Analyses): Added subsections and new subgroups for the primary endpoint and subgroup analyses for annualized incidence rate of VOC - Section 8.2.8 (Sensitivity Analyses): Added sensitivity analysis for change from basleine in TCD which excludes data after dose pause for ongoing subjects - Section 8.3 (Safety): Added statement to exclude data collected more than 28 days after last dose from safety summaries; all reported adverse events will be inlcuded in listings - Section 8.3.3 (Adverse Events): Added summary of TEAEs and SAEs that occurred during the first 28 days after last dose for ongoing subjects at time of dose pause - Minor edits for clarity and consistency

1. GLOSSARY OF ABBREVIATIONS

AE	adverse event
ALK	alkaline phosphatase
ALT	alanine aminotransferase
AST	aspartate aminotransferase
BTCD	baseline transcranial doppler
CCI	[redacted]
CMH	Cochran-Mantel-Haenszel
CRF	case report form
CTCAE	Common Terminology Criteria for Adverse Events
CV	coefficient of variation
DSMB	Data Safety Monitoring Board
CCI	[redacted]
EOS	End of Study
EOT	End of Treatment
GBT	Global Blood Therapeutics
Hb	hemoglobin
HbF	hemoglobin fetal
HU	hydroxyurea
ITT	intent to treat
IXRS	interactive response system
LDH	lactate dehydrogrenase
LOCF	last observation carried forward
MAR	missing at random
MedDRA	Medical Dictionary for Regulatory Activities
mg	milligram
MI	multiple imputation
MMRM	mixed model for repeated measures
CCI	[redacted]
PK	pharmacokinetic(s)
PPK	population pharmacokinetic(s)
RBC	red blood cell
SAE	serious adverse event

SAP	statistical analysis plan
SCD	sickle cell disease
SD	standard deviation
SOC	System Organ Class
TAMMV	time-averaged mean of the maximum velocity
TCD	transcranial doppler
TEAE	treatment-emergent adverse event
TIA	transient ischemic attack
VOC	vaso-occlusive crisis
WHO	World Health Organization
WOCF	worst observed value carried forward

2. INTRODUCTION

The purpose of Study GBT440-032 (C5341021) is to evaluate the effect of voxelotor on reducing the risk of stroke, as measured by transcranial doppler (TCD) flow velocity in subjects 2 to < 15 years of age with sickle cell disease (SCD) who have conditional TCD flow velocity at baseline.

This document describes the statistical methods to be used in the summary and analysis of data from Study GBT440-032. Population pharmacokinetic (PK) analyses will be described in a separate document. Descriptive statistics for CCI [redacted] may be summarized in a separate report. Where this document differs from the study protocol, the methodology described in the SAP is considered the latest and supersedes the corresponding section(s) in the protocol.

As of 1 May 2024, study drug dosing in GBT440-032 was voluntarily paused by the Sponsor due to a safety imbalance. Study visits and other study activities continued in accordance with the protocol. Subsequently, a decision was made by the Sponsor on 25 September 2024 to terminate the study based on the totality of clinical data for Oxybryta (voxelotor) and the last subject last visit occurred subsequently.

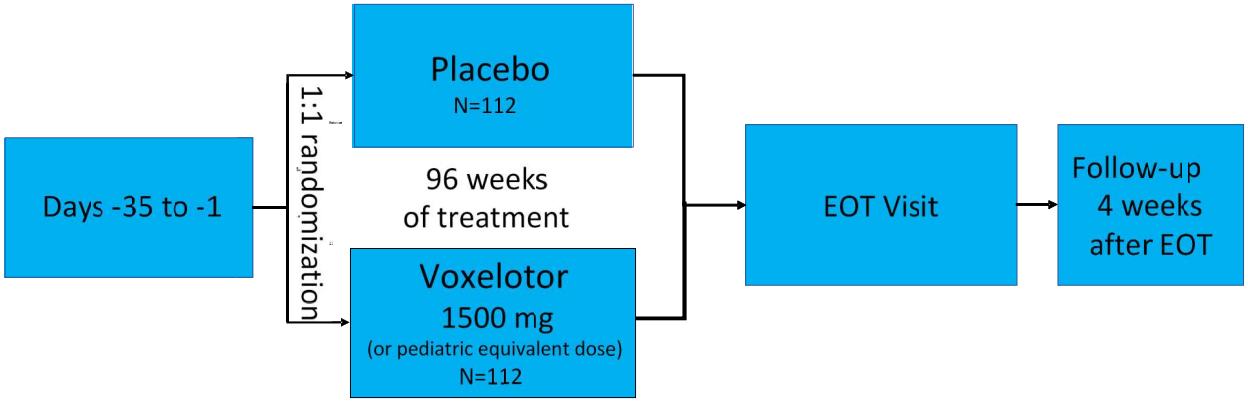
2.1. Study Overview

GBT440-032 is a multicenter, randomized, placebo-controlled, double-blind, prospective study evaluating the efficacy and safety of voxelotor in pediatric subjects with SCD, 2 to < 15 years of age. Approximately 224 subjects will be randomized 1:1 to voxelotor or placebo. The study schema is presented in Figure 1.

The duration of study involvement for an individual subject is expected to be approximately 105 weeks.

Subjects ≥ 12 years old will be administered a fixed dose of voxelotor 1500 mg per day. Subjects younger than 12 years of age will receive voxelotor at a weight-based (1500-mg equivalent) dose, as described in Section 1.6 of the protocol.

Figure 1: Study Schema



Abbreviations: EOT, end of treatment.
Note: Primary endpoint will be assessed at Week 24.

At least 15 subjects from 2 years to < 4 years of age will be enrolled. Administration of voxelotor will be initiated in subjects < 4 years of age after the Data Safety Monitoring Board (DSMB) has reviewed safety and PK data from at least 12 subjects < 30 kg treated with voxelotor for at least 28 days. The data for at least 12 subjects may be generated from any of the voxelotor studies.

Following completion of study treatment, eligible subjects will be given the option to enroll in an open-label extension study (under a separate protocol) to receive voxelotor for a minimum of 3 years (3 years or until commercial product or a managed access program becomes available, whichever is longer).

2.2. Study Measurements and Visit Schedule

Please refer to the study protocol amendment 4, Appendix A, Appendix B, and Appendix C for schedule of assessments.

3. STUDY OBJECTIVES

Primary Objective:

The primary objective is to evaluate the effect of voxelotor compared to placebo on the TCD time-averaged mean of the maximum velocity (TAMMV) arterial cerebral blood flow at 24 weeks in SCD subjects ≥ 2 to < 15 years of age with conditional (170 to < 200 cm/sec) TCD flow velocity.

Secondary Objectives:

The secondary objectives are to evaluate the effects of voxelotor compared to placebo on:

- Change in TCD flow velocity at Week 48 and Week 96
- Conversion of TCD category from conditional to abnormal
- Reversion of TCD category from conditional to normal
- Proportion of subjects with TCD response
- Change in Hb over time
- Change in clinical measures of hemolysis
- Annualized incidence rate of vaso-occlusive crises (VOCs)

CCI

[redacted]

CCI

Safety Objective:

The safety objective is to assess the safety and tolerability of voxelotor compared to placebo.

Pharmacokinetic Objective:

The pharmacokinetic objective is to assess the PK of voxelotor as evaluated by population PK (PPK) analysis.

4. STUDY ENDPOINTS

4.1. Primary Endpoint

The primary efficacy endpoint is the change from Baseline at 24 weeks in TAMMV arterial cerebral blood flow, as measured by TCD.

4.2. Secondary Endpoints

The secondary endpoints are:

- Change from Baseline in TCD flow velocity at Week 48 and Week 96
- Time to conversion to abnormal TCD flow velocity (≥ 200 cm/sec)
- Time to reversion to normal TCD flow velocity (< 170 cm/sec)
- Proportion of subjects with TCD flow velocity reduction ≥ 15 cm/sec at Week 24, Week 48, and Week 96
- Change in Hb from Baseline at Week 24, Week 48, and Week 96
- Change in clinical measures of hemolysis (unconjugated bilirubin, % reticulocyte, absolute reticulocyte, and lactate dehydrogenase [LDH]) from Baseline at Week 24, Week 48, and Week 96
- Annualized incidence rate of reported VOCs

CCI

4.4. Safety Endpoints

The safety endpoints include adverse events (AEs), clinical laboratory assessments, and vital signs.

4.5. Pharmacokinetic Endpoints

The endpoint is whole blood and plasma voxelotor PK as evaluated by PPK analysis using nonlinear mixed effects modeling.

5. DEFINITIONS AND TERMINOLOGY

Study Treatment

The term study treatment refers to either voxelotor or placebo.

Baseline Value

Baseline measurements for efficacy and safety assessments will be the average of all pre-randomization or pre-first dose values, respectively, collected on or after Study Day -35 (see Table 1) with the exception of TCD flow velocity. For TCD flow velocity, the baseline value will be the value at Screening.

Day 1

Day 1 is the date of randomization.

Study Day

Study Day is defined relative to the date of randomization.

For study assessments or events that occur on or after the date of randomization, the study day of an assessment/event is calculated as:

Study Day = Event date – Date of randomization + 1.

For pre-randomization events, the study day is calculated as:

Study Day = Event date – Date of randomization.

Study Visit

Study Visit is the visit as defined in Table 1.

Treatment Day

Treatment Day is defined relative to the date of first dose.

Treatment day of an event that occurs on or after date of first dose is calculated as:

Treatment Day = Event date – Date of first dose + 1.

Treatment day of an event that occurs before date of first dose is calculated as:

Treatment Day = Event date – Date of first dose.

Exposure to Study Drug

Duration of exposure is defined as the number of weeks from date of first dose to date of last dose and is calculated as:

$$\frac{(Date\ of\ last\ dose - Date\ of\ first\ dose + 1)}{7}$$

Actual exposure is defined as the number of weeks from date of first dose to date of last dose excluding days where treatment was entirely missed or intermittently stopped as recorded in the case report form (CRF) and is calculated as:

$$\frac{(Date\ of\ last\ dose\ -\ Date\ of\ first\ dose - Number\ of\ days\ study\ drug\ was\ not\ taken\ +\ 1)}{7}$$

Completion of Treatment

Completion of study treatment is as specified on the End of Treatment (EOT) CRF and includes all subjects who did not permanently discontinue study drug early (ie, prior to the end of the 96-week treatment period).

Completion of Study

Completion of study is a specified on the End of Study (EOS) CRF. Subjects are considered to have completed the study if (1) the subject completed the study through the Week 96 visit and elected to enroll in the open-label extension study, or (2) the subject did not enroll in the open-label extension study and completed the study through 96 weeks plus the EOS follow-up visit (28 days after the last dose of study drug).

Change from Baseline

Change from baseline for a given endpoint is defined as the Study Visit value minus the Baseline value.

CCI

CCI

[redacted]

Treatment-emergent Adverse Event

A treatment-emergent adverse event (TEAE) is defined as an adverse event (AE) with an onset date between date of first dose and 28 days after last dose, inclusive.

Concomitant Medications

Concomitant medications are those medications taken on or after the initiation of study drug. This definition includes medications started prior to the initiation of study drug but continue concurrently with study drug.

Prior Medications

Prior medications are those medications taken prior to date of first dose.

6. GENERAL STATISTICAL CONSIDERATIONS

6.1. Sample Size and Power

The sample size is planned to provide sufficient statistical power for the primary efficacy analysis and for a robust safety database. In addition, the number of events required to assess the secondary efficacy endpoint of time to conversion to abnormal TCD category is also considered.

For the primary endpoint (change from Baseline in TCD flow velocity [cm/sec]), assuming equal allocation across treatment groups and a common variance, an estimate of the number of subjects in each of 2 treatment groups used the following assumptions in the power calculation:

- Targeted treatment effect: difference in mean change from Baseline in TCD flow velocity is 15 cm/sec
- Common standard deviation is 24 cm/sec
- A significance level of 5% (2 -sided)

Assuming a 20% drop out rate, a sample size of 112 subjects per group provides > 95% power to detect the targeted treatment effect for the following hypothesis test using a Student's t-test:

$$H_0: \mu_1=\mu_2 \qquad \text{vs.} \qquad H_A: \mu_1\neq\mu_2,$$

where μ_1 and μ_2 stand for mean change from Baseline in TCD flow velocity at Week 24 in voxelotor and placebo groups, respectively.

For the secondary efficacy endpoint of time to conversion to abnormal TCD category, assuming 35% of subjects in the placebo group have experienced a TCD flow velocity > 200 cm/sec by Week 96, a hazard ratio of 0.5 between voxelotor group and placebo, a fixed follow up time of up to 96 weeks for each subject, and an overall 20% drop out rate, a total of 49 events is expected. This provides approximately 65% power for a log-rank test at a two-sided significance level of 5% (PASS version 11).

6.2. Randomization

Randomization will be carried out centrally through an interactive response system (IXRS). Permuted blocks within randomization strata will be used. Eligibility of the subject should be confirmed by the Investigator prior to randomization.

At the time of randomization, subjects will be stratified by HU use (yes; no), age (2 years to ≤ 8 years; > 8 years to < 15 years), Screening TCD value (170 cm/sec to < 185 cm/sec;

185 cm/sec to < 200 cm/sec), and geographic region (Africa including Middle East and North Africa [MENA]; rest of world). Stratification of subjects by age was chosen to evenly balance by age distribution using the expected median age of 8 years.

6.3. Analysis Timing

The primary study analysis was performed after all randomized subjects completed the study through Week 48 or discontinued early, and all data through a corresponding data cut-off date (determined based on the last subject's Week 48 visit date) was entered in the database, reviewed, and verified. This included formal evaluation of the primary Week 24 efficacy endpoint, secondary Week 24 and Week 48 efficacy endpoints, and summary of all safety data through the data cut-off date. Summaries of Week 96 and time to event efficacy endpoints based on partial data available at this time were considered exploratory.

A final study analysis will be performed after the last subject's last visit and final study database lock. Summaries of Week 96 and time to event efficacy endpoints based on partial data will be considered exploratory due to the early termination of the study. The estimated treatment effects (voxelotor vs. placebo) and corresponding 95% CIs will be presented for these endpoints, but without formal hypothesis testing.

6.4. Handling of Data

6.4.1. Visit Windows

The conventions outlined below will be used to associate data measures, according to the study day corresponding to the date the data were collected, to an analysis visit window based on nominal (ie, planned) visits, using the measure's study day relative to Day 1. Data collected before Study Day 1 will be considered screening information with regard to assignment of visit windows. Target study days, protocol-specified ranges of study days for each scheduled visit, and the analysis visit windows defined for each nominal visit day are shown in Table 1.

Table 1: Nominal Analysis Visit Windows

Nominal Visit Time	Target Study Day	Protocol-Specified Study Day Range	Study Day Range for Statistical Analysis
Screening	-	[-35, -1]	[-35, -1]*
Day 1	1	[1, 1]	[1, 1]
Week 2	14	[7, 21]	[2, 21]
Week 4	28	[21, 35]	[22, 42]
Week 8	56	[49, 63]	[43, 70]
Week 12	84	[77, 91]	[71, 98]
Week 16	112	[105, 119]	[99, 140]
Week 24	168	[161, 175]	[141, 182]
Week 28	196	[189, 203]	[183, 224]
Week 36	252	[245, 259]	[225, 294]
Week 48	336	[329, 343]	[295, 378]
Week 60	420	[413, 427]	[379, 462]
Week 72	504	[497, 511]	[463, 546]
Week 84	588	[581, 595]	[547, 630]
Week 96	672	[665, 679]	[631, 714]

* For TCD, all TCD data collected at Screening will be used regardless of study day.

6.4.2. Imputation of Missing Data for Evaluation of Efficacy

Guidelines regarding how missing data will be handled for each efficacy endpoint are described below.

6.4.2.1. TCD Change from Baseline

- For subjects discontinued from the study due to conversion to abnormal TCD flow velocity, missing TCD values post abnormal TCD assessment will be imputed using multiple imputation (MI). All post-baseline TCD values ≥ 200 cm/sec observed in the study for subjects with abnormal TCD will be used as the reference distribution for MI. This approach avoids the need to make a parametric assumption on the missing data given observed values, and appropriately accounts for variation due to "fill-in" missing values. See Section 8.2.1 for details.
  - A sensitivity analysis using worst observed value carried forward (WOCF) for subjects that discontinue the study early due to conversion to abnormal TCD flow velocity will also be performed (Section 8.2.8).

- For all other subjects that discontinue the study early, missing TCD data will not be imputed.
- Subjects initiating HU postbaseline are to be discontinued from the study. TCD assessments obtained after HU initiation will be excluded from analysis.
- RBC transfusions: TCD results within 8 weeks after the transfusion date will be imputed for the MMRM analysis of TCD change from baseline. Post transfusion results will be imputed by data from the most recent study visit prior to the transfusion (including Baseline) with a non-missing TCD value.

6.4.2.2. Conversion to Abnormal TCD

- For subjects who withdraw from the study or reach analysis data cutoff date without an event (ie, TCD flow velocity reaching abnormal range), their time to conversion to abnormal will be censored at the date of last TCD assessment prior to study withdrawal or analysis data cutoff date, as applicable (Section 8.2.2.2, Table 2).

6.4.2.3. TCD Responder

A TCD responder is defined as a TCD flow velocity reduction from Baseline ≥ 15 cm/sec, and will be assessed separately at Week 24, Week 48, and Week 96.

- Subjects with a missing TCD flow velocity value at the assessment timepoint will be classified as non-responders.
- Subjects initiating HU postbaseline are to be discontinued from the study. TCD assessments obtained after HU initiation will be excluded from analysis (considered missing).
- RBC transfusions: TCD results within 8 weeks after the transfusion date will be imputed for the analysis of TCD response. Post transfusion results will be imputed by data from the most recent study visit prior to the transfusion (including Baseline) with a non-missing TCD value.

6.4.2.4. Hemoglobin and Hemolysis Change from Baseline

- For the primary method of analysis of these endpoints (MMRM), missing at random (MAR) will be assumed.
  - For change from baseline in Hb, a sensitivity analysis using last observation carried forward (LOCF) for subjects that discontinue the study early due to conversion to abnormal TCD flow velocity will also be performed.
- Subjects initiating HU postbaseline are to be discontinued from the study. Assessments of Hb and clinical measures of hemolysis obtained after HU initiation will be excluded from analysis.
- RBC transfusions: all laboratory results within 8 weeks after the transfusion date will be imputed for the MMRM change from baseline analysis of Hb and hemolysis measures. Post transfusion results will be imputed by data from the most recent Study

Visit prior to the transfusion (including Baseline) with a non-missing laboratory value.

6.4.3. Imputation of Incomplete Dates

An incomplete date is any date for which either the day, month, or year is unknown, but all three fields are not unknown. For analyses which require a complete date, the following imputation will be used. For data listings, any partial dates will be displayed in the listing without imputation (eg, JAN2023, 2021).

6.4.3.1. Start Dates for Concomitant Medications

- For missing start day only – Day will be imputed as the first day of the month (ie, 1) with the following exception: If the partial date falls in the same month and year as the treatment start date then the concomitant medication start date will be imputed as the treatment start date.
- For missing start day and month – Day and month will be imputed as the first day of the year (ie, 1 January) with the following exception: If the partial date falls in the same year as the treatment start date then the concomitant medication start date will be imputed as the treatment start date.

6.4.3.2. Start Dates for Adverse Events when the Event Occurs Prior to the First Dose

- For missing start day only – Day will be imputed as the first day of the month (ie, 1) with the following exception: If the partial date falls in the same month and year as the treatment start date then the AE start date will be imputed as the day prior to the treatment start date.

6.4.3.3. Start Dates for Adverse Events when the Event is Not Prior to the First Dose

- For missing start day only – If the partial date falls in the same month and year as the treatment start date then the adverse event start date will be imputed as the treatment start date. Otherwise, the day will be imputed as the first day of the month (ie, 1) as long as the imputed date does not occur before the treatment start date.

7. ANALYSIS POPULATIONS

The following populations will be considered in the analysis of data. Assignment or exclusion of subjects from the analysis populations will be performed prior to study unblinding.

- Intent-to-Treat (ITT) Population: the ITT population includes all randomized subjects. This is the primary analysis population for efficacy, demographics, baseline characteristics, and disposition data. Subjects will be grouped according to the treatment group to which they are randomized.
- Safety Population: All subjects who receive at least 1 dose of study treatment will be included in the safety population. This is the primary analysis population for safety and exposure data. Subjects will be grouped according to the study treatment they receive.

- PK Population: The PK population will consist of all subjects who receive active study drug and have at least 1 measured concentration at a scheduled PK time point after the start of dosing. If any subjects are found to be noncompliant with respect to dosing or have incomplete data, protocol violations, or clinical events that affect PK, a decision will be made on a case-by-case basis as to their inclusion in the analysis. Subjects in this population will be used for all PK summaries.

8. STATISTICAL METHODS

8.1. Subject Disposition, Demographic, and Baseline Characteristics

The number of subjects randomized will be tabulated by country, study site, and treatment group. Subject disposition will be presented for all subjects in the ITT population. The number of subjects who completed the study and discontinued early from the study will be provided. The reasons for early discontinuation at any point also will be presented by treatment group. Additionally, the number of weeks on study will be summarized for all treated subjects.

Demographic data and baseline disease characteristics including age, sex, race, ethnicity, SCD genotype, Baseline Hb, Baseline HbF, Baseline hemolysis measures, Baseline (Screening) TCD flow velocity, Baseline HU use, and Baseline EQ-5D-Y (self-assessment, proxy) will be summarized using descriptive statistics. Subjects who record more than one race will be grouped into a single category denoted as multi-racial. This information will be reviewed for baseline differences between treatment groups, but no formal statistical testing will be performed.

Medical/surgical history will be coded using the Medical Dictionary for Regulatory Activities (MedDRA), and summarized descriptively.

Prior and concomitant medications (see Section 5 for definitions) will be coded using the World Health Organization (WHO) Drug Dictionary for the entire period of the study, and summarized descriptively.

8.2. Efficacy Analyses

Efficacy analyses will be based on the ITT patient population (Section 7), with subjects grouped according to the treatment assigned at randomization. Data from all randomized subjects, regardless of adherence to study drug or to the protocol will be included in the efficacy analyses. This includes data from subjects who discontinued study drug early but continued with study assessments.

Covariate Adjustment
Unless otherwise noted, analyses of primary and secondary efficacy endpoints will be adjusted (see Section 8.2.1 and Section 8.2.2) for the following randomization stratification variables:

- Baseline HU use (yes; no)
- Age (2 years to ≤ 8 years; > 8 years to < 15 years)
- Screening TCD value (170 cm/sec to < 185 cm/sec; 185 cm/sec to < 200 cm/sec)

Note: The randomization stratification variable of geographic region (Africa including MENA; rest of world) as stated in the protocol is not included as a covariate as the vast majority of enrolled subjects are from a single region (Africa including MENA).

For stratification variables, the value recorded in the clinical database (ie, per case report form [baseline HU use and age] or by the central reading center [Screening TCD flow velocity]) will be compared to value captured in the IXRS. If differences are observed, the values from the clinical database will be used in the analysis.

Statistical Tests
All statistical tests will be conducted at a two-sided alpha level of 0.05. Adjustment for multiple comparisons is described in Section 8.2.6.

8.2.1. Primary Endpoint

The primary analysis of change from Baseline in TCD flow velocity to Week 24 will be performed using mixed model for repeated measures (MMRM). Independent variables will include treatment, study visit, treatment by study visit interaction, and randomization stratification factors (ie, HU use, age, and Screening TCD flow velocity). Within-subject variability will be modeled using an unstructured covariance matrix. A single model will be fit incorporating available TCD flow velocity data from all timepoints, including to Week 96.

Missing data will be handled per Section 6.4.2.1. Specifically for subjects who discontinued from the study due to conversion to abnormal TCD flow velocity, missing TCD values post abnormal TCD assessment will be randomly imputed. All post-baseline TCD values $\geq$ 200 cm/sec observed in the study for subjects with abnormal TCD will be used as the reference distribution. Multiple imputation based on re-sampling from the reference distribution will be used. For each iteration, the imputed TCD data will be combined with the observed data, and the combined augmented dataset analyzed using the primary analysis MMRM methodology described in this section. This will be repeated 100 times and the results will be combined using Rubin's rules (Little, et al [2012] and Little and Rubin [2002]). A seed of 440032 will be used.

The primary analysis is the comparison of treatments at Week 24 for voxelotor versus placebo. The Kenward-Roger correction for denominator degrees of freedom, standard errors, and test statistics will be utilized. Sample SAS code for the MMRM will be similar to the following:

```
PROC MIXED;
  CLASS subject treatment visit agegrp HU BTCD;
  MODEL ChgTCD = treatment visit treatment*visit agegrp HU BTCD/ DDFM=KR;
  REPEATED visit / SUB=subject TYPE=un;
  LSMEANS treatment*visit / DIFF CL E OM;
RUN;
```

Note: agegrp, HU, and BTCD are per definition of stratification factors for randomization. If there are convergence issues with the statistical model, a compound symmetry covariance matrix will be used.

The adjusted mean (lsmean) of change from baseline in TCD at each visit, with the estimated standard error and 95% confidence interval (CI), will be estimated from the MMRM model for each imputed augmented dataset. Results from the multiple imputed datasets will be combined using Rubin's rules, and results will be presented by treatment group in tabular and graphic

format. The difference in the adjusted means between treatment groups and the associated 95% CI of the difference will be provided.

For the primary endpoint, subgroup analyses will be performed to evaluate the consistency of results across pre-specified subgroups (see Section 8.2.7).

In addition, sensitivity analyses will be performed to evaluate the robustness of results to missing data assumptions (see Section 8.2.8).

8.2.2. Secondary Endpoints

8.2.2.1. Change from Baseline in TCD Flow Velocity at Week 48 and Week 96

Change in TCD flow velocity from Baseline to Week 48 and Week 96 will be analyzed with the same MMRM as described for the primary endpoint (change from Baseline in TCD flow velocity at Week 24). A single model will be fit incorporating data from all timepoints.

8.2.2.2. Time to Event Endpoints

TCD flow velocities are categorized as follows:

Normal: < 170 cm/sec

Conditional: 170 to < 200 cm/sec

Abnormal: ≥ 200 cm/sec

An abnormal TCD observation in the range of 200 to < 220 cm/sec will require a repeated assessment within 2 to 4 weeks for confirmation of abnormal status (timing based on Principal Investigator clinical judgment). If the confirmatory assessment is < 200 cm/sec, the subject will not be categorized as abnormal for purposes of analysis. An abnormal TCD observation ≥ 220 cm/sec will not require a confirmation assessment.

Time to conversion to abnormal will be calculated as number of weeks from date of randomization to the date of TCD assessment when an abnormal TCD flow velocity (≥ 200 cm/sec) is first determined. If the abnormal TCD requires a confirmation assessment, the date of the first abnormal TCD assessment will be used.

For each of the subjects, the time to conversion to abnormal will be right-censored according to the conventions in Table 2.

Summary statistics will include number of subjects with event, number of subjects censored, median time to event along with 95% CI for the median and quartiles. The CIs for the median will be calculated according to Brookmeyer and Crowley (1982) and the CIs for the survival function estimates at fixed time points will be derived using the log-log transformation according to Kalbfleisch and Prentice (2002) (conftype=loglog default option in SAS Proc LIFETEST) with back transformation to a CI on the untransformed scale. The estimate of the standard error will be computed using Greenwood's formula.

For time to conversion to abnormal TCD flow velocity, a log-rank test, stratified for the randomization stratification factors, will be used to test the difference between voxelotor and placebo groups. A Cox proportional hazard model will be used to estimate the hazard ratio, as appropriate. The Kaplan-Meier method will be used to estimate the landmark rates, and the

associated 95% confidence intervals at Week 24, Week 48, and Week 96 in voxelotor and placebo groups.

Time to reversion to normal TCD flow velocity (< 170 cm/sec) will be summarized descriptively by treatment group for subjects who achieve a normal TCD flow velocity.

Table 2: Date of Abnormal TCD Event or Censoring

Situation	Date of Abnormal TCD or Censoring	Time to Conversion
Confirmed abnormal TCD event: TCD ≥ 200 cm/sec but < 220 cm/sec for first assessment and ≥ 200 cm/sec for repeat TCD assessment	Date of first TCD value ≥ 200 cm/sec but < 220 cm/sec	Event
Abnormal TCD not requiring confirmation: TCD value ≥ 220 cm/sec	Date of TCD assessment ≥ 220 cm/sec	Event
No abnormal TCD event observed: All TCD < 200 cm/sec as of data cutoff date TCD ≥ 200 cm/sec but < 220 cm/sec for first assessment and < 200 cm/sec for repeat TCD assessment	Date of last TCD assessment with a non-missing value prior to data cutoff date	Censored
Confounding event prior to abnormal TCD: Initiation of postbaseline HU without having experienced an abnormal TCD (either confirmed abnormal TCD event or abnormal TCD event not requiring confirmation) Withdraw from study without having experienced an abnormal TCD (either confirmed abnormal TCD event or abnormal TCD event not requiring confirmation)	Date of last TCD assessment with a non-missing value prior to start of HU or study termination date	Censored
Missing two or more consecutive TCD assessments immediately prior to an observed abnormal TCD: TCD ≥ 220 cm/sec immediately after two or more consecutively missed TCD assessments TCD ≥ 200 cm/sec but < 220 cm/sec for first assessment and ≥ 200 cm/sec for repeat TCD assessment immediately after two or more consecutively missed TCD assessments	Date of last TCD assessment with a non-missing value prior to the two or more consecutively missed TCD assessments	Censored

8.2.2.3. TCD Response Endpoint

The TCD responder endpoint, defined as a TCD flow velocity reduction from Baseline ≥ 15 cm/sec at Week 24, Week 48, or Week 96, will be analyzed using an exact Cochran-Mantel-Haenszel (CMH) general association test stratified by the randomization stratification factors for the treatment difference between voxelotor and placebo. The difference in percent

responders between treatment groups with corresponding 95% CI using the Mantel-Haenszel stratum weight method will be presented. The analysis will be performed separately for each timepoint of interest. Missing data will be handled per Section 6.4.2.3, with subjects with a missing value at the assessment timepoint classified as non-responders.

For the analysis of TCD response at Week 24, the average of TCD values at Week 24 and Week 28 will be used as a sensitivity analysis. If the TCD assessment is missed at one time point, the TCD value at the other (non-missing) time point will be used. Subjects with missing TCD values at both Week 24 and Week 28 timepoints will be classified as non-responders.

In addition, a sensitivity analysis will be performed to assess an alternative TCD response definition based on a TCD flow velocity reduction from Baseline ≥ 30 cm/sec. Analyses will be performed separately for the Week 24, Week 48, and Week 96 time points.

8.2.3. Change from Baseline in Hemoglobin and Measures of Hemolysis

Change in hemoglobin and percent change in clinical measures of hemolysis (unconjugated bilirubin, % reticulocytes, absolute reticulocyte, and LDH) from Baseline at Week 24, Week 48, and Week 96 will be analyzed with a similar MMRM as described for the primary efficacy endpoint, change from Baseline in TCD flow velocity, with the exception that the baseline lab value will be included as a covariate. For each parameter, a single model will be fit incorporating available data from all timepoints, including to Week 96. Missing data will be handled per Section 6.4.2.4.

Figures displaying the relationship between change from baseline in TCD flow velocity and change from baseline in Hb at Weeks 24, 48, and 96 will be generated.

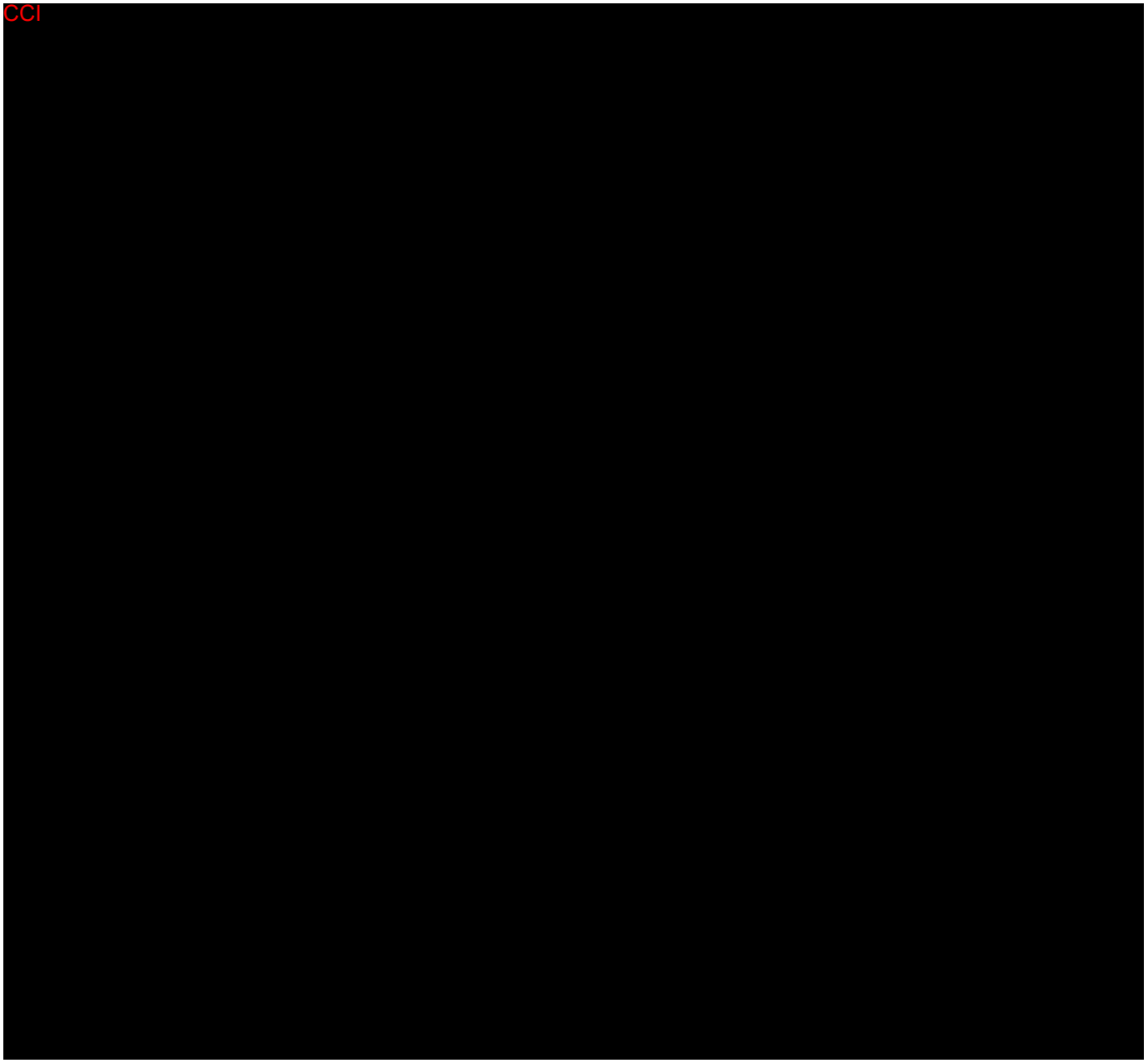
In addition, the percentage of subjects with a Hb response (defined as > 1 g/dL increase from Baseline) at Week 24 will be calculated by treatment arm for all randomized subjects (ITT population). The Study Visit Hb value for Week 24 will be the average of Hb values at Week 24 and Week 28. If the Hb assessment is missed at one time point, the Hb value at the non-missing time point will be used. Subjects with missing Hb assessment at both Week 24 and Week 28, initiation of HU post randomization and prior to Week 24, or a transfusion within 8 weeks of the Week 24 Hb assessment will be considered a Hb non-responder.

8.2.4. Annualized Incidence Rate of VOCs

Annualized incidence rate of VOC will be calculated for each treatment group (total number of events)/(total subject-years on treatment) and associated 95% exact Poisson confidence limits will be reported. For any subjects that initiating HU postbaseline, data up to the time of HU initiation will be used. The incidence rate ratio (voxelotor vs placebo) will be provided. A figure of the mean cumulative functions for VOC incidences for each treatment group will be presented.

In addition, for subjects ongoing in the study at the time of the dose pause, the rate of VOCs observed on-treatment and within the first 28 days after their last dose of study drug will be calculated. The rates for each time period will be standardized to 28-day rates and 95% CIs provided.

CCI



8.2.6. Adjustment for Multiple Comparisons

To control the overall significance level (alpha) for the study at a 2-sided 5% level, a fixed sequence hierarchical test procedure will be used to formally evaluate voxelotor treatment effect in comparison to placebo at Weeks 24 and 48.

The primary and key secondary endpoints will be tested in the following order:

1. TCD flow velocity change from Baseline at Week 24 (primary endpoint)
2. TCD flow velocity change from Baseline at Week 48
3. TCD flow velocity response rate at Week 24

Each endpoint will be tested at a 2-sided alpha level of 5%. Testing of endpoints subsequent to a non-significant result will be considered exploratory in nature. Per Section 6.3 (Analysis

Timing), summaries of Week 96 or time to event efficacy endpoints based on partial data will be considered exploratory.

8.2.7. Subgroup Analyses

8.2.7.1. Change from Baseline in TCD at Week 24

Subgroup analyses will be performed to evaluate the consistency of results across pre-specified subgroups defined by demographic and baseline characteristics.

The following subgroups will be analyzed with respect to the primary efficacy endpoint (change from baseline in TCD flow velocity at Week 24):

- Age (2 to ≤ 8 years; > 8 to < 15 years)
- Age (2 to < 4 years; 4 to < 12 years; 12 to < 15 years)
- Sex (male; female)
- Geographic region (sub-Saharan Africa; MENA/US/UK)
- Baseline HU use (yes; no)
- Baseline TCD value (170 to < 185 cm/sec; 185 to < 200 cm/sec)

Subgroups by SCD genotype and race are not included as the vast majority of subjects belong to a single category (ie, HbSS for SCD genotype and African or Black/African American for race).

For the primary efficacy endpoint, a MMRM similar to that specified for the primary analysis (Section 8.2.1) will be used for each subgroup analysis based on data subset for the patient subgroup of interest. Baseline covariates included in the primary analysis but no longer relevant given the subgroup of interest will be excluded from the model.

If convergence problems with the statistical models arise due to a small number of subjects per subgroup, the analysis may be simplified by combining some of the subgroups, by excluding baseline stratification variables from the model, and/or by using a compound symmetry covariance matrix as appropriate.

The estimated treatment effects (voxelotor vs. placebo) and corresponding 95% CIs will be displayed graphically for each level of the subgroups specified.

8.2.7.2. Annualized Incidence Rate of VOC

The annualized incidence rate of VOCs will be summarized for the below subgroups and as described in Section 8.2.4. The incidence rate ratio (voxelotor vs. placebo) and corresponding 95% CIs will be displayed graphically for each level of the subgroups specified.

- Age (2 to < 4 years; 4 to < 12 years; 12 to < 15 years)
- Sex (male; female)
- Geographic region (sub-Saharan Africa; MENA/US/UK)
- Baseline HU use (yes; no)
- Number of VOCs in past 12 months (0 events; 1 event; 2 to 10 events)

- Baseline Hb (< 7 g/dL; 7 g/dL to < 9 g /dL; ≥9 g/dL)
- Malaria TEAE on study (yes; no)

8.2.8. Sensitivity Analyses

The following sensitivity analyses will be performed to evaluate the robustness of the primary efficacy endpoint results to missing data assumptions:

- WOCF: For subjects discontinued early from the study due to conversion to abnormal TCD flow velocity, missing TCD values post abnormal TCD assessment will be imputed using WOCF. All other data handling rules in Section 6.4.2.1 and the MMRM model as described in Section 8.2.1 will be applied.
- Tipping point: A tipping point analysis will be conducted by adding (or subtracting) a constant delta (or so-called shift parameter) to the post-discontinuation imputed TCD values for subjects that discontinue early due to conversion to abnormal TCD flow velocity. A separate delta will be applied to each treatment group (voxelotor and placebo) in the direction of lack of efficacy (TCD flow velocity increase for voxelotor group and TCD flow velocity decrease for the placebo group). A range of values will be used to adjust the imputed TCD flow velocity values independently for each treatment group based on a grid of delta adjustments. The resulting point estimate for the treatment effect and corresponding p-value under each pair of deltas will be tabulated. If the delta parameter needed to overturn the conclusion is so extreme that it is considered clinically implausible, then this indicates robustness to missing data assumptions.

In addition, as study drug dosing was paused prior to study termination, the following sensitivity analysis will also be performed:

- Exclusion of TCD data post dose pause: For subjects ongoing in the study at the time of the dose pause, TCD data collected after their last dose of study drug will be excluded from analysis and change from baseline in TCD assessed based on the MMRM model as described in Section 8.2.1.

8.3. Safety Analyses

Safety analysis will be performed on all subjects receiving at least one dose of study drug (Safety Population; Section 7). Adverse events will be coded using the MedDRA version at the time of the analysis data cutoff.

All safety summaries will include results up to 28 days after last dose of study drug. Data collected more than 28 days after the last dose of study drug will be excluded from table summaries. All reported adverse events will be included in the listings.

8.3.1. Exposure to Study Drug

The duration of study drug exposure and actual study drug exposure will be summarized with descriptive statistics. The number and percentage of subjects who missed at least 1 dose of study drug or had at least 1 dose reduction will be summarized. The summary will include the number and percentage of subjects who had a dose change (reduction or increase) due to adverse event, age/weight change, or other, and dose missed due to an adverse event, non-compliance, or other. Study drug adherence will be summarized using descriptive statistics.

8.3.2. Concomitant Medications

Concomitant medications will be coded using the WHO Drug Dictionary and summarized descriptively.

8.3.3. Adverse Events

AEs will be mapped to system organ class (SOC) and preferred term using the current version of Medical Dictionary for Regulatory Activities (MedDRA). AEs will be summarized according to SCD relationship.

SCD-related AEs will be summarized separately (including but not limited to sickle cell anemia with crisis, acute chest syndrome, pneumonia, priapism, dactylitis, splenic sequestration, hepatic sequestration and osteonecrosis). All other TEAEs are considered non-SCD-related TEAEs.

The total number and percentage of subjects with TEAEs will be presented by SOC and preferred term for each treatment group. TEAEs will be presented by severity. TEAEs will be tabulated presenting the SOC alphabetically, and within each system, the preferred term will be presented in decreasing order of the number of voxelotor-treated subjects who experienced each TEAE. Treatment-related AEs will be similarly presented.

The frequency of subjects who experience each TEAE or treatment-related TEAE will be determined as follows: A subject experiencing the same AE multiple times will only be counted once for the preferred term. Similarly, if a subject experiences multiple AEs within the same SOC, that subject will be counted only once for that system. If changes in the severity of an AE are recorded in the eCRF, only the most severe incidence of the AE will be counted. If a subject experiences multiple occurrences of a TEAE, only the related event or the worst severity (table dependent) will be counted for each subject within each SOC or preferred term for the summaries of treatment-related TEAEs or TEAEs by maximum severity.

Missing onset dates will be imputed as previously outlined in Section 6.4.3 as required to determine treatment-emergent events.

Listings of all AEs, TEAEs leading to discontinuation of treatment, SAEs, Grade ≥ 3 TEAEs, and deaths will be provided.

In addition, for subjects ongoing in the study at the time of the dose pause, TEAEs and SAEs that occurred during the first 28 days after the last dose of study drug will be summarized.

8.3.4. Clinical Laboratory Assessments

All hematology and chemistry laboratory assessments will be performed by a central laboratory. Laboratory data will be converted into System International (SI) units for data analysis.

The following protocol-specified clinical laboratory findings will be summarized using descriptive statistics by treatment group:

1. Hematology:
    - RBCs
    - Hematocrit
    - Platelets
    - White blood cells (WBCs) with differential (basophils, eosinophils, neutrophils, monocytes, and lymphocytes)
    - RBC distribution width (RDW)
    - Mean corpuscular volume (MCV)
    - Mean corpuscular hemoglobin concentration (MCHC)
2. Chemistry:
    - Alanine aminotransferase (ALT)
    - Albumin
    - Aspartate aminotransferase (AST)
    - Blood urea nitrogen (BUN)
    - Calcium
    - Creatinine
    - Glucose
    - LDH
    - Bilirubin (total and direct)

Note: Planned summaries of change from baseline in hemoglobin and clinical measures of hemolysis (indirect bilirubin, % reticulocytes, absolute reticulocytes, and LDH) are described in Section 8.2.3.

Descriptive statistics will be presented for baseline, each evaluation post baseline, and change from baseline for each post baseline evaluation. If any of the clinical laboratory assessments are below the lower limit of quantitation or above the upper limit of quantitation, then the numerical limit will be used in the tabulation of the descriptive statistics.

Laboratory abnormalities will be graded via the Common Terminology Criteria for Adverse Event (CTCAE) using the current version at the time of analysis database lock. The number and percentage of subjects experiencing treatment-emergent graded abnormalities will be summarized by treatment group. Laboratory abnormality shifts from baseline through Week 24, Week 48, and Week 96 will be summarized by treatment group.

8.3.5. Vital Signs

Vital signs (oral temperature, heart rate, and systolic and diastolic blood pressure) will be summarized using descriptive statistics for Baseline, each scheduled visit, and change from Baseline for each postbaseline visit.

The average of each vital sign parameter collected at Screening and Day 1 (pre-dose) will be the baseline.

8.4. Pharmacokinetic Analyses

PPK analysis will consist of all subjects who receive active study drug and have at least 1 measured concentration at a scheduled PK time point after the start of dosing. If any subjects are found to be noncompliant with respect to dosing or have incomplete data, protocol deviations, or clinical events that affect PK, a decision will be made on a case-by-case basis as to their inclusion in the analysis. Subjects in this population will be used for all PK summaries. PPK analyses using nonlinear mixed effects modeling will be performed to characterize voxelotor PK in plasma and whole blood. The influence of demographic covariates (such as body weight, height, age, gender) on voxelotor PK parameters (ie, clearance [CL] and volume of distribution) will be investigated. If appropriate, the voxelotor PK data may be pooled with PK data from other studies.

Population PK analyses will be described in a separate document.

9. PROTOCOL DEVIATIONS

Important protocol deviations (defined as deviations classified as 'critical' in severity) and protocol deviations related to COVID-19 will be summarized by presenting the number and percentage of subjects with at least one deviation of each deviation type. Protocol deviations will also be displayed in a data listing and sorted by treatment group, subject number, and then by date (where applicable) within each subject number. The type of deviation along with a description and any additional comments about the deviation will be listed.

10. CHANGES IN THE PLANNED ANALYSES

Additional analyses that are included in the clinical study report but are not mentioned in the SAP will be identified as such.

11. REFERENCES

1. Brookmeyer, R, and Crowley, JJ. (1982). A Confidence Interval for the Median Survival Time. Biometrics, 38: 29–41.
2. Kalbfleisch, JD, and Prentice,, RL. (2002). Statistical Analysis of Failure Time Data. 2nd Edition. Hoboken, Wiley Interscience.
3. Kenward, MG, and Roger, JH. (1997). Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood. Biometrics, 53: 983–997.
4. Little, RJ, D'Agostino, R, Cohen, ML, Dickersin, K, Emerson, SS, Farrar, JT, Frangakis, C, Hogan, JW, Molenberghs, G, Murphy, SA, Neaton, JD, Rotnitzky, A, Scharfstein, D, Shih, WJ, Siegel, JP, and Stern, H. (2012). The Prevention and Treatment of Missing Data in Clinical Trials. New England Journal of Medicine, 367(14): 1355–1360.
5. Little, RJ, and Rubin, DB. (2002). Statistical Analysis with Missing Data. Wiley Series in Probability and Statistics.
6. Penman, AD, and Johnson, WD. (2008). A SAS program for calculating cumulative incidence of events (with confidence limits) and number at risk at specified time intervals with partially censored data. Computer Methods and Programs in Biomedicine, 89: 50–55.

12. TABLES, FIGURES, AND LISTINGS

A separate document will provide mockups of the tables, figures, and listings that supports the analyses proposed in this SAP.