# Statistical Analysis Plan for Study M19-063

# A Randomized, Open Label Phase 3 Study Evaluating Safety and Efficacy of Venetoclax in combination with Azacitidine after allogeneic Stem Cell Transplantation in Subjects with Acute Myeloid Leukemia (AML)

## Date: 16 November 2023

## Version 2.0

## Table of Contents

## List of Tables

## List of Figures



## List of Appendices

# 1.0 Introduction

This Statistical Analysis Plan (SAP) describes the statistical analyses for Venetoclax (ABT-199) Study M19-063 titled "A Randomized, Open Label Phase 3 Study Evaluating Safety and Efficacy of Venetoclax in combination with Azacitidine after allogeneic Stem Cell Transplantation in Subjects with Acute Myeloid Leukemia (AML)" Version 7.0.

Study M19-063 examines the efficacy and safety of venetoclax in combination with azacitidine after allogeneic stem cell transplantation in subjects with AML.

The analyses of pharmacokinetic endpoints and biomarker exploratory endpoints will not be covered in this SAP.

The SAP will not be updated in case of administrative changes or amendments to the protocol unless the changes impact the analysis.

Unless noted otherwise, all analyses will be performed using SAS Version 9.4 (SAS Institute Inc., Cary, NC 27513) or later under the UNIX operating system.

# 2.0 Study Design and Objectives

## 2.1 Objectives and Hypotheses

The primary objectives of the study are to determine the recommended Phase 3 dose of venetoclax in combination with azacitidine in AML patients when given as maintenance therapy following allogeneic stem cell transplantation (Part 1) and to determine efficacy of venetoclax in combination with azacitidine on overall survival (OS) in AML patients compared to Best Supportive Care (BSC) when given as maintenance therapy following allogeneic stem cell transplantation (Part 2).

The secondary objectives of the study (Part 2) are:

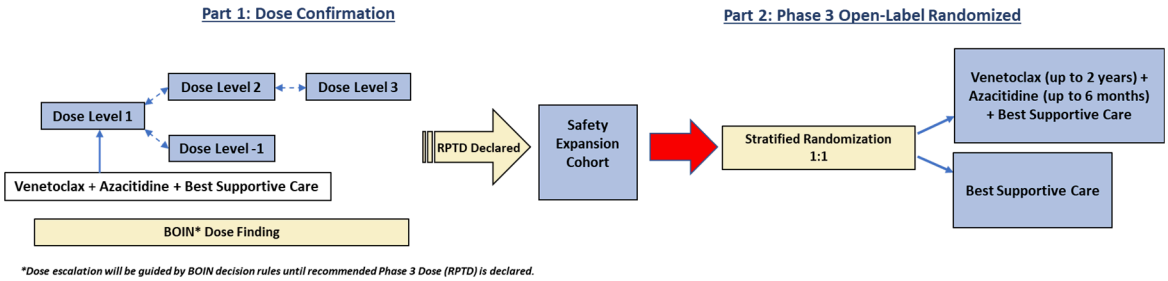
1. To confirm the safety of venetoclax in combination with azacitidine after allogeneic transplantation in subjects diagnosed with AML.

2. To determine the efficacy of venetoclax in combination with azacitidine on morphologic relapse-free survival (RFS) and composite RFS.
3. To determine the effect of venetoclax in combination with azacitidine on the frequency and severity of graft-versus-host disease (GvHD).
4. To determine the effect of venetoclax in combination with azacitidine on quality of life (QoL).
5. To determine the effect of venetoclax in combination with azacitidine on measurable residual disease levels.

## 2.2 Study Design Overview

The schematic of the study is shown below (Figure 1).

**Figure 1. Study Schematic**




## 2.3 Treatment Assignment and Blinding

In Part 1 of the study, dose escalation and de-escalation will be guided by a Bayesian Optimal Interval (BOIN) design based on the cumulative number of dose limiting toxicities (DLTs) observed at the current dose level. Additional information regarding the BOIN design is provided in Appendix D.

In Part 2 of the study, which is open label, subjects will be randomized 1:1 to one of the following two arms:

- Arm A: venetoclax and azacitidine: 6 cycles treatment with azacitidine and up to 24 cycles treatment with venetoclax in addition to best supportive care.
- Arm B: best supportive care only.

Randomization will be stratified by:

- < 18 years of age: no further stratification in this group
- ≥ 18 years of age, stratified further by:
  - Previous treatment with either venetoclax or azacitidine (yes/no)
  - High risk of recurrence (yes/no), where high risk is defined as one or more of the following features
    - Minimal residual disease (MRD) positive (≥ $10^{-3}$), per investigator prior to transplantation, or
    - High risk AML as defined in Appendix F of the protocol, or
    - Previous recurrence (CR2 or later), or
    - Previous refractory disease at the end of induction
  - Geographic region (North and South America + Australia + Europe versus China versus Japan versus Other Asian Countries)

## 2.4 Sample Size Determination

This study has two parts. Part 1 is a dose confirmation study and Part 2 is a randomized study.

The sample size for Part 1 is dependent upon the dose levels utilized.

The below assumed hazard ratio of 0.65 is based upon structured interviews with qualified health care providers, which included the question for clinically meaningful treatment improvement. The assumed hazard ratio of 0.65 describes a relevant and clinically

meaningful improvement over BSC. In Study M14-358, a CR + CRi rate of 71% was observed for venetoclax [redacted] mg in combination with azacitidine[1] as compared to a historical 28% CR + CRi rate in azacitidine alone.[2] The observed median OS for venetoclax [redacted] mg in combination with azacitidine in Study M14-358 was 16.9 months[1] as compared to a historical 10.4 months for azacitidine monotherapy.[2] The assumed hazard ratio below appears to be reasonable based on what was observed in the treatment effect of venetoclax [redacted] mg in combination with azacitidine in Study M14-358. The sample size calculation for Part 2 is based on the following assumptions:

- 2-year OS for the best supportive care arm is 50%.
- Hazard ratio (venetoclax plus azacitidine versus best supportive care) is 0.65.
- An efficacy interim analysis of OS at 75% of total OS events with O'Brien-Fleming boundary.
- 1:1 randomization ratio to venetoclax plus azacitidine arm and the best supportive care arm.

With the above assumptions, a total of 231 OS events will provide approximately 90% overall power to detect a statistically significant difference between treatment arms at two-sided alpha level of 0.05 using the log-rank test. A total of approximately 400 subjects will be randomized into the study to obtain the 231 OS events.

# 3.0 Endpoints

## 3.1 Primary Endpoint(s)

The primary endpoint of Part 1 of the study is the frequency of DLTs of venetoclax in combination with azacitidine and to confirm the dose for the randomized portion, Part 2, of the study. The primary endpoint of Part 2 of the study is OS, defined as the time from randomization to death from any cause.

## 3.2 Secondary Endpoint(s)

Key secondary endpoints (Part 2) include:

- Independent Review Committee (IRC)-assessed morphologic RFS, defined as time from randomization to morphologic relapse from AML or death from any cause.
- Independent Review Committee (IRC)-assessed composite RFS, defined as time from randomization to morphologic relapse from AML, non-morphologic relapse from AML or death from any cause.
- GvHD-free, relapse-free survival (GRFS), defined as the time from randomization to disease relapse, incidence of GvHD, or death from any cause. Incidence of GvHD is defined as grade 3 or 4 acute GvHD, or chronic GvHD requiring immunosuppressive therapy (IST).
- The rate of subjects without higher grade GvHD at 90 days after randomization (Arm B) or initiation of treatment (Arm A). Higher grade GvHD is defined as Grade 2 or higher acute GvHD, or moderate or severe chronic GvHD.
- Change from baseline in physical functioning at 6 months after randomization (Arm B) or initiation of treatment (Arm A) in adult subjects, as measured by the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire (EORTC QLQ-C30) physical functioning domain.
- Fatigue change from baseline in adult subjects as measured by Patient Reported Outcomes Measurement Information System (PROMIS) Cancer Fatigue SF 7a at 6 months after randomization (Arm B) or initiation of treatment (Arm A).
- Measurable residual disease (MRD) conversion rate, defined as percentage of subjects with MRD $\geq 10^{-3}$ at randomization who convert to MRD $< 10^{-3}$ after randomization (Arm B) or initiation of study treatment (arm A).
- Time to deterioration in Global Health Status (GHS)/QoL in adult subjects, defined as the time from randomization to death from any cause or deterioration of $\geq 10$ points from baseline in GHS/QoL score as measured by EORTC QLQ-C30.

Additional secondary endpoints (Part 2) include:

- Fatigue change at time points other than 6 months after initiation of treatment (Arm A) or randomization (Arm B).
- Patient reported outcomes: Change in subject reported signs, symptoms and impact of AML as measured by the EORTC-QLQ-C30, and European Quality-of-Life-5 dimesional-5-level (EQ-5D-5L).

## 3.3 Additional Efficacy Endpoint(s)

The additional efficacy endpoints (Part 2) are:

- Patient reported outcomes for subjects with chronic GvHD.
- Patient reported outcomes for adolescent subjects: Pediatric Quality of Life Inventory (PedsQL) (for Subjects 12 - 17 years old).
- Health Care Resource Utilization: Information will be collected on each hospitalization including reason for admission (e.g., disease relapse, AML-related illness, treatment-related AE) and days of hospitalization by treatment setting (e.g., inpatient unit, special care unit, etc.).
- Immunosuppressive therapy as a secondary indicator of GvHD severity.
- Frequency and type of infections related morbidity and mortality, and anti-infective medications used.
- Rate of AML recurrence.

## 3.4 Safety Endpoints

Safety and tolerability will be assessed by evaluating adverse events (AEs), physical examinations and changes in laboratory data and vital signs for the entire study treatment duration.

# 4.0 Analysis Populations

The following population sets will be used for the analyses.

The Intent-to-Treat (ITT) Population includes all randomized subjects in Part 2. The ITT Population will be used for all efficacy analyses. Subjects will be included in the analysis according to the treatment arms that they are randomized to.

The Safety Analysis Set consists of all subjects who received at least 1 dose of study drug in Part 1 or Arm A and all randomized subjects in Arm B. Subjects will be included in the analysis according to the assigned treatment arm.

# 5.0 Subject Disposition

The total number of subjects who were screened, enrolled, and treated will be summarized. Reasons for exclusion, including screen failure, will be summarized.

A summary of subject accountability will be provided where the number of subjects in each of the following categories will be summarized:

- Subjects enrolled in the study;
- Subjects who took at least one dose of study drug;
- Subjects who discontinued study drug (all reasons and primary reason);
- Subjects who discontinued study (all reasons and primary reason).

# 6.0 Study Drug Duration and Compliance

For the Safety Analysis Set, duration of treatment will be summarized for subjects enrolled in Part 1 and in Arm A of Part 2. Duration of treatment is defined for each subject as last dose date minus first dose date + 1. Duration of treatment will be summarized using the number of subjects treated, mean, standard deviation, median, minimum and maximum. In addition, the number and percentage of subjects in the following treatment duration intervals will be summarized:

- 0 to 4 weeks (0 to 28 days),
- > 4 weeks to 8 weeks (29 to 56 days),
- > 8 weeks to 12 weeks (57 to 84 days),
- > 12 weeks to 16 weeks (85 to 112 days),
- > 16 weeks to 20 weeks (113 to 140 days),
- > 20 weeks to 24 weeks (141 to 168 days),
- > 24 weeks to 28 weeks (169 to 196 days),
- > 28 weeks to 32 weeks (197 to 224 days),
- > 32 weeks to 36 weeks (225 to 252 days),
- > 36 weeks to 52 weeks (253 to 364 days),
- > 52 weeks (> 364 days).

The number of cycles that subjects are exposed to study drug will be summarized for Part 1 and Arm A of Part 2. There will be no statistical comparison for the study treatment exposure.

# 7.0 Demographics, Baseline Characteristics, Medical History, and Prior/Concomitant Medications

Demographics, baseline or disease characteristics, medical history, and prior and concomitant medications will be summarized for the ITT population overall and by treatment arm. Categorical variables will be summarized with the number and percentage of subjects; percentages will be calculated based on all observations (including missing observations). Continuous variables will be summarized with descriptive statistics (number of non-missing observations, mean and standard deviation, median, minimum and maximum).

## 7.1 Demographics and Baseline Characteristics

All baseline characteristic summary statistics and analyses are based on characteristics prior to the first dose of study drug or on or prior to date of randomization for subjects who did not receive any study drug or was randomized to Arm B.

Distributions of the continuous demographic and baseline characteristic variables will be summarized by treatment arm with the number of non-missing observations, mean, standard deviation, and median, as well as the minimum and maximum values.

For the categorical demographic and baseline characteristic variables, the frequency and percentages of subjects within each category will be summarized by treatment arm. The number of subjects with missing information will also be summarized.

There will be no statistical comparison for the demographic and baseline characteristics.

The following demographic and baseline characteristics will be summarized:

Demographics:

- Age (years) and Age Categories (< 18, ≥ 18, ≥ 18 - < 65, ≥ 65, ≥ 75 years)
- Gender (Male/Female)
- Race (White, Black or African American, Asian, and Other)
- Region by Stratification (North and South America + Australia + Europe, China, Japan, Other Asian countries/territories)
- Region by Country (US, Europe, China, Japan, Rest of World)
- Height (cm)
- Weight (kg)

Baseline and Disease-Related Characteristics:

- Previous treatment with venetoclax or azacitidine (Yes, No)
- Previous treatment with venetoclax (Yes, No)
- Previous treatment with azacitidine (Yes, No)
- High risk of recurrence (Yes, No)
- MRD status prior to transplantation ($\geq 10^{-3}$, $< 10^{-3}$)
- Cytogenetic risk at initial diagnosis (Favorable, Intermediate, Poor)
- ELN risk category at initial diagnosis (Favorable, Intermediate, Adverse)

- Response to the initial treatment for AML (CR, CRi, Morphologic leukemia-free state, PR, Treatment failure)
- AML recurrence prior to transplant (Yes, No)
- Remission status prior to transplant (CR1, CR2 or later, non-CR)
- Baseline remission status (CR, CRi)
- Previous refractory disease at the end of induction (Yes, No)
- Time from initial AML diagnosis to study entry
- Time from transplant to study entry
- Pre-transplant conditioning regimen (Reduced intensity or non-myeloablative, Myeloablative)
- Stem cell donor type (Matched related, Other)
- Stem cell graft source (Bone marrow, Cord blood stem cell, Peripheral blood stem cell)
- Donor-recipient HLA match (6/6, 10/10, 12/12, Other)
- Use of ex-vivo T-cell depletion (Yes, No)
- Acute GvHD at baseline (Grade 0, Grade 1, Grade 2, Grade 3, Grade 4)
- Baseline hemoglobin (g/dL)
- Baseline CTCAE grade of anemia
- Baseline platelet count ($10^9$/L)
- Baseline CTCAE grade of thrombocytopenia
- Baseline neutrophil count ($10^9$/L)
- Baseline CTCAE grade of neutropenia
- Hepatic function (Normal, Mild impairment, Moderate impairment, Severe impairment)
- Renal function (Normal, Mild impairment, Moderate impairment, Severe impairment)

## 7.2 Medical History

Medical history data will be coded using the Medical Dictionary for Regulatory Activities (MedDRA). The actual version of the MedDRA coding dictionary will be noted in the statistical tables and clinical study report. The number and percentage of subjects in each medical history category (by MedDRA system organ class and preferred term) will be summarized overall and by treatment arm. The system organ class (SOC) will be presented in alphabetical order, and the preferred terms will be presented in alphabetical order within each SOC. Subjects reporting more than one condition/diagnosis will be counted only once in each row (SOC or preferred term). There will be no statistical comparison for the medical history.

## 7.3 Prior and Concomitant Medications

Prior and concomitant medications will be summarized by generic name. For subjects enrolled in Part 1 or Arm A in Part 2, a prior medication is defined as any medication taken prior to the date of the first dose of study drug. A concomitant medication is defined as any medication that started prior to the date of the first dose of study drug and continued to be taken after the first dose of study drug or any medication that started on or after the date of the first dose of study drug, but not after the date of the last dose of study drug. For subjects enrolled in Arm B in Part 2 or Arm A in Part 2 without receiving any study drug, a prior medication is defined as any medication taken prior to the date of randomization. A concomitant medication is defined as any medication that started prior to the date of randomization and continued to be taken after randomization or any medication that started on or after the date of randomization, but not after the End Visit, disease relapse (assessed by investigator) or study discontinuation, whichever is the earlier.

The number and percentage of subjects taking medications will be summarized by generic drug name based on the World Health Organization (WHO) Drug Dictionary for both prior and concomitant medications. A subject who reports the use of two or more medications will be counted only once in the summary of any prior or concomitant

medication. A subject who reports two or more uses of the same medication will be counted only once in the total for the associated generic drug name.

If an incomplete or missing start date was collected for a medication, the medication will be assumed to be a concomitant medication, unless there is evidence to the contrary.

There will be no statistical comparison for the summaries on prior or concomitant medications.

# 8.0 Efficacy Analyses

## 8.1 General Considerations

Unless otherwise specified, all efficacy analyses will be conducted in the ITT population and all tests will be 2-sided at an alpha level of 0.05 (when rounded to four decimal places).

The data cutoff date for the primary analysis of OS will be the projected date that the 231$^{st}$ OS event has occurred in the ITT population. When the 231$^{st}$ OS events in the ITT population occurs, there will be a final review of the eCRF data. After the data collection is completed and reviewed for completeness and all data management quality assurance (QA) and quality control (QC) procedures are performed, the clinical database will be extracted for documentation and statistical analyses of the efficacy and safety. Unless otherwise specified, data after the cutoff date will be excluded from statistical analyses.

Although the study is open label, the Sponsor will not conduct any by-treatment analysis or summary of the data using the actual treatment information until the database lock for the final primary analysis of OS, regulatory submissions based on interim results, or termination of the study, whichever occurs earlier.

Unless otherwise specified, the primary analyses of the disease relapse component in the related endpoints (e.g., morphologic RFS, composite RFS, GRFS) will be based on IRC assessment. Supplementary analyses will be performed based on the investigator assessment.

Unless otherwise specified, age (< 18, ≥ 18), previous treatment with either venetoclax or azacitidine (yes, no) and high risk of AML recurrence (yes, no) will be used in all stratified analyses of efficacy endpoints in the ITT population. The strata to be used in the efficacy analyses are defined as follows:

- Stratum 1: age < 18
- Stratum 2: age ≥ 18, previous treatment with either venetoclax or azacitidine (yes), high risk of AML recurrence (yes)
- Stratum 3: age ≥ 18, previous treatment with either venetoclax or azacitidine (yes), high risk of AML recurrence (no)
- Stratum 4: age ≥ 18, previous treatment with either venetoclax or azacitidine (no), high risk of AML recurrence (yes)
- Stratum 5: age ≥ 18, previous treatment with either venetoclax or azacitidine (no), high risk of AML recurrence (no)

In the analyses of endpoints that only includes adult subjects, such as time to deterioration in GHS/QoL, stratum 1 (age < 18) will not be included.

The stratification factor values under which the subject is randomized by the IVRS/IWRS will be used in the efficacy analyses.

Unless otherwise specified, time-to-event endpoints will be analyzed using Kaplan-Meier methodology and compared between treatment arms using the log-rank test, stratified by the strata defined above. The hazard ratio between treatment arms will be estimated using the Cox proportional hazards model, stratified by the strata defined above. In addition, results from stratified and unstratified Wilcoxon test, unstratified log-rank test and Cox proportional hazards model will also be provided.

For subjects randomized to Arm A, baseline refers to the last non-missing observation before the first dose of study drug administration or randomization if no study drug is given. For subjects randomized to Arm B, baseline refers to the last non-missing

observation on or prior to the Cycle 1 Day 1 visit or randomization if the Cycle 1 Day 1 visit is not performed.

## 8.2 Handling of Missing Data

Details on the handling of missing data are described in the analysis section corresponding to each endpoint.

## 8.3 Primary Efficacy Endpoint(s) and Analyses

### 8.3.1 Primary Efficacy Endpoint(s)

The primary efficacy endpoint is OS. OS is defined as the time from randomization to death from any cause. The primary analysis of OS will occur when the 231$^{st}$ OS event is observed in the ITT population. All subjects in the ITT population will be included in the OS analysis.

### 8.3.2 Handling of Missing Data for the Primary Efficacy Endpoint(s)

Subjects who have not died will be censored at the last known alive date on or before the cutoff date. The last known alive date will be determined by selecting the last available date of the following study procedures: randomization, start date of adverse event, bone marrow collection, disease assessment, cytogenetic test, molecular marker test, flow cytometry, MRD assessment, clinical GvHD assessment, vital signs assessment, clinical laboratory collection, microbiology test, study drug administration, start date of concomitant medicine, concomitant procedures, start date of follow-up therapies, follow-up procedures, survival follow-up, COVID-19 status, hospitalization, biospecimen sample collection, quality of life assessments, and performance status.

### 8.3.3 Primary Efficacy Analysis

The distribution of OS will be estimated for each treatment arm using Kaplan-Meier methodology and compared between treatment arms using the log-rank test, stratified by

strata defined in Section 8.1. The hazard ratio between treatment arms will be estimated using the Cox proportional hazards model, stratified by strata defined in Section 8.1.

The attributes of the estimand corresponding to the primary efficacy endpoint are summarized in Table 1.

**Table 1. Summary of the Estimand Attributes of the Primary Efficacy Endpoint(s)**

| | Attributes of the Estimand | | | |
|---|---|---|---|---|
| **Endpoint** | **Treatment** | **Population** | **Handling of Intercurrent Events** | **Population-level Summary** |
| OS | Best supportive care with and without venetoclax plus azacitidine | All randomized subjects | Initiation of new anti-leukemic therapy<br>• Treatment Policy Strategy: this intercurrent event is irrelevant; OS data will be used in the analysis regardless of whether the subject is on new anti-leukemic therapy or not<br>Discontinuation (either premature discontinuation or completion of protocol-specified treatment duration) of study drug<br>• Treatment Policy Strategy: this intercurrent event is irrelevant; OS data will be used in the analysis regardless of whether the subject is on study drug or not | Hazard ratio |

### 8.3.4 Additional Analyses of the Primary Efficacy Endpoint(s)

The following supplementary analyses of OS will be performed:

- Analysis of OS by censoring subjects at initiation of new anti-leukemic therapies.

Covariate-adjusted Cox proportional hazards model incorporating additional prognostic factors may also be performed.

## 8.4 Secondary Efficacy Analyses

### 8.4.1 Key Secondary Efficacy Analyses

The key secondary endpoints include IRC-assessed morphologic RFS, IRC-assessed composite RFS, GRFS, time to deterioration in GHS/QoL, rate of subjects without higher grade of GvHD at 90 days after treatment or randomization, change in physical functioning from baseline at 6 months after treatment or randomization, change in fatigue from baseline at 6 months after treatment or randomization, and MRD conversion rate.

The attributes of estimands corresponding to the key secondary efficacy endpoints are summarized in Table 2.

**Table 2. Summary of the Estimand Attributes of the Key Secondary Efficacy Endpoints**

| | Attributes of the Estimand | | | |
|---|---|---|---|---|
| **Endpoint** | **Treatment** | **Population** | **Handling of Intercurrent Events** | **Population-level Summary** |
| Morphologic RFS | Best supportive care with and without venetoclax plus azacitidine | All randomized subjects | Initiation of new anti-leukemic therapy<br>• Treatment Policy Strategy: this intercurrent event is irrelevant; morphologic RFS data will be used in the analysis regardless of whether the subject is on new treatment or not<br>Discontinuation (either premature discontinuation or completion of protocol-specified treatment duration) of study drug<br>• Treatment Policy Strategy: this intercurrent event is irrelevant; morphologic RFS data will be used in the analysis regardless of whether the subject is on study drug or not | Hazard ratio |

| | Attributes of the Estimand | | | |
|---|---|---|---|---|
| **Endpoint** | **Treatment** | **Population** | **Handling of Intercurrent Events** | **Population-level Summary** |
| Composite RFS | Best supportive care with and without venetoclax plus azacitidine | All randomized subjects | Initiation of new anti-leukemic therapy<br>• Treatment Policy Strategy: this intercurrent event is irrelevant; composite RFS data will be used in the analysis regardless of whether the subject is on new treatment or not<br>Discontinuation (either premature discontinuation or completion of protocol-specified treatment duration) of study drug<br>• Treatment Policy Strategy: this intercurrent event is irrelevant; composite RFS data will be used in the analysis regardless of whether the subject is on study drug or not | Hazard ratio |
| GRFS | Best supportive care with and without venetoclax plus azacitidine | All randomized subjects | Initiation of new anti-leukemic therapy<br>• Treatment Policy Strategy: this intercurrent event is irrelevant; GRFS data will be used in the analysis regardless of whether the subject is on new treatment or not<br>Discontinuation (either premature discontinuation or completion of protocol-specified treatment duration) of study drug<br>• Treatment Policy Strategy: this intercurrent event is irrelevant; GRFS data will be used in the analysis regardless of whether the subject is on study drug or not | Hazard ratio |

| | Attributes of the Estimand | | | |
|---|---|---|---|---|
| **Endpoint** | **Treatment** | **Population** | **Handling of Intercurrent Events** | **Population-level Summary** |
| Time to deterioration in QoL | Best supportive care with and without venetoclax plus azacitidine | All randomized adult subjects (≥ 18 years old) | Initiation of new anti-leukemic therapy<br>• While On Treatment Strategy: data collected after the initiation of new anti-leukemic therapy will not be included in the analysis<br>Discontinuation (either premature discontinuation or completion of protocol-specified treatment duration) of study drug<br>• Treatment Policy Strategy: this intercurrent event is irrelevant; data will be used in the analysis regardless of whether the subject is on study drug or not | Hazard ratio |
| Subjects without higher grade GvHD | Best supportive care with and without venetoclax plus azacitidine | All randomized subjects | Initiation of new anti-leukemic therapy<br>• While On Treatment Strategy: data after the initiation of new anti-leukemic therapy will not be used in the analysis<br>Discontinuation (either premature discontinuation or completion of protocol-specified treatment duration) of study drug<br>• Treatment Policy Strategy: this intercurrent event is irrelevant; data will be used in the analysis regardless of whether the subject is on study drug or not | Difference in proportions of subjects without higher grade GvHD |

| | Attributes of the Estimand | | | |
|---|---|---|---|---|
| **Endpoint** | **Treatment** | **Population** | **Handling of Intercurrent Events** | **Population-level Summary** |
| Physical functioning | Best supportive care with and without venetoclax plus azacitidine | All randomized adult subjects (≥ 18 years old) | Initiation of new anti-leukemic therapy<br>• While On Treatment Strategy: data after the initiation of new anti-leukemic therapy will not be used in the analysis<br>Discontinuation (either premature discontinuation or completion of protocol-specified treatment duration) of study drug<br>• Treatment Policy Strategy: this intercurrent event is irrelevant; data will be used in the analysis regardless of whether the subject is on study drug or not | Difference in change from baseline in physical functioning score |
| PROMIS Cancer Fatigue SF 7a | Best supportive care with and without venetoclax plus azacitidine | All randomized adult subjects (≥ 18 years old) | Initiation of new anti-leukemic therapy<br>• While On Treatment Strategy: data after the initiation of new anti-leukemic therapy will not be used in the analysis<br>Discontinuation (either premature discontinuation or completion of protocol-specified treatment duration) of study drug<br>• Treatment Policy Strategy: this intercurrent event is irrelevant; data will be used in the analysis regardless of whether the subject is on study drug or not | Difference in change from baseline in fatigue score |

| | Attributes of the Estimand | | | |
|---|---|---|---|---|
| **Endpoint** | **Treatment** | **Population** | **Handling of Intercurrent Events** | **Population-level Summary** |
| MRD Conversion | Best supportive care with and without venetoclax plus azacitidine | All randomized subjects with MRD $\geq 10^{-3}$ at baseline | Initiation of new anti-leukemic therapy<br>• While On Treatment Strategy: data after the initiation of new anti-leukemic therapy will not be used in the analysis<br>Discontinuation (either premature discontinuation or completion of protocol-specified treatment duration) of study drug<br>• While On Treatment Strategy: data collected after discontinuation of study drug will not be used in the analysis | Difference in proportions of subjects who convert to MRD $< 10^{-3}$ |

## IRC-Assessed Morphologic RFS

IRC-assessed morphologic RFS is defined as time from randomization to morphologic relapse as assessed by IRC or death from any cause, whichever occurs earlier. Subjects without any morphologic RFS event will be censored at the date of the last disease assessment on or prior to the cutoff date. The date of the last assessment will be determined by the date of the last bone marrow assessment or complete blood count (CBC) assessment (white blood cell count, platelet count, neutrophil count, hemoglobin, blasts), whichever occurs later. Subjects without any morphologic RFS event and without any disease assessment after randomization will be censored at the date of randomization.

The analysis of IRC-assessed morphologic RFS will be done according to the statistical method for time-to-event endpoints and the strata described in Section 8.1. All subjects in the ITT population will be included in the analysis.

The following supplementary analyses of morphologic RFS will be performed:

- Analysis of morphologic RFS by treating new anti-leukemic therapy as events.
- Analysis of morphologic RFS using investigator assessment of morphologic relapse. A cross tabulation of investigator and IRC assessments will be provided.

**IRC-Assessed Composite RFS**

IRC-assessed composite RFS is defined as time from randomization to morphologic relapse as assessed by IRC, non-morphologic relapse as assessed by IRC, or death from any cause, whichever occurs earlier. Subjects without any composite RFS event will be censored at the date of the last disease assessment on or prior to the cutoff date. The date of the last assessment will be determined by the date of the last bone marrow assessment or complete blood count (CBC) assessment (white blood cell count, platelet count, neutrophil count, hemoglobin, blasts), cytogenetic assessment, molecular marker test, and MRD assessment, whichever occurs later. Subjects without any composite RFS event and without any disease assessment after randomization will be censored at the date of randomization.

The analysis of IRC-assessed composite RFS will be done according to the statistical method for time-to-event endpoints and the strata described in Section 8.1. All subjects in the ITT population will be included in the analysis.

The following supplementary analyses of composite RFS will be performed:

- Analysis of composite RFS by treating new anti-leukemic therapy as events.
- Analysis of composite RFS using investigator assessment of morphologic and non-morphologic relapse. A cross tabulation of investigator and IRC assessments will be provided.

**GvHD-Free, Relapse-Free Survival (GRFS)**

GRFS is defined as the time from randomization to disease relapse (morphologic or non-morphologic) as assessed by IRC, incidence of GvHD, or death from any cause. Incidence of GvHD is defined as grade 3 or 4 acute GvHD, or chronic GvHD requiring immunosuppressive therapy. All subjects in the ITT population will be included in the analysis of GRFS. A supplementary analysis of GRFS will be performed based on disease relapse as assessed by investigator.

Subjects without any of the specified events will be censored at the censoring date of composite RFS or the last clinical GvHD assessment, whichever is earlier. Subjects without any of the specified events and without any disease assessment or any clinical GvHD assessment after randomization will be censored at the date of randomization.

The analysis of GRFS will be done according to the convention described in Section 8.1 for time-to-event endpoints with the exception that the strata for the analysis will be defined as follows:

- Stratum 1: age < 18
- Stratum 2: age ≥ 18, previous treatment with either venetoclax or azacitidine (yes), high risk of AML recurrence (yes), type of donor (matched related)
- Stratum 3: age ≥ 18, previous treatment with either venetoclax or azacitidine (yes), high risk of AML recurrence (yes), type of donor (other)
- Stratum 4: age ≥ 18, previous treatment with either venetoclax or azacitidine (yes), high risk of AML recurrence (no), type of donor (matched related)
- Stratum 5: age ≥ 18, previous treatment with either venetoclax or azacitidine (yes), high risk of AML recurrence (no), type of donor (other)
- Stratum 6: age ≥ 18, previous treatment with either venetoclax or azacitidine (no), high risk of AML recurrence (yes), type of donor (matched related)
- Stratum 7: age ≥ 18, previous treatment with either venetoclax or azacitidine (no), high risk of AML recurrence (yes), type of donor (other)

- Stratum 8: age ≥ 18, previous treatment with either venetoclax or azacitidine (no), high risk of AML recurrence (no), type of donor (matched related)
- Stratum 9: age ≥ 18, previous treatment with either venetoclax or azacitidine (no), high risk of AML recurrence (no), type of donor (other)

**Rate of Subjects Without Higher Grade of GvHD**

The rate of subjects without higher grade GvHD at 90 days after randomization (Arm B) or initiation of treatment (Arm A) will be compared between the treatment arms using Cochran-Mantel-Haenszel test stratified by type of donor (matched related, other). All subjects in the ITT population will be included in the analysis. Subjects without GvHD assessment at 90 days will be included in the denominator in the analysis. The 95% confidence interval based on the binomial distribution (Clopper-Pearson exact method) by treatment arms will be provided. Higher grade GvHD is defined as Grade 2 or higher acute GvHD, or moderate or severe chronic GvHD. Data after subjects start new anti-leukemic therapy will not be included in the analysis.

Rate of subjects without higher grade of GvHD at 90 days after randomization (Arm B) or initiation of treatment (Arm A) will be compared between the treatment arms using Cochran-Mantel-Haenszel test stratified by type of donor (matched related, other). The 95% confidence interval for GvHD rate based on the binomial distribution (Clopper-Pearson exact method) by treatment arms will be provided.

**Physical Functioning**

Physical functioning will be assessed using the physical functioning score as assessed by EORTC QLQ-C30 in adult subjects. Scores will be computed according to the EORTC QLQ-C30 scoring manual. Data after initiation of new anti-leukemic therapy will be excluded from the analysis. A linear mixed effects regression model will be used to fit the longitudinal data. In the linear mixed effects regression model, previous treatment with either venetoclax or azacitidine (Yes, No), high risk of recurrence (Yes, No), baseline score, time, treatment arm and treatment arm by time interaction will be included as fixed

factors. The following covariance structures in the order of: unstructured, compound symmetry and first-order autoregressive, will be used until the model converges. Change from baseline scores after 6 months of treatment (Arm A) or randomization (Arm B) will be compared between the treatment arms within the framework of the linear mixed effects model. The least-squares mean change from baseline and the corresponding 95% confidence interval will be provided for each treatment arm. In addition, the least-squares mean of the difference between the treatment arms and the corresponding 95% confidence interval will be provided.

**PROMIS Cancer Fatigue SF 7a**

PROMIS is a system of highly reliable, precise measures of patient-reported health status for physical, mental, and social well-being.[4] PROMIS instruments measure concepts such as pain, fatigue, physical function, depression, anxiety and social function. Fatigue will be assessed using the PROMIS Fatigue SF 7a, a 7-item questionnaire that assesses the impact and experience of fatigue over the past 7 days. All questions employ the following five response options: 1 = Never, 2 = Rarely, 3 = Sometimes, 4 = Often, and 5 = Always. For Question 7, scores will be reversed. The recommended minimum important difference range is 3 to 5 points.[5]

Patient reported outcome on fatigue will be assessed using the total fatigue score as measured by PROMIS Cancer Fatigue SF 7a. Scores will be computed according to the PROMIS Cancer Fatigue SF 7a scoring manual. Data after initiation of new anti-leukemic therapy will be excluded from the analysis. A linear mixed effects regression model will be used to fit the longitudinal data. In the linear mixed effects regression model, previous treatment with either venetoclax or azacitidine (Yes, No), high risk of recurrence (Yes, No), baseline score, time, treatment arm and treatment arm by time interaction will be included as fixed factors. The following covariance structures in the order of: unstructured, compound symmetry and first-order autoregressive, will be used until the model converges. Change from baseline scores after 6 months of treatment (Arm A) or randomization (Arm B) will be compared between the treatment arms within the framework of the linear mixed effects model. The least-squares mean change from

baseline and the corresponding 95% confidence interval will be provided for each treatment arm. In addition, the least-squares mean of the difference between the treatment arms and the corresponding 95% confidence interval will be provided.

**MRD Conversion Rate**

MRD conversion rate is defined as percentage of subjects with MRD $\geq 10^{-3}$ at randomization who convert to MRD $< 10^{-3}$ after randomization (Arm B) or initiation of treatment (Arm B). Subjects whose MRD is $< 10^{-3}$ at randomization will be excluded from the analysis. Subjects who have MRD $\geq 10^{-3}$ at randomization but do not have any MRD assessment after randomization will be assumed to be MRD positive. For subjects in Arm A, the analysis window is defined as the period from the first dose of study drug to the last dose of study drug + 7 days, disease relapse (morphologic or non-morphologic) as assessed by IRC, or start date of new anti-leukemic therapy, whichever is earlier. Subjects in Arm A who have MRD $\geq 10^{-3}$ at baseline and do not receive any study drug will be treated as non-responders. For subjects in Arm B, the analysis window is defined as the period from the date of randomization to the End Visit, disease relapse as assessed by IRC, discontinuation from study, or start date of new anti-leukemic therapy, whichever is earlier.

The MRD response rates will be compared between treatment arms using Fisher's exact test. The 95% confidence interval for MRD conversion rate based on the binomial distribution (Clopper-Pearson exact method) by treatment arms will be provided.

**Time to Deterioration in GHS/QoL**

The EORTC-QLQ-C30 is a 30-item patient-reported questionnaire composed of both multi item and single scales including 5 functional scales (physical, role, emotional, social, and cognitive), 3 symptom scales (fatigue, nausea and vomiting, and pain), a global health status/QoL scale, and 6 single items (dyspnea, insomnia, appetite loss, constipation, diarrhea, and financial difficulties).[3]

Time to deterioration in GHS/QoL is defined as the time from randomization to death from any cause or the first-time deterioration of ≥ 10 points in GHS/QoL score as measured by EORTC QLQ-C30, on or prior to the initiation of new anti-leukemic therapy. All adult subjects in the ITT population will be included in the analysis.

Subjects without any of the specified events will be censored at the date of the last GHS/QoL assessment on or prior to the initiation of new anti-leukemic therapy. Subjects without any of the specified events and without any post-baseline GHS/QoL assessment will be censored at the date of randomization.

The analysis of time to deterioration in GHS/QoL will be done according to the convention described in Section 8.1 for time-to-event endpoints.

### 8.4.2 Supportive Secondary Efficacy Analyses

**PROMIS Cancer Fatigue SF 7a**

Change from baseline scores at time points other than 6 months after treatment (Arm A) or randomization (Arm B) will be analyzed using linear mixed effects model described in Section 8.4.1. Data after initiation of new anti-leukemic therapy will be excluded from the analysis.

**EORTC QLQ-C30**

Analyses will be performed on the subscales/items from the EORTC QLQ-C30. Scores will be calculated as per the scoring manuals. Linear mixed effects regression models similar to that described for the analysis of PROMIS Cancer Fatigue SF 7a will be used to compare treatment arms. Subjects without baseline score or any post-baseline score will be excluded from the linear mixed effects regression analysis. Data after initiation of new anti-leukemic therapy will be excluded from the analysis.

**EQ-5D-5L**

The EQ-5D-5L is a generic preference instrument that has been validated in numerous populations.[6] The EQ-5D-5L has 5 dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. These dimensions are measured on a 5-point scale: no problems, slight problems, moderate problems, severe problems, and extreme problems. The scores for the 5 dimensions are used to compute a single utility index score ranging from zero to 1 representing the general health status of the individual. The EQ-5D-5L also contains a visual analog scale (VAS) to assess the subject's overall health. The minimum important difference for the EQ-5D-5L utility index score in cancer patients is 0.08, and the minimum important difference for EQ-5D-5L VAS is 7.[7]

Longitudinal analyses will be performed for EQ-5D-5L. Scores will be calculated as per the scoring manuals. Linear mixed effects regression models similar to that described for the analysis of PROMIS Cancer Fatigue SF 7a will be used to compare treatment arms. Subjects without baseline score or any post-baseline score will be excluded from the linear mixed effects regression analysis. Data after initiation of new anti-leukemic therapy will be excluded from the analysis.

## 8.5 Additional Efficacy Analyses

**Pediatric Quality of Life Inventory (PedsQL)**

Pediatric patients' health-related and cancer specific QoL will be assessed using the Pediatric Quality of Life Inventory (PedsQL) and PedsQL Cancer Module. The PedsQL is a 23-item patient self-reported multidimensional questionnaire composed of 4 functioning domains (physical, emotional, social and school). The PedsQL Cancer Module is a 27-item patient self-reported questionnaire composed of 8 domains (pain and hurt, nausea, procedural anxiety, treatment anxiety, worry, cognitive problems, perceived physical appearance, and communication). For both the PedsQL and PedsQL Cancer Module, subject rate their problem with functioning and cancer specific QoL, respectively, in the past one month on a 5-point Likert scale ranging from 0 to 4

(0 = Never, 1 = Almost Never, 2 = Sometimes, 3 = Often, and 4 = Almost always). Scores will be reversed and are transformed on a scale from 0 to 100. Higher scores indicate better QoL from the PedsQL and lower problems in the PedsQL Cancer Module.

Exploratory analyses will be performed on PedsQL and PedsQL cancer module. Scores will be calculated as per the scoring manuals. Linear mixed effects regression models similar to that described for the analysis of PROMIS Cancer Fatigue SF 7a may be used to compare treatment arms. If the sample size of the pediatric patients is too small to fit a linear mixed effects model, descriptive statistics (such as mean and standard deviation) may be used to summarize the scores. Data after initiation of new anti-leukemic therapy will be excluded from the analysis.

**Chronic GvHD PRO**

Chronic GvHD (cGvHD) related symptoms and overall patient ratings will be assessed following form B of Lee 2015 for chronic GvHD.[8] Subjects rate their symptoms on an 11-point scale. The 11-point scale is measured from "0" (not present) to "10" (as bad as you can imagine) when symptoms are at their worst. Subjects rate their overall cGvHD as 1-mild, 2-moderate or 3-severe. Subjects rate their overall cGvHD severity on an 11-point scale as 0 (cGvHD symptoms not severe at all) through 10 (most severe cGvHD symptoms possible). Subjects rate change in cGvHD over the past month using a 7-point scale as +3 (very much better) through –3 (very much worse).

Overall cGvHD severity rating and change in cGvHD (7-point scale) will be summarized over time.

The analysis of cGvHD PRO will only include subjects with cGvHD PRO data. Data after initiation of new anti-leukemic therapy will be excluded from the analysis.

**Health Care Resource Utilization**

Number and percentage of subjects admitted to hospital during the treatment period (for Am B: time from randomization to End Visit, morphologic or non-morphologic relapse

assessed by investigator, or study discontinuation, whichever is the earliest) and duration of hospitalization will be summarized descriptively for each treatment arm.

**Immunosuppressive Therapy**

Immunosuppressive therapies used for GvHD prophylaxis or treatment will be summarized by generic name. The number and percentage of subjects who take immunosuppressive therapies for GvHD will be summarized by generic name for each treatment arm. The number of therapies taken by each subject and the total duration of the therapies will be descriptively summarized for each treatment arm.

The following tabulations of GvHD severity (overall clinical grading for aGvHD and severity score for cGvHD) will be provided to explore the relationship between the use of immunosuppressive therapy and GvHD severity:

- GvHD severity at initiation of the first immunosuppressive therapy and at discontinuation of the first immunosuppressive therapy
- GvHD severity at initiation of the first immunosuppressive therapy and the lowest GvHD severity on or prior to discontinuation of the first immunosuppressive therapy
- GvHD severity at initiation of the first immunosuppressive therapy and at discontinuation of the last immunosuppressive therapy
- GvHD severity at initiation of the first immunosuppressive therapy and the lowest GvHD severity on or prior to the discontinuation of the last immunosuppressive therapy.

**Infectious-Related Morbidity and Mortality**

Number and percentage of death due to infection will be summarized descriptively for each treatment arm. Data of infecting micro-organism, treatment, and the outcome of the treated infections, if available, will be summarized.

**Rate of AML Recurrence**

Rate of AML recurrence at the time of data cutoff will be estimated using the Kaplan-Meier methodology based on the definition of composite RFS. The definition and analysis of composite RFS are described in Section 8.4.1.

## 8.6 Efficacy Subgroup Analyses

To evaluate of the impact of demographics and baseline characteristics on efficacy, subgroup analyses may be performed on efficacy endpoints including, but not limited to, RFS, OS, and GRFS and will be performed for subgroups including, but not limited to, those defined below:

- Age (< 18, ≥ 18, ≥ 18 - < 65, ≥ 65 years)
- Gender (Male, Female)
- Race (White, Black or African American, Asian, Other)
- Ethnicity (Hispanic or Latino, Not Hispanic or Latino)
- Previous treatment with venetoclax (Yes, No)
- Previous treatment with azacitidine (Yes, No)
- MRD prior to transplant (≥ $10^{-3}$, < $10^{-3}$)
- Remission status prior to transplant (CR1, CR2 or later, non-CR)
- MRD after transplant at time of randomization (≥ $10^{-3}$, < $10^{-3}$)
- Baseline remission status (CR, CRi)
- Conditioning regimen (Reduced intensity, Myeloablative)
- Type of donor (Matched related, Other)
- Cytogenetic risk at initial diagnosis (Favorable, Intermediate, Poor)
- ELN risk category at initial diagnosis (Favorable, Intermediate, Poor)
- Geographic regions (US, Europe, China, Japan, Rest of World)
- Dose modifications during Cycles 1 or 2 (Yes, No)
- Acute GvHD at baseline (Grade 0, Grade 1, Grade 2, Grade 3, Grade 4)

- Other influencing factors of GvHD including stem cell graft source (bone marrow, cord blood stem cell, peripheral blood stem cell), HLA match (10/10 match, other matches) and use of ex-vivo T-cell depletion (yes, no)

For subgroup analyses of categorical variables, the difference in the proportions and the corresponding 95% confidence interval (exact unconditional confidence limits) for the difference will also be provided. For subgroup analyses of time-to-event variables, the hazard ratio and the 95% confidence interval based on unstratified Cox proportional hazards model will be provided.

# 9.0 Safety Analyses

## 9.1 General Considerations

Safety analyses will only include subjects in the Safety Analysis Set as defined in Section 4.0. Safety summaries will be presented separately for Part 1 and Part 2 by treatment arm based on the assigned treatment arm. There will be no statistical comparison for any safety analysis between treatment arms. Unless otherwise specified, data after the cutoff date will be excluded from statistical analyses.

## 9.2 Adverse Events

Adverse events (AEs) will be summarized and presented using primary MedDRA System Organ Classes (SOCs) and preferred terms (PTs) according to the version of the MedDRA coding dictionary used for the study at the time of database lock. The actual version of the MedDRA coding dictionary used will be noted in the AE tables and in the clinical study report. Specific adverse events will be counted once for each subject for calculating percentages, unless stated otherwise. In addition, if the same adverse event occurs multiple times within a subject, the highest severity and level of relationship to investigational product will be reported.

### 9.2.1 Treatment-Emergent Adverse Events

For Part 1 and Part 2 Arm A, treatment-emergent AEs are defined as any AE with the onset that is after the first dose of study drug until 30 days after the last dose of the study drug. Events where the onset date is the same as the study drug start date are assumed to be treatment-emergent. If an incomplete or missing onset date was collected for an AE, the AE will be assumed to be treatment-emergent unless there is evidence that confirms that the AE was not treatment-emergent (e.g., the AE end date was prior to the date of the first dose of study drug). For Part 2 Arm B, treatment-emergent AEs are defined as AEs with an onset date on or after the date of randomization until 30 days after the End Visit, disease relapse (morphologic or non-morphologic relapse assessed by investigator) or study discontinuation, whichever is the earliest.

All treatment-emergent AEs will be summarized overall, as well as by primary MedDRA SOC and Preferred Term. The SOCs will be presented in alphabetical order, and the PTs will be presented in descending frequency order for the active treatment arm within each SOC. For venetoclax + azacitidine arms, treatment-emergent AEs (including SAEs and deaths) will be summarized separately for the venetoclax + azacitidine combination treatment period and the venetoclax monotherapy treatment period as well as for the entire treatment period.

### 9.2.2 Adverse Event Overview

An overview of AEs will be presented consisting of the number and percentage of subjects experiencing at least one event for each of the following AE categories:

- Any treatment-emergent AE
- Any treatment-emergent AE with NCI toxicity grade $\geq 3$
- Any treatment-emergent AE with NCI toxicity grade 3 or 4
- Any treatment-emergent AE with reasonable possibility related to venetoclax as assessed by the investigator (for Part 1 and Part 2 Arm A only)
- Any treatment-emergent AE with reasonable possibility related to azacitidine as assessed by the investigator (for Part 1 and Part 2 Arm A only)

- Any treatment-emergent serious AE
- Any treatment-emergent AE leading to discontinuation of venetoclax (for Part 1 and Part 2 Arm A only)
- Any treatment-emergent AE leading to dose interruption of venetoclax (for Part 1 and Part 2 Arm A only)
- Any treatment-emergent AE leading to dose reduction of venetoclax (for Part 1 and Part 2 Arm A only)
- Any treatment-emergent AE leading to discontinuation of azacitidine (for Part 1 and Part 2 Arm A only)
- Any treatment-emergent AE leading to dose interruption of azacitidine (for Part 1 and Part 2 Arm A only)
- Any treatment-emergent AE leading to dose reduction of azacitidine (for Part 1 and Part 2 Arm A only)
- Any treatment-emergent AE leading to death
- All deaths

### 9.2.3 Treatment-Emergent Adverse Events by SOC and/or PT

Treatment-emergent adverse events will be summarized by SOC and PT. Specific adverse events will be counted once for each subject for calculating percentages, unless stated otherwise. In addition, if the same adverse event occurs multiple times within a subject, the highest severity and level of relationship to investigational product will be reported. Treatment-emergent adverse events will be summarized by PT and sorted by decreasing frequency for the total active group.

### 9.2.4 All Deaths, SAEs and Adverse Events Leading to Study Drug Discontinuation

SAEs and AEs leading to study drug discontinuation will be summarized by SOC and PT and in listing format.

The number of subject deaths and causes of death will be summarized (1) for all deaths, (2) for deaths occurring within 30 days of the first dose of study drug (or the

randomization for subjects in Arm B), (3) for deaths occurring within 60 days of the first dose of study drug (or the randomization for subjects in Arm B), (4) for deaths occurring within 30 days after the last dose of study drug (or the Final Visit: End Visit, disease relapse assessed by investigator, or study discontinuation, whichever is the earlier, for subjects in Arm B), and (5) for deaths occurring more than 30 days after the last dose of study drug (or the Final Visit: End Visit, disease relapse assessed by investigator, or study discontinuation, whichever is the earlier, for subjects in Arm B).

### 9.2.5 Adverse Events of Special Interest

Adverse events of special interest will be summarized by SOC and PT and will be based on standardized or company MedDRA queries (SMQs or CMQs). Adverse events of special interest are categorized as follows:

- Tumor Lysis Syndrome (TLS).

Details on the search criteria for TLS are provided in Appendix B.

In addition, treatment-emergent AEs and serious AEs will be summarized for selected grouped preferred terms (PTs) described in Appendix B.

## 9.3 Analysis of Laboratory Data

Data collected from central and local laboratories, including additional laboratory testing due to an SAE, will be used in all analyses. The clinical laboratory tests defined in the protocol operations manual (e.g., hematology and clinical chemistry) will be summarized. Change from baseline will be summarized and presented for each lab test at scheduled post-baseline visits. For subjects enrolled in Part 1 or Part 2 Arm A, baseline is defined as the last non-missing observation before the first administration of study drug or randomization if no study drug is given. For subjects enrolled in Part 2 Arm B, baseline is defined as the last non-missing observation on or prior to the Cycle 1 Day 1 visit or randomization if the Cycle 1 Day 1 visit is not performed.

For shifts relative to National Cancer Institute Common Toxicity Criteria for Adverse Events, baseline and post-baseline laboratory observations will be categorized as grade 0, grade 1, grade 2, grade 3, or grade 4. The baseline grade is defined as the grade of the last measurement collected on or prior to the first dose of study drug (or randomization for non-treated subjects). In cases where multiple values are collected on the same day, the maximum grade value will be selected as the value for that day for post-baseline, the minimum grade value will be selected as the value for that day for baseline. If a subject had missing baseline and non-missing post-baseline for a given lab, the baseline grade will be assumed to be grade 0. The maximum NCI toxicity grade value is the value with highest NCI toxicity grade collected after the first dose of study drug.

For each variable, shift tables will be generated that cross tabulate the number of subjects with baseline values of grade 0, grade 1, grade 2, grade 3, or grade 4 versus maximum post-baseline values of grade 0, grade 1, grade 2, grade 3, or grade 4. Subjects with missing measurements will be counted in the "missing" category. All treated subjects in Part 1 and Part 2 Arm A and all randomized subjects in Part 2 Arm B will be included in the cross tabulation regardless whether baseline or post-baseline measurements are collected.

The separate laboratory shifts tables based on the two criteria below will be generated for each laboratory tests related to CTCAE:

1. Shifts from Grade 0 (Normal) at baseline to grade 1 - 4 post-baseline (maximum) and worsening from an abnormal baseline value of at least one grade up post-baseline (maximum)
2. Shifts from Grade 0 - 2 at baseline to grade 3 or 4 post-baseline (maximum) and from grade 3 at baseline value to Grade 4 post-baseline (maximum).

The analyses of laboratory data will include post-baseline measurements within 30 days after the last dose of the study drug for subjects in Part 1 or Part 2 Arm A or within 30 days after the End Visit, disease relapse assessed by investigator, or study

discontinuation, whichever is earlier, for subjects in Arm B. Detailed listings of all data for subjects experiencing NCI CTCAE grade 3 or 4 blood chemistry and hematology values will be provided.

Number and percentage of subjects with liver enzyme value meeting the criteria for potential drug-induced liver injury (ALT > 3 × ULN or AST > 3 × ULN and Bilirubin > 2 × ULN within 72 hours of each other) will be presented.

Number and percentage of subjects meeting the Howard criteria for Laboratory TLS will be presented. Laboratory TLS requires two or more metabolic abnormalities must be present during the same 24-hour period. The evaluation period for TLS is from the first dose of study drug (or date of randomization for subjects in Arm B) until 7 days after the first dose of study drug (or date of randomization for subjects in Arm B).

- uric acid > 476 mcmol/l,
- potassium > 6 mmol/l,
- inorganic phosphorus > 1.5 mmol/l,
- calcium < 1.75 mmol/l.

**Conversion of Leukocyte Differential Count from (%) to Absolute Count**

The following conversion process will be performed, if a lab test only provides a percent (%) value for leukocyte differential count without an absolute count. The converted absolute counts will be used in all efficacy and safety analyses.

1. The conversion will only be performed when the reported leukocyte differential count and the total leukocyte count have the same collection date and the same collection time.
2. The converted absolute count ($\times 10^9$/L) = Total leukocyte count ($\times 10^9$/L) × leukocyte differential count (%)/100.

The following normal ranges will be used to calculate the normal ranges for the converted absolute neutrophil count (ANC) and the converted absolute lymphocyte count (ALC).

| Lab Test | Normal Ranges (SI units)[11] |
|---|---|
| Total leukocyte count ($\times 10^9$/L) | 4.5 – 11 |
| Neutrophil differential count (%) | 40 – 70 |
| Lymphocyte differential count (%) | 22 – 44 |

Based on the conversion formula, the lower limit of normal (LLN) for the converted ANC is $1.8 \times 10^9$/L and the LLN for the converted ALC is $0.99 \times 10^9$/L. These LLN values will be used in the safety analyses of the converted ANC and ALC.

## 9.4 Analysis of Vital Signs

Change from baseline will be summarized and presented for each scheduled post-baseline visit for the vital sign parameters. For subjects enrolled in Part 1 or Part 2 Arm A, baseline is defined as the last non-missing observation before the first administration of study drug or randomization if no study drug is given. For subjects enrolled in Arm B, baseline is defined as the last non-missing observation on or prior to the Cycle 1 Day 1 visit or randomization if the Cycle 1 Day 1 visit is not performed.

## 9.5 Safety Subgroup Analyses

Safety analyses described in Section 9.2, Section 9.3 and Section 9.4 may be summarized for the subgroups including, but not limited to, those defined below:

- Age (< 18, ≥ 18, ≥ 18 - < 65, ≥ 65 years)
- Gender (Male, Female)
- Race (White, Black or African American, Asian, Other)
- Ethnicity (Hispanic or Latino, Not Hispanic or Latino)
- Region (US, Europe, China, Japan, Rest of World)
- Hepatic impairment at baseline (Yes, No)
- Renal impairment at baseline (Yes, No)

# 10.0 Interim Analyses

Interim review of safety data from the ongoing study (Part 2) at the following proposed time points will be conducted:

- The first analysis to review the safety data will occur when 20 subjects have been on study drug or best supportive care for 3 cycles in Part 2 of the study
- The subsequent reviews of safety data will occur approximately every 6 months after the first review of safety data.

In addition, an efficacy interim analysis will be performed once 173 OS events as assessed by IRC (75% of the total 231 events) are observed. The Lan-DeMets alpha spending function with O'Brien-Fleming boundary will be used at the interim analysis to ensure that the false positive rate for OS the primary efficacy endpoint is controlled at the two-sided 0.05 level. The planned stopping boundaries for OS at each interim and final analysis are described in Table 3 below.

**Table 3. Planned Stopping Boundaries at the Interim Analysis and Final Analysis of OS**

| | | Stopping Boundaries |
|---|---|---|
| **Analysis** | **# of Events** | **P-Value (two-sided)** |
| IA (75%) | 173 | 0.0192 |
| FA | 231 | 0.0443 |

FA = final analysis; IA = interim analysis

The actual stopping boundaries at IA and FA of OS will be derived using Lan-DeMets alpha spending function based on the observed number of OS events in the extracted database.

At the 75% OS efficacy interim, the probability of declaring success is 70% assuming a hazard ratio of 0.65 and is 85% assuming a hazard ratio of 0.6.

### 10.1 Data Monitoring Committee

The interim review of safety data and the interim efficacy analyses will be performed by an independent Statistical Data Analysis Center. An independent data monitoring committee (IDMC) composed of persons independent of AbbVie and with relevant expertise in their field will review the above interim analyses.

A separate IDMC charter will be prepared outside of the protocol and will describe the roles and responsibilities of the IDMC members, frequency of data reviews, and relevant safety data to be assessed.

The IDMC will make a recommendation based on OS to either declare efficacy at the interim analysis or continue the study as planned. The planned stopping boundaries are described in Table 3 for OS. The actual stopping boundaries will be determined based on the actual number of observed OS events in the extracted database. At OS IA, if the IDMC informs the Sponsor that the OS data is statistically significant and in favor of the venetoclax arm, upon notification from the Sponsor's Internal Executive Review Committee and after consultation with regulatory authorities as required to ensure maturity of data for registration, the Sponsor will then prepare regulatory submissions globally based on the available data. If the results of the OS analysis are not statistically significant in favor of the venetoclax arm at OS IA, the study will continue to follow protocol-specified procedures until the final analysis.

Details of the IDMC review will be presented in the IDMC charter.

## 11.0 Overall Type-I Error Control

The fixed sequence testing procedure will be performed with a significance level of 0.05 (two-sided) for the primary efficacy endpoint and key secondary efficacy endpoints sequentially. The ranking and alpha spending of the key secondary efficacy endpoints are described in Table 4. If the statistical test is not significant for the primary efficacy endpoint, then statistical significance will not to be declared for any of the key secondary endpoints. The Lan-DeMets alpha spending function with O'Brien Fleming boundary will

be used at the efficacy interim analysis to ensure that the false positive rate for OS is controlled at the two-sided 0.05 level. The detailed description of alpha spending for OS is described in Table 4.

**Table 4. Testing Sequence and Alpha-Spending Boundaries (Two-Sided P-Value) for Primary and Key Secondary Endpoints**

| Testing Sequence | Endpoint | Timing of Analysis | |
|---|---|---|---|
| | | OS IA | OS FA |
| 1 | OS | As specified in Table 3 | As specified in Table 3 |
| 2 | Morphologic RFS | As specified in Table 3 | As specified in Table 3[a] |
| 3 | Composite RFS | As specified in Table 3 | As specified in Table 3[a] |
| 4 | GRFS[a] | As specified in Table 3 | As specified in Table 3[a] |
| 5 | Rate of subjects without higher grade GvHD[b] | 0.05 | NA |
| 6 | MRD[b] | 0.05 | NA |
| 7 | Physical functioning[b] | 0.05 | NA |
| 8 | PROMIS Cancer Fatigue SF 7a[b] | 0.05 | NA |
| 9 | Time to deterioration in QoL[b] | 0.05 | NA |

RFS = relapse-free survival; FA = final analysis; IA = interim analysis; NA = not applicable; OS = overall survival; OBF = O'Brien-Fleming

a. The actual stopping boundary will be recalibrated based on the actual number of observed events.

b. Will be tested one time using the data at 75% OS interim (IA).

# 12.0 Version History

**Table 5. SAP Version History Summary**

| Version | Date | Summary |
|---|---|---|
| 1.0 Draft | 29 Aug 2019 | Original version |
| 1.0 Draft | 15 Jan 2020 | The following changes were made in response to FDA statistical comments<br>a. Changed the secondary endpoint "GvHD rate" to "rate of subjects without higher grade GvHD." The definition and analysis of the endpoint were updated in Section 3.2 and Section 8.4.1.<br>b. Added two additional sensitivity analyses for RFS in Section 8.3.4.<br>The following changes were made based on protocol Amendment 2<br>a. Added the exploratory endpoint "recurrence rate in AML" to Section 3.3. The analysis of the endpoint was added to Section 8.5.<br>b. The alpha spending method for OS was changed to O'Brien-Fleming method in Table 3 in Section 11.0.<br>c. Corrected a typo in the analysis description of MRD conversion rate in Section 8.4.1.<br>d. Description of study design of Part 1 was added to Section 2.0 and Appendix D. |
| 1.0 | 27 Mar 2020 | The following changes were made<br>a. Removed ECOG from baseline summary. Data are not collected according to the protocol.<br>b. Removed molecular marker from baseline summary. Data will not be available from central lab. Availability of data from local lab at the time of original diagnosis is uncertain.<br>c. Removed the Unknown category from subgroup analysis.<br>d. Added calculation of albumin-corrected calcium.<br>e. Reformat the presentation of stopping boundaries for interim analyses in Table 2.<br>f. Corrected the text descriptions of the alpha spending function for OS in Section 11.0.<br>g. Other editorial edits throughout the document.<br>This version of SAP includes contents that address FDA statistical comments but are not described in the current protocol Amendment 2. These contents will be implemented in a future protocol amendment. |

| Version | Date | Summary |
|---|---|---|
| 2.0 Draft | 28 Feb 2022 | The following changes were made based on protocol amendments 3, 4 and 5<br>a. Added additional (unranked) secondary endpoints and updated exploratory endpoints to align with the protocol.<br>b. Added additional summaries of baseline characteristics based on the new subgroups defined in the protocol.<br>c. Added the definitions of estimands for primary and key secondary efficacy endpoints to align with the protocol.<br>d. Clarified the wording on the descriptions of the analyses to align with the definition of the estimands.<br>e. Added donor type (matched related, other) to the stratified analysis of GRFS.<br>f. Replaced Miettinen-Nurminen test with Cochran-Mantel-Haenszel test for the analyses of rate of subjects with and without higher grade of GvHD at 90 days after randomization.<br>g. Added additional subgroups for efficacy analyses.<br>h. Removed the RFS futility analysis.<br>i. Updated the statistical significance level from one-sided to two-sided.<br>j. Removed chronic GvHD from baseline and subgroup analyses.<br>k. Clarified the analyses of chronic GvHD PRO data.<br>l. Removed DILI and GvHD from adverse events of special interest.<br>The following additional changes were made<br>a. Removed the sensitivity analysis of RFS by censoring all new anti-leukemic therapy.<br>b. Clarified that the analysis of GRFS is based on IRC-assessed disease relapse and a sensitivity analysis based on investigator-assessed disease relapse.<br>c. Removed statistical tests for exploratory efficacy endpoints, which will be summarized descriptively.<br>d. Clarified that PTs within SOC will be summarized in descending frequency order for the active treatment arm.<br>e. Clarified that, for the active treatment arms, the treatment-emergent AEs will be summarized by treatment period and for the entire treatment period.<br>f. Removed the NCI CTCAE version number throughout the document. At the time this version of SAP is being amended, AbbVie is not fully ready to implement CTCTA version 5. The actual CTCAE version used in the safety analyses will be noted in the summary table and in the clinical study report.<br>g. Updated the list of selected adverse events. |

| Version | Date | Summary |
|---|---|---|
| | | h. Added lab TLS summary based on Howard criteria.<br>i. Clarified wording on IDMC recommendation at the interim analysis.<br>j. Clarified the summary for protocol deviation.<br>k. Other editorial edits throughout the document. |
| 2.0 | 16 Nov 2023 | The following changes were made based on Protocol Amendment 7:<br>a. The primary endpoint was changed to OS and the corresponding sample size assumptions, estimand, statistical analysis method and data cutoffs for the primary endpoint were updated.<br>b. Morphologic RFS and non-morphologic RFS were added to the key secondary endpoints and the corresponding estimands and statistical analysis methods were updated.<br>c. The definition of GRFS and the corresponding estimand and statistical analysis method was updated.<br>d. Change from baseline in physical functioning was added to the key secondary endpoints and the corresponding estimand and statistical analysis method were added.<br>e. The threshold of deterioration for the key secondary endpoint of time to deterioration in GHS/QoL was updated.<br>f. The key secondary endpoint of minimal residual disease was renamed to measurable residual disease.<br>g. The description of the exploratory endpoint of immunosuppressive therapy and the corresponding analysis were updated.<br>h. Age categories for subgroup analyses were updated.<br>i. Baseline remission status was added to the baseline characteristics and efficacy subgroup analyses.<br>j. The timing of the efficacy interim analysis and the corresponding stopping boundaries were updated.<br>k. The ranking of the key secondary endpoints was updated.<br>The following additional changes were made:<br>a. Added the statement that "Sponsor will not conduct any by-treatment analysis or summary of the data using the actual treatment information until the database lock for the final primary analysis of OS, regulatory submissions based on interim results, or termination of the study, whichever occurs earlier."<br>b. Added additional variables to baseline characteristics.<br>c. Clarified the definition of prior medications for untreated subjects in Part 1 or Part 2 Arm A.<br>d. Clarified what and how variance covariance matrices will be used in the linear mixed model analysis of longitudinal data (e.g., PROMIS Fatigue). |

| Version | Date | Summary |
|---|---|---|
| | | e. Removed sensitivity analysis of GvHD rate.<br>f. Clarified the definitions of baseline for each arm.<br>g. Updated analysis of immunosuppressive therapy.<br>h. Updated the SAS version for conducting analyses to 9.4 or higher.<br>i. Other editorial edits throughout the document. |

# 13.0 References

1. Pollyea DA, Pratz KW, Jonas BA, et al. Venetoclax in combination with hypomethylating agents induces rapid, deep, and durable responses in patients with AML ineligible for intensive therapy. 2018 ASH Annual Meeting & Exposition. Abstract 285. Presented 02 December 2018.
2. Dombret H, Seymour JF, Butrym A, et al. International Phase 3 study of azacitidine vs conventional care regimens in older patients with newly diagnosed AML with > 30% blasts. Blood. 2015;126(3):291-9.
3. Aaronson NK, Ahmedzai S, Bergman B, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: A quality-of-life instrument for use in international clinical trials oncology. J Natl Cancer Inst. 1993;85(5):365-76.
4. Cella D, Yount S, Rothrock N, et al. The patient-reported outcomes measurement information system (PROMIS): progress of an NIH roadmap cooperative group during its first 2 years. Med Care. 2007;45(5 Suppl 1):S3-11.
5. Yost KJ, Eton DT, Garcia SF, et al. Minimally important differences were estimated for six PROMIS Cancer scales in advanced-stage cancer patients. J Clin Epidemiol. 2011;64(4):507-16.
6. Oppe M, Devlin NJ, van Hout B, et al. A program of methodological research to arrive at the new international EQ-5D-5L valuation protocol. Value Health. 2014;17(4):445-53.
7. Pickard AS, Neary MP, Cella D. Estimation of minimally important differences in EQ-5D utility and VAS scores in cancer. Health Qual Life Outcomes. 2007;5:70.

8. Jagasia MH, Greinix HT, Arora M, et al. National Institutes of Health Consensus Development Project on criteria for clinical trials in chronic-versus-host disease: I. The 2014 diagnosis and staging working group report. Biol Blood Marrow Transplant. 2015;21(3):389-401.
9. Buttgereit F, da Silva JA, Boers M, et al. Standardised nomenclature for glucocorticoid dosages and glucocorticoid treatment regimens: current questions and tentative answers in rheumatology. Ann Rheum Dis. 2002;61(8):718-22.
10. Liu D, Ahmet A, Ward L, et al. A practical guide to the monitoring and management of the complications of systemic corticosteroid therapy. Allergy Asthma Clin Immunol. 2013;9(1):30.
11. Kratz A, Ferraro M, Sluss PM, et al. Case records of the Massachusetts General Hospital. Weekly clinicopathological exercises. Laboratory reference values. N Engl J Med. 2004;351(15):1548-63.
12. Lin R, Yin G. Bayesian optimal interval design for dose finding in drug-combination trials. Stat Methods Med Res. 2017;26(5):2155-67.

## Appendix A. Protocol Deviations

The number and percentage of subjects who reported at least one of the following protocol deviation categories will be provided.

- Subject entered into the study even though s/he did not satisfy entry criteria.
- Subject developed withdrawal criteria during the study and was not withdrawn.
- Subject received wrong treatment or incorrect dose of study.
- Subject took prohibited concomitant medication.

## Appendix B. Definition of Adverse Events of Special Interest

Adverse Events of Special Interest (AESI) will be identified using the following search criteria:

| Area of Safety Interest | Search Criteria |
|---|---|
| Tumor Lysis Syndrome (AE) | SMQ – "Tumor Lysis Syndrome" (narrow) |

Additional selected adverse events will be identified using the following search criteria:

| Selected Adverse Events | Search Criteria |
|---|---|
| Neutropenia | PT terms – "Neutropenia" and "Neutrophil count decreased" |
| Neutropenia | PT terms – "Neutropenia," "Neutrophil count decreased," "Febrile neutropenia," "Agranulocytosis," "Neutropenic infection," and "Neutropenic sepsis" |
| Serious Infections | SAEs in the SOC of "Infections and infestations" |
| Second Primary Malignancy | SMQ – "Malignant tumours" (narrow), "Myelodysplastic syndromes" (narrow) |
| Hemorrhages | SMQ – "Haemorrhages" (narrow) |
| Thrombocytopenia | PT terms – "Thrombocytopenia" and "Platelet count decreased" |

## Appendix C. Renal and Hepatic Function

Renal function is classified by the creatinine clearance (mL/min) as follows.

| Renal Function | CrCL (mL/min) |
|---|---|
| Normal | ≥ 90 |
| Mild Impairment | 60 ≤ CrCL < 90 |
| Moderate Impairment | 30 ≤ CrCL < 60 |
| Severe Impairment | CrCL < 30 |

Creatinine clearance is calculated by the Cockcroft Gault formula.

$$\text{CrCL} = \frac{(140 - \text{Age}) \bullet (\text{Weight in kg}) \bullet [0.85 \text{ if Female}]}{72 \bullet \text{Serum Creatinine (mg/dL)}}$$

Or, if serum creatinine is in µmol/L:

$$\text{CrCL} = \frac{(140 - \text{Age}) \bullet (\text{Weight in kg}) \bullet [1.23 \text{ if Male, } 1.04 \text{ if Female}]}{\text{Serum Creatinine (µmol/L)}}$$

Hepatic function is classified according to the bilirubin and AST values as follows.

| Hepatic Function | Bilirubin (mg/dL) | AST (IU/L) |
|---|---|---|
| Normal | ≤ 1.0 | ≤ 40 |
| Mild Impairment | ≤ 1.0 | > 40 |
| | > 1.0 and ≤ 1.5 | Any* |
| Moderate Impairment | > 1.5 and ≤ 3.0 | Any* |
| Severe Impairment | > 3.0 | Any* |

* Subjects with total bilirubin in the corresponding range do not require an observed AST value to be classified in this category (i.e., AST may be missing).

## Appendix D. Bayesian Optimal Interval (BOIN) Design

This document contains the statistical methodology of the BOIN design[12] for Part 1 (Dose Confirmation) of Study M19-063. The goal of Part 1 is to identify the recommended Phase 3 dose (RPTD) for Part 2 (Randomization) of the study.

**Methodology**

Let $\phi$ denote the prespecified target toxicity rate. Let $\Delta_L > 0$ and $\Delta_U > 0$ be prespecified lower and upper cutoffs, respectively, which satisfy $0 < \phi - \Delta_L < \phi + \Delta_U < 1$. The BOIN design consists of determining $\Delta_L$ and $\Delta_U$ such that the interval $(\phi - \Delta_L, \phi + \Delta_U)$ minimizes the probability of incorrect dose assignment decisions. To determine $\Delta_L$ and $\Delta_U$, consider three hypotheses at dose level $j$:

$$H_{0j}: p_j = \phi, H_{1j}: p_j = \phi_1, H_{2j}: p_j = \phi_2$$

where $p_j$ denotes the true toxicity rate at dose level $j$, $\phi_1$ denotes the highest toxicity rate that is deemed subtherapeutic such that dose escalation should be made and $\phi_2$ denotes the lowest toxicity rate that is deemed overly toxic such that dose de-escalation should be made. Thus, the correct dose assignment decisions under $H_{0j}$, $H_{1j}$ and $H_{2j}$ are retainment (i.e., stay at current dose), escalation and de-escalation, respectively.

Under the assumption of a noninformative prior probability for the three hypotheses (i.e., $\Pr(H_{0j}) = \Pr(H_{1j}) = \Pr(H_{2j}) = 1/3$), the optimal interval $(\phi - \Delta_L, \phi + \Delta_U)$ which minimizes the probability of incorrect dose assignment decisions is given by:

$$\phi - \Delta_L = \frac{\log\left(\frac{1-\phi_1}{1-\phi}\right)}{\log\left(\frac{\phi(1-\phi_1)}{\phi_1(1-\phi)}\right)} \text{ and } \phi + \Delta_U = \frac{\log\left(\frac{1-\phi}{1-\phi_2}\right)}{\log\left(\frac{\phi_2(1-\phi)}{\phi(1-\phi_2)}\right)}$$

Suppose the current dose level is $j$, where $j = 1, ..., J$ Let $\hat{p}_j = \left(y_j/n_j\right)$denote the estimated toxicity rate based on the accumulated information on dose level $j$ with $y_j$ toxicities and $n_k$ subjects. Define an admissible dose escalation set as $A_E = \{(j + 1)\}$ and an admissible dose de-escalation set as $A_D = \{(j - 1)\}$.

**Dose Assignment Decisions**

Dose assignment decisions under the BOIN design proceeds as follows:

1. Treat the first cohort at the starting dose.
2. Suppose the current cohort is treated at dose $j$; then for the next cohort of subjects:
   a. If $\hat{p}_j \leq \phi - \Delta_L$, escalate to the next higher dose level that belongs to $A_E$.
   b. If $\hat{p}_j \geq \phi + \Delta_U$, de-escalate to the next lower dose level that belongs to $A_D$.
   c. Otherwise, if $\phi - \Delta_L < \hat{p}_j < \phi + \Delta_U$, the current dose is retained.
3. This process continues until the RPTD is declared or the study is terminated due to excessive toxicity.

The BOIN design also imposes the following safety rule: dose levels which satisfy $Pr\{p_j > \phi | y_j\} \geq \lambda$ are eliminated, where $\lambda$ is a prespecified threshold probability.

**Application of BOIN Design to Study M19-063**

Table 6 gives the dose levels of venetoclax in combination with azacitidine to be explored.

**Table 6. BOIN Dose Levels of Venetoclax and Azacitidine**

| Dose Level | Venetoclax Dose (mg) | Venetoclax duration/cycle (days) | Azacitidine Dose (mg/m$^2$) |
|---|---|---|---|
| –1[a] | [redacted] | 14 | [redacted] |
| 1[b] | | 28 | |
| 2 | | 28 | |
| 3 | | 28 | |

BOIN = Bayesian Optimal Interval Design

a. Dose level –1 may be explored if the starting dose is not tolerated.

b. Starting dose.

For Study M19-063, the toxicity rates and safety threshold are prespecified as follows:

$\phi = 0.20$ (i.e., target toxicity rate)

$\phi 1 = 0.6\phi$ (i.e., highest toxicity rate which is subtherapeutic)

$\phi 2 = 1.4\phi$ (i.e., lowest toxicity rate which is overly toxic)

$\lambda = 0.95$ (i.e., safety threshold for dose combination elimination)

The above prespecified values yield an optimal interval of (0.157, 0.238). Up to approximately 20 subjects will be enrolled.

Table 7 below gives the decision rules for the BOIN design with a target DLT rate of 20% and optimal interval specified above.

**Table 7. Dose Level Decision Rules for BOIN**

| | **# of Subjects Evaluable at Current Venetoclax Dose Level** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Action** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** |
| Escalate if # of subjects with DLT ≤ | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 |
| Stay at current combination if # of subjects with DLT = | - | - | 1 | 1 | - | - | 2 | 2 | 2 | 2 |
| De-escalate if # of subjects with DLT ≥ | 1 | 1 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 3 |
| Eliminate* if # of subjects with DLT ≥ | 2 | 3 | 3 | 3 | 4 | 4 | 4 | 5 | 5 | 5 |

BOIN = Bayesian Optimal Interval Design

a. Eliminate current and higher doses (i.e., venetoclax dose and azacytidine dose ≥ current dose level).

# Document Approval

Study M19063 - Statistical Analysis Plan Version 2 - 16Nov2023 (E3 16.1.9)

**Version:** 1.0 **Date:** 27-Nov-2023 **Company ID:** 20231127-0900f9f686b340f1-1.0-en

| Signed by: | Date: | Meaning of Signature: |
|---|---|---|
| [redacted] | [redacted] | Approver - Statistics |
| | | Approver - Statistics |
| | | Author |
| | | Approver |
| | | Approver - Product Safety Team (PST) Lead |
| | | Approver - Statistics |