Official Title: A Phase 3, Double-blind, Randomized, Placebo-Controlled, Multicenter Study Evaluating the Efficacy and Safety of Mitapivat in Subjects With Non-Transfusion-Dependent Alpha- or Beta-Thalassemia (ENERGIZE)

NCT Number: NCT04770753

Document Date: SAP Version 2.0: 14-Nov-2022

STATISTICAL ANALYSIS PLAN

A Phase 3, Double-blind, Randomized, Placebo-Controlled, Multicenter Study Evaluating the Efficacy and Safety of Mitapivat in Subjects With Non-Transfusion-Dependent Alpha- or Beta-Thalassemia (ENERGIZE)

AG348-C-017

Version: v2.0

Date: 14-Nov-2022

[redacted]

Prepared by:

[redacted] PhD
Study Statistician

[redacted]

[redacted]

Name and Title (Printed)	Signature	Date (DD MMM YYYY)

Approved by:

[redacted] MD
Medical Director, Clinical Development

[redacted]

[redacted]

Name and Title (Printed)	Signature	Date (DD MMM YYYY)

[redacted] PhD
Head of Clinical Pharmacology and DMPK

[redacted]

[redacted]

Name and Title (Printed)	Signature	Date (DD MMM YYYY)

[redacted] PhD
Head of Biometrics

[redacted]

[redacted]

Name and Title (Printed)	Signature	Date (DD MMM YYYY)

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND DEFINITIONS OF TERMS

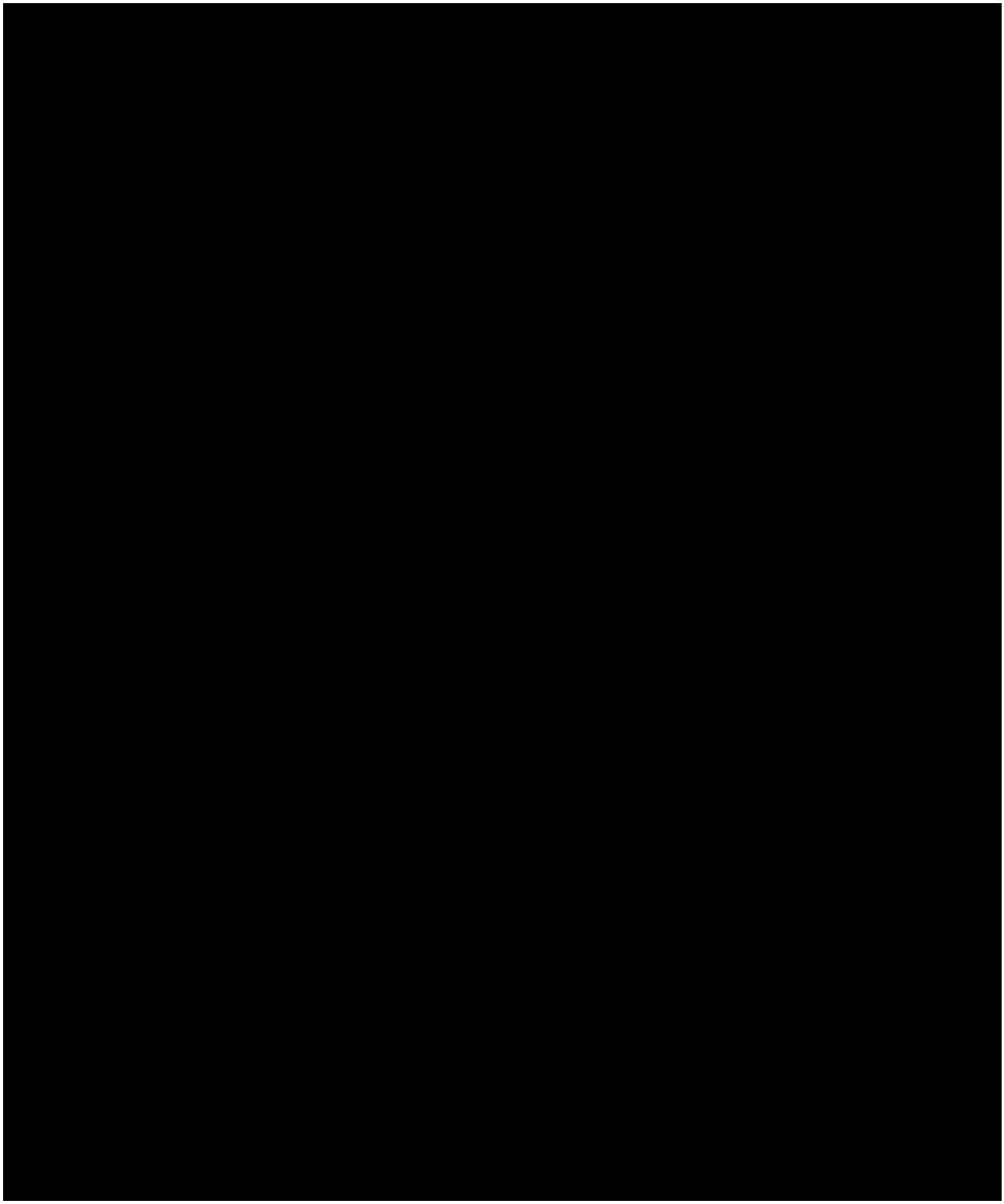
Abbreviation	Definition
2,3-DPG	2,3-Diphosphoglycerate
6MWT	Six-minute walk test
AE	Adverse event
ALP	Alkaline Phosphatase
ALT	Alanine Aminotransferase
ANCOVA	Analysis of covariance
AST	Aspartate Aminotransferase
ATC	Anatomical Therapeutic Chemical
ATP	Adenosine Triphosphate
AUC	Area under the curve
BID	Twice daily
BMD	Bone mineral density
BMI	Body mass index
CI	Confidence Interval
C_{max}	Maximum observed concentration
CSR	Clinical Study Report
CTCAE	Common Terminology Criteria for Adverse Events
CV	Coefficient of variation
DXA	Dual-energy x-ray absorptiometry
ECG	Electrocardiogram
eCRF	Electronic Case Report Form
eDISH	Evaluation of Drug-Induced Serious Hepatotoxicity
EODB	End of double-blind
EOS	End of study
EOT	End of treatment
[redacted]	[redacted]
FACIT	Functional Assessment of Chronic Illness Therapy
[redacted]	[redacted]
FAS	Full Analysis Set
Hb	Hemoglobin
HRQOL	Health-related quality of life

ICF	Informed consent form
IDMC	Independent data monitoring committee
IQR	Interquartile range
IXRS	Interactive voice/web response system
LDH	Lactate dehydrogenase
LLN	Lower limit of normal
LLQ	Lower limit of quantification
LS	Least squares
Max	Maximum
MedDRA	Medical Dictionary for Regulatory Activities
MCMC	Markov Chain Monte Carlo
Min	Minimum
MMRM	Mixed-effect model for repeated measures
[redacted]	[redacted]
NTDT	Non-transfusion dependent thalassemia
OLE	Open-label extension
PD	Pharmacodynamic
PerfO	Performance outcome
PGIC	Patient Global Impression of Change
PGIS	Patient Global Impression of Severity
PK	Pharmacokinetic
PPS	Per-Protocol Set
PRO	Patient Reported Outcome
[redacted]	[redacted]
PT	Preferred Term
QD	Once-daily
QOL	Quality of Life
RBC	Red blood cell
RSD	Relative standard deviation
SAE	Serious adverse event
SAP	Statistical analysis plan
SD	Standard deviation
SOC	System organ class

TEAE	Treatment-emergent adverse event
T_{max}	Time of maximum observed concentration
ULN	Upper limit of normal
WBC	White blood cell
WHO	World Health Organization

1. VERSION HISTORY

This statistical analysis plan (SAP) describes the analysis associated with protocol AG348-C-017 Amendment 2 (dated 28-September-2022).



2. INTRODUCTION

This SAP provides the detailed methodology for summary and statistical analyses of the data collected in Study AG348-C-017. This document may modify the plans outlined in the protocol; however, any major modifications of the primary endpoint definition or its analysis will also be reflected in a protocol amendment.

The primary clinical study report (CSR) will include all data for each subject up to the end of Double-blind (EODB) Period. The EODB period is defined as the end of the safety follow-up (for subjects who discontinue study treatment during the Double-blind Period) or before the start of study treatment in the Open-label Extension (OLE) Period (for subjects who continue onto the OLE Period after completing the Double-blind Period).

A CSR addendum will include data (associated with objectives that encompass OLE Period evaluations) through the End of Study (EOS). The EOS is defined as the time at which all subjects complete all study visits, are lost to follow-up, have withdrawn consent for further participation in the study, or when the Sponsor terminates the study. Study completion is the date of the last visit of the last subject in the study.

In the following sections, references to "EODB period date/EOS date" are meant to indicate that the EODB period subject-specific dates will be used for analyses to be reported in the primary CSR and the EOS date will be used for analyses to be reported in the addendum CSR.

3. OBJECTIVES AND ENDPOINTS

The study objectives and endpoints are provided in Table 2.

Table 2: Objectives and Endpoints

Primary Objective	Primary Endpoint
• To compare the effect of mitapivat versus placebo on anemia in subjects with α- or β-non-transfusion dependent thalassemia (NTDT)	• Hemoglobin (Hb) response, defined as a ≥1.0 g/dL increase in average Hb concentration from Week 12 through Week 24 compared with baseline
Key Secondary Objectives	**Key Secondary Endpoints**
• To compare the effect of mitapivat versus placebo on fatigue	• Change from baseline in average Functional Assessment of Chronic Illness Therapy (FACIT) -Fatigue subscale score from Week 12 through Week 24
• To compare the effect of mitapivat versus placebo on additional measures of anemia	• Change from baseline in average Hb concentration from Week 12 through Week 24
Secondary Objectives	**Secondary Endpoints**
• To evaluate the effect of mitapivat versus placebo on anemia and markers of hemolysis and erythropoiesis	• Hb 1.5+ response, defined as a ≥1.5 g/dL increase in average Hb concentration from Week 12 through Week 24 compared with baseline • Change from baseline in indirect bilirubin, lactate dehydrogenase (LDH), and haptoglobin at Week 24 • Change from baseline in reticulocytes and erythropoietin at Week 24
• To evaluate the effect of mitapivat versus placebo on additional measures of fatigue	• Improvement in the Patient Global Impression of Severity (PGIS) -Fatigue by at least 1 category at Week 12, 16, 20, and 24 compared with baseline, or "no change" if no or mild fatigue at baseline • Improvement in the Patient Global Impression of Change (PGIC) -Fatigue at Week 12, 16, 20, and 24, or "no change" if no or mild fatigue at baseline
• To evaluate the effect of mitapivat versus placebo on physical activity	• Change from baseline in the 6-minute walk test (6MWT) distance at Week 24
• To evaluate the effect of mitapivat versus placebo on iron metabolism	• Change from baseline in markers of iron metabolism, including serum ferritin and transferrin saturation at Week 24
• To evaluate the safety of mitapivat	• Type, severity, and relationship of adverse events (AEs) and serious adverse events (SAEs)
• To evaluate the PK and PD effects of mitapivat	• Plasma or blood concentrations and PK parameters of mitapivat and PD parameters, including adenosine triphosphate (ATP) and 2,3-diphosphoglycerate (2,3-DPG)
[redacted]	

[redacted]

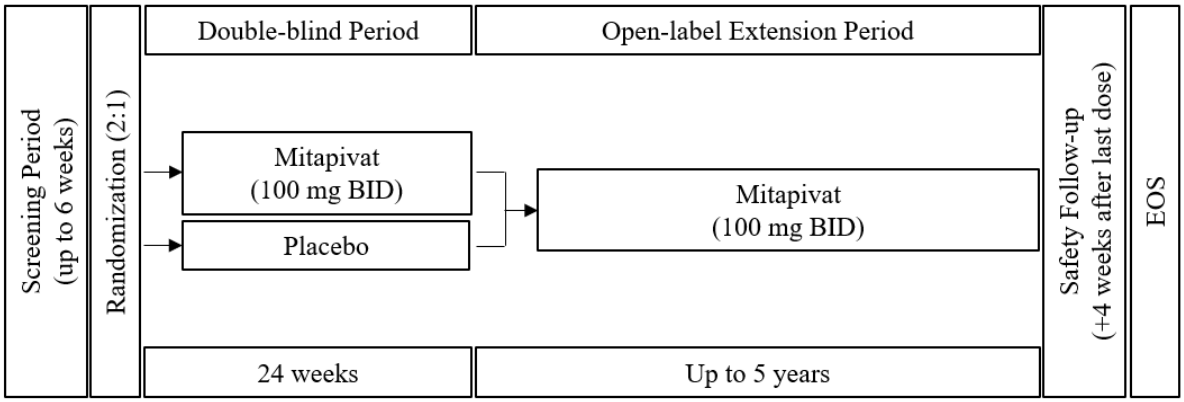
[a] Analyses associated with this objective will be described in a separate SAP.

4. STUDY DESIGN

This is a Phase 3, double-blind, randomized, placebo-controlled, multicenter study evaluating the efficacy and safety of mitapivat versus placebo in adult subjects with α- or β-NTDT followed by an OLE Period. Approximately 171 subjects are planned to be randomized in this study. This multicenter study will be conducted internationally.

An overview of the study design is presented in Figure 1.

Figure 1: Study Design



Abbreviations: BID = twice daily; EOS = End of Study.

5. ANALYSIS DATA SETS

The following analysis data sets will be evaluated and used for presentation of the data:

Population	Description
All Screened Subjects	All subjects who sign the informed consent form (ICF).
Full Analysis Set (FAS)	All subjects who are randomized. Subjects will be classified according to the randomized treatment arm according to the intent-to-treat (ITT) principle. The FAS is the primary analysis set for the evaluation of efficacy endpoints, unless otherwise specified.
Per-Protocol Set (PPS)	The PPS is a subset of the FAS. Subjects who meet any of the following criteria will be excluded from the PPS: • Do not receive at least 1 dose of the randomized treatment • Do not meet Inclusion Criterion #2 (ie, do not have documented thalassemia) • Do not meet Inclusion Criterion #3 (ie, Hb concentration >10.0 g/dL based on an average of at least 2 Hb concentration measurements [separated by ≥7 days] collected during the Screening Period) • Do not meet Inclusion Criterion #4 (ie, do not meet criteria for NTDT) • Meet Exclusion Criterion #3 (ie, prior exposure to gene therapy or prior bone marrow or stem cell transplantation) • Meet Exclusion Criteria #4 or #5 (ie, currently receiving treatment with luspatercept or hematopoietic stimulating agents)
Safety Analysis Set	All subjects who receive at least 1 dose of study treatment. Subjects will be classified according to the treatment received. If a subject randomized to placebo receives at least 1 dose of mitapivat in the Double-blind Period, then the subject will be classified to the mitapivat arm.
Open-label Extension (OLE) Analysis Set	All subjects who receive at least 1 dose of open-label mitapivat after the Double-blind Period.
Pharmacokinetic (PK) Analysis Set	A subset of the safety analysis set including all subjects with at least 1 mitapivat plasma concentration measurement ≥ LLQ.
Pharmacodynamic (PD) Analysis Set	A subset of the safety analysis set including all subjects with at least 1 blood 2,3-DPG or ATP concentration measurement ≥ LLQ.
PK/PD Analysis Set	A subset of the PK analysis set including all subjects in the PK and PD analysis sets, with at least 1 time-matched ATP or 2,3-DPG concentration to a mitapivat plasma concentration.

Abbreviations: 2,3-DPG =2,3-diphosphoglycerate; ATP = adenosine triphosphate; ICF = informed consent form; ITT = intent-to-treat; LLQ = lower limit of quantification.

Table 3 summarizes the use of the analysis sets.

Table 3: Analysis Data Sets for Each Endpoint

Endpoints	FAS	PPS	Safety Analysis Set	OLE Analysis Set	PK Analysis Set	PD Analysis Set	PK/PD Analysis Set
Demographic and other baseline characteristics	✓			✓			
Disposition	✓			✓			
Major protocol deviations	✓			✓			
Exposure and concomitant therapies			✓	✓			
Efficacy [redacted]	✓	✓ (primary and key secondary only*)		✓			
Safety			✓	✓			
PK					✓		
PD						✓	
PK/PD							✓

* Key secondary endpoints are defined in Section 7.6.2.
Abbreviations: FAS = full analysis set; OLE = open-label extension; PD = pharmacodynamic; PK = pharmacokinetic; PPS = per-protocol set.

6. GENERAL STATISTICAL CONSIDERATIONS

6.1. Randomization, Blinding, and Unblinding

Eligible subjects will be randomized in a 2:1 ratio to mitapivat or matched-placebo.

Randomization assignment will be implemented by an interactive voice/web response system (IXRS) and stratified by:

- Baseline Hb concentration (≤9.0 g/dL or 9.1-10.0 g/dL)
- Thalassemia genotype (α-thalassemia/HbH disease or β-thalassemia)

This study includes a Double-blind Period followed by an OLE Period. Study subjects, Investigators, clinical study center personnel, pharmacists, and the Sponsor will be blinded to the subject's treatment assignment. After completing the Double-blind Period, subjects will be provided the opportunity to receive mitapivat in the OLE Period. At the last study visit of the Double-blind Period, subjects who continue in the OLE Period will be provided with active mitapivat; however, study subjects, Investigators, clinical study center personnel, and the Sponsor will continue to remain blinded to the randomized treatment assignment during the previous Double-blind Period until the study is unblinded for the analysis of the primary endpoint.

6.2. Sample Size Determination and Decision Rules

6.2.1. Sample Size Determination

The following statistical hypothesis will be tested to address the primary objective:

$H_{01}: p_{t1} - p_{c1} = 0$ vs $H_{11}: p_{t1} - p_{c1} \neq 0$

where p_{t1} and p_{c1} are the proportion of subjects with a Hb response in the mitapivat arm and placebo arm, respectively.

Assuming an Hb response rate of 10% in the placebo arm, 171 subjects (114 subjects randomized to mitapivat and 57 subjects randomized to placebo) are needed to provide a 95% power to detect an increase in Hb response rate from 10% in the placebo arm to 35% in the mitapivat arm based on a 2-sided significance level of 0.05.

Given the large variability of Hb concentration in patients with anemia, the SD of change from baseline can be as high as 1.3 g/dL (Kalantar-Zadeh and Aronoff, 2009). Assuming no changes in Hb concentration in subjects receiving placebo, approximately 10% of subjects could achieve an Hb response simply because of the variability in Hb concentration measurements. Therefore, a 10% Hb response rate is being assumed for the placebo arm.

The assumption of a 35% Hb response rate was informed by the results from Study AG348 C-010. As of 20 August 2020, 70% (14/20 overall, 10/15 in those with β-thalassemia, and 4/5 in those with α-thalassemia) subjects achieved a ≥1.0 g/dL increase in average Hb concentration from Week 12 through Week 24 compared with baseline. Considering a more heterogeneous population in this large, double-blind, Phase 3 study and a commonly noted decrease in efficacy between Phase 2 and Phase 3 studies, a 35% response rate for mitapivat arm is deemed to be a reasonable assumption.

Additionally, the following statistical hypotheses will be tested to address the key secondary objectives:

$H_{02}: \mu_{t1} - \mu_{c1} = 0$ vs $H_{12}: \mu_{t1} - \mu_{c1} \neq 0$

$H_{03}: \mu_{t2} - \mu_{c2} = 0$ vs $H_{13}: \mu_{t2} - \mu_{c2} \neq 0$

where

- μ_{t1} and μ_{c1} are the mean change from baseline in average FACIT-Fatigue subscale score from Week 12 through Week 24 in the mitapivat arm and placebo arm respectively
- μ_{t2} and μ_{c2} are the mean change from baseline in average Hb concentration from Week 12 through Week 24 in the mitapivat arm and placebo arm, respectively

The sample size of 171 subjects will also provide 80% power at a 2-sided significance level of 0.05 to detect a 5.5 difference in the change from baseline in average FACIT-Fatigue subscale score from Week 12 through Week 24 between the mitapivat arm and the placebo arm. These calculations assume a change from baseline in average FACIT-Fatigue subscale score from Week 12 through Week 24 in the mitapivat arm of 4.5 and in the placebo arm of -1 and a common standard deviation (SD) of 12.

Assumptions of the FACIT-Fatigue score change from baseline are from several sources. FACIT-Fatigue results from placebo-controlled studies in NTDT have not yet been published. The available literature and feedback from the thalassemia experts suggest an increase of 3 points in FACIT-Fatigue score for every 1 g/dL increase in Hb. Results of the open-label luspatercept NTDT study with similar eligibility criteria included mean FACIT-Fatigue score change from baseline to Week 24 of approximately +2.8 (N=17) and mean Hb change from baseline to Week 24 of approximately +1.2 g/dL (N=~25) (Piga et al, 2016). In a recent randomized, open-label, controlled trial in paroxysmal nocturnal hemoglobinuria, the observed FACIT-Fatigue score change from baseline to Week 16 was +9.2 in the pegcetacoplan treatment arm, and Hb change over the same period was +2.4 g/dL (N=41 patients randomized) (Hillmen et al, 2020). Based on the Hb data observed in the Phase 2 Study AG348-C-010 and these results from trials, an increase in the FACIT-Fatigue score of approximately +4.5 in the mitapivat arm over Weeks 12 to 24 and a small decrease in the score of -1 in the placebo group are expected, with an estimated SD of 12 for the difference of changes.

Overall type I error will be maintained at or below 2-sided α-level of 0.05 by a fixed -sequence testing procedure (Westfall and Krishen, 2001) . H_{01} will be tested first; if H_{01} is rejected, then H_{02} will be tested; if H_{02} is rejected, then H_{03} will be tested. All tests will be conducted at the 2-sided 0.05 significance level.

6.2.2. Decision Rules

The study will have demonstrated that mitapivat is statistically significantly superior to placebo for Hb response if the 2-sided p-value for the test of the difference in the Hb response rate is ≤0.05 and the Hb response rate is higher for the mitapivat arm than for the placebo arm.

To protect the integrity of the study and to preserve the type I error at or below 2-sided α=0.05, the testing strategy outlined in Section 6.2.1 will be followed and statistical significance will be achieved for mitapivat compared to placebo for each key secondary endpoint that can be tested per this strategy at the associated significance level.

6.3. Definitions

6.3.1. Study Drug and Study Treatment

In the Double-blind Period, both study drug and study treatment are defined as mitapivat or matched placebo.

In the OLE Period, both study drug and study treatment are defined as mitapivat.

6.3.2. Start and End Dates of Study Treatment

The start of study treatment is the earliest date/time of administration of a non-zero dose of the study drug.

The end of study treatment is the latest date/time of administration of a non-zero dose of the study drug on or before the EODB period date/ EOS date.

6.3.3. Study Day

The study day for assessments or events occurring on or after the start of study treatment (eg, AE onset, disease assessment) will be calculated as:

Study day=Date of the assessment or event−start of study treatment+1.

The study day for assessments or events occurring before the start of study treatment (eg, baseline characteristics, medical history) will be negative and calculated as:

Study day=Date of the assessment or event−start of study treatment.

There is no study day 0. The study day will be displayed in data listings.

6.3.4. Baseline

6.3.4.1. Double-blind Period

Throughout the SAP for analyses based on the Double-blind Period, 'reference date' will be used to refer to randomization date for subjects randomized and not dosed, or the start of study treatment for subjects randomized and dosed.

Efficacy [redacted] Evaluations

For markers of iron metabolism (Section 7.6.3.5), markers of hemolysis and erythropoiesis (Section 7.6.3.2), [redacted] baseline is defined as the average of all assessments within 42 days before the 'reference date'. Assessments collected within 8 weeks after an RBC transfusion will be excluded from the baseline derivation for these efficacy parameters and for endpoints associated with measures of fatigue. Baseline will be derived based on central laboratory data; if no central laboratory data are available, then local laboratory data will be used to derive the baseline.

For other efficacy [redacted] endpoints baseline is defined as the last assessment before the 'reference date'.

Demographic and Other Baseline Characteristics

For summaries of demographic and other baseline characteristics based on the FAS, baseline is defined as the last assessment before the 'reference date'.

Safety Evaluations

For clinical laboratory assessments the following applies:

- Before deriving the baseline:
  - If there are multiple records with the same assessment day and time from the same laboratory, the average value will be used.
  - If there are multiple records with the same assessment day and time from different laboratories, the value from the central laboratory will be used.
- The baseline will then be the last value before the start of study treatment.

For all other safety parameters, the last assessment before the start of study treatment will be used as the baseline.

PD Evaluations

The baseline is defined as the pre-dose whole blood concentration of 2,3-DPG and ATP on study day 1. If this concentration is <LLQ, baseline will be set to missing.

If, per protocol, an assessment is to be performed on study day 1, before the first dose of study treatment, and the assessment time, time of first dose of study treatment, or both are missing or not collected, it will be assumed that the assessment is performed before the first dose of study treatment. Unscheduled assessments will be used in the determination of baseline; however, an unscheduled assessment on study day 1 for which the time relative to start of study treatment cannot be determined, will be considered to have been obtained after administration of study treatment.

If no assessment meets the definition of baseline for an evaluation, the baseline will be set to missing.

6.3.4.2. Open-label Extension Period

For subjects randomized to mitapivat, the baseline for all assessments in the OLE Period will be that derived in the Double-blind Period (Section 6.3.4.1).

For subjects randomized to placebo, the baseline is derived following the same rules as for the Double-blind Period but using start of mitapivat in the OLE Period instead of randomization date or 'reference date'.

Age at baseline will be calculated based on the date of informed consent for the study as described in Section 6.4.2.

In addition, for evaluations that are unique to the OLE Period (bone mineral density [BMD] by DXA [Section 7.7.5], [redacted] baseline is derived as the last assessment before first dose of study treatment (mitapivat or placebo).

6.3.5. On-Treatment Period

The on-treatment period starts on the date of the start of study treatment and ends 28 days after the end of study treatment (including dose taper period).

Within the on-treatment period, there are two distinct treatment periods:

- Double-blind on-treatment period starts on the date of the start of study treatment and ends
  - 28 days after the end of study treatment, for subjects who discontinue study treatment during the Double-blind Period of the study
  - Before the start of the OLE treatment period, for subjects who continue onto the OLE Period of the study

- OLE on-treatment period applies only to subjects who complete the Double-blind Period of the study, and starts with the first dose of open-label mitapivat in the OLE Period and ends 28 days after the end of study treatment

For subjects who receive mitapivat in the OLE Period, assessments or events occurring on the date of first dose of mitapivat in the OLE Period will be included in the analysis of the data from the double-blind on-treatment period, except for AEs and concomitant medications reported for subjects in the placebo arm which will be included only in the analysis of the data from the OLE on-treatment period.

Data listings will include all assessments and events, with those that occur outside of the on-treatment period flagged.

6.4. General Methods

6.4.1. Data Handling After End of Double-blind Period

Data after the EODB period may not undergo the cleaning process and will not be displayed in any listings or used for summary statistics, statistical analyses, or imputations for the primary CSR.

6.4.2. Standard Derivations and Reporting Conventions

The following conversion factors will be used to convert days into weeks, months or years: 1 week=7 days, 1 month=30.4375 days, and 1 year=365.25 days.

The following derivations will be implemented.

- Age (years) = (date of informed consent−date of birth+1)/365.25

  If only day of birth is missing: Age (years): (year/month of informed consent−year/month of birth)

  If day and month of birth are missing: Age (years): (year of informed consent−year of birth)

  The integer part of the calculated age will be used for reporting purposes.
- Body mass index (BMI, kg/m^2)=weight (kg)/height $(m)^2$
- Duration (in days) from a reference date (eg, randomization date, start date of study treatment) =

  date of event−reference date + 1, if the date of the event is on or after the reference date

  date of event−reference date, if the date of the event is before the reference date

Reporting conventions will be as follows:

- Mean (including arithmetic mean and geometric mean) and median will be displayed to one more decimal place than the raw data.

- Standard deviation (SD) will be displayed to two more decimal places than the raw data.
- Percentages (including percent changes from baseline for whole blood concentration of 2,3-DPG and ATP, relative standard deviation [RSD%], arithmetic coefficient of variation [CV%], and geometric coefficient of variation [CV% geometric mean]) will be displayed to 1 decimal place (however, percentages corresponding to 0 counts will be reported as 0 rather than 0.0 and 100 percent will be reported as 100 rather than 100.0).
- p-values will be reported with 4 decimal places; all p-values should be specified to be 1-sided or 2-sided.
- Unless otherwise specified, rounding will be performed to the closest integer/first decimal using the common mid-point between the two consecutive values, eg, 5.11 to 5.14 will be rounded to 5.1, and 5.15 to 5.19 will be rounded to 5.2.
  - Non-zero percentages that are <0.1 before rounding will be displayed as "<0.1", eg, 0.09 will be reported as <0.1 rather than as 0.1.
  - p-values<0.0001 before rounding will be displayed as "<0.0001", eg, a p-value of 0.00009 will be displayed as <0.0001 rather than as 0.0001.
- PK parameters will be reported to 3 significant figures, except for T_{max} and T_{last} which will be reported to 2 significant digits.

Note that some PK parameters are observed values (eg, C_{max}), while other PK parameters are derived from the intensive PK profiles (eg, $AUC_{0\text{-}last}$). For all the PK parameters, regardless of whether they are observed or derived, the individual PK parameter values are considered "raw data" when implementing the above reporting rules.

6.4.3. Pooling of Data Across Sites

In order to provide overall estimates of treatment effects, data will be pooled across sites. The "site" factor will not be considered in statistical models or subgroup analyses given the high number of participating sites in contrast to the anticipated small number of subjects randomized at each site.

6.4.4. Continuous and Categorical Variables

Continuous variables will be summarized using descriptive statistics, ie, number of non-missing values, mean, SD, median, quartiles, minimum, and maximum. PK concentrations and parameters will be further summarized using arithmetic mean together with the CV% and 90% CI, geometric mean together with CV% geometric mean and 90% CI. PD concentration and parameters will be further summarized using arithmetic mean together with the CV% and 95% CI, geometric mean together with the CV% and 95% CI, and RSD% (applicable to PD concentration only).

Categorical variables will be summarized by frequency distributions (number and percentage of subjects within a given category in the analysis data set). Unless otherwise specified, the calculation of percentages will include the "missing" category. Therefore, counts of missing observations will be included in the denominator and presented as a

separate category. For summaries by visit, percentages will be based on the number of subjects with data available for that visit, unless otherwise specified.

6.4.5. Unscheduled Visits

Generally, data collected at unscheduled visits will be included and summarized for both safety and efficacy analyses in the same manner as the data collected at scheduled visits. PK and PD data collected at unscheduled visits will be listed and may not be summarized. Descriptive statistics (mean, SD, median, quartiles, minimum, and maximum) by nominal visit or time point for safety endpoints such as clinical laboratory measurements and vital signs will include only data from scheduled visits.

Summaries of outliers (eg, worst value, worst change from baseline, worst Common Terminology Criteria for Adverse Events [CTCAE] grade) during the on-treatment period for safety endpoints such as AEs, vital signs, and clinical laboratory measurements will include data from both scheduled and unscheduled visits.

For efficacy endpoints (Section 7.6), [redacted] BMD by DXA scan (Section 7.7.5), data collected at unscheduled and scheduled postbaseline visits will be mapped to scheduled visits using analysis visit windows, and then values at scheduled postbaseline visits will be derived based on the rules described below. Descriptive statistics by nominal visit and longitudinal plots will be provided using the derived values at scheduled visits.

Analysis Visit Windows

For the evaluation of efficacy endpoints (Section 7.6), [redacted] and BMD by DXA scan (Section 7.7.5), the analysis visit windows will be derived based on the target study day for the scheduled visits as follows. Note that based on the schedule of assessments, a Week 4 Visit, for example, will have a target study day of 1+(4×7)=29.

For the Double-blind Period:

- Visit windows will be implemented for scheduled visits after study day 1 (ie, starting at Week 4 through the end of the Double-blind Period on Week 24).
- For analysis visit(n):
  - Start day of visit window =1+end day of window for visit(n−1). If n=1, start day of the visit window is study day 2
  - End day of visit window = [(target day for analysis visit(n) + target day for analysis visit(n+1))/2]–1 for Week < 24. The end day of visit window for Week 24 is the end of the Double-blind Period
  - Not applicable to DXA scan results [redacted] which are only scheduled in the OLE Period.

For the OLE Period:

- Visit windows will be implemented for all scheduled visits after the first dose of mitapivat in the OLE Period. Week 288 is the last scheduled visit in the OLE Period.

- For analysis visit(n):
  - Start day of visit window = 1+ end day of window for visit(n−1). If n=1, start day of the visit is the day after the first dose of mitapivat in the OLE Period.
  - End day of visit window = [(target day for analysis visit(n) + target day for analysis visit(n+1))/2]–1 for Week < 288. The end day of visit window for Week 288 is the EOS date.

Derivation of Values at Scheduled Postbaseline Visits Based on Analysis Visit Windows

For laboratory parameters associated with efficacy endpoints (Section 7.6), [redacted] if multiple assessments are identified within a visit window for a parameter, then

- Central laboratory assessment(s) (scheduled or unscheduled) within the visit window will be used
- If no central laboratory value is within the visit window, local laboratory assessment(s) within the visit window will be used

For the evaluation of efficacy endpoints (Section 7.6), [redacted] and BMD by DXA scan (Section 7.7.5), if multiple assessments are identified within a visit window for a parameter and after selecting the assessments for laboratory parameters as described above, the following rules will be applied:

- The assessment measured closest to the target study day of the scheduled visit will be used
- If there are multiple assessments equidistant to the target study day
  - the average value will be used for laboratory parameters
  - the later assessment will be used for PerfO, PRO, HRQOL, and BMD by DXA scan

6.5. Methods for Handling Missing Data

6.5.1. Adverse Event and Concomitant Medication Start Dates

If the end date is non-missing and the imputed start date is after the end date, the end date will be used as the start date.

(1) Missing day only

- If the month and year are the same as the month and year of the date of the start of study treatment, the date of the start of study treatment will be used.
- If the month and year are before the month and year of the date of the start of study treatment, the last day of the month will be used.
- If the month and year are after the month and year of the date of the start of study treatment, the first day of the month will be used.

(2) Missing day and month

- If the year is the same as the year of the date of the start of study treatment, the date of the start of study treatment will be used.
- If the year is before the year of the date of the start of study treatment, 31 December will be used.
- If the year is after the year of the date of the start of study treatment, 01 January will be used.

(3) Missing day, month, and year

- The date of the start of study treatment will be used.

6.5.2. Adverse Event and Concomitant Medication End Dates

If the start date is non-missing and the imputed end date is before the start date, the start date will be used as the end date. If the death date is available and the imputed end date is after the death date, the death date will be used as the end date. If an imputation for an AE end date results in an AE end date that is after the EODB period date/EOS date, the AE will be considered as ongoing at the EODB period date/EOS date.

(1) Missing day only

- The last day of the month will be used.

(2) Missing day and month

- 31 December will be used.

(3) Missing day, month, and year

- The event will be regarded as ongoing.

6.5.3. Exposure

No imputation will be done for the date of the first dose of study drug.

If the date of the last dose of study drug is missing or partially missing, it will be imputed as follows (separately for each period):

- If the last date of study drug is completely missing and there is no End of Treatment Disposition electronic case report form (eCRF) page for the study drug AND there is no death date, the subject should be considered to be ongoing and the EODB period date/EOS date, will be used as the last dosing date.
- If the last date of study drug is completely or partially missing and there is EITHER an End of Treatment Disposition eCRF page for the study drug OR a death date (on or before the EODB period date/EOS date), then the imputed last dose date is:

  =Last day of the year, if only the year is available and Year <Year of min(EOT date, death date)

=Last day of the month, if both the year and month are available and Year=Year of min(EOT date, death date) and Month <Month of min(EOT date, death date)

=min(EOT date, death date), for all other cases

6.5.4. Death Date

Missing or partial death dates will not be imputed.

6.5.5. Pharmacokinetic Data

6.5.5.1. Handling of Missing Concentrations

Missing concentrations before the first dose of study treatment will be set to 0 in the analysis. A concentration is deemed missing if the PK sample is not collected, or the PK sample is collected but the concentration is not reportable.

In addition:

- If ≤2 samples have concentrations ≥LLQ, none of the PK parameters will be derived.
- If pre-dose concentrations are missing at a steady-state visit, PK parameters associated with that visit will not be derived with the exception of C_{max} and T_{max}.

Since the protocol allows a ±30 minutes window for the 7-hour PK sample collection, if the end time of the predefined area under the concentration-time curve falls outside the range of the available data then:

- the $AUC_{0\text{-}last}$ will be used to represent $AUC_{0\text{-}7,}$ if T_{last} ≥6.5 h.
- $AUC_{0\text{-}7}$ will not be reported, if T_{last} <6.5 h.

6.5.5.2. Incomplete Dosing Information

Samples associated with incomplete dosing information (ie, missing dosing date, dosing time, actual dose received, and/or dosing schedule) will be excluded from the analysis. The missing dosing information will not be imputed.

6.5.5.3. Missing Sampling Dates or Times

Samples associated with missing sample dates and/or times will be excluded from the analysis. The missing sampling dates and/or times will not be imputed.

6.5.5.4. Incomplete Concentration-time Data

Subjects who have incomplete concentration-time profiles (ie, missing concentrations at more than one scheduled time point) may be excluded from the PK analysis set. The decision will be made by the unblinded clinical pharmacologist before the study is unblinded for the analysis of the primary endpoint.

7. STATISTICAL ANALYSES

7.1. Subject Disposition

For all subjects screened in the study, the following will be summarized:

- Number of subjects screened in the study
- Frequency (number and percentage) of subjects who discontinued the study before randomization, overall and by reason for discontinuation. Percentages will be calculated based on the number of subjects screened in the study.

In addition, the frequency of subjects in each of the analysis sets described in Section 5 will be summarized by treatment arm. Percentages will be calculated only for analysis sets that are a subset of the FAS or a subset of the safety analysis set.

The following summaries will be presented by treatment arm based on the FAS or (for the summaries associated with the OLE Period only) the OLE analysis set:

- Frequency of subjects in each randomization strata and combination of randomization strata (per IXRS)
- Frequency of subjects randomized in each geographic region, country, and site
- Frequency of subjects randomized and not treated, overall and by reason for discontinuation
- Frequency of subjects who completed the Double-blind Period
- Frequency of subjects with study drug ongoing in the Double-blind Period
- Frequency of subjects who discontinued study drug in the Double-blind Period, overall and by the reason for discontinuation of study drug
- Frequency of subjects who completed the OLE Period
- Frequency of subjects with study drug ongoing during the OLE Period
- Frequency of subjects who discontinued study drug in the OLE Period, overall and by the reason for discontinuation of study drug
- Frequency of subjects who completed the safety follow-up period
- Frequency of subjects ongoing in the safety follow-up period
- Frequency of subjects who discontinued the safety follow-up period, overall and by the reason for discontinuation
- Frequency of subjects who completed the study
- Frequency of subjects ongoing in the study
- Frequency of subjects who discontinued the study, overall and by reason for discontinuation; the reason for discontinuation will be derived as the reason associated with the earliest epoch that was discontinued.

The frequency of subjects with disposition reason, in each epoch, due to reasons associated with COVID-19 will further be summarized under the main reason for discontinuation.

In addition, cross-tabulation of subjects randomized (mitapivat, matched placebo) vs subjects who have received at least 1 dose of study drug (mitapivat, matched placebo) in the Double-blind Period will be performed.

Disposition for all screened subjects and randomization data will be provided in by-subject listings.

7.2. Protocol Deviations

All major protocol deviations that impact the safety of the subjects, the conduct of the study, or the evaluation of the study results will be reported by treatment arm based on the FAS or (for the summaries associated with the OLE Period only) the OLE analysis set. These will include:

- Subjects randomized despite not satisfying the eligibility criteria
- Subjects who develop withdrawal criteria while on the study but are not withdrawn
- Subjects who receive a study drug different from that assigned at randomization
- Subjects who are randomized under the wrong stratification factor(s)
- Subjects who receive an excluded concomitant medication

In addition, for each category of major protocol deviations, those related to COVID-19 will be summarized.

Major protocol deviations will be provided in a by-subject listing.

7.3. Demographic and Other Baseline Characteristics

The following summaries will be presented by treatment arm and overall based on the FAS or (for the summaries associated with the OLE Period only) the OLE analysis set, unless otherwise specified.

7.3.1. Demographics and Physical Measurements

Demographic characteristics and physical measurements at baseline will be summarized as follows:

- Demographic characteristics
  - Sex: male, female (child-bearing potential status will be summarized for female subjects), unknown
  - Race: American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or other Pacific Islander, White, multiracial, unknown, not reported
  - Ethnic origin: Hispanic or Latino, not Hispanic or Latino, not reported

  - Age (years): summary statistics
  - Age categories:
    - <65 years, ≥65 years
    - <35 years, ≥35 years
- Physical measurements
  - Height (cm)
  - Weight (kg)
  - BMI (kg/m^2)

Demographic data for all screened subjects will be provided in a by-subject listing.

7.3.2. Disease Characteristics

In addition to randomization stratification factors (Section 7.1) and baseline for disease-related clinical laboratory assessments, the following baseline characteristics of the underlying disease will be summarized based on the data entered in the eCRF:

- Transfusion burden (number of RBC units transfused in the 24-week period before randomization) in categories: 0, 1-2, 3-5, >5.
- Splenectomy status (Yes, No; if Yes, age of splenectomy)
- Prior cholecystectomy status (Yes, No; if Yes, age of cholecystectomy)
- Prior Iron chelation status (Yes, No); the status is "Yes" if a subject has received chelation therapy within 1 year (365 days) before randomization
- Prior hydroxyurea status (Yes, No)
- DXA scan results by location (hip and spine): BMD and their corresponding T-scores and Z-scores. Frequency of subjects with worst T-score in 3 categories (≤-2.5, >-2.5-<-1, ≥-1.0) by location and across the two locations.

Data on disease characteristics will be provided in by-subject listings.

7.3.3. Medical and Surgical History

Medical and surgical history will be summarized in frequency tabulations according to the latest version of the Medical Dictionary for Regulatory Activities (MedDRA) by System Organ Class (SOC) and Preferred Term (PT).

Medical and surgical history will be provided in by-subject listings.

7.3.4. Prior Therapies

The following summaries will be presented by treatment arm and overall based on the safety analysis set.

Prior medications are defined as medications (from the "Prior and Concomitant Medications" eCRF) that are started before the start of study treatment.

All non-study medications will be coded according to the Anatomical Therapeutic Chemical (ATC) code and PT using the latest version of the World Health Organization (WHO) Drug Dictionary. All prior medications will be summarized in frequency tabulations according to the WHO ATC third level and PT.

Prior procedures are defined as procedures (from the "Prior and Concomitant Procedures" eCRF) that are started before the start of study treatment.

The prior procedures will be coded according to the latest version of the MedDRA by SOC and PT and will be summarized in frequency tabulations by SOC and PT.

Prior medications and prior procedures will be provided in by-subject listings.

7.4. Exposure to Study Drug and Compliance

The following summaries will be presented by treatment arm based on the safety analysis set for the Double-blind Period using "Prescribed Dose", "First Dose of Study Medication (Double-blind Period)", and "Study Drug Dispensation and Return" eCRFs; and the OLE analysis set for the OLE Period using the "Prescribed Dose", "First Dose of Study Medication (Open-label Extension Period)", and "Study Drug Dispensation and Return" eCRFs.

Subjects will receive 100 mg BID mitapivat or matched placebo for oral administration. Subjects who discontinue study drug should undergo the recommended dose taper. Study drug dose and recommended dose taper are provided in Table 4.

Table 4: Study Drug Dose and Recommended Dose Taper Regimen

Level	Dose	Schedule
Full dose	100 mg	BID
Taper dose	100 mg	QD for 7 days, then discontinue study drug

Abbreviations: BID = twice daily; QD = once daily

7.4.1. Treatment Duration and Exposure

Duration of exposure to study drug will be summarized as a continuous variable as well as in categories.

- Duration of exposure (days) = (last dose date−first dose date + 1)

  Duration of exposure (weeks) = (last dose date−first dose date + 1) / 7

  Duration of exposure (years) = (last dose date−first dose date + 1) / 365.25
- Double-blind Period: the categories will be >0-6, >6-12, >12-24, and >24 weeks

- OLE Period: the categories will be >0-0.5, >0.5-1, >1-2, >2-3, >3-4, >4-≤5, and >5 years.

Compliance will be summarized based on percentage of tablets taken, where

- Percentage of tablets taken =100×(total number of tablets administered)/(total number of tablets intended)
- Total number of tablets administered = total number of tablets dispensed – tablets returned
- Duration of prescription = end date of prescription – start date of prescription +1
- Total number of tablets intended = sum of number of tablets intended over all prescription periods
- Number of tablets intended during each prescription period = prescribed dosing frequency × number of tablets intended for each administration × duration of the prescription. Prescribed dosing frequency takes value of 1 and 2 for QD and BID, respectively

Percentage of tablets taken will be summarized by treatment arm. The frequency of subjects whose compliance is <80%, 80-100%, >100-120%, and >120% will be summarized.

7.4.2. Dose Modifications

The summary of dose modifications will include:

- The frequency of subjects with at least 1 dose reduction
- The reason for dose reduction
- The frequency of subjects with at least 1 interruption of study drug
- The reason for interruption of study drug

Dose reduction is defined as a prescribed non-zero daily dose that is lower than the planned dose (200 mg daily). An interruption of study drug is defined as a prescribed daily dose of 0 mg.

The reason for an interruption of study drug or dose reduction will be derived based on the "Reason for prescribed dose change" in the "Prescribed Dose" eCRF.

7.5. Concomitant Therapies

The following summaries will be presented by treatment arm based on the safety analysis set for the Double-blind Period and the OLE analysis set for the OLE Period.

Concomitant medications are defined as non-study medications (from the "Prior and Concomitant Medications" eCRF) that are started during the on-treatment period or are started before the start of the study treatment and end or remain ongoing during the on-treatment period.

All non-study medications will be coded according to ATC code and PT using the latest version of the WHO Drug Dictionary. All concomitant medications will be summarized in frequency tabulations according to WHO ATC third level and PT.

Concomitant procedures are defined as procedures (from the "Prior and Concomitant Procedures" eCRF) that are started during the on-treatment period or are started before the start of the study treatment and end or remain ongoing during the on-treatment period.

The concomitant procedures will be coded by the latest version of MedDRA by SOC and PT and will be summarized in frequency tabulations by SOC and PT.

Concomitant medications and concomitant procedures will be provided in by-subject listings.

7.6. Efficacy Analyses

The following analyses will be based on the FAS using the IXRS randomization stratification factors, unless otherwise specified. The efficacy analyses will include all data from the Double-blind Period. For subjects who discontinue treatment or become unblinded prior to completion of the 24-week Double-blind Period, their efficacy data prior to start of the OLE Period will be included in this evaluation based on the ITT principle.

7.6.1. Primary Endpoint (Hb Response)

The primary endpoint is Hb response, defined as a ≥1.0 g/dL increase in average Hb concentration from Week 12 through Week 24 compared with baseline (defined in Section 6.3.4.1).

Hb concentrations assessed within 8 weeks after an RBC transfusion will be excluded from the analysis of the primary endpoint. Subjects who do not have at least 2 on-treatment Hb concentration assessments from Week 12 to Week 24 will be considered non-responders.

7.6.1.1. Primary Analyses

The primary endpoint of Hb response will be tested using the Mantel-Haenszel stratum weighted method adjusting for randomization stratification factors. The proportion of subjects who achieved a Hb response will be summarized for each treatment arm. The adjusted difference in Hb response rate between the mitapivat arm and placebo arm, along with the 95% CI and the 2-sided p-value, will be provided.

In addition, for subjects who achieve an Hb response, mean Hb concentration increase and duration of Hb response will be summarized by treatment arm. For subjects who achieve Hb response, duration of Hb response will be calculated as the number of days from start of Hb response through the date of loss of Hb response, where:

- Start of Hb response is defined as the earliest date on or after the 'reference date' (as defined in Section 6.3.4.1) with Hb change from baseline ≥1.0 g/dL
- Loss of Hb response is defined to have occurred on the first date with Hb change from baseline <1.0 g/dL after the last assessment with Hb change from baseline ≥1.0 g/dL

- For subjects who do not have loss of Hb response, the duration of Hb response will be censored at the end of the 24-week Double-blind Period.

7.6.1.2. Sensitivity Analyses

The following sensitivity analyses will be performed:

- Hb response will be summarized using the methodology described for the primary analysis but based on PPS.
- Hb response will be analyzed using the methodology described for the primary analysis but including only those subjects who completed 24 weeks of study treatment and who did not receive any concomitant medications (prior to completion of 24 weeks of study treatment) that could affect the Hb concentrations. The list of associated concomitant medications will be finalized prior to the database lock and unblinding of the study for the primary analysis.

7.6.2. Key Secondary Endpoints

7.6.2.1. Change from Baseline in Average FACIT-Fatigue Subscale Score from Week 12 through Week 24

The FACIT-Fatigue subscale includes a 13-item self-reported fatigue subscale, which assesses the severity and impact of fatigue (including the impact on daily activities and functioning). The subscale has a 7-day recall period and is scored on a 5-point scale: 0 (not at all) to 4 (very much). The total FACIT-Fatigue subscale score ranges from 0 to 52, with a higher score indicating better HRQOL. FACIT-Fatigue subscale scores within 8 weeks after an RBC transfusion will be excluded from the analyses described below.

The primary analysis for change from baseline in average FACIT-Fatigue subscale score from Week 12 through Week 24 will be based on an analysis of covariance (ANCOVA) with treatment arm as the independent variable, and covariates for randomization stratification factors and baseline FACIT-Fatigue subscale score. Treatment effect will be evaluated as a contrast of mitapivat versus placebo. The least square (LS) means will be presented by treatment arm and the LS mean for the treatment difference between mitapivat and placebo will be presented with the associated 95% CI and 2-sided p-value.

The frequency of subjects with each response level will be summarized by treatment arm and study visit. The total subscale score will be summarized by treatment arm and study visit.

The following sensitivity analyses will also be performed:

- Mixed-effect Model for Repeated Measures (MMRM)

  The change from baseline in average FACIT-Fatigue subscale score from Week 12 through Week 24 will be compared between the mitapivat arm and the placebo arm by the MMRM method. The model will include change from baseline in FACIT-Fatigue subscale score as the dependent variable; baseline FACIT-Fatigue subscale score as a covariate, treatment, visit, and treatment-by-

visit interaction as fixed factors; and subject as the random effect with adjustment for the randomization stratification factors.

The MMRM model will be based on the restricted maximum likelihood method assuming an unstructured covariance structure to model the within-subject errors. If the model does not convergence, the compound symmetry structure will be used instead. A Kenward-Roger approximation will be used for the denominator degrees of freedom.

The estimated treatment difference between the mitapivat arm and the placebo arm in the change from baseline in average FACIT-Fatigue subscale score from Week 12 through Week 24 based on the LS Means will be provided with 95% CI and the 2-sided p-value. The estimated change from baseline for each treatment arm and the estimated treatment difference at each visit will be provided with 95% CI. Two-sided p-values will be provided for the estimated treatment difference at each visit.

- To assess the durability of effect of the study treatment and longitudinal trending, the FACIT-Fatigue score changes from baseline at each planned visit starting from Week 12 (ie, Week 12, Week 16, Week 20, Week 24) will be summarized using an ANCOVA model as described for the primary analysis.
- Pattern-mixture model with control-based pattern imputation for missing data

  Missing FACIT-Fatigue data will be imputed using multiple imputation methodology (Rubin, 1987) by replacing each missing value with a set of plausible values that represent the uncertainty about the correct value to impute. A Markov Chain Monte Carlo (MCMC) method will first be used to generate data sets following monotone missing patterns, and a regression-based pattern-mixture method will then be used to impute the remaining missing values based on the distribution data from the subjects in the placebo arm. This conservative approach assumes that after discontinuation from study treatment, subjects in the mitapivat arm will exhibit the same future evolution of the disease as subjects in the placebo arm; this approach will tend to move the estimate of the treatment effect toward a smaller value compared with missing-at-random-based analysis methods, such as mixed models with repeated measures or standard multiple imputation models. The same ANCOVA model as described for the primary analysis will be used on each of the multiply-imputed datasets and the statistics will be combined according to Rubin's rules (Rubin, 1987).

7.6.2.2. Change from Baseline in Average Hb Concentration from Week 12 through Week 24

The change from baseline in average Hb concentrations from Week 12 through Week 24 will be analyzed using an ANCOVA with treatment arm as the independent variable, and covariates for randomization stratification factors and baseline Hb concentration. Treatment effect will be evaluated as a contrast of mitapivat versus placebo. The LS means will be presented by treatment arm and the LS mean for the treatment difference between mitapivat and placebo will be presented with the associated 95% CI and 2-sided p-value.

A sensitivity analysis for change from baseline in average Hb concentrations from Week 12 through Week 24 will be performed based on the MMRM method as described in Section 7.6.2.1.

Hb concentrations assessed within 8 weeks after an RBC transfusion will be excluded from these analyses.

7.6.3. Additional Secondary Efficacy Endpoints

7.6.3.1. Hb 1.5+ Response

Hb 1.5+ response is defined as a ≥1.5 g/dL increase in average Hb concentration from Week 12 through Week 24 compared with baseline (defined in Section 6.3.4.1).

Hb concentrations assessed within 8 weeks after an RBC transfusion will be excluded from the analysis of this endpoint. Subjects who do not have at least 2 on-treatment Hb concentration assessments from Week 12 to Week 24 will be considered non-responders.

Hb 1.5+ response will be summarized using the Mantel-Haenszel stratum weighted method adjusting for randomization stratification factors. The proportion of subjects who achieved a Hb 1.5+ response will be summarized for each treatment arm. The adjusted difference in Hb 1.5+ response rate between the mitapivat arm and placebo arm, along with the 95% CI will be provided.

7.6.3.2. Change from Baseline in Markers of Hemolysis and Erythropoiesis at Week 24

Markers of hemolysis include indirect bilirubin, LDH, and haptoglobin. Markers of erythropoiesis include reticulocytes and erythropoietin.

Values and changes from baseline will be summarized based on descriptive statistics, by treatment arm and study visit, through Week 24 of the Double-blind Period. Change from baseline to Week 24 will be compared between treatment arms following the same methodology as described in Section 7.6.2.2.

By-subject longitudinal plots will be presented with values at baseline and postbaseline, transfusions, and prescribed dose over time. The plots will further include Hb response status, treatment arm, age, sex, race, randomization stratification factors, 24-week baseline transfusion burden, and hydroxyurea status.

7.6.3.3. PGIS-Fatigue and PGIC-Fatigue

The PGIS and PGIC questions are patient-reported prospective and retrospective ratings, respectively, of disease-related concepts or overall status. The PGIS is a validated, single-item measure used to assess the HRQOL impact or severity of a specific health condition (US FDA, 2018; Viktrup et al, 2012). The PGIC is a validated, single-item measure used to assess the change over time in a concept or overall status (Guy, 1976; US FDA, 2018). The PGIS questions will be administered to assess severity (on a 4-point scale ranging from "None" to "Severe") of the following 3 concepts as perceived by the subject over the recall periods noted below:

- Fatigue (7-day recall)

[redacted]

Similarly, the PGIC questions will be used to assess subject-perceived change compared with baseline in the same concepts on a 5-point scale ranging from "Much better" to "Much worse."

Assessments within 8 weeks after an RBC transfusion will be excluded from the analysis of the PGIS-Fatigue and PGIC-Fatigue endpoints.

The frequency of subjects with each response level of PGIS-Fatigue and PGIC-Fatigue will be summarized by treatment arm and study visit. Missing data will also be summarized.

A subject will be considered to have met the PGIS-Fatigue endpoint at Weeks 12, 16, 20, or 24 (improvement in PGIS-Fatigue by at least 1 category at the corresponding visit week compared with baseline, or "no change" if no or mild fatigue at baseline) if the severity at baseline compared to post-baseline (at the corresponding study visit) meets one of the following:

- None (baseline) -> None (post-baseline)
- Mild -> Mild or None
- Moderate -> Mild or None
- Severe -> Moderate, Mild or None

At each of Week 12, 16, 20 and 24 visits, the PGIS-Fatigue endpoint will be compared between the mitapivat arm and the placebo arm using the Mantel-Haenszel stratum weighted method adjusting for randomization stratification factors. The proportion of subjects who achieved the PGIS-Fatigue endpoint will be summarized for each treatment arm and study visit. The adjusted difference in the proportion of subjects who met the PGIS-Fatigue endpoint between the mitapivat arm and placebo arm, along with the 95% CI will be provided for each study visit.

A subject will be considered to have met the PGIC-Fatigue endpoint at Weeks 12, 16, 20, or 24 (improvement PGIC-Fatigue or "no change" if no or mild fatigue at baseline as assessed by PGIS-Fatigue) if the subject-perceived change compared with baseline meets one of the following at the corresponding study visit

- PGIS at baseline is None or Mild and PGIC is "No Change", "A Little Better", or "Much Better"
- PGIS at baseline is Moderate or Severe and PGIC is "A Little Better" or "Much Better"

At each of Week 12, 16, 20 and 24 visits, the PGIC-Fatigue endpoint will be compared between the mitapivat arm and the placebo arm using the Mantel-Haenszel stratum weighted method adjusting for randomization stratification factors. The proportion of subjects who achieved the PGIC-Fatigue endpoint will be summarized for each treatment arm and study visit. The adjusted difference in the proportion of subjects who met the PGIC-Fatigue

endpoint between the mitapivat arm and placebo arm, along with the 95% CI will be provided for each study visit.

7.6.3.4. Change from Baseline in the 6MWT Distance at Week 24

The 6MWT is a well-established performance outcome (PerfO) measure that is widely used to evaluate physical activity in terms of distance walked in patients with a variety of conditions. The test measures the distance an individual can walk on a hard, flat surface in 6 minutes. A lower score, based on less distance covered in 6 minutes, indicates worse level of physical activity.

Values and changes from baseline will be summarized based on descriptive statistics, by treatment arm and study visit, through Week 24 of the Double-blind Period. Change from baseline to Week 24 will be compared between treatment arms following the ANCOVA methodology as described in Section 7.6.2.2.

7.6.3.5. Changes from Baseline in Markers of Iron Metabolism at Week 24

Iron metabolism will be assessed based on serum ferritin and transferrin saturation.

Values and changes from baseline will be summarized based on descriptive statistics, by treatment arm and study visit, through Week 24 of the Double-blind Period. Change from baseline to Week 24 will be compared between treatment arms following the ANCOVA methodology as described in Section 7.6.2.2.

By-subject longitudinal plots will be presented with values at baseline and postbaseline, transfusions, and prescribed dose over time. The plots will further include Hb response status, treatment arm, age, sex, race, randomization stratification factors, 24-week baseline transfusion burden, and hydroxyurea status.

7.6.4. Subgroup Analyses

Subgroup analyses to be performed for the primary and key secondary endpoints based on the FAS are presented in Table 5.

Table 5: Subgroup Analyses for the Primary and the Key Secondary Endpoints

Subgroup	Categories
Randomization stratification factor (per IXRS): Baseline Hb concentration	≤9.0 g/dL; 9.1-10.0 g/dL
Randomization stratification factor (per IXRS): Thalassemia genotype	α-thalassemia/HbH disease; β-thalassemia
Age at screening	<35 years; ≥35 years
Sex	Male; Female
Race	Asian; Black or African American; White; other (other includes American Indian or Alaska Native, Native Hawaiian or other Pacific Islander, multiracial, unknown, and not reported)
Geographical region	North America and Europe; Asia-Pacific; rest of the world

Efficacy analyses in subgroups will be purely exploratory and are intended to evaluate the consistency of treatment effect. If there is a low number of subjects within a category (≤10% of the subjects in the FAS), the categories will be pooled (if 3 or more categories are pre-specified for the subgroup) or the subgroup will not be analyzed (if only 2 pre-specified categories in the subgroup).

For each category within each subgroup the following unstratified analyses will be performed and presented in a forest plot, separately for the primary endpoint and each of the key secondary endpoints:

- Hb response will be summarized for each treatment arm (number of responders and response rate), along with the 95% exact CI using the Clopper-Pearson method. The difference in response rates between the mitapivat arm and the placebo arm will be summarized together with the 95% CI based on the exact Clopper-Pearson confidence limits.
- Each of the key secondary endpoints will be summarized for each treatment arm using LS means, and the difference between the mitapivat arm and the placebo arm will be presented with the associated 95% CI based on the ANCOVA model described in Section 7.6.2.1 and Section 7.6.2.2 without considering the randomization stratification factors as covariates.

7.7. Safety Analyses

Safety data from the double-blind on-treatment period (Section 6.3.5) will be summarized based on the safety analysis set, by treatment arm.

Safety data from the OLE on-treatment period alone (Section 6.3.5) will be summarized separately based on the OLE analysis set, by treatment arm and overall.

7.7.1. Adverse Events

Treatment-emergent adverse events (TEAEs) are AEs with a first onset date during the on-treatment period or worsening from baseline. All summaries described below will be based on TEAEs, if not otherwise specified.

All AEs will be listed by subject and AEs with onset outside of the on-treatment period will be flagged in the listings. Unless otherwise specified, TEAEs will be summarized according to the latest version of MedDRA by SOC and/or PT, severity (based on CTCAE v4.03 grading), seriousness, and relation to study treatment in decreasing frequency based on the frequencies observed for the mitapivat arm.

Each subject will be counted only once within each SOC or PT. If a subject experiences multiple TEAEs under the same PT within a SOC for the same summary period, only the TEAE assessed as related or with the worst severity, as applicable, will be included in the summaries of relationship and severity. If a subject has TEAEs with missing and non-missing grades, the maximum of the non-missing grades will be displayed. No imputation of missing grades will be performed.

The following will be summarized:

- TEAEs by SOC and PT

- TEAEs by SOC, PT, and worst grade
- Most common TEAEs and Grade ≥3 TEAEs by PT; these will include TEAEs (any grade) reported in ≥10% of subjects in either treatment arm or Grade ≥3 TEAEs reported in ≥5% of subjects in either treatment arm. These thresholds may be changed based on the observed data without an amendment to this SAP.
- Treatment-related TEAEs, by SOC and PT
- Treatment-related TEAEs, by SOC, PT, and worst grade
- Grade ≥3 TEAEs, by SOC and PT
- Treatment-related Grade ≥3 TEAEs, by SOC and PT
- Serious TEAEs, by SOC and PT
- Treatment-related serious TEAEs, by SOC and PT
- TEAEs leading to discontinuation of study drug, by SOC and PT
- TEAEs leading to interruption of study drug, by SOC and PT
- TEAEs leading to dose reduction, by SOC and PT
- TEAEs leading to death, by SOC and PT
- Treatment-related TEAEs leading to death, by SOC and PT

7.7.1.1. Adverse Events Associated with COVID-19

The selection of AEs associated with COVID-19 will be based on the MedDRA MSSO list of PTs. The following will be summarized:

- TEAEs associated with COVID-19, by SOC and PT
- Grade ≥3 TEAEs associated with COVID-19, by SOC and PT
- Serious TEAEs associated with COVID-19, by SOC and PT
- TEAEs associated with COVID-19 leading to discontinuation of study drug, by SOC and PT
- TEAEs associated with COVID-19 leading to interruption of study drug, by SOC and PT
- TEAEs associated with COVID-19 leading to dose reduction, by SOC and PT
- TEAEs associated with COVID-19 leading to death, by SOC and PT

7.7.2. Death

The frequency of subjects who died, along with the cause of death, will be summarized for the following categories:

- On-treatment death: Deaths during the on-treatment period
- Post-treatment death: Deaths after the end of the on-treatment period

- Overall: All deaths

In addition, deaths related to COVID-19 will be summarized.

Deaths for all screened subjects will be provided in a by-subject listing.

7.7.3. Clinical Laboratory Data

Clinical laboratory test results will be expressed in SI units. For chemistry and hematology laboratory tests, the actual values and the changes from baseline will be summarized by study visit.

For each laboratory test performed in the study, a by-subject listing of laboratory test results will be presented with the corresponding CTCAE grades (if applicable), laboratory normal ranges, and flags for values below lower limit of normal (LLN) or above upper limit of normal (ULN).

Parameters with CTCAE grades available:

Clinical laboratory test results will be graded according to CTCAE v4.03 as applicable. Grading will be derived based on the numerical thresholds defined by the CTCAE criteria. Non-numerical qualifiers will not be taken into consideration in the derivation of CTCAE grading.

Laboratory test results classified according to CTCAE will be described using the worst grade. For parameters graded with 2 separate toxicity criteria, such as potassium (hypokalemia/hyperkalemia), the toxicities will be summarized separately. Low direction toxicity (eg, hypokalemia) grades at baseline and postbaseline will be set to 0 when the variables are derived for summarizing high direction toxicity (eg, hyperkalemia), and vice versa.

The frequency of subjects with laboratory toxicities during the on-treatment period will be tabulated as follows. The denominator used to calculate percentages for each laboratory test is the number of subjects evaluable for CTCAE grading for that parameter (ie, those subjects for whom a Grade of 0, 1, 2, 3 or 4 can be derived).

- The summary of laboratory parameters by CTCAE grade will include the number and percentage of subjects with Grade 1, 2, 3, 4; Grade 3-4; and Any Grade (Grades 1-4) during the on-treatment period. The highest CTCAE grade during the on-treatment period is considered the worst grade.
- The shift table will summarize baseline CTCAE grade versus worst CTCAE grade during the on-treatment period. The highest CTCAE grade during the on-treatment period is considered the worst grade.
- Newly occurring or worsening laboratory abnormalities (Any Grade, Grade 3-4) during the on-treatment period will also be summarized.

Parameters with CTCAE grades not available:

Results of laboratory tests that are not part of CTCAE will be presented according to the following categories: below the LLN, within normal limits, and above the ULN according to the laboratory normal ranges.

Shift tables will display the frequency of subjects with shifts from baseline missing, <LLN, normal, or >ULN to each of <LLN, normal or >ULN during the on-treatment period.

7.7.3.1. Hematology

For **WBC differential counts** [total neutrophil (including bands), lymphocyte, monocyte, eosinophil, and basophil counts], the absolute value will be used when reported. When only percentages are available (relevant primarily for neutrophils and lymphocytes, because the CTCAE grading is based on the absolute counts), the absolute value is derived as follows:

Derived differential absolute count=(WBC count)×(Differential % value/100)

If the range for the differential absolute count is not available (ie, the range is only available for the percentage) then Grade 1 will be attributed as follows:

- Lymphocyte count decreased:
  - Derived absolute count does not meet Grade 2-4 criteria, and
  - % value <% LLN value, and
  - Derived absolute count ≥800/mm^3
- Neutrophil count decreased:
  - Derived absolute count does not meet Grade 2-4 criteria, and
  - % value<% LLN value, and
  - Derived absolute count ≥1,500/mm^3

7.7.3.2. Chemistry

Liver function tests: Alanine aminotransferase (ALT), aspartate aminotransferase (AST), alkaline phosphatase (ALP), and total bilirubin are used to assess possible drug-induced liver toxicity. The ratios of test result to ULN will be calculated and categorized for these parameters during the on-treatment period.

[redacted]

[redacted]

7.7.3.3. Sex Hormone Tests

For sex hormone test results, shift tables will display the frequency of subjects with shifts from baseline missing, < LLN, normal, > ULN to each of < LLN, normal or > ULN during the on-treatment period.

In addition, individual longitudinal plots will be provided for each sex hormone by sex.

7.7.3.4. Pregnancy Tests

Pregnancy test results will be presented in a by-subject listing.

7.7.4. Vital Signs and Physical Measurements

For all physical measurements and vital sign assessments (height, weight, BMI, systolic blood pressure, diastolic blood pressure, pulse rate, temperature) the actual values and the changes from baseline will be summarized by treatment arm and study visit. Change from baseline does not apply to height.

Vital signs and physical measurements will be presented in a by-subject listing.

7.7.5. Dual-Energy X-ray Absorptiometry Scans

Results from bone mineral density on DXA scan (BMD and their T-scores and Z-scores) will be summarized by treatment arm and study visit, for each location.

In addition, shift tables will summarize the frequency of subjects with shifts from baseline T-scores missing, ≤-2.5, >-2.5-<-1, ≥-1.0 to worst T-score ≤-2.5, >-2.5-<-1, ≥-1.0 during the on-treatment period, by location and across the two locations.

7.7.6. Electrocardiograms

ECGs are collected during the screening period only and will be presented in a by-subject listing.

7.8. Pharmacokinetic Analyses

Only data from subjects in the mitapivat arm will be included in summary tables and plots. If a subject randomized to placebo receives at least 1 dose of mitapivat and has a PK concentration ≥LLQ, the data will be listed but will be excluded from the summary tables and plots.

All listings, summaries and plots will be presented by treatment arm based on the PK analysis set.

Concentrations <LLQ will be set to 0 for summary tables and plots with the following exceptions:

- Concentrations <LLQ between 2 quantifiable concentrations will be set to missing.
- Concentrations <LLQ following the last quantifiable concentration in a profile will be set to missing.

The following data will be excluded from summary tables and plots, unless otherwise specified:

- Pre-dose samples collected post-dose
- Post-dose samples collected pre-dose
- Concentrations >5% of C_{max}, for sample collected before the first dose of mitapivat
- Multiple samples collected at the same actual time point
- Samples not collected within the protocol-specified sampling windows will be excluded from the analysis but will be included in by-subject concentration-time plots.

In addition, concentrations ≥LLQ following 2 or more consecutive concentrations <LLQ and samples not collected within the scheduled visit window will be flagged in listings. The decision as to whether these will be included or excluded in whole or in part from the analyses will be made by the unblinded clinical pharmacologist before the study is unblinded for the analysis of the primary endpoint.

7.8.1. Concentration-Time Data

PK concentrations will be summarized by treatment arm, study visit and nominal time point. All the concentration-time data will be listed.

Arithmetic mean (±SD) of PK concentrations associated with the intensive PK sampling collection visit (Week 20), will be plotted vs nominal sampling time on both linear and log-linear scales.

By-subject PK concentrations associated with the intensive PK sampling collection visit will be plotted vs actual sampling time on both linear and log-linear scales (Week 20) by treatment arm.

By-subject PK concentrations associated with sparse PK sampling collection visits (pre-dose on Day 1 and Week 12) will be plotted vs visit using box plots. The lower and upper boundaries of the box correspond to the interquartile range (IQR). The solid line represents the median value. The whiskers represent the minimum or maximum values within 1.5×IQR.

7.8.2. Pharmacokinetic Parameters

Concentration-time data associated with the intensive PK sampling collection visit (Week 20) will be used to derive the PK parameters defined in Table 6 using noncompartmental analysis methodology (Gabrielsson J and Weiner D, 2016) in Phoenix™ WinNonlin® version 8.2 or later. The derivations will be based on the actual sampling time points calculated relative to the actual time of the most recent administration of mitapivat.

PK parameters will be listed and summarized by treatment arm and study visit. AUC0-last and T_{last} will be reported in the listings, but not included in the table summary if $T_{last} < 6.5$h or $T_{last} > 7.5$h.

Table 6: Pharmacokinetic Parameters of Mitapivat in Plasma

Parameter	Description
$AUC_{0\text{-}last}$	Area under the concentration-time curve from time "0" to T_{last} on dosing day, calculated using the linear-log trapezoidal rule
T_{last}	Time of the last quantifiable concentration
C_{max}	Maximum observed concentration.
T_{max}	Time of maximum observed concentration
C_{last}	Last quantifiable concentration after a single dose or within the dosing interval (tau) for multiple doses

7.9. Pharmacodynamic Analyses

All listings, summaries and plots for PD data will be presented by treatment arm based on the PD analysis set. Concentrations <LLQ will be set to missing.

The following data will be excluded from summary tables and plots, unless otherwise specified:

- Pre-dose blood samples collected post-dose
- Post-dose blood samples collected pre-dose
- Multiple samples collected at the same actual time point
- Samples not collected within the protocol-specified sampling windows

7.9.1. Whole Blood Concentrations of 2,3-DPG and ATP

Whole blood concentrations of 2,3-DPG and ATP, and their change from baseline and percent change from baseline will be summarized by treatment arm and study visit. Individual data will be listed.

Whole blood concentrations of 2,3-DPG and ATP measured on Week 20 will be used to derive the PD parameters defined in Table 7. The PD parameters will be listed and summarized by treatment arm.

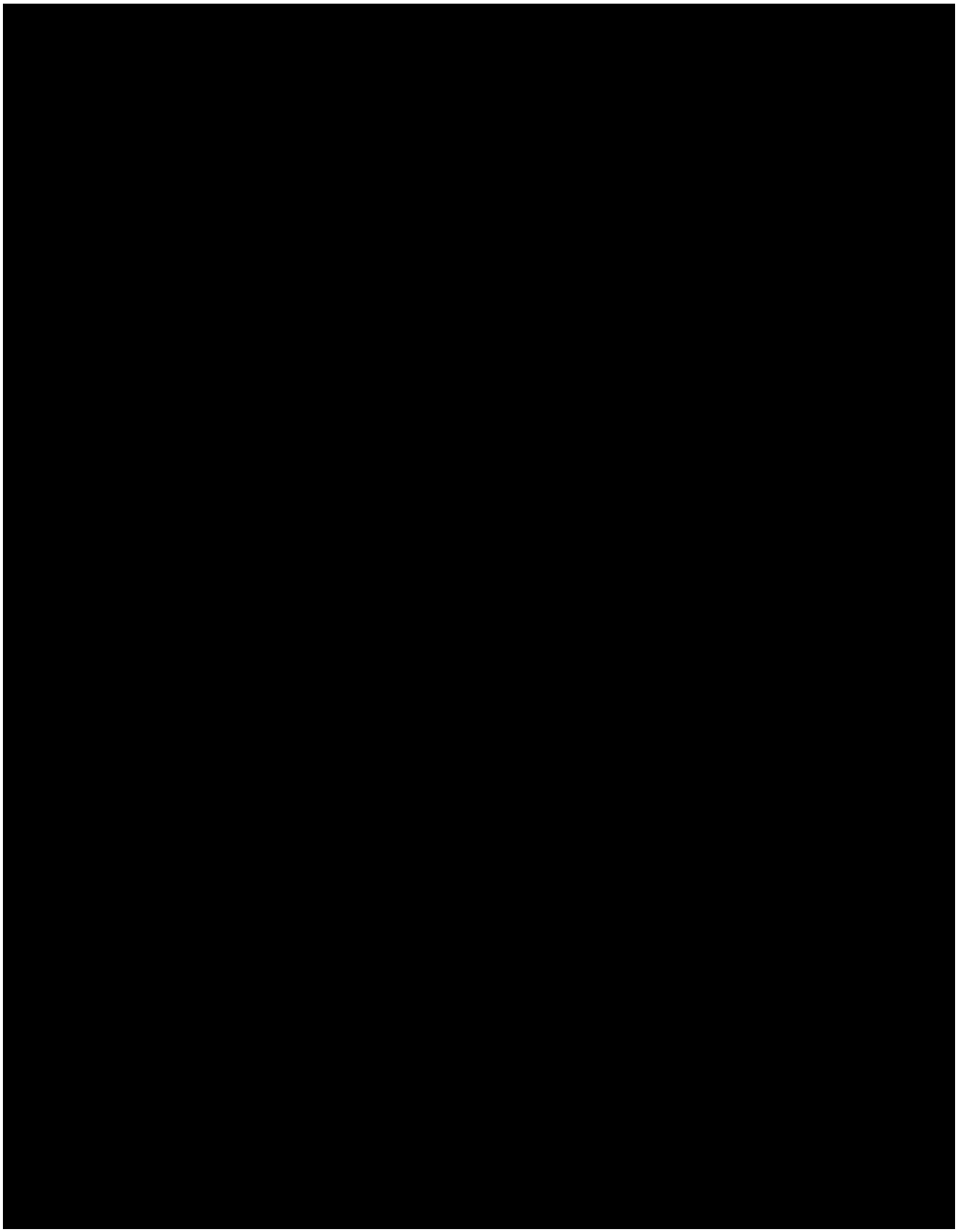
Table 7: Pharmacodynamic Parameters of Whole Blood Concentrations of 2,3-DPG and ATP

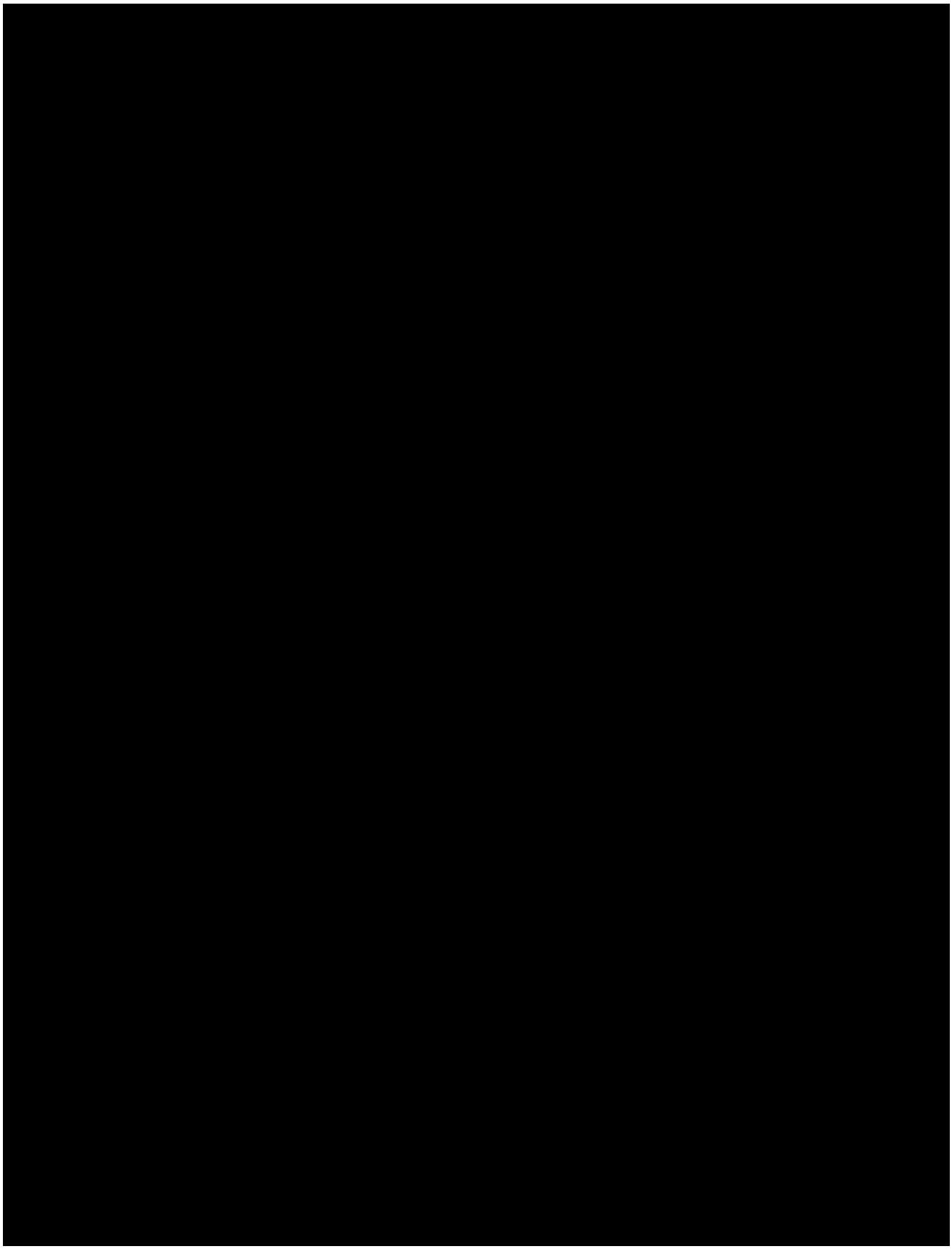
Parameters	Description
B	Baseline value as defined in Section 6.3.4
AUC_Above_B_{0-7}	Area of the response curve that is above the baseline value from time point zero (pre-dose) up to 7 hours (nominal time), using the linear trapezoid rule
AUC_Below_B_{0-7}	Area of the response curve that is below the baseline value from time point zero (pre-dose) up to 7 hours (nominal time), using the linear trapezoid rule.
AUC_Net_B_{0-7}	Net area of the response curve above and below the baseline value, calculated as AUC_Above_B_{0-7} - AUC_Below_B_{0-7}
R_{max}	Maximum observed response value over 7 hours post-dose
BR_{max}	Maximum change from baseline value over 7 hours post-dose, calculated as R_{max} - B
$\%BR_{max}$	Maximum percent change from baseline value over 7 hours post-dose, calculated as (R_{max} - B) / B × 100
R_{min}	Minimum observed response value over 7 hours post-dose
BR_{min}	Minimum change from baseline value over 7 hours post-dose, calculated as R_{min} - B
$\%BR_{min}$	Minimum percent change from baseline value over 7 hours post-dose, calculated as (R_{min} - B) / B × 100
T_{max}	Time of observed R_{max}
T_{min}	Time of observed R_{min}

7.10. Pharmacokinetic and Pharmacodynamic Analyses

Longitudinal boxplots will be presented side-by-side for mitapivat plasma trough levels and for whole blood concentrations of ATP and 2,3-DPG vs time. The analysis will be based on the PK/PD analysis set including only time points associated with time-matched mitapivat plasma concentrations and ATP/2,3-DPG whole blood concentrations.

[redacted]





[redacted]

7.12. Interim Analyses

There is no planned interim analysis in this study. An Independent Data Monitoring Committee (IDMC) will be responsible for ongoing monitoring of the safety of subjects according to the IDMC Charter.

8. REFERENCES

Cella DF, Tulsky DS, Gray G, et al. The Functional Assessment of Cancer Therapy scale: development and validation of the general measure. J Clin Oncol. 1993;11(3):570-579. doi:10.1200/JCO.1993.11.3.570

Gabrielsson J, Weiner D. (2016). Pharmacokinetic & Pharmacodynamic Data Analysis: Concepts and Applications. (5th Edition). Swedish Pharmaceutical Society, Stockholm, Sweden

Guy W. ECDEU Assessment Manual for Psychopharmacology. U.S. Dept. of Health, Education, and Welfare, Public Health Service, Alcohol, Drug Abuse, and Mental Health Administration, National Institute of Mental Health, Psychopharmacology Research Branch, Division of Extramural Research Programs; 1976.

Hillmen P, Szer J, Weitz I, et al. Results of the PEGASUS phase III randomized trial demonstrating superiority of the C3 inhibitor, pegcetacoplan, compared to eculizumab in patients with paroxysmal nocturnal hemoglobinuria. HemaSphere. 2020;4(1)(suppl):52. doi:10.1097/HS9.0000000000000404

Kalantar-Zadeh K, Aronoff GR. Hemoglobin variability in anemia of chronic kidney disease. J Am Soc Nephrol. 2009;20(3):479-487. doi:10.1681/ASN.2007070728

Piga A, Tartaglione I, Gamberini R, et al. Luspatercept Increases Hemoglobin, Decreases Transfusion Burden, and Improves Patient-Reported Outcomes in Adults with Beta-Thalassemia. Acceleron Pharma. 05 December 2016. https://acceleronpharma.com/wp-content/uploads/2017/03/20161205-Luspatercept-Increases-Hemoglobin.pdf

Rubin D. Introduction. In: Multiple Imputation for Nonresponse in Surveys. John Wiley and Sons, Inc.; 1987:1-26:chap 1.

US FDA. Methods to identify what is important to patients and select, develop or modify fit-for-purpose clinical outcomes assessments. United States Food and Drug Administration; 2018. Discussion document for patient-focused drug development public workshop on guidance 3.

Viktrup L, Hayes RP, Wang P, Shen W. Construct validation of patient global impression of severity (PGI-S) and improvement (PGI-I) questionnaires in the treatment of men with lower urinary tract symptoms secondary to benign prostatic hyperplasia. BMC Urol. 2012;12:30. doi:10.1186/1471-2490-12-30

Westfall PH, Krishen A. Optimally weighted, fixed sequence, and gatekeeping multiple testing procedures. J Stat Plan Inference. 2001;99(1):25-40. doi:10.1016/S0378-3758(01)00077-5