Official Title: A Phase 3, Randomized, Double-Blind, Placebo-Controlled Study to Evaluate the Efficacy and Safety of AG-348 in not Regularly Transfused Adult Subjects With Pyruvate Kinase Deficiency

NCT Number: NCT03548220

Document Date(s): SAP Version 1.0 – 30 June 2020

STATISTICAL ANALYSIS PLAN

A Phase 3, Randomized, Double-Blind, Placebo-Controlled Study to Evaluate the Efficacy and Safety of AG-348 in not Regularly Transfused Adult Subjects With Pyruvate Kinase Deficiency

AG348-C-006

Version: 1.0

Date: 30-Jun-2020

Prepared by:

[redacted], PhD
Study Statistician

DocuSigned by:
[redacted]
Signer Name: [redacted]
Signing Reason: I am the author of this document
Signing Time: 30-Jun-2020 | 2:36 PM EDT
F0943499CC9B4D90BB78456F4D8F7C1D

Name and Title (Printed)	Signature	Date (DD MMM YYYY)
		30-Jun-2020 | 2:36 PM EDT

Approved by:

[redacted], MD
[redacted]

DocuSigned by:
[redacted]
Signer Name: [redacted]
Signing Reason: I approve this document
Signing Time: 30-Jun-2020 | 4:48 PM EDT
5F9CA1B6F96E458C9A168748145B8447

Name and Title (Printed)	Signature	Date (DD MMM YYYY)
		30-Jun-2020 | 4:48 PM EDT

[redacted], PhD
[redacted]

DocuSigned by:
[redacted]
Signer Name: [redacted]
Signing Reason: I approve this document
Signing Time: 30-Jun-2020 | 1:07 PM EDT
1733C50FC95F44F39DEF81FD0E9BEA2A

Name and Title (Printed)	Signature	Date (DD MMM YYYY)
		30-Jun-2020 | 1:07 PM EDT

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND DEFINITIONS OF TERMS

Abbreviation	Definition
AE	Adverse event
AESI	Adverse event of special Interest
ALP	Alkaline Phosphatase
ALT	Alanine Aminotransferase
ANCOVA	Analysis of Covariance
AST	Aspartate Aminotransferase
ATC	Anatomical Therapeutic Chemical
BID	Twice daily
BMI	Body mass index
CGIC	Clinician Global Impression of Change
CI	Confidence Interval
CMH	Cochran-Mantel-Haenszel
CTCAE	Common Terminology Criteria for Adverse Events
DXA	Dual-energy x-ray absorptiometry
ECG	Electrocardiogram
EOS	End of study
EOT	End of treatment
EPO	Erythropoietin
EQ-5D-5L	European quality of life five-dimensional descriptive system
FACT-An	Functional Assessment of Cancer Therapy-Anemia
FAS	Full Analysis Set
Hb	Hemoglobin
HLT	High Level Term
HRQOL	Health-related quality of life
I-DMC	Independent data monitoring committee
IXRS	Interactive response system
LDH	Lactate dehydrogenase
LFT	Liver function test
LIC	Liver Iron Concentration
LLN	Lower limit of normal
LS	Least Square

Abbreviation	Definition
Max	Maximum
MedDRA	Medical Dictionary for Regulatory Activities
MI	Multiple imputation
Min	Minimum
MMRM	Mixed-Effect Model Repeated Measure
MRI	Magnetic resonance imaging
PGIC	Patient Global Impression of Change
PGIS	Patient Global Impression of Severity
PKD	Pyruvate kinase deficiency
PKDD	Pyruvate kinase deficiency diary
PKDIA	Pyruvate kinase deficiency impact assessment
PKR	Pyruvate kinase isoform R
PPS	Per-Protocol Set
PRO	Patient Reported Outcome
PT	Preferred Term
QD	Once-daily
QOD	Every Other Day
QTc	Heart-rate corrected QT interval
QTcB	Heart rate-corrected QT interval using the Bazett's formula
QTcF	Heart rate-corrected QT interval using the Fridericia's formula
RBC	Red blood cell
ROW	Rest of the world
SAE	Serious adverse event
SAP	Statistical analysis plan
SD	Standard deviation
SE	Standard error
SF-12	12-item Short Form Healthy Survey
SOC	System organ class
TEAE	Treatment-emergent adverse event
ULN	Upper limit of normal
WHO	World Health Organization

1. VERSION HISTORY

This statistical analysis plan (SAP) describes the analysis associated with protocol AG348-C-006 Amendment 3, Version 4.0 (dated 14 August 2019).

Table 1: Summary of Major Changes in Statistical Analysis Plan Amendments

Version	Version Date	Summary of Changes
1.0	30-Jun-2020	Original version.

2. INTRODUCTION

This SAP provides the detailed methodology for summary and statistical analyses of the data collected in Study AG348-C-006 except for pharmacokinetic data which will be described in a separate SAP. This document may modify the plans outlined in the protocol.

The clinical study report (CSR) will include all data up to the End of Study (EOS) for all subjects, which is defined as the time at which all subjects have completed the study or have been lost to follow-up.

3. TRIAL OBJECTIVES AND ENDPOINTS

3.1. Objectives

3.1.1. Primary Objective

The primary objective of the study is to evaluate the efficacy of treatment with AG-348 compared with placebo in increasing hemoglobin (Hb) concentrations.

3.1.2. Secondary Objectives

- To evaluate the safety of AG-348
- To determine the effect of the study treatment regimens on markers of hemolysis, hematopoietic activity, and other indicators of clinical activity
- To determine the effect of the study treatment regimens on health-related quality of life (HRQOL), as determined using patient-reported outcomes (PROs)
- To evaluate the pharmacokinetics of AG-348 after oral administration
- To evaluate the relationship between AG-348 pharmacokinetics and safety parameters

3.1.3. Exploratory Objectives

- To evaluate the relationship of AG-348 pharmacokinetics to indicators of clinical activity
- To evaluate the pharmacodynamic (PD) markers of pyruvate kinase deficiency (PK deficiency) and how they are affected by study treatment

- To determine the effect of the study treatment regimens on:
  - Number of transfusion events and number of red blood cell (RBC) units transfused
  - Markers of iron metabolism and indicators of iron overload

3.2. Endpoints

3.2.1. Primary Endpoint

The primary endpoint is the Hb response, defined as a ≥1.5 g/dL (0.93 mmol/L) increase in Hb concentration from baseline that is sustained at 2 or more scheduled assessments at Weeks 16, 20, and 24 during the Fixed Dose Period. The individual subject's baseline Hb concentration is defined as the average of all available Hb concentrations collected for that subject during the Screening Period up to the first dose of study treatment.

3.2.2. Key Secondary Endpoint

The key secondary endpoint is the average change from baseline in Hb concentration at Weeks 16, 20, and 24.

3.2.3. Other Secondary Endpoints

- Maximal Hb concentration increase from baseline
- Time to first achieve an increase in Hb concentration of 1.5 g/dL (0.93 mmol/L) or more from baseline
- Average change from baseline at Weeks 16, 20, and 24 in markers of hemolysis: bilirubin, lactate dehydrogenase (LDH), and haptoglobin levels
- Average change from baseline at Weeks 16, 20, and 24 in markers of hematopoietic activity: reticulocyte percentages
- Change from baseline in HRQOL PRO scores: Pyruvate Kinase Deficiency Diary (PKDD) and Pyruvate Kinase Deficiency Impact Assessment (PKDIA)
- Safety endpoints, including: the type, incidence, severity, and relationship to study treatment of adverse events (AEs) and serious adverse events (SAEs); number of discontinuations due to AEs; results of clinical laboratory tests over time (eg, serum chemistry, liver function tests [LFTs], hematology, lipids, sex steroids, urinalysis, coagulation); physical examination findings; dual-energy x-ray absorption (DXA) scans; vital signs; 12-lead electrocardiogram (ECG) data
- Pharmacokinetic endpoints, including plasma concentrations over time and pharmacokinetic parameters of AG-348 (eg, area under the concentration-time curve [AUC], maximum concentration [C_{max}], others as applicable)
- Exposure-response relationship between safety parameters and AG-348 concentration and relevant AG-348 pharmacokinetic parameters

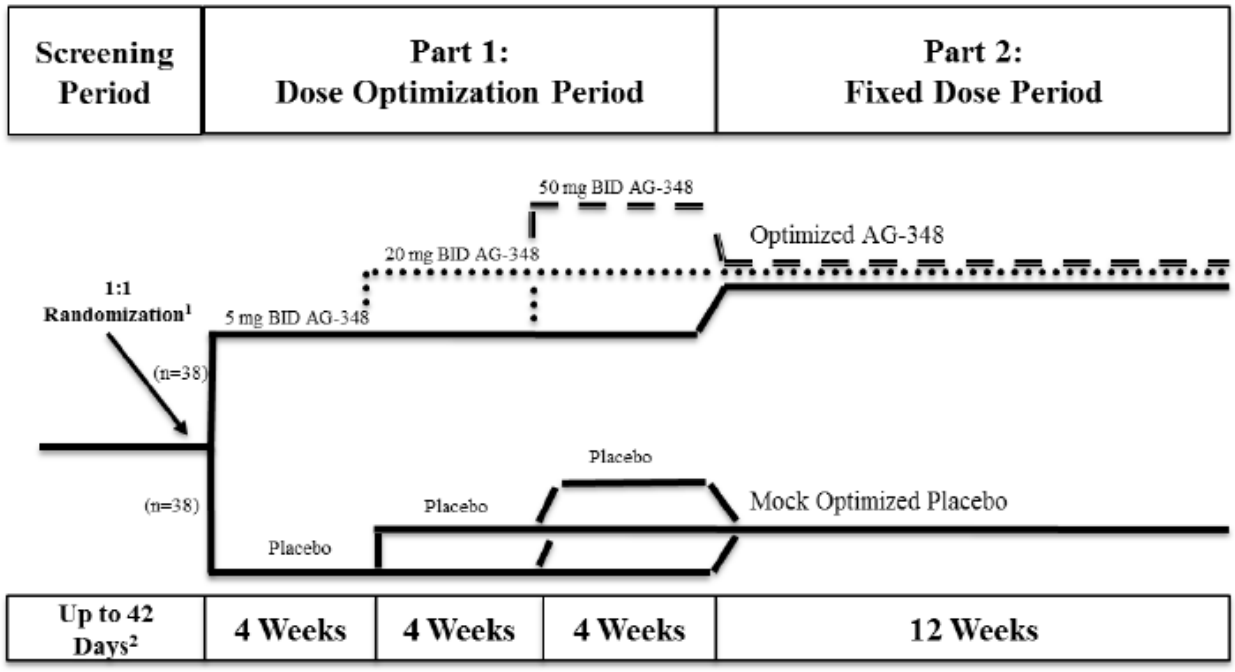
3.2.4. Exploratory Endpoints

- Exposure-response (or pharmacokinetic-pharmacodynamic) relationship between relevant pharmacokinetic parameters and endpoints that are indicators of clinical activity
- Change from baseline in additional markers of hematopoietic activity
- Change from baseline in markers of iron metabolism and indicators of iron overload
- Change from baseline in RBC-specific form of pyruvate kinase (PKR) protein level
- Relationship between baseline PKR protein level and Hb response status
- Change from baseline in HRQOL PRO scores: European quality of life five-dimensional descriptive system (EQ-5D-5L)
- Change from baseline in PKR flux assay results
- Proportion of subjects requiring transfusions and the total number of RBC units transfused

4. STUDY DESIGN

AG348-C-006 is a Phase 3, randomized, multicenter, double-blind, placebo-controlled study consisting of a Dose Optimization Period (Part 1) followed by a Fixed Dose Period (Part 2). An overview of the study design is provided in Figure 1. This study will evaluate the efficacy and safety of orally administered AG-348 (hereinafter referred to as mitapivat) as compared with placebo in subjects with PK deficiency who are not regularly receiving blood transfusions.

Figure 1: Overview of Study Design for Study AG348-C-006



Abbreviations: BID = twice daily

[1]Stratified by average of screening Hb concentrations (Hb <8.5 g/dL vs Hb ≥8.5 g/dL [5.28 mmol/L]) and the *PKLR* gene mutation category (missense/missense vs missense/non-missense). In rare instances where *PKLR* gene mutation category cannot be made definitively (eg, if a subject harbors 3 mutant PKLR alleles), the subject will be assigned to the missense/non missense category.
[2]A subject's Screening Period duration may be extended beyond 42 days upon the Medical Monitor's, or designee's, approval.

5. ANALYSIS DATA SETS

Only subjects who sign informed consent will be included in the analysis sets below. The following analysis sets will be evaluated and used for presentation of the data:

- The Full Analysis Set (FAS) will include all subjects who are randomized. Subjects will be classified according to the randomized treatment arm.
- The Safety Analysis Set will include all subjects who receive at least 1 dose of study treatment. Subjects will be classified according to the treatment actually received. If a subject randomized to placebo receives at least one dose of mitapivat then the subject will be classified to the mitapivat arm.
- The Per-Protocol Set (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 have Hb assessments at Weeks 16, 20, and 24 during the Fixed Dose Period

Table 2 summarizes the use of the analysis sets.

Table 2: Analysis Sets for Each Endpoint

Endpoints	Full Analysis Set	Per Protocol Analysis Set	Safety Analysis Set
Demographic and other baseline characteristics	✓		✓
Disposition	✓		
Major protocol deviations	✓		
Exposure and concomitant therapies			✓
Efficacy	✓	✓ (primary endpoint only)	
Safety			✓

6. GENERAL STATISTICAL CONSIDERATIONS

6.1. Randomization, Blinding, Unblinding, and Crossover

Eligible subjects will be randomized in a 1:1 ratio to mitapivat or matching placebo. Randomization assignment will be implemented by an interactive response system (IXRS) and stratified by:

- The average of screening Hb concentrations (<8.5, ≥8.5 g/dL [5.28 mmol/L])

- The *PKLR* gene mutation category (missense/missense, missense/nonmissense). In rare instances where *PKLR* gene mutation category cannot be made definitively (eg, if the subject harbors 3 mutant *PKLR* alleles), the subject will be assigned to the missense/nonmissense category.

This is a double-blind study; the subject, Investigators, and site personnel will all be blinded to the subject's study treatment assignment until database lock, if that subject is not entering the open label extension study AG348-C-011. If the subject expresses his/her intention to enter the extension study, the subject, Investigators, and site personnel will be unblinded to the subject's study treatment assignment, but only after the subject completes the Week 24 assessments. The Sponsor study team will be blinded to study treatment assignment until the database has been locked. The subject, Investigators, site personnel, and Sponsor study team will not have access to the pharmacokinetic data until database lock.

In the event that a subject's treatment assignment is unblinded to the subject, Investigator, and/or the Sponsor study team, either accidentally or in the case of emergency unblinding, the subject will be allowed to continue study treatment. Although the protocol specified that data from the unblinded subject after unblinding would be excluded from efficacy analyses all data will be included in all analyses following the intention-to-treat principle, irrespective of unblinding.

Crossover is not permitted in this study.

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}: $Odds\ Ratio = 1$ vs H_{11}: $Odds\ Ratio \neq 1$

where the odds ratio of Hb response rate of the mitapivat arm over that of the placebo arm is tested.

Approximately 76 subjects will be randomized in the study in a 1:1 ratio to mitapivat or matched placebo. Assuming a response rate of 35% in the mitapivat arm and 5% in the placebo arm, 76 subjects (38 per arm) are needed to have 90% power to reject H_{01} based on a 2-sided Fisher's Exact test with 0.05 significance level. The sample size calculation was performed in EAST© Cytel Inc. Version 6.4.

In addition, the following statistical hypotheses will be tested to address the secondary objectives:

H_{02}: $\beta_{t2} - \beta_{c2} = 0$ vs H_{12}: $\beta_{t2} - \beta_{c2} \neq 0$

H_{03}: $\beta_{t3} - \beta_{c3} = 0$ vs H_{13}: $\beta_{t3} - \beta_{c3} \neq 0$

H_{04}: $\beta_{t4} - \beta_{c4} = 0$ vs H_{14}: $\beta_{t4} - \beta_{c4} \neq 0$

H_{05}: $\beta_{t5} - \beta_{c5} = 0$ vs H_{15}: $\beta_{t5} - \beta_{c5} \neq 0$

H_{06}: $\beta_{t6} - \beta_{c6} = 0$ vs H_{16}: $\beta_{t6} - \beta_{c6} \neq 0$

H_{07}: $\beta_{t7} - \beta_{c7} = 0$ vs H_{17}: $\beta_{t7} - \beta_{c7} \neq 0$

H_{08}: $\beta_{t8} - \beta_{c8} = 0$ vs H_{18}: $\beta_{t8} - \beta_{c8} \neq 0$

where

β_{t2} and β_{c2} are the average of mean change from baseline in Hb at Weeks 16, 20, and 24 during the Fixed Dose Period in the mitapivat arm and the placebo arm, respectively.

β_{t3} and β_{c3} are the average of mean change from baseline in indirect bilirubin at Weeks 16, 20, and 24 during the Fixed Dose Period in the mitapivat arm and the placebo arm, respectively.

β_{t4} and β_{c4} are the average of mean change from baseline in reticulocyte percentage at Weeks 16, 20, and 24 during the Fixed Dose Period in the mitapivat arm and the placebo arm, respectively.

β_{t5} and β_{c5} are the average of mean change from baseline in LDH at Weeks 16, 20, and 24 during the Fixed Dose Period in the mitapivat arm and the placebo arm, respectively.

β_{t6} and β_{c6} are the average of mean change from baseline in haptoglobin at Weeks 16, 20, and 24 during the Fixed Dose Period in the mitapivat arm and the placebo arm, respectively.

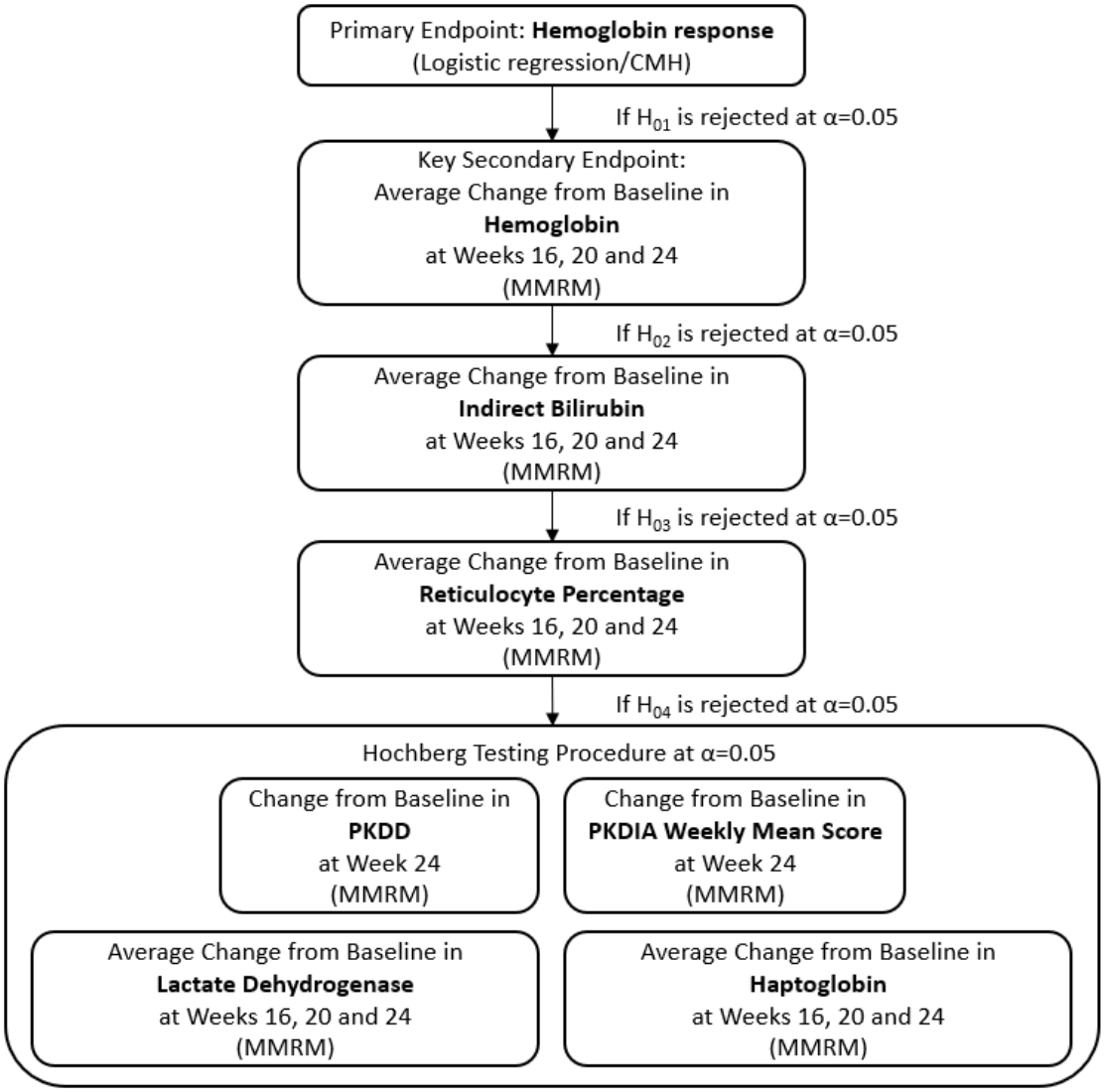
β_{t7} and β_{c7} are the mean change from baseline in PKDD weekly mean score at Week 24 during the Fixed Dose Period in the mitapivat arm and the placebo arm, respectively.

β_{t8} and β_{c8} are the mean change from baseline in PKDIA score at Week 24 during the Fixed Dose Period in the mitapivat arm and the placebo arm, respectively.

The sample size of 76 subjects (38 per arm) also provides >80% power to detect a difference of 1.4 g/dL between mitapivat and placebo in the average of mean change from baseline in Hb at Weeks 16, 20, and 24, based on a 2-sample t-test, assuming a standard deviation (SD) of 1.5 g/dL (98% power) or 2.1 g/dL (81% power).

To control the overall type I error in the study at the 2-sided 5% level, the fixed sequence testing procedure (Mauer, 1995) will be used to adjust for multiple statistical testing of the primary and secondary efficacy endpoints. These endpoints will be tested following the testing strategy outlined in Figure 2.

Figure 2: Testing Strategy



Abbreviations: CMH = Cochran-Mantel-Haenszel test; MMRM = Mixed-Effect Model Repeated Measure; PKDD = Pyruvate Kinase Deficiency Diary; PKDIA = Pyruvate Kinase Deficiency Impact Assessment.

For PKDD, PKDIA, LDH, and haptoglobin endpoints, the Hochberg testing procedure (Hochberg, 1988) will be used with 2-sided α=0.05 if all 4 endpoints higher in the hierarchy achieved statistical significance. Four 2-sided p-values for PKDD, PKDIA, LDH, and haptoglobin endpoints will be ordered from largest to smallest and compared to a set of critical values. The comparison will be conducted sequentially by comparing the largest p-value to α, then the second-largest p-value to α/2, then the third largest p-value to α/3, and the fourth largest p-value to α/4, until a p-value for an endpoint is smaller than the corresponding critical value, whereupon the Hochberg procedure provides a conclusion of statistically significant effect of mitapivat compared to placebo for that endpoint and **all** endpoints with smaller p-values.

No interim analyses for efficacy are planned.

6.2.2. Decision Rules

The study will have demonstrated that mitapivat is statistically significantly superior to placebo if the 2-sided p-value for Hb response is <0.05 in favor of the mitapivat 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 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

Both study drug and study treatment are defined as mitapivat or matched placebo.

6.3.2. Start and End Dates of Study Treatment

The start of study treatment (mitapivat or matched placebo) is the earliest date/time of administration of a non-zero dose of the study treatment.

The end of study treatment (mitapivat or matched placebo) is the latest date/time of administration of a non-zero dose of the study treatment on or before the 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, laboratory assessment) will be calculated as:

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

The study day for assessments occurring before the start of study treatment (eg, lab assessment during the Screening Period, 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

Efficacy Evaluations

For efficacy and exploratory laboratory parameters [Hb, hemolysis markers (indirect bilirubin, LDH, haptoglobin), reticulocyte percentage, iron markers, and erythropoietic markers], baseline is defined as the average of all screening assessments within 45 (42+3) days before the date of randomization (including assessments on the date of randomization) for subjects randomized and not dosed or before the start of study treatment for subjects randomized and dosed. Assessments collected within 61 days after a transfusion will be excluded from the baseline derivation.

Baseline for efficacy laboratory parameters will be derived based on central laboratory data; if no central laboratory data are available before the start of study treatment, then local laboratory data will be used to derive the baseline.

For HRQOL assessments (except for PKDD), the last measurement before the start of study treatment will be used as the baseline. For PKDD, baseline of weekly mean score is defined as the average of daily scores collected within 7 days before the start of study treatment.

Baseline Characteristics

For summaries of baseline characteristics based on the FAS, baseline will be defined as follows:

- For subjects randomized and not dosed: the last assessment on or before the date of randomization
- For subjects randomized and dosed: the last assessment on or before the start of study treatment

Safety Evaluations

For alanine aminotransferase (ALT) and aspartate aminotransferase (AST), baseline is defined as the average of all screening assessments collected within 45 (42+3) days before the start of study treatment. Baseline will be derived based on central laboratory data; if no central laboratory data are available before the start of study treatment, then local laboratory data will be used to derive the baseline.

For other laboratory assessments:

- Prior to 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 on or before the start of study treatment.

Triplicate ECGs are collected in the study; the baseline for each ECG measurement is the average of the last predose replicate measurements on or before the start of study treatment. Unscheduled assessments will not be included in the calculation of the average. QTcP will be derived based on RR and QT. The average of the replicate measurements will be determined after the derivation of the individual parameter at each time point.

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

If, per protocol, an assessment (efficacy, baseline characteristic, or safety) 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, is missing (or not collected), it will be assumed that the assessment is performed before study treatment administration. Unscheduled assessments will be used in the determination of baseline; however, an unscheduled assessment on study day 1 will be considered to have been obtained after study treatment administration.

If no assessment meets the definition of baseline for an evaluation (efficacy, baseline characteristic, or safety), the baseline will be set to missing.

6.3.5. On-Treatment Period and Optimized Dose

The **on-treatment period** starts on the date of start of study treatment and ends 28 days after the end of study treatment.

Within the on-treatment period the following dosing periods are defined:

- **Dose Optimization Period** starts on the date of start of study treatment and ends on the date of the dose administered at the Week 12 Visit or the earliest date of EOS, the end of the on-treatment period, and the first day of the dose taper prescription if a subject discontinued the study before reaching the Week 12 Visit
- **Fixed Dose Period** starts 1 day after the end of the Dose Optimization Period and ends on the first day of the dose taper prescription if the subject enters the dose taper period or the date 4 days after the end of study treatment otherwise
- **Dose Taper Period** starts 1 day after the first day of the dose taper prescription and ends on the date of the earlier date of EOS and the end of on-treatment period. This period is only applicable to subjects who went through the dose taper

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

The **optimized dose** is defined as the dose prescribed at the Week 12 visit (5 mg BID, 20 mg BID, or 50 mg BID) or the last prescribed dose if the subject discontinues study drug before the Week 12 visit.

6.4. General Methods

6.4.1. Data Handling After Cutoff Date

Not applicable.

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)=(year of given informed consent − year of birth), since only year of birth is collected in the eCRF.

  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 and median will be displayed to one more decimal place than the raw data.
- SD will be displayed to two more decimal places than the raw data.
- Percentages 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.

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. Time-to-event endpoints in the presence of censoring will be estimated using Kaplan-Meier (KM) methodology.

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. Data collected at unscheduled visits will be included in by-subject listings together with the data collected at 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 laboratory measurements and ECG parameters will include data from both scheduled and unscheduled visits.

Individual longitudinal plots for laboratory measurements during the on-treatment period will include data from both scheduled and unscheduled visits.

Descriptive statistics (mean, SD, median, quartiles, minimum, maximum) by nominal visit will only be provided for DXA scans results and liver iron concentrations (LIC) by MRI. 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.

For efficacy and exploratory endpoints [Hb, hemolysis markers (indirect bilirubin, LDH, and haptoglobin), reticulocyte percentage, iron markers, erythropoietic markers, and HRQOL], 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 during the on-treatment period for efficacy endpoints such as Hb concentration will be provided using the derived values at scheduled visits.

Analysis Visit Windows

For the evaluation of Hb, hemolysis markers (indirect bilirubin, LDH, and haptoglobin), reticulocyte percentage, iron markers, erythropoietic markers, and HRQOL endpoints, the analysis visit windows will be derived based on the target study day for the scheduled visits as follows. Note that based on the scheduled of assessments, a Week 4 Visit, for example, will have a target study day of 1+(4×7)=29.

- Visit windows will be implemented for scheduled visits after Day 1.
- 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 at Week 24 is the end of the Fixed Dose Period for Hb, hemolysis markers (indirect bilirubin, LDH, and haptoglobin), reticulocyte percentage, iron markers, erythropoietic markers, and HRQOL endpoints
- For DXA scan results and LIC by MRI, the analysis visit window for the Week 24 Visit will start on study day 86 and end on the EOS. Results outside this window will not be summarized.

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

For efficacy laboratory parameters (Hb, hemolysis markers, reticulocyte percentage, iron markers, and erythropoietic markers), any assessments obtained within 61 days after a transfusion will be excluded. In addition:

- Central laboratory assessment(s) (scheduled or unscheduled) within the visit windows 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 HRQOL data, assessments (scheduled or unscheduled) within the visit windows and on or before the end of the Fixed Dose Period will be used.

If multiple assessments are identified within a visit window for a parameter, 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 efficacy laboratory parameters
  - the later assessment will be used for HRQOL endpoints

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 an imputation for an AE end date results in an AE end date that is after the EOS date, the AE will be considered as ongoing at the 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 study drug):

- If the last date of study drug is completely missing and there is no End of Treatment Disposition eCRF page for the study drug AND there is no death date, the subject should be considered to be ongoing and the EOS date for the analysis 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 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

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:

- 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 but not treated, overall and by reason for discontinuation
- Frequency of subjects with study drug ongoing
- Frequency of subjects who discontinued study drug, overall and by the reason for discontinuation of study drug
- Frequency of subjects who completed the study
- Frequency of subjects ongoing in the study
- Frequency of subjects who discontinued the study, overall and by the reason for study discontinuation

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, a cross-tabulation of subjects randomized (mitapivat, matched placebo, none) vs subjects who have received at least one dose of study drug (mitapivat, matched placebo, none) will be presented.

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. These will include:

- Subjects randomized despite not satisfying the eligibility criteria
- Subjects who develop withdrawal criteria whilst 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, 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)
  - Race: American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, White, Other, Unknown
  - Ethnic origin: Hispanic or Latino, Not Hispanic or Latino, Not reported
  - Age (years): summary statistics
  - Age categories:
    - <65, ≥65 years
    - <35, ≥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

The following baseline characteristics of the underlying disease will be summarized based on the data entered in the eCRF:

- Baseline Hb concentration [both continuously and by categories (<8.5 g/dL, ≥8.5 g/dL)]
- Prior splenectomy status (Yes, No; if Yes, age of splenectomy)
- Prior chelation status (Yes, No); the status is "Yes" if a subject has received chelation therapy within 52 weeks (364 days) prior to the first dose of study treatment.
- Prior cholecystectomy status (Yes, No; if Yes, age of cholecystectomy)
- DXA scan results by location (femoral total and adjusted spine): Bone mineral density (BMD) and corresponding T-scores and Z-scores. Frequency of subjects with T-scores in 3 categories (≤−2.5, >−2.5-<−1.0, ≥−1.0)

Disease characteristics will be provided in by-subject listings.

7.3.3. Medical History

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

Medical history and surgical history of splenectomy 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 medications will be provided in a by-subject listing.

Prior transfusions are collected from the Transfusion History eCRF.

Transfusion history will be provided in a by-subject listing.

7.4. Exposure to Study Drug and Compliance

7.4.1. Treatment Duration and Exposure

The frequency of subjects with an optimized dose of 5 mg BID, 20 mg BID or 50 mg BID will be summarized by treatment arm.

Duration of exposure will be summarized as a continuous variable as well as in categories (>0-≤4, >4-≤8, >8-≤12, >12-≤20, >20-≤24, and >24 weeks), where

Duration of exposure = end date of study drug – start date of study drug +1.

Study drug 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
- Number of tablets intended during each prescription: for each new prescription, prescribed dosing frequency × duration of the prescription ×3. Prescribed dosing frequency takes value of 0.5, 1, and 2 for every other day (QOD), once daily (QD) and twice daily (BID), respectively.
- Total number of tablets intended = sum of number of tablets intended over all prescriptions.

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
- Summary of reasons for dose reduction

Dose reduction is defined as the prescribed dose being decreased from the previous dose collected in the Prescribed Dose eCRF. Dose prescriptions and modifications will be provided in a by-subject listing.

7.5. Concomitant Therapies

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

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 transfusions are collected in the "On Study Transfusions" eCRF page.

Concomitant transfusions will be provided in a by-subject listing.

7.6. Efficacy Analyses

The following analyses will be based on the FAS using the IXRS randomization stratification factors, unless otherwise specified.

7.6.1. Primary Endpoint

The primary endpoint is Hb Response, defined as a ≥1.5 g/dL increase in Hb concentration from baseline that is sustained at 2 or more scheduled assessments at Weeks 16, 20, and 24 during the Fixed Dose Period. Subjects with missing Hb assessment(s) over Weeks 16, 20, and 24 and who did not obtain 2 assessments with ≥1.5 g/dL increase from baseline will be considered as non-responders. The details for derivation of Hb at baseline and postbaseline visits are provided in Sections 6.3.4 and 6.4.5.

7.6.1.1. Primary Analysis

Subjects' Hb response status (Yes vs No) will be analyzed using a logistic regression model. The model includes treatment as an independent variable and adjusts for the randomization stratification factors. The estimated odds ratio between the mitapivat arm and the placebo arm with the corresponding 95% confidence interval (CI) and the 2-sided p-value will be provided.

If the logistic regression model fails to converge, or its maximum likelihood estimate does not exist due to quasi-complete separation, the non-parametric CMH test will be used for the primary analysis.

The frequency of subjects with Hb response (Hb responders) will be presented by treatment arm, and within each treatment arm by optimized dose. If the number of subjects with an optimized dose is small (<5% of the randomized subjects), the results by optimized dose may not be interpretable.

7.6.1.2. Sensitivity Analyses

Cochran-Mantel-Haenszel Test

Hb response rate will be compared between the mitapivat arm and the placebo arm based on the CMH test without continuity correction and based on the randomization stratification factors. The common odds ratio will be provided with the corresponding 95% CI and the 2-sided p-value.

Logistic Regression based on PPS

The primary analysis will be repeated based on the PPS.

Multiple Imputation (MI)

The Hb response status is indeterminate if a subject has missing Hb assessment(s) over weeks 16, 20, and 24 in the following scenarios:

- Missing 1 visit and only 1 out of the 2 remaining assessments has ≥1.5 g/dL change from baseline; or
- Missing ≥2 visits

In the primary analysis, these subjects will be considered non-responders. If >10% of randomized subjects are non-responders due to missing Hb assessments, a sensitivity analysis based on multiple imputation (MI) will be performed.

Missing Hb concentrations at Weeks 16, 20, and 24 will be multiply imputed for each treatment arm using the Markov Chain Monte Carlo (MCMC) method following the missing-at-random (MAR) assumption. In the imputation model, treatment arm, the randomization stratification factors, baseline Hb, and all scheduled postbaseline visits will be included. One hundred imputed datasets will be generated using the SAS PROC MI procedure. Hb response rate will be compared between the mitapivat arm and the placebo arm in each imputed dataset using the CMH test. CMH statistics will be normalized using the Wilson-Hilferty transformation (Wilson, 1931) (Goria, 1992), and then combined following Rubin's rules (Rubin, 1987) to generate a p-value for the combined CMH test. Sample code is provided in Appendix A (Ratitch B., 2013).

7.6.2. Key Secondary Endpoint

7.6.2.1. Definition

The key secondary endpoint is the average change from baseline in Hb concentration at Weeks 16, 20, and 24 during the Fixed Dose Period.

7.6.2.2. Primary Analysis

The average change from baseline in Hb concentrations at Week 16, 20, and 24 will be compared between the mitapivat arm and the placebo arm by the Mixed-effect Model Repeat Measurement (MMRM) method. The model will include change from baseline in Hb as the dependent variable; baseline Hb 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 average change from baseline at Weeks 16, 20, and 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. A longitudinal plot of estimated change from baseline (+/-SE) at each visit by treatment arm (mitapivat vs Placebo) will be provided.

In addition, the following summaries and figures will be provided based on Hb data collected up to the end of the Fixed Dose Period:

- Summary of Hb concentrations by visit for each treatment arm
- Summary of Hb data by visit and prior splenectomy status for each treatment arm
- A waterfall plot of the subject-level average Hb over Weeks 16, 20, and 24 with treatment arm and optimized dose, PKLR gene mutation category, baseline Hb concentration (<8.5 g/dL, ≥8.5 g/dL), and Hb response status indicated for each subject.

7.6.2.3. Sensitivity Analysis

Although the primary analysis using the MMRM model does not use formal imputation for missing Hb concentrations at Weeks 16, 20, and 24, it assumes missing at random (MAR) when handling missing data. To evaluate the robustness of the results, an ANCOVA model will be fitted with the average change from baseline in Hb at Weeks 16, 20 and 24 as the response variable; fixed effects include baseline Hb concentration, treatment arm and the randomization stratification factors. The estimated treatment difference between the mitapivat arm and the placebo arm will be provided with the 95% CIs and the 2-sided p-values.

7.6.3. Additional Secondary Efficacy Endpoints

By-subject longitudinal plots will be presented with Hb concentrations, hemolysis markers, and reticulocyte percentage at baseline and scheduled visits, and prescribed dose over time; the plots will further include treatment arm and optimized dose, age, sex, race, PKLR gene mutation category, baseline Hb concentration, prior splenectomy status, baseline chelation status, and postbaseline chelation status.

7.6.3.1. Maximum Hb Change from Baseline

Maximum Hb change from baseline up to the end of the Fixed Dose Period will be summarized by treatment arm.

A waterfall plot of the subject-level maximum Hb change from baseline will be provided with treatment arm and optimized dose, PKLR gene mutation category, baseline Hb concentration (<8.5 g/dL, ≥8.5 g/dL) and Hb response status indicated for each subject; reference lines will be included at 1.5 g/dL and 1.0 g/dL.

7.6.3.2. Hemolysis Markers

Each hemolysis marker (indirect bilirubin, LDH, and haptoglobin) will be analyzed using the same method as described for Hb in Section 7.6.2.2 The estimated treatment difference between the mitapivat arm and the placebo arm in the average change from baseline at Weeks 16, 20, and 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. A longitudinal plot of estimated change from baseline (+/-SE) at each visit will be provided for each treatment arm.

In addition, for each hemolysis marker, the following summaries and figures will be provided based on data collected up to the end of the Fixed Dose Period:

- Summary of data by visit for each treatment arm
- Summary of data by visit and prior splenectomy status for each treatment arm
- A longitudinal plot of mean value (+/-SD) at each visit by treatment arm and prior splenectomy status
- Summary of data by visit and baseline Hb concentration (<8.5 g/dL, ≥ 8.5 g/dL) for each treatment arm
- A longitudinal plot of mean value (+/-SD) at each visit by treatment arm and baseline Hb concentration (<8.5 g/dL, ≥ 8.5 g/dL)

7.6.3.3. Hematopoietic Activity Marker

The hematopoietic activity marker reticulocyte percentage will be analyzed using the same methodology as described for Hb in Section 7.6.2.2.

The summaries and figures listed for hemolysis markers in Section 7.6.3.2. will also be provided for reticulocyte percentage.

7.6.3.4. Pyruvate Kinase Deficiency Diary (PKDD)

PKDD is a self-administered, daily, 7-item PRO measure of the core signs and symptoms of PKD in adults. The PKDD daily scores will be calculated and provided by the instrument developer (Endpoint Outcomes) based on subjects' response to the PKDD questionnaire. For summary and analysis, the daily PKDD scores summarized into weekly mean scores.

The change from baseline in PKDD weekly mean score at Week 24 will be evaluated and compared between the mitapivat arm and the placebo arm using the MMRM method. The model will include change from baseline in weekly mean scores as the dependent variable, baseline PKDD weekly mean score as a covariate, treatment, week, treatment-by-week interactions, and the stratification factors as fixed factors; and subject as the random effect.

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 has convergence issue, the compound symmetry structure will be used instead. A Kenward-Roger approximation will be used for the denominator degrees of freedom.

The estimated treatment effect of change from baseline in PKDD weekly mean score for each arm and the treatment difference between the mitapivat arm and the placebo arm at each week will be provided with corresponding 95% CIs. Two-sided p-values will be provided for the estimated treatment difference at each week.

In addition, PKDD weekly mean scores up to the end of the Fixed Dose Period will be summarized by treatment arm at each visit.

7.6.3.5. Pyruvate Kinase Deficiency Impact Assessment (PKDIA)

PKDIA is a 12-item PRO measure of the impacts of PKD experienced by adults. The PKDIA score at each visit will be calculated and provided by the instrument developer (Endpoint Outcomes) based on subjects' response to the PKDIA questionnaire.

The change from baseline in PKDIA score at Week 24 will be evaluated and compared between the mitapivat arm and the placebo arm using the MMRM method. The model will include change from baseline in PKDIA score as the dependent variable, baseline PKDIA score as a covariate, treatment arm, visit, treatment-by-visit interaction, and the randomization stratification factors as fixed factors, and subject as the random effect.

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 has convergence issue, the compound symmetry structure will be used instead. A Kenward-Roger approximation will be used for the denominator degrees of freedom.

The estimated treatment effect of change from baseline in PKDIA score for each treatment arm and the treatment difference between the mitapivat arm and the placebo arm at each week will be provided with corresponding 95% CIs. Two-sided p-values will be provided for the estimated treatment difference at each visit.

In addition, the following summaries will be provided for PKDIA scores up to the end of the Fixed Dose Period by treatment arm at each visit, including:

- Summaries of scores for each item. For questions 9a, 11a and 12, the frequency of subjects with answers in each category will be summarized. For questions 9b and 11b, scores from patients who answered "Yes" will be summarized. For all the other questions, response will be treated as continuous variables and summarized
- PKDIA score and change from baseline

7.6.4. Subgroup Analyses

Subgroup analyses to be performed for the primary endpoint (Hb response) and the key secondary endpoint (average change from baseline in Hb at Weeks 16, 20, and 24) based on the FAS are presented in Table 3.

Table 3: Subgroup Analyses Primary and the Key Secondary Endpoints

Subgroup	Categories
Randomization stratification factor (per IXRS): Average of screening Hb concentration	<8.5 g/dL, ≥8.5 g/dL
Randomization stratification factor (per IXRS): *PKLR* gene mutation	Missense/Missense, Missense/Non-missense
Baseline Hb concentration	<8.5 g/dL, ≥8.5 g/dL
Age at screening	<35, ≥35 years
Sex	Female, Male
Race	White, Asian, Black or African American, Other (Other includes American Indian or Alaska Native, Native Hawaiian or other Pacific Islander, and not reported)
Geographic region	North America, Western Europe, ROW
Prior splenectomy status	Yes, No
Prior cholecystectomy status	Yes, No
Prior chelation status	Yes, No

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 the key secondary endpoint:

- Hb response will be summarized for each treatment arm (number of responders and response rate), and the odds ratio and 95% exact confidence interval will be provided.
- The average of change from baseline in Hb at Weeks 16, 20 and 24 during the Fixed Dose Period will be summarized for each treatment arm using LS means and SE, and

the difference between the mitapivat arm and the placebo arm will be presented with the associated 95% CI based on the MMRM model.

7.7. Safety Analysis

Summaries of safety data will be presented by treatment arm based on the safety analysis set.

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

In addition, the following will be summarized by treatment arm and prescribed dose at TEAE onset.

- First occurrence of TEAEs by PT
- First occurrence of serious TEAEs by PT
- TEAEs by PT
- Serious TEAEs by PT

7.7.1.1. Adverse Events of Special Interest

Transaminase increase is an AESI for mitapivat and will be reported by the investigator in the AESI eCRF page if there is a transaminase increase of >2.5×baseline or an increase in AST or ALT to Grade ≥2 in severity, whichever is lower.

Additional TEAEs of interest for mitapivat are as follows:

- AEs of endocrinological interest (identified based on the criteria outlined in the mitapivat program specified Safety Search Criteria)
- Insomnia (PTs under HLT of "Disturbances in Initiating and Maintaining Sleep" or identified based on the criteria outlined in the mitapivat program specified Safety Search Criteria)

The following will be summarized for AESIs and the additional TEAEs of interest:

- AESIs/TEAEs of interest by PT
- AESIs/TEAEs of interest by PT and worst grade
- Grade ≥3 AESIs/TEAEs of interest by PT
- AESIs/TEAEs of interest leading to discontinuation of study drug by PT
- Serious AESIs/TEAEs of interest by PT
- AESIs/TEAEs of interest leading to death by PT

In addition, the following will be summarized by treatment arm and prescribed dose at TEAE onset for the additional AE of interest "Insomnia":

- First occurrence of TEAEs by PT
- First occurrence of serious TEAEs by PT
- TEAEs by PT
- Serious TEAEs by PT

7.7.1.2. 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 in the safety analysis set who died will be tabulated based on information from the EOS eCRF. Deaths will be summarized for the following categories:

- On-treatment death: Deaths within 28 days after the last dose of study treatment (ie, deaths during the on-treatment period)
- Post-treatment death: Deaths more than 28 days after the last dose of study treatment (ie, 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. Preferred unit (g/dL) will also be used for Hb in efficacy analysis.

For each laboratory test (chemistry, hematology, coagulation) 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, 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.

The summary of liver function tests will include the following categories. The frequency of subjects with each of the following during the on-treatment period will be summarized by treatment arm:

- ALT >3×ULN, ALT >5×ULN, ALT >10×ULN, ALT >20×ULN
- AST >3×ULN, AST >5×ULN, AST >10×ULN, AST >20×ULN
- (ALT or AST) >3×ULN, (ALT or AST) >5×ULN, (ALT or AST) >10×ULN, (ALT or AST) >20×ULN
- total bilirubin >2×ULN
- Concurrent ALT >3×ULN and total bilirubin >2×ULN
- Concurrent AST >3×ULN and total bilirubin >2×ULN
- Concurrent (ALT or AST) >3×ULN and total bilirubin >2×ULN
- Concurrent (ALT or AST) >3×ULN and total bilirubin >2×ULN and ALP ≥2×ULN
- Concurrent (ALT or AST) >3×ULN and total bilirubin >2×ULN and (ALP <2×ULN or missing)

Concurrent measurements are those occurring on the same date.

Categories will be cumulative, ie, a subject with an AST >10×ULN will also appear in the categories >5×ULN and >3×ULN. Liver function test elevation and possible Hy's Law cases will be summarized using frequency counts and percentages.

An evaluation of Drug-Induced Serious Hepatotoxicity (eDISH) plot will be created, with different symbols for different treatment arms, by graphically displaying:

- Peak serum ALT (/ULN) vs peak total bilirubin (/ULN) including reference lines at ALT=3×ULN and total bilirubin=2×ULN
- Peak serum AST (/ULN) vs peak total bilirubin (/ULN) including reference lines at AST=3×ULN and total bilirubin=2×ULN

In addition, the following individual longitudinal plots and by-subject listings will be provided:

- Individual longitudinal plot of ALT including subjects with at least one ALT during the on-treatment period >2.5×baseline or worsening to CTCAE Grade ≥2 during the on-treatment period
- Individual longitudinal plot of AST including subjects with at least one AST during the on-treatment period >2.5×baseline or worsening to CTCAE Grade ≥2 during the on-treatment period

- Listing of all total bilirubin, ALT, AST, and ALP values for subjects with a postbaseline total bilirubin >2×ULN, ALT >3×ULN, or AST >3×ULN
- Listing of all total bilirubin, indirect bilirubin, ALT, AST and ALP values for subjects with a postbaseline ALT >ULN or AST >ULN

In addition, a shift table from baseline to the worst CTCAE grade of ALT and AST during the on-treatment period will be provided. For each subject:

- If the worst CTCAE grade of ALT is worse than that of AST during the on-treatment period, the baseline CTCAE grade of ALT will be used
- If the worst CTCAE grade of AST is worse than that of ALT during the on-treatment period, the baseline CTCAE grade of AST will be used
- If AST and ALT have the same worst CTCAE grade during the on-treatment period, the lower baseline CTCAE grade of ALT and AST will be used

For **calcium**, CTCAE grading is based on corrected calcium and ionized calcium. Corrected Calcium is calculated from albumin and calcium as follows:

Corrected calcium (mmol/L)=measured total calcium (mmol/L)+0.02×[40−serum albumin (g/L)]

7.7.3.3. Sex Steroid Tests

For sex steroid 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 Test

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

7.7.4. Vital Signs and Physical Measurements

All physical measurements and vital sign assessments (height, weight, BMI, systolic blood pressure, diastolic blood pressure, pulse rate, temperature) will be presented in a by-subject listing.

7.7.5. Electrocardiograms

All ECG summaries and listings will be based on the central reading results.

ECG summaries will include all ECG assessments from the on-treatment period. QTcB and QTcF interval will be derived based on RR and QT interval (see below), if not collected in the eCRF. The average of the replicate measurements should be determined after the derivation of the individual parameter at each time point.

Selecting Primary QT Interval Correction for Heart Rate

The analysis of QT interval data is complicated by the fact that the QT interval is highly correlated with heart rate. Because of this correlation, formulas are routinely used to obtain a corrected QT interval, denoted QTc, which is independent of heart rate. This QTc is intended to

represent the QT interval at a standardized heart rate. Several correction formulas have been proposed in the literature. For this analysis several of those methods of correction will be used, as described below. The QT interval corrected for heart rate by the Bazett's formula, QTcB, is defined as

$$QTcB = \frac{QT}{\sqrt{RR}},$$

and the QT interval corrected for heart rate by the Fridericia's formula, QTcF, is defined as

$$QTcF = \frac{QT}{\sqrt[3]{RR}},$$

where RR represents the RR interval of the ECG, in seconds.

Although Bazett's correction is the historical standard, it does not perform well when heart rate fluctuates. Fridericia's formula may perform better under these conditions. If QTcB and QTcF methods do not adequately correct for heart rate and there are a sufficient number of subjects (>30) with baseline ECGs, an alternative correction (QTcP) to achieve the goal of getting uncorrelated QTc and RR is based on a linear regression method which yields, theoretically, uncorrelated QTc and RR.

Linear regression method:

- Fit a model QT (ms)=a+b×RR (sec) to baseline data
- Use the estimated slope, $\hat{b}$, to correct QT
- Corrected QT for heart rate will be derived as follows:

  QTcP (ms)=QT (ms)+ $\hat{b}$×[1−RR (sec)]

Data will be summarized using QTcF and QTcB. However, if QTcF is not appropriate for the data set because of an observed large correlation between corrected QT and heart rate using the baseline assessments, the results will also be summarized using QTcP.

ECG Summaries

The following analyses will be performed for each applicable ECG parameter (RR, PR, QRS, QT, and QTc) during the on-treatment period. The denominator to calculate percentages for each category is the number of subjects evaluable for the category.

- Pearson correlation between QT and RR interval, QTc (QTcF, QTcB, and, if applicable, QTcP) and RR interval using individual (non-averaged) baseline assessments
- Frequency of subjects with notable ECG values, defined as those in the following categories:
  - QT/QTc interval increase from baseline >30 ms, >60 ms
  - QT/QTc interval > 450 ms, > 480 ms, > 500 ms
  - PR interval >200 ms

- QRS duration >120 ms

All ECG assessments and qualitative ECG abnormalities will be presented in by-subject listings.

7.7.6. DXA Scans

DXA scan results including bone mineral density (BMD), T-scores, Z-scores during the on-treatment period will be summarized by treatment arm, location (total femur and adjusted spine), and visit. For T-scores, shift from baseline to Week 24 by category (≤-2.5, >-2.5 to <-1.0, ≥-1.0) will be provided.

All DXA scan results will be presented in a by-subject listing.

7.7.7. Menstrual Cycle Diary

Menstrual cycle diary data collected from women of childbearing potential during the on-treatment period will be summarized by treatment arm and by regular contraceptive status (oral contraceptives or depot injection). The following summaries will be included:

- Total number of menstrual cycles reported
- Total number of abnormal menstrual cycles in the following categories: heavier, lighter, longer, shorter, sooner and later than usual.

Menstrual cycle diary data will be presented in a by-subject listing with regular contraceptive status flagged.

7.8. Exploratory Analysis

Unless otherwise specified, summaries for exploratory endpoints will be presented by treatment arm based on the full analysis set.

7.8.1. Use of Iron Chelation Therapy

For each subject, chelation status (Yes vs No) prior to and during the on-treatment period will be determined based on data collected in the Prior and Concomitant Medications eCRF. Shift in chelation status will be summarized.

7.8.2. Iron Markers

For iron markers including serum iron, serum ferritin, total iron-binding capacity, transferrin saturation and hepcidin, the following summaries will be provided for each parameter based on data collected up to the end of the Fixed Dose Period:

- Summaries at each visit
- Longitudinal plots of mean value (+/-SD) at each visit by treatment arm

In addition, by-subject longitudinal plots will be presented with iron markers at baseline and scheduled visits, and prescribed dose over time; the plots will further include treatment arm and optimized dose, age, sex, race, PKLR gene mutation category, baseline Hb concentration, prior splenectomy status, baseline chelation status, and postbaseline chelation status.

Iron markers including non-transferrin bound iron (NTBI), C-reactive protein (CRP) will be presented in a by-subject listing.

7.8.3. Erythropoietic Activity Markers

Erythropoietic markers include erythropoietin (EPO), soluble transferrin receptor, erythroferrone, and absolute reticulocyte count.

By-subject longitudinal plots will be presented with erythropoietic markers at baseline and scheduled visits, and prescribed dose over time; the plots will further include treatment arm and optimized dose, age, sex, race, PKLR gene mutation category, baseline Hb concentration, prior splenectomy status, baseline chelation status, and postbaseline chelation status

Erythropoietic markers will be presented in a by-subject listing.

7.8.4. Liver Iron Concentration (LIC) by Magnetic Resonance Imaging (MRI)

Liver MRI results will be summarized at each visit.

7.8.5. HRQOL Measurements

7.8.5.1. EQ-5D-5L

For EQ-5D-5L results, the following descriptive summaries will be provided at each visit, including:

- Index score and change from baseline.
- VAS score and change from baseline.
- Frequency of subjects in each level for each of the 5 dimensions.

7.8.5.2. SF-12

For SF-12 results, the following descriptive summaries will be provided at each visit, including:

- Physical Component Summary (PCS) score and change from baseline.
- Mental Component Summary (MCS) score and change from baseline.
- Frequency of subjects with answers in each category for each question.

7.8.5.3. FACT-An Subscale

For FACT-An subscale results, the following descriptive summaries will be provided at each visit.

- Subscale score and change from baseline.
- Maximum change from baseline in subscale score.
- Frequency of subjects with at least 4 points increase from baseline in subscale score.
- Frequency of subjects with answers in each category for each question.

7.8.5.4. PGIC, PGIS and CGIC

For PGIC, PGIS and CGIC results, the frequency of subjects with answers in each category for each question will be summarized at each visit.

7.9. Interim Analysis

There are no interim analyses for efficacy planned for this study.

Safety data will be reviewed regularly by an I-DMC. The last I-DMC data review will be no later than 4 months after the last subject is randomized in the study. Details are provided in the I-DMC Charter.

8. REFERENCES

Goria, M. N. (1992). On the Fourth Root Transformation of Chi-Square. *Australian Journal of Statistics, 34*(1), 55-64.

Hochberg, Y. (1988). A Sharper Bonferroni Procedure for Multiple Tests of Significance. *Biometrika, 75*(4), 800-802.

Mauer, W. H. (1995). Multiple Comparisons in Drug Clinical Trials and Preclinical Assays: a-priori Ordered Hypotheses. *Biometrie in der Chemisch-harmazeutischen Industrie, 6*, 3-18.

Ratitch B., L. I. (2013). Combining Analysis Results from Multply Imputed Categorical Data. *SAS Conference Proceedings: Pharmaceutical SAS User Group*.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys.* New York: John Wiley & Sons.

Wilson, E. B. (1931). The Distribution of Chi-Square. *Proceedings of the National Academy of Sciences of the United States of America, 17*(12), 684-688.

9. APPENDICES

Appendix A. Sensitivity Analysis of Primary Endpoint on Hemoglobin Response Using Multiple Imputation Sample Code

The sensitivity analysis specified in Section 7.6.1.2 based on MI will be performed in the following 3 steps with the sample code provided.

Step 1. 100 datasets will be imputed.

```
proc MI data=adhb_in seed=12345 NIMPUTE= 100 out=adhb_mi;
     by trt;
     em maxiter=5000 converge=1e-4 itprint outem=outem;
     var mutcat hgbcat wk0 wk4 wk8 wk12 wk16 wk20 wk24;
     mcmc chain=multiple initial=EM;
run;
```

Step 2. For each imputed dataset, Hb response status will be derived and analyzed using the CMH test.

```
proc freq data=adhb_mi(where=(paramcd='HGBRESP'));
     table mutcat*hgbcat*trt*aval /cmh;
     by _imputation_;
     ods output cmh=cmh;
run;
```

Step 3. Apply Wilson-Hilferty transformation (Wilson, 1931) (Goria, 1992) to the CMH statistics. Combine results and compute two-sided p-value following Rubin's rules (Rubin, 1987) (Ratitch B., 2013)

```
Data cmh_wh;
Set cmh(WHERE=(AltHypothesis="General Association"));
    cmh_value_wh=((VALUE/DF)**(1/3) - (1-2/(9*DF)))/SQRT(2/(9*DF));
    cmh_sterr_wh = 1.0;
RUN;

PROC MIANALYZE DATA=cmh_wh;
     ODS OUTPUT PARAMETERESTIMATES=mian_cmh_wh;
     MODELEFFECTS cmh_value_wh;
     STDERR cmh_sterr_wh;
RUN;

DATA mian_cmh_wh_p;
SET mian_cmh_wh;
    pval = Probt;
RUN;
```