1.0 Title Page

Statistical Analysis Plan

Study M15-656

A Randomized, Double-Blind, Placebo Controlled Phase 3 Study of Venetoclax in Combination with Azacitidine Versus Azacitidine in Treatment Naïve Subjects with Acute Myeloid Leukemia Who Are Ineligible for Standard Induction Therapy

Date: 03 March 2020

Version 7.0

2.0 Table of Contents

List of Tables

3.0 Introduction

This statistical analysis plan (SAP) describes the full efficacy and safety statistical analyses for venetoclax Study Protocol M15-656 (Viale-a) Amendment 7 dated August 21, 2019. It will provide details of statistical methods and describe analysis conventions to guide the statistical programming work.

Unless noted otherwise, all analyses will be performed using SAS® version 9.3 or later (SAS Institute Inc., Cary, NC 27513) under the UNIX operating system.

4.0 Study Objectives, Design, and Procedures

4.1 Objectives

The primary objectives of this study are to evaluate if venetoclax in combination with azacitidine will improve overall survival (OS) and composite complete remission rate (complete remission + complete remission with incomplete marrow recovery; CR + CRi) versus placebo in combination with azacitidine, in treatment naïve subjects with acute myeloid leukemia (AML).

The secondary objectives of the study are:

- To evaluate if venetoclax in combination with azacitidine will improve the rate of CR.
- To evaluate if venetoclax in combination with azacitidine will improve the rate of CR and complete remission with partial hematologic recovery rate (CRh).
- To evaluate if venetoclax in combination with azacitidine will improve the proportion of subjects achieving composite complete remission (CR or CRi) by Cycle 2.
- To evaluate if venetoclax in combination with azacitidine will improve the proportion of subjects achieving CR or CRh by Cycle 2.
- To evaluate if venetoclax in combination with azacitidine will improve the transfusion independence rate

- To evaluate if venetoclax in combination with azacitidine will improve event-free survival (EFS)
- To evaluate if venetoclax in combination with azacitidine will improve the MRD response rate
- To evaluate if venetoclax in combination with azacitidine will improve the response rates and overall survival in molecular subgroups
- To evaluate if venetoclax in combination with azacitidine reduces fatigue and improves global health status/quality of life (GHS/QoL) based on patient reported outcome (PRO) assessments (Patient Reported Outcomes Measurement Information System [PROMIS] Cancer Fatigue Short Form [SF] 7a and European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Core [EORTC QLQ-C30]).

The exploratory objectives of the study are:

- Exploration of biomarkers predictive of venetoclax activity and duration of response may be performed. These analyses maybe part of a multi-study assessment to compare responses to the therapies and/or disease state. Potential analyses may include, but will not be limited:
  - To evaluate BCL2 expression and outcome measures of overall survival and complete remission rate
- To evaluate the impact of venetoclax on remaining subscales/items from the EORTC QLQ-C30 and EQ-5D-5L.

4.2 Study Design and Plan

This is a Phase 3, randomized, double-blind, placebo controlled, multicenter trial evaluating efficacy and safety of venetoclax in combination with azacitidine versus placebo in combination with azacitidine in treatment naïve subjects with AML who are ≥ 18 years of age and not eligible for standard induction therapy due to age or comorbidities. Subjects will be randomized to one of the two treatment Arms in a 2:1 ratio, both of which will have treatment cycles of 28 days.

- Arm A: Venetoclax 400 mg orally QD on Days 1 – 28 plus azacitidine 75 mg/m^2 SC or IV Daily for 7 days
- Arm B: Placebo for venetoclax 400 mg orally QD on Days 1 – 28 plus azacitidine 75 mg/m^2 SC or IV Daily for 7 days

Approximately 400 subjects will be randomized to meet scientific and regulatory objectives without enrolling an undue number of subjects in alignment with ethical considerations.

An open-label safety cohort in China of up to 12 subjects: The objective of this cohort is to enroll subjects who meet the same eligibility criteria at sites in mainland China to evaluate the safety and PK profile of venetoclax at 400 mg daily dose in combination with azacitidine prior to allowing enrollment into the double blind, randomized portion of the study as per agreement with National Medical Products Administration of China.

4.3 Sample Size

For US and countries using the US as the reference country, the study includes a single primary endpoint of overall survival. For Japan, EU and countries using EU as reference countries, the study includes dual-primary endpoints of overall survival (OS) and CR + CRi rate. The sample size calculation is based on the following assumptions:

- The significance level (two-sided 0.05) will be split to give a 0.01 significance level to the CR + CRi rate analysis and a 0.04 significance level to the OS analysis.
- CR + CRi rate of 28% for placebo plus azacitidine arm
- CR + CRi rate of 55% for venetoclax plus azacitidine arm
- Median OS of 10.4 months for placebo plus azacitidine arm
- Median OS of 14.9 months for venetoclax plus azacitidine arm (hazard ratio of 0.7)
- Interim analysis of OS at 75% of death events with O'Brien-Fleming boundary
- 2:1 randomization ratio to venetoclax plus azacitidine, and placebo plus azacitidine arm

With the above assumptions, a total of 225 subjects (150 in venetoclax in combination with azacitidine arm, and 75 in placebo with azacitidine arm) will provide 88% power to detect statistically significant difference in CR + CRi rate between treatment arms at two-sided alpha level of 0.01; a total of 360 death events will provide 86.7% power to detect statistically significant difference in OS between treatment arms at two-sided alpha level of 0.04. A total of approximately 400 subjects (267 in venetoclax in combination with azacitidine arm, and 133 in placebo with azacitidine arm) will be randomized into the study to obtain the 360 death events.

Additional up to twelve subjects will be enrolled in the open label Chinese safety cohort to receive venetoclax and azacitidine, as required per the China Food and Drug Administration (CFDA). The total sample size will be approximately 412. The subjects from Chinese safety cohort will not be included in the efficacy and safety analysis of the double blind portion of the study.

4.4 Interim Analysis

To ensure subject safety, an Independent Data Monitoring Committee (IDMC) will review unblinded safety data at the following proposed time points:

- The first meeting to review the unblinded safety data will be occur approximately 3 months after the 20th subjects has been randomized and dosed
- The subsequent reviews of unblinded safety data will occur approximately every 3 months after the first review of unblinded safety data.

The members of the IDMC are comprised of an independent multidisciplinary group who are responsible for safeguarding the interests of trial subjects, assessing the safety, when appropriate, the efficacy during the trial, as well as for monitoring the quality and integrity of the clinical trial, as outlined in the IDMC Charter. Additional ad hoc meetings may occur as needed for the IDMC to fulfill its responsibilities as detailed in the IDMC charter.

Further, two interim analyses of OS will be performed. The first interim analysis (IA1) will occur concurrently with the primary analysis of CR + CRi rate at least 6 months after the 225th subject in the full analysis set is randomized. An administrative significance level 0.0001 will be allocated to the first interim analysis for OS (at the time of CR + CRi primary analysis).

All analysis for the unblinded IDMC review will be done by statisticians from Axio (independent of the Sponsors) and the Sponsors will remain blinded for this data. For IA1, the IDMC will review unblinded data from the first ~225 subjects randomized in the study. The primary analysis of CR + CRi rate will occur when the 225th subject in the full analysis set has been followed for at least 6 months since being randomized.

The IDMC will inform the Sponsor Steering Committee if CR + CRi rate data is statistically significant in favor of Arm A at the significance level 0.01 (two-sided). Then an unblinded Filing Team will be formed to prepare regulatory submissions in Japan, EU and EU reference countries. All of the investigators, study sites will remain blinded at this time as well as any Sponsor Study Team members who will have any interactions with the sites going forward. The study will continue to follow all protocol specified procedures. Refer to the Filing Team Charter for the detailed unblinding procedures. The CR + CRi rate will be the only endpoint to be statistically tested. Descriptive analyses will be performed for other endpoints.

The second interim analysis (IA2) will be performed once approximately 270 death events (75% of the total 360 events) in the full analysis set are observed. The Lan-DeMets alpha spending function with O'Brien-Fleming boundary will be used to ensure that the one-sided false positive rate will be strongly controlled for overall survival. The IDMC will make a recommendation based on overall survival to either to stop for success at IA2 or proceed to the final analysis (FA) using pre-specified stopping rule. The planned stopping boundaries for efficacy are described in Table 1 (EU and EU reference countries) and Table 2 (US and US reference countries) below for IA2 and final analysis of OS endpoint. The actual stopping boundaries at IA2 and FA of OS endpoint will be derived using Lan-DeMets alpha spending function, based on the observed number of death events in the

extracted database. The actual stopping boundaries will be presented in the IDMC charter addendum. Details of the IDMC review will be presented in the IDMC charter.

At IA2, if the IDMC informs the Sponsor that the OS data is statistically significant and in favor of Arm A, upon notification from the Sponsor Steering Committee after consultation with regulatory authorities as required, the Sponsor will then unblind the study to prepare regulatory submissions globally. Otherwise, the study will remain blinded and continue to follow protocol specified procedures until the final analysis (FA).

In addition, at IA2, the DMC will make a recommendation based on OS to stop the study due to a potential detrimental OS effect which is measured by HR > 1.

Table 1. Planned Efficacy Stopping Boundaries at the IA2 and FA of OS Endpoint (EU and EU Reference Countries)

Scenario	Look	# of Events	Efficacy Stopping Boundaries (HR/one-sided p-value)
CR + CRi rate is positive at IA1 (two-sided overall alpha for OS is 0.05)	IA2	270	0.739/0.010
	FA	360	0.799/0.022
CR + CRi rate is not positive at IA1 (two-sided overall alpha for OS is 0.04)	IA2	270	0.729/0.007
	FA	360	0.791/0.018

Table 2. Planned Efficacy Stopping Boundaries at the IA2 and FA of OS Endpoint (US and US Reference Countries)

Look	# of Events	Efficacy Stopping Boundaries (HR/one-sided p-value)
IA2	270	0.739/0.010
FA	360	0.799/0.022

Interim statistical analyses and summaries for presentations to the IDMC will be prepared by Axios (independent of the Sponsors). AbbVie/GNE personnel will remain blinded and not have access to the interim analyses prepared for the IDMC.

5.0 Analysis Sets

5.1 Definition for Analysis Sets

There are two groups of subjects randomized under different randomization schedules. Group 1 consists of two subjects randomized under original protocol. Group 2 consists of subjects randomized under Protocol Amendment 1 and later versions.

Two study populations will be analyzed, defined as follows:

- The full analysis set consists of all Group 2 subjects randomized by IVRS/IWRS (exclude the open-label China safety cohort). The full analysis set will be used for efficacy analyses. The data from full analysis set will be analyzed by the treatment arm assignment given at the time of randomization, even if the subject takes the incorrect drugs that do not match the assigned treatment, or does not receive any treatment, or does not follow the protocol until completion.
- The safety analysis set consists of all Group 1 and Group 2 subjects (exclude the open-label China safety cohort) who take at least one dose of venetoclax/placebo and azacitidine combination. The safety analysis set will be used for safety analyses. The data from safety analysis set will be analyzed by the actual treatment that subject received.

5.2 Variables Used for Stratification of Randomization

For Group 1, the subject randomization is stratified by age (18 – < 75, ≥ 75), and region (US, EU, Japan (JP), Rest of world (ROW)). For Group 2, the subject randomization is stratified by age (18 – < 75, ≥ 75), cytogenetic risk (intermediate, poor) and region (US, EU, China, JP, ROW). Cytogenetic risk is based on National Comprehensive Cancer Network (NCCN) Risk Categorization (Guidelines for AML version 2 2016).

6.0 Analysis Conventions

General Considerations

Unless otherwise noted, for all statistical analyses, statistical significance will be determined by a two-sided *P* value ≤ 0.05. The date of randomization is defined as the date that the IVRS/IWRS issues a randomization number.

Definition of Study Drug

Unless otherwise specified, the study drug in this document refers to venetoclax/placebo and azacitidine. The first dose date of study drug is defined as the date of the 1st dose of venetoclax/placebo or azacitidine, whichever occurs first, is administrated. The last dose date of study drug is defined as the date of the last dose of venetoclax/placebo or azacitidine, whichever occurs later, is administrated.

Definition of Baseline

All baseline summary statistics and analyses will be based on characteristics prior to the initiation of any component of study drug (or randomization for non-treated subjects). Unless otherwise stated, baseline for a given variable will be defined as the last value for that variable obtained prior to the first dose of any component of study drug or randomization for non-treated subjects.

Definition of Final Observation

The final observation for the analyses listed in Table 3 and Table 4 is defined as the last non-missing observation collected after the first dose of study drug to the last dose of study drug + 30 days.

Definition of Data Cutoff Date

The data cutoff date for IA1 will be the projected date when the 225th subject in the full analysis set has been followed for at least 6 months since being randomized.

The data cutoff date for IA2 will be the projected date when the 270th death occurred in the full analysis set.

The data cutoff date for final analysis will be the projected date when the 360th death occurred in the full analysis set.

The same data cutoff date will be applied for both efficacy and safety analyses.

Stratification Factor for Efficacy Analyses

Age (18 – < 75, ≥ 75) and cytogenetic risk (intermediate, poor) will be used in all stratified analyses of the efficacy endpoints. The stratification factor value under which the subject is randomized by the IVRS/IWRS will be used in the efficacy analyses.

In order to perform a region-specific subgroup analysis to support regional filings, region is used as a stratification factor in the randomization. However, region will not be included in the stratified efficacy analysis since it is not considered as a prognostic factor.

Definition of Cycle Rx Days in Each Cycle

Cycle Rx Days for each cycle are calculated for each time point relative to first dose of study drug in each cycle. The day of the first dose of study drug in each cycle is defined as Cycle Rx Day 1, while the day prior to the first dose of study drug is defined as Cycle Rx Day –1 (there is no Cycle Rx Day 0).

Definition of Analysis Windows

All time points and corresponding time windows are defined for each cycle are based on Cycle Rx Day 1 of each cycle.

For visit wise clinical laboratory analyses, vital signs, and PRO analyses, the time windows specified in Table 3 and Table 4 describe how efficacy and safety data are assigned to protocol specified visits respectively. Analysis time windows are constructed using the following algorithm:

- Determine the nominal Cycle Rx day for each scheduled visit.
- Determine the window around a specific nominal Cycle Rx day as in Table 3 and Table 4.
- If more than one assessment is included in a time window of the assessment, the one closest to the nominal day should be used. If there are two observations with equal distance to the nominal day, the later one will be used in analyses. If multiple values are collected on the same day, the arithmetic average will be calculated and used as the value of that day for the mean change analyses.
- For analysis of tumor lysis syndrome (TLS) laboratory variables meeting the Howard criteria where multiple values were collected over the course of a day (within 4 hours prior to dosing and 6 – 8 hours post dose) during ramp up in 7 days of Cycle 1 will be used. All values will be used and no averages will be taken.
- For analysis of liver enzyme laboratory variables meeting criteria for potential drug-induced liver injury, all values will be used and no averages will be taken.
- For laboratory shift tables, the Common Terminology Criteria (CTC) grade will be assigned based on all observed laboratory values and no averages will be taken. If multiple grades are collected on the same day, the worst grade will be used as the value of that day for the shift table analyses.

Table 3. Time Windows for Analysis of Hematology, Chemistry, and Vital Signs Parameters

Scheduled Visit	Nominal Cycle Rx Day	Time Window (Cycle Rx Day Range)
Baseline	≤ 1	See the baseline definition (Section 6.0)
Cycle 1 Day 2	2	[2]
Cycle 1 Day 3	3	[3]
Cycle 1 Day 4	4	[4]
Cycle 1 Day 5	5	[5]
Cycle 1 Day 8	8	[6, 11]
Cycle 1 Day 15	15	[12, 18]
Cycle 1 Day 21	21	[19, 25]
Cycle 2 Day 1	1	[–3, 10]
Every Cycle from Cycle 3 Day 1	1	[–10, 10]
Final Observation	NA	The last non-missing observation collected after the first dose of study drug to the last dose of study drug + 30 days

Note: Hematology and chemistry samples collection are specified in the study protocol. Please refer to the study protocol for details.

Table 4. Time Windows for Analysis of PROMIS Cancer Fatigue SF 7a, EORTC QLQ-C30, and EQ-5D-5L

Scheduled Visit	Nominal Cycle Rx Day	Time Window (Cycle Rx Day Range)
Baseline	≤ 1	See the baseline definition (Section 6.0)
Every Other Cycle from Cycle 3 Day 1	1	[–10, 10]

Note: The schedule of assessment for PROMIS Cancer Fatigue SF 7a, EORTC QLQ-C30, and EQ-5D-5L is specified in the study protocol. Please refer to the study protocol for details.

Conversion of leukocyte (white blood cell [WBC]) differential counts from percentage (%) value to absolute value

All efficacy and safety analyses will include leukocyte differential values in % and absolute values. The following conversion process will be performed, if sites only

provided % values for leukocyte differential counts without absolute counts on the same collection date and time of the total leukocyte counts.

1. The conversion is only for leukocyte differential counts in % values which were collected on the same date and at same time as total leukocyte counts.
2. Leukocyte differential absolute value ($\times 10^9$ L) = Total leukocyte count ($\times 10^9$ L) × leukocyte differential value (%)/100.
3. The following low normal ranges will be used for leukocyte differential counts in % value converted to absolute count values for the shift tables.

Lab Test: **Total Leukocytes (WBC) and differential leukocyte counts**	**Normal Range (SI units)**
Total Leukocyte ($\times 10^9$/L)	4.5 - 11
Neutrophils (%)	40 - 70
Lymphocytes (%)	22 - 44

* Source: Kratz A. Laboratory Reference Values. N Engl J Med. 2004;351:1548-63.

Lower limit of normal of 1.8×10^9/L for neutrophil is calculated from total leukocyte count of 4.5×10^9/L, according to formula on item 2 above. Lower limit of normal of 0.99×10^9/L for lymphocytes is calculated from total leukocyte count of 4.5×10^9/L. Above the lower limit of normal range values are used to classify CTCAE grade for neutrophil and lymphocyte counts.

7.0 Subject Disposition

The screen failure reasons will be summarized for the screen failure subjects. Analyses for the subject disposition will be performed on the full analysis set. A separate summary for the 2 subjects in the Group 1 will also be provided.

The number and percentage of subjects will be summarized overall, by treatment arm, by country and by site, for each of the following categories:

- Subjects randomized into the study

- Subjects who received at least one dose of study drug
- Subjects who were randomized but never received any component of study drug

Additionally, the number and percentage of subjects will be summarized overall, and by treatment arm for each of the following categories:

- Subjects who discontinued venetoclax/placebo due to all reasons and primary reason
- Subjects who discontinued azacitidine due to all reasons and primary reason
- Subjects who discontinued study due to all reasons and primary reason
- Subjects randomized under each Protocol Amendment

There will be no statistical comparison for the subject disposition between the two treatment arms.

The median of time on study using reverse Kaplan-Meier approach and range (min and max) of time on study will be provided. Subjects alive up to data cut-off date will be considered as events and death of subjects on or prior to data cut-off date will be censored in the reverse Kaplan-Meier approach for estimation of median of time on study.

8.0 Demographics, Baseline Characteristics, Medical History, Previous Concomitant Medications, and Post Treatment Therapies

8.1 General Consideration

All demographic, baseline characteristics, medical history, prior, concomitant medication and post treatment therapy summaries will be performed by treatment arm on the full analysis set as described in Section 5.1. A separate summary for the 2 subjects in the Group 1 will also be provided.

8.2 Demographic and Baseline Characteristics

All baseline characteristic summary statistics and analyses are based on characteristics prior to the first dose of study drug or date of randomization for non-treated subjects.

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 between the two treatment arms.

The following demographic and baseline characteristics will be summarized:

Demographics:

- age (years) and age categories (18 – < 65, 65 – < 75 years, ≥ 75 years)
- gender (male, female)
- race (white, black or African American, Asian, other)
- region (US, EU, China, JP, ROW)
- height (cm)
- weight (kg)

Baseline and Disease-Related Characteristics:

- ECOG performance status (Grade 0, 1, 2, 3)
- Cytogenetic risk (intermediate, poor)
- Bone marrow blast count (< 30%, ≥ 30% – < 50%, ≥ 50%)
- Bone marrow blast count (%)

- CTC grade of neutropenia
- Neutrophils value ($\times 10^9$/L)
- CTC grade of anemia
- Hemoglobin value (G/L)
- CTC grade of thrombocytopenia
- Platelet count ($\times 10^9$/L)
- Type of AML (de novo, secondary, therapy-related AML)
- AML with Myelodysplasia related changes (AML-MRC)
- Reasons for being ineligible for standard induction therapy
- Antecedent hematologic history of MDS (Yes, No)
- Baseline RBC transfusion dependence (Yes, No)
- Baseline platelet transfusion dependence (Yes, No)
- Molecular marker via central lab (FLT3, IDH1/2, TP53, NPM1)
- AML with myelodysplasia related changes (Yes, No)
- Hepatic impairment (Yes, No)
- Renal impairment (Yes, No)

The cross tabulation of the following demographic and baseline characteristics captured in IVRS/IWRS and CRF will be generated:

- Age categories (18 – < 75, ≥ 75 years)
- Cytogenetic risk (intermediate, poor)
- Region (US, EU, China, JP, and ROW)

8.3 Medical History

Medical history data will be coded for conditions/diagnoses by body system organ class and preferred term. The number and percentage of subjects with a particular condition/diagnosis will be summarized by body system organ class and preferred term. Subjects reporting more than one condition/diagnosis within a system organ class will be

counted only once for that system organ class. There will be no statistical comparison for the medical history between the two treatment arms.

8.4 Previous Treatment, Concomitant Medications, and Post-Treatment Therapies

A prior medication is defined as any medication taken prior to the first dose of study drug. A concomitant medication is defined as any medication that started prior to the first dose of study drug, and continued to be taken after the first dose of study drug, or any medication that started after the first dose of study drug, but not after the last dose of study drug. All medications are considered prior medications for randomized subjects who have not received any study drug. A post-treatment medication for the treatment of AML is defined as any medications taken after the discontinuation of study drug and entered via Post Treatment eCRFs.

The number and percentage of subjects who have taken medications will be summarized by generic drug name coded by WHO Drug dictionary 2017Q1 or a recent version for prior medications, prior oncology therapies for previous malignancies, concomitant medication, and post-treatment therapies. The number and percentage of subjects who have taken TLS prophylaxis agents including intravenous hydration on or prior to the first dose of study drug and concomitant to the study treatment will be summarized separately. The number and percentage of subjects who have received transfusions on or prior to first dose of study drug will be summarized by transfusion type. There will be no statistical comparison between the two treatment arms performed for the aforementioned summaries.

For summaries of concomitant medications, if an incomplete or missing start date was collected for a medication, the medication will be assumed to be concomitant medication and prior medication unless there is evidence that confirms that the medication was not a concomitant medication (e.g., the medication end date was prior to the first dose of study drug).

A subject who reports the use of two or more medications will be counted only once in the summary of "Any 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. Similar rules apply to prior medications and post-treatment therapy as well.

9.0 Study Treatment Exposure and Compliance

The duration of exposure to venetoclax/placebo will be summarized by treatment arm for safety analysis set as described in Section 5.1. Duration of exposure is defined for each subject as (last dose date – first dose date) + 1. Duration of exposure will be summarized using the following statistics (months): sample size (N), mean, standard deviation, median, and range. In addition, the number and percentage of subjects exposed to venetoclax/placebo will be summarized for the following categories of exposure duration:

- 0 to 4 weeks (0 to 28 days),
- > 4 weeks to 8 weeks (28 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 by treatment arm. There will be no statistical comparison for the study treatment exposure between the two treatment arms.

10.0 Efficacy Analyses

10.1 General Considerations

Unless otherwise noted, for all statistical analysis, statistical significance will be determined by a two-sided *P* value ≤ 0.05 (when rounded to three decimal places).

When the approximately 360th death for the OS FA in the full analysis set 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 study will be unblinded to all and not limited to the previously defined Sponsor Filling Team and clinical database data will be extracted for documentation and statistical analyses of the efficacy and safety data. Efficacy analyses will be performed in the full analysis set.

Unless otherwise specified, the primary analysis of all response/progression related endpoints (e.g., CR/CRi rate) will be based on the investigator assessment. Sensitivity analyses will be performed based on Independent Review Committee's assessment as specified in Section 10.5.

For the analysis of all response/progression related endpoints, MRD response rate, and transfusion independence rate, the data collected after the data cutoff date or initiation of post treatment therapy, or confirmed PD, or confirmed MR, whichever comes earlier, will be excluded. The definition of confirmed PD and confirmed MR are described in Section 10.3.7.

Censoring Dates for Prematurely Blind Broken Subjects

For all efficacy endpoints except OS, if the subject's blind has prematurely broken by sites before "Cutoff" date, the date of premature blind break will be set as "Cutoff" date for the subject.

10.2 Primary Efficacy Analysis

10.2.1 Primary Efficacy Endpoints in EU and EU Reference Countries

For Japan, EU and EU reference countries, this study has dual primary endpoints of CR + CRi rate (as assessed by investigator) and overall survival (OS). The significance level of 0.05 (two sided) will be split between the dual primary endpoints to give a 0.01 significance level to the CR + CRi rate analysis (based on the investigator assessment) and an overall 0.04 significance level to the OS analysis.

All disease assessments use the criteria as defined in the Protocol Section 5.3.3.

Composite Complete Remission

Composite Complete Remission (CR + CRi) rate will be defined as the proportion of subjects who achieve a complete remission (CR) or complete remission with incomplete blood count recovery (CRi) at any time point during the study per the modified IWG criteria for AML. Subjects who are randomized but have no IWG disease assessment will be considered as non-responders in the calculation of CR + CRi rate.

CR + CRi rate will be compared between treatment arms using Cochran-Mantel-Haenszel (CMH) test stratified by age (18 – < 75, ≥ 75) and cytogenetic risk (intermediate, poor). In addition, the 95% confidence interval for CR + CRi rate based on the binomial distribution (Clopper-Pearson exact method) by treatment arms will be provided. The analysis of CR + CRi rate will be performed with the first 225 subjects in the full analysis set. The 95% confidence interval for the risk difference (exact unconditional confidence limits) will also be provided.

Overall Survival

Overall survival will be defined as the number of days from the date of randomization to the date of death. All events of death will be included, regardless of whether the event occurred while the subject was still taking study drug or after the subject discontinued

study drug. If a subject has not died, then the data will be censored at the date the subject was last known to be alive on or before the cutoff date. The date of the last known alive will be determined by selecting the last available date of the following study procedures for a subject: adverse event start date, bone marrow collection, disease assessment, vital signs assessment, clinical laboratory collection, study drug administration, concomitant medicine start date, biospecimen sample collection, transfusion, survival follow-up, quality of life assessments, and performance status. All subjects in the full analysis set will be included in the analysis.

The distribution of overall survival will be estimated for each treatment arm using Kaplan-Meier methodology and compared between treatment arms using the log-rank test stratified by age (18 – < 75, ≥ 75) and cytogenetic risk (intermediate, poor). The hazard ratio between treatment arms will be estimated using the Cox proportional hazards model stratified by age (18 – < 75, ≥ 75) and cytogenetic risk (intermediate, poor).

10.2.2 Primary Efficacy Endpoints in US and US Reference Countries

For US and US reference countries, this study has a single primary efficacy endpoint of overall survival (OS), with significance level of 0.05 (two-sided). The analysis method for OS is described in Section 10.2.1.

10.3 Secondary Efficacy Analyses

The secondary efficacy endpoints for US and US reference countries and for Japan, EU and EU reference countries are listed in Table 5.

Table 5. List of Secondary Efficacy Endpoints

Secondary Endpoint	US and US reference countries	Japan, EU and EU reference countries
CR + CRi rate	X	
CR + CRh rate	X	
CR + CRi rate by initiation of Cycle 2		X
CR + CRh rate by initiation of Cycle 2	X	
CR rate	X	X
Post-baseline RBC transfusion independence	X	X
Post-baseline platelet transfusion independence	X	X
RBC transfusion independence conversion rate	X	X
Platelet transfusion independence conversion rate	X	X
CR + CRi MRD response rate		X
MRD Negative Remission Rate[a]	X	
CR + CRi rate in FLT3 subgroup		X
CR + CRh rate in FLT3 subgroup	X	
CR + CRi rate in IDH1/2 subgroup		X
CR + CRh rate in IDH1/2 subgroup	X	
OS in FLT3 subgroup	X	X
OS in IDH1/2 subgroup	X	X
PROMIS Cancer Fatigue SF 7a	X	X
EORTC QLQ-C30	X	X
Event free survival (EFS)	X	X

a. The threshold to determine MRD Negative remission rate will be defined separately using methods described in the external SAP for MRD threshold validation.

10.3.1 CR + CRh Rate

CRh (Complete remission with partial hematologic recovery) is a derived response based on bone marrow blast and hematology lab values. A subject achieves a CRh when meeting the following criteria:

- Bone marrow with < 5% blasts and
- Peripheral blood neutrophil count of > $0.5 \times 10^3/\mu L$ and

- Peripheral blood platelet count of > $0.5 \times 10^5/\mu L$ and
- A one week (≥ 7 days) platelet transfusion-free period prior to the hematology lab collection.

For a bone marrow sample collected before the last cycle of study treatment, the hematology lab results collected from the date of the bone marrow sample collection up to the Day 1 of a subsequent cycle of study treatment will be used for CRh analysis. For a bone marrow sample collected during or after the last cycle of study treatment, the hematology lab results collected within 14 days after bone marrow sample collection date will be used for CRh analysis.

CR + CRh rate will be defined as the proportion of subjects who achieve a complete remission (CR) or complete remission with partial hematologic recovery (CRh) at any time point during the study. Subjects who are randomized but have no disease assessment will be considered as non-responders in the calculation of CR + CRh rate. All subjects in the full analysis set will be included in the analysis.

CR + CRh rate will be compared between treatment arms using Cochran-Mantel-Haenszel (CMH) test stratified by age (18 – < 75, ≥ 75) and cytogenetic risk (intermediate, poor). In addition, the 95% confidence interval for CR + CRh rate based on the binomial distribution (Clopper-Pearson exact method) by treatment arm will be provided. The risk difference and 95% confidence interval for the risk difference (exact unconditional confidence limits) will also be provided.

10.3.2 CR + CRi and CR + CRh Rate by the Initiation of Cycle 2

CR + CRi rate by the initiation of Cycle 2 will be defined as the proportion of subjects who achieved a complete remission (CR) or complete remission with incomplete blood count recovery (CRi) by the initiation of Cycle 2 per the modified IWG criteria for AML. Subjects who are randomized but have no IWG disease assessment by the initiation of Cycle 2 will be considered to be non-responders in the calculation of CR + CRi rate by the initiation of Cycle 2. For subjects who discontinue treatment before initiation of

Cycle 2, all disease assessments evaluated before the cutoff date and the initiation of post treatment therapy, whichever occurs earlier, will be included. All subjects in the full analysis set will be included in the analysis.

CR + CRi rate by the initiation of Cycle 2 will be compared between treatment arms using Cochran-Mantel-Haenszel (CMH) test stratified by age (18 – < 75, ≥ 75) and cytogenetic risk (intermediate, poor). In addition, the 95% confidence interval for CR + CRi rate by the initiation of Cycle 2 based on the binomial distribution (Clopper-Pearson exact method) by treatment arm will be provided. The 95% confidence interval for the risk difference (exact unconditional confidence limits) will also be provided.

Similar analyses will be performed for the CR + CRh rate by the initiation of Cycle 2.

10.3.3 Complete Remission Rate

CR rate will be defined as the proportion of subjects who achieved a complete remission (CR) at any time point during the study per the modified IWG criteria for AML (refer to protocol Section 5.3.3). Subjects who are randomized but have no IWG disease assessment will be considered to be non- responders in the calculation of CR rate. All subjects in the full analysis set will be included in the analysis.

CR rate will be compared between treatment arms using Cochran-Mantel-Haenszel (CMH) test stratified by age (18 – < 75, ≥ 75) and cytogenetic risk (intermediate, poor). In addition, the 95% confidence interval for CR rate based on the binomial distribution (Clopper-Pearson exact method) by treatment arm will be provided. The risk difference and 95% confidence interval for the risk difference (exact unconditional confidence limits) will also be provided.

10.3.4 Transfusion Independence Rates

Post baseline transfusion independence rate will be calculated as the proportion of subjects who achieved transfusion independence post baseline. Transfusion independence is defined as a period of at least 56 days (≥ 56 days) with no transfusion after the first dose

of study drug and on or before the last dose of study drug + 30 days or before death or before the initiation of post-treatment therapy whichever is earliest. In addition, the rate of conversion will be calculated as proportion of subjects being post-baseline transfusion independent from baseline transfusion dependence. The transfusion independence rate will be evaluated for 1) red blood cells (RBC) and 2) platelets. All subjects in the full analysis set will be included in the analysis. Subjects who did not have any dose of study drug are considered as transfusion dependent.

The post-baseline transfusion independence rates will be compared between two arms using CMH test stratified by age (18 – < 75, ≥ 75) and cytogenetic risk (intermediate, poor). In addition, 95% confidence interval based on the binomial distribution (Clopper-Pearson exact method) by treatment arms will be provided for the post-baseline transfusion independence rate. The risk difference and 95% confidence interval for the risk difference (exact unconditional confidence limits) will also be provided.

The rates of conversion from baseline transfusion dependence to post-baseline transfusion independence for 1) RBC and 2) platelets will be estimated and the 95% confidence interval based on the binomial distribution (Clopper-Pearson exact method) by treatment arms will be provided. The risk difference and 95% confidence interval for the risk difference (exact unconditional confidence limits) will also be provided.

The duration of transfusion independence will also be summarized. The duration of transfusion independence is defined as the 1st time period that a subject received no RBC/platelet transfusions for at least 56 days. The descriptive statistics (median and range) will be provided for the duration of transfusion independence.

Time to first transfusion independence will be summarized as duration from the date of randomization to the start date of first transfusion independence. The descriptive statistics (median and range) will be provided for the time to first transfusion summary.

10.3.5 CR + CRi MRD Response Rate

For EU and EU reference countries, MRD response rate is defined as the best MRD value which is less than 10^{-3} of residual blasts per leukocytes as measured in bone marrow among subjects who achieved CR + CRi.

MRD Negative remission rate is defined as the proportion of subjects who achieve MRD negativity and remission (CR and/or CRh as determined in the external SAP for MRD threshold validation) for all subjects in the full analysis. Subjects who are randomized but have no MRD assessment will be considered as non-responder for the calculation of MRD negative remission rate. The CMH test stratified by age (18 – < 75, ≥ 75) and cytogenetic risk (intermediate, poor) will be used to compare MRD Negative remission rate between two treatment arms. In addition, 95% confidence interval based on the binomial distribution (Clopper-Pearson exact method) by treatment arms will be provided. The risk difference and 95% confidence interval for the risk difference (exact unconditional confidence limits) will also be provided.

10.3.6 Health Related Quality of Life (HRQoL)

Subjects in the full analysis set who do not have baseline measurement or any post-baseline measurements will not be included in PRO analyses. Post-baseline measurements will be obtained according to the visit window as in Table 4.

PROMIS Cancer Fatigue SF 7a

PROMIS instruments measure concepts such as pain, fatigue, physical function, depression, anxiety and social function (https://www.assessmentcenter.net/documents/PROMIS%20Fatigue%20Scoring%20Manual.pdf). Fatigue will be assessed using the PROMIS Cancer Fatigue SF 7a that has been developed for use in oncology populations. PROMIS Cancer Fatigue SF 7a is a seven item questionnaire that assesses the impact and experience of fatigue over the past 7 days.

Scores will be computed according to the procedures outlined in the PROMIS Fatigue scoring manual, available at https://www.assessmentcenter.net/Manuals.aspx. Change in the PROMIS Fatigue score from baseline will be compared between the two treatment arms. A linear mixed effects regression model with a covariance structure will be fitted to the longitudinal data to test for differences between the two treatment arms. For the linear mixed effects regression model, stratification factors and treatment arm will be included as fixed factors. Furthermore, time and treatment by time interaction will be included in the model. The repeated correlation structure in the timepoints will be assessed by using the Bayesian Information Criterion (BIC). The following covariance structures will be explored: Unstructured (TYPE = UN), compound symmetry (TYPE = CS) and first-order autoregressive (TYPE = AR(1)). The type resulting in model convergence and the lowest BIC will be used for analysis. Change from baseline the PROMIS Fatigue score will be compared between the two treatment arms at each post-baseline visit excluding the final observation.

EORTC QLQ-C30

The QLQ-C30 is a 30-item subject self-report questionnaire composed of both multi-item and single scales, including five functional scales (physical, role, emotional, social, and cognitive), three symptom scales (fatigue, nausea and vomiting, and pain), a global health status/quality of life scale, and six single items (dyspnea, insomnia, appetite loss, constipation, diarrhea, and financial difficulties). Subjects rate items on a four-point scale, with 1 as "not at all" and 4 as "very much."

Scores will be computed according to procedures outlined in the EORTC QLQ-C30 scoring manual, available at http://groups.eortc.be/qol/manuals. Change in the EORTC QLQ-C30 GHS/QoL score will be compared between the treatment arms. A linear mixed effects regression model with a covariance structure will be fitted to the longitudinal data to test for differences between the two treatment arms. For the linear mixed effects regression model, stratification factors and treatment arm will be included as fixed factors. Furthermore, time and treatment by time interaction will be included in the model. The repeated correlation structure in the timepoints will be assessed by using the Bayesian

Information Criterion (BIC). The following covariance structures will be explored: Unstructured (TYPE = UN), compound symmetry (TYPE = CS) and first-order autoregressive (TYPE = AR(1)). The type resulting in the lowest BIC will be used for analysis. The change from baseline in the QLQ-C30 Fatigue score will be compared between the two treatment arms at each post-baseline visit excluding the final observation.

10.3.7 Event-Free Survival

Event-free survival (EFS) will be defined as the number of days from randomization to the date of confirmed progressive disease, confirmed morphological relapse from CR or CRi, treatment failure defined as failure to achieve CR, CRi or MLFS after at least 6 cycles of study treatment collected on study drug completion eCRF or death from any cause. The confirmation is required for morphologic relapse (MR) and progressive disease (PD). Based on the IWG criteria as implemented in the protocol, if the bone marrow blasts are between 5 – 10%, a repeat bone marrow should be performed to distinguish between bone marrow regeneration and relapse. If there is an assessment of MR with bone marrow blasts ≤ 10% followed by a non-PD/non-MR response prior to start of post-treatment therapy, this assessment is considered as unconfirmed MR.

Confirmation PD is required by 2 consecutive assessments. Un-confirmed PD is defined as a PD followed by a non-PD/Non-MR prior to the post-treatment therapy. Otherwise, PD is considered as confirmed at the first observation date of PD.

If the specified event does not occur, subjects will be censored. The detailed censoring rule is described in Table 6. Data for subjects without any disease assessments performed after randomization will be censored at the time of randomization. All subjects in the full analysis set will be included in the analysis.

Table 6. Event/Censor and Corresponding Event/Censor Time for EFS

Situation	Event/Censor	Event Time/Censor Time
Confirmed PD, confirmed morphologic relapse (MR), and treatment failure events on or prior to any post-treatment therapy and data cutoff date, whichever is earliest	Event	Earliest event date of confirmed PD, MR, and treatment failure
Death on or prior to data cutoff date regardless whether it occurred after any post-treatment therapy initiated	Event	Death date
No confirmed MR, confirmed PD, treatment failure, or death events, and without post-treatment therapy initiated on or prior to the data cutoff date	Censor	Last disease assessment date (bone marrow or hematology lab collection date) on or prior to the data cutoff date
No confirmed MR, confirmed PD, or treatment failure on or prior to post-treatment therapy initiated on or prior to the data cutoff date, and no death on or prior to data cutoff	Censor	Start date of post-treatment therapy

Un-confirmed MR defined as a MR with bone marrow blast ≤ 10% and followed by non-PD/Non-MR prior to the post-treatment therapy.

Un-confirmed PD is defined as a PD followed by non-PD/Non-MR prior to the post-treatment therapy.

The distribution of EFS will be estimated for each treatment arm using Kaplan-Meier methodology and compared between treatment arms using the log-rank test stratified by age (18 – < 75, ≥ 75) and cytogenetic risk (intermediate, poor). Median EFS will be calculated and 95% confidence interval for median EFS will be presented by treatment arm. The hazard ratio between treatment arms will be estimated using the Cox proportional hazards model stratified by age (18 – < 75, ≥ 75) and cytogenetic risk (intermediate, poor).

10.4 Exploratory Efficacy Analysis

10.4.1 Additional Quality of Life Analyses

EQ-5D-5L

The EQ-5D-5L has five dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. These dimensions are measured on a five level 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 (0.0) to 1 (1.0) 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.

Additional exploratory analyses comparing the effects of venetoclax + azacitidine versus placebo + azacitidine will be performed on the following PRO measures: EQ-5D-5L, and the subscales/items from the EORTC QLQ-C30. The EQ-5D-5L status will be converted into a single preference-weighted health utility index score by applying country-specific weights (if available) or US weights (if not available). Descriptive statistics will be calculated as per the scoring manuals for all scales/items of the EORTC QLQ-C30, the EQ-5D-5L utility score, and the EQ-5D VAS score at each assessment. Linear mixed effects regression models will be used to test for differences between treatment arms, and mean change in values at each assessment will be calculated to identify any statistically significant differences versus baseline. Within-group changes from baseline at each assessment will also be assessed. The change from baseline in score will be compared between the two treatment arms at each post-baseline visit excluding the final observation. Additional analyses may include an assessment of time to deterioration and time to improvement.

10.4.2 Duration of CR + CRi

Duration of CR + CRi will be defined as the number of days from the date of first response (CR or CRi) per the modified IWG criteria for AML to the earliest evidence of confirmed morphologic relapse, confirmed progressive disease or death due to disease progression. If a subject is still responding at the data cutoff date, then the subject will be censored. The detailed censoring rule is described in Table 7. Duration of CR + CRi will be analyzed for all randomized subjects who achieve the best response of CR or CRi.

Table 7. Event/Censor and Corresponding Event/Censor Time for Duration of CR + CRi

Situation	Event/Censor	Event Time/Censor Time
Confirmed morphologic relapse (MR) or Confirmed progressive disease (PD) on or prior to any post-treatment therapy and data cutoff date, whichever is earliest	Event	Earliest MR or confirmed PD event date
Death due to disease progression on or prior to data cutoff date regardless whether it occurred after any post-treatment therapy initiated	Event	Death date
None of the above events	Censor	Last adequate disease assessment date (bone marrow or hematology lab collection date) on or prior to the earliest post-treatment therapy or data cutoff date

Un-confirmed MR defined as a MR with bone marrow blast ≤ 10% and followed by non-PD/Non-MR prior to the post-treatment therapy.

Un-confirmed PD is defined as a PD followed by non-PD/Non-MR prior to the post-treatment therapy.

The distribution of duration of CR + CRi will be estimated for each treatment arm using Kaplan-Meier methodology. Median duration of CR + CRi will be calculated and 95% confidence interval for median duration of CR + CRi will be presented by treatment arm.

Duration of CR will be defined as the number of days from the date of first CR per the modified IWG criteria for AML to the earliest evidence of confirmed morphologic relapse, confirmed progressive disease or death due to disease progression.

A sensitivity analysis will be performed by considering post-treatment therapy. If a subject received a post-treatment therapy prior to the data cutoff date and did not experience an event of duration of CR + CRi prior to starting the post-treatment therapy, then the subject's data will be censored at the start of the post-treatment therapy.

A sensitivity analysis will be performed by considering all deaths in the duration of CR + CRi.

10.4.3 Duration of CR + CRh and Duration of CR

Duration of CR + CRh will be defined as the number of days from the date of first response (CR or CRh) to the earliest evidence of confirmed morphologic relapse, confirmed progressive disease or death due to disease progression.

Similar analysis as duration of CR + CRi will be performed for duration of CR and duration of CR + CRh.

10.5 Efficacy Analysis per Independent Review Committee

An Independent Review Committee (IRC) will evaluate disease assessment data on the full analysis set as defined in the IRC Charter. The following efficacy endpoints, described in Section 10.2 and Section 10.3, will be analyzed based on this IRC review:

- CR rate
- CR + CRi rate
- CR + CRi rate by the initiation of Cycle 2
- EFS

10.6 Additional Efficacy Analyses

In addition to the stratified log-rank test for the primary and secondary efficacy endpoints, the following analyses may be performed for the comparison of OS, CR rate, CR + CRi rate, CR + CRh rate, CR + CRi rate by the initiation of Cycle 2, and CR + CRh rate by the initiation of Cycle 2 between the two treatment arms.

- Stratified Wilcoxon test, Un-stratified log-rank test and the Cox proportional hazards model for OS
- Stratified log-rank test for OS, adding potential prognostic factors that are not used in the randomization stratification (e.g., baseline bone marrow blast count)
- Censoring OS and DoR at the start of post-study treatment (e.g., intensive chemotherapy) if occurred prior to an event.

- CMH test for CR rate, CR + CRi rate, and CR + CRi rate by the initiation of Cycle 2 using IRC data.
- Fisher's exact test for CR rate, CR + CRi rate, CR + CRh rate, CR + CRi rate by the initiation of Cycle 2, and CR + CRh rate by the initiation of Cycle 2 using investigator disease assessment data and IRC data
- Modified OS, CR rate, CR + CRi rate, CR + CRh rate, CR + CRi rate by the initiation of Cycle 2, and CR + CRh rate by the initiation of Cycle 2 analyses to use all data available in the extracted database.

10.7 Handling of Multiplicity

10.7.1 General Consideration

For CR + CRi rate, CR + CRh rate, CR + CRi rate by the initiation of Cycle 2, CR + CRh rate by the initiation of Cycle 2, CR rate, transfusion independence rates, and MRD response rate, the information fraction to be used at IA2 (270 death events) is calculated as detailed in Table 8. The Lan DeMets alpha spending function and hierarchical testing strategy to control the familywise error rate (FWER).

Table 8. Calculation of Information Fraction (IF)

Endpoint	Information Fraction
CR + CRi rate	Portion of subjects with at least 6 months follow up since randomization
CR + CRh rate	Portion of subjects with at least 6 months follow up since randomization
EFS	The Lan-DeMets alpha spending function with O'Brien-Fleming boundary will be applied, assuming that 100% information fraction corresponds 360 EFS events
CR + CRi MRD response rate	Portion of subjects with at least 6 months follow up since randomization
MRD Negative Remission Rate	Portion of subjects with at least 9 months follow up since randomization
CR + CRi rate by the initiation of Cycle 2	Portion of subjects who started Cycle 2 Day 1 dosing or discontinued treatment by the end of Cycle 1
CR + CRh rate by the initiation of Cycle 2	Portion of subjects who started Cycle 2 Day 1 dosing or discontinued treatment by the end of Cycle 1
CR rate	Portion of subjects with at least 9 months follow up since randomization
Transfusion independence rates	Portion of subjects with at least 9 months follow up since randomization

10.7.2 Testing Strategy for EU and EU Reference Countries

For primary efficacy endpoints, a significance level of 0.01 (two-sided) will be allocated for the analysis of CR + CRi rate and a significance level of 0.04 (two-sided) will be allocated for the analysis of overall survival to ensure the control of the familywise error rate (FWER). If statistical test is significant for CR + CRi rate, the significance level 0.01 allocated to CR + CRi rate analysis will be recycled to overall survival analysis.

If statistical test is significant for the primary efficacy endpoint of OS, then the fixed sequence testing procedure will be performed with a significance level of 0.05 for key secondary efficacy endpoints sequentially. If statistical test is not significant for the primary efficacy endpoint of OS, then statistical significance will not be declared for any of the secondary efficacy endpoints. Table 9 presents the hierarchical ranking and alpha spending for each primary and key secondary endpoint at IA1, IA2 and FA.

Table 9. Alpha-spending Boundary (One-Sided p-value) for Each Ranked Endpoint for EU and EU Reference Countries

Endpoint		Timing of Analysis		
		IA1	IA2	FA
1	CR + CRi rate	0.005	No test	No test
1	OS	0.0001	As specified in Table 1	As specified in Table 1
3	CR + CRi rate by the initiation of Cycle 2[a]	No test	To be calculated with IF as specified in Table 8	To be calculated with IF as specified in Table 8
4	Post-baseline RBC transfusion independence[a]	No test	To be calculated with IF as specified in Table 8	To be calculated with IF as specified in Table 8
5	CR + CRi rate in IDH1/2 subgroup[a,b]	No test	To be calculated with IF as specified in Table 8	To be calculated with IF as specified in Table 8
6	CR rate[a]	No test	To be calculated with IF as specified in Table 8	To be calculated with IF as specified in Table 8
7	CR + CRi rate in FLT3 subgroup[a,b]	No test	To be calculated with IF as specified in Table 8	To be calculated with IF as specified in Table 8
8	Post-baseline platelet transfusion independence[a]	No test	To be calculated with IF as specified in Table 8	To be calculated with IF as specified in Table 8
9	EFS[a]	No test	To be calculated with IF as specified in Table 8	To be calculated with IF as specified in Table 8
10	CR + CRi MRD response rate[a]	No test	To be calculated with IF as specified in Table 8	To be calculated with IF as specified in Table 8
11	OS in IDH1/2 subgroup[b]	No test	0.0001	0.025 if CR + CRi is rejected; 0.02 otherwise
12	OS in FLT3 subgroup[b]	No test	0.0001	0.025 if CR + CRi is rejected; 0.02 otherwise
13	EORTC QLQ-C30 GHS/QoL	No test	0.0001	0.025 if CR + CRi is rejected; 0.02 otherwise
14	PROMIS Cancer Fatigue SF 7a	No test	0.0001	0.025 if CR + CRi is rejected; 0.02 otherwise

a. The Lan DeMets with O'Brien-Fleming approach will be applied to determine the efficacy boundaries of the following ranked secondary endpoints with the information fraction defined in Table 8.

b. If the size of identified biomarker subgroup population (FLT3, IDH1/2) is less than 50, or randomization ratio between two arms in each biomarker subgroup (FLT3, IDH1/2), or in each allelic ratio group of FLT3-ITD (≥ 0.5, < 0.5), or in each FLT3 mutation subtype (FLT3-ITD, FLT3-TKD), or in each IDH mutation subtype (IDH1, IDH2) is higher than 3:1, the endpoint will be ranked below PROMIS Cancer Fatigue SF 7a in the fixed sequence testing procedure. The biomarker subgroups are identified via central lab results.

10.7.3 Testing Strategy for US and US Reference Countries

For the primary endpoint OS, a significance level of 0.05 (two-sided) will be used for the analysis of OS. If statistical test is significant for the primary efficacy endpoint of OS, then the fixed sequence testing procedure will be performed with a significance level of 0.05 for key secondary efficacy endpoints sequentially. If statistical test is not significant for the primary efficacy endpoint of OS, then statistical significance will not be declared for any of the secondary efficacy endpoints. Table 10 presents the hierarchical ranking and alpha spending for each primary and key secondary endpoint at IA1, IA2 and FA.

The threshold to determine the MRD Negative remission rate will be defined separately using methods described in external SAP for MRD threshold validation. If the threshold of MRD Negative remission rate is not determined at time of analysis, this endpoint will be ranked below PROMIS Cancer Fatigue SF 7a and will be analyzed once the threshold is determined.

Table 10. Alpha-spending Boundary (One-Sided p-value) for Each Ranked Endpoint for US and US Reference Countries

Endpoint		Timing of Analysis		
		IA1	IA2	FA
1	OS	0.0001	As specified in Table 2	As specified in Table 2
2	CR + CRh rate[b]	No test	To be calculated with IF as specified in Table 8	To be calculated with IF as specified in Table 8
3	CR + CRi rate	0.025[a]		
4	CR + CRh rate by the initiation of Cycle 2[b]	No test	To be calculated with IF as specified in Table 8	To be calculated with IF as specified in Table 8
5	Post-baseline RBC transfusion independence[b]	No test	To be calculated with IF as specified in Table 8	To be calculated with IF as specified in Table 8
6	CR + CRh rate in IDH1/2 subgroup[a,b]	No test	To be calculated with IF as specified in Table 8	To be calculated with IF as specified in Table 8
7	CR rate[b]	No test	To be calculated with IF as specified in Table 8	To be calculated with IF as specified in Table 8
8	CR + CRh rate in FLT3 subgroup[b,c]	No test	To be calculated with IF as specified in Table 8	To be calculated with IF as specified in Table 8
9	Post-baseline platelet transfusion independence[b]	No test	To be calculated with IF as specified in Table 8	To be calculated with IF as specified in Table 8
10	EFS[b]	No test	To be calculated with IF as specified in Table 8	To be calculated with IF as specified in Table 8
11	MRD Negative remission rate[b,d]	No test	To be calculated with IF as specified in Table 8	To be calculated with IF as specified in Table 8
12	OS in IDH1/2 subgroup[c]	No test	0.0001	0.025
13	OS in FLT3 subgroup[c]	No test	0.0001	0.025
14	EORTC QLQ-C30 GHS/QoL	No test	0.0001	0.025
15	PROMIS Cancer Fatigue SF 7a	No test	0.0001	0.025

a. The testing of CR + CRi rate will be only based on the data used in primary analysis of CR + CRi rate (N = 225) at IA1. The full alpha (0.025) is utilized.

b. The Lan DeMets with O'Brien-Fleming approach will be applied to determine the efficacy boundaries of the following ranked secondary endpoints with the information fraction defined in Table 8.

c. If the size of identified biomarker subgroup population (FLT3, IDH1/2) is less than 50, or randomization ratio between two arms in each biomarker subgroup (FLT3, IDH1/2), or in each allelic ratio group of FLT3-ITD (≥ 0.5, < 0.5), or in each FLT3 mutation subtype (FLT3-ITD, FLT3-TKD), or in each IDH mutation subtype (IDH1, IDH2) is higher than 3:1, the endpoint will be ranked below PROMIS Cancer Fatigue SF 7a in the fixed sequence testing procedure. The biomarker subgroups are identified via central lab results.

d. The threshold to determine the MRD Negative remission rate will be defined separately using methods described in the external SAP for MRD threshold validation by Biomarker team. If the threshold of MRD Negative remission rate is not determined at time of analysis, this endpoint will be ranked below PROMIS Cancer Fatigue SF 7a and will be analyzed once the threshold is determined.

10.8 Efficacy Subgroup Analyses

To evaluate the impact of demographic and baseline characteristics on efficacy, subgroup analyses will be performed for CR rate, CR + CRi rate, CR + CRh rate, CR + CRi rate by the initiation of Cycle 2, CR + CRh rate by the initiation of Cycle 2, and overall survival for the full analysis set defined in Section 5.1. The subgroups defined below, not limited to, will be used for these analyses:

1. Gender (Male, Female)
2. Age (18 – < 65 years, 65 – < 75 years, ≥ 75 years)
3. Region (US, EU, China, JP, Asian, ROW)
4. Baseline ECOG (Grade < 2, ≥ 2)
5. Type of AML (De Novo, Secondary and therapy-related AML)
6. Cytogenetic risk (Intermediate, Poor)
7. Molecular marker by central lab (FLT3, IDH1/2, TP53, NPM1)
8. Antecedent hematologic history of MDS (Yes, No)
9. AML with Myelodysplasia related changes (AML-MRC)
10. Post-study treatment (Yes, No).

11.0 Safety Analyses

Safety assessments will only include subjects in the safety analysis set as described in Section 5.1. Safety will be assessed by treatment received for the safety analysis set.

11.1 Analysis of Treatment-Emergent Adverse Events

All summaries/analyses involving AEs will include treatment-emergent adverse events (TEAE) only, unless otherwise specified. TEAE are defined as any adverse event with onset or worsening from the day of that the first dose of study drug is administrated and

no more than 30 days after the last dose of 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 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, whichever comes first).

For summaries of AEs related (reasonable possibility) to study drug, at each level of summation (overall, SOC, and PT) each subject is counted only once. If a subject has an AE with unknown relationship, then the subject will be counted in the relationship category of "unknown," even if the subject has another occurrence of the same event with a relationship of "no reasonable possibility" present. The only exception is if the subject has another occurrence of the same AE with the relationship of "reasonable possibility." In this case, the subject will be counted under the reasonable probability category.

Adverse event data will be summarized and presented using primary MedDRA system organ classes (SOCs) and preferred terms (PTs) according to the MedDRA coding dictionary version 21.0 or higher.

The number and percentage of subjects experiencing treatment-emergent adverse events will be summarized for the following adverse event summaries:

- Any treatment-emergent adverse event.
- Any treatment-emergent adverse event with NCI toxicity (CTCAE V4.03) Grade 3, 4, or 5 adverse events.
- Any treatment-emergent serious adverse event.
- Any treatment-emergent adverse event with reasonable possibility related to venetoclax/placebo by the investigator.
- Any treatment-emergent adverse event leading to discontinuation of venetoclax/placebo.
- Any treatment-emergent adverse event leading to venetoclax/placebo interruption.

- Any treatment-emergent adverse event leading to venetoclax/placebo reduction.
- Any treatment-emergent adverse event with reasonable possibility related to azacitidine by the investigator.
- Any treatment-emergent adverse event leading to discontinuation of azacitidine.
- Any treatment-emergent adverse event leading to azacitidine interruption.
- Any treatment-emergent adverse event leading to azacitidine reduction.
- Any treatment-emergent adverse event leading to death.
- Any treatment-emergent adverse event leading to dose limiting toxicities (for China subjects in the AML safety cohort)

In addition, the treatment emergent AEs and serious AEs for, but not limited to, selected grouped preferred terms (PTs) listed in Table 11, will be summarized.

Table 11. Selected Adverse Events (but not limited to)

Selected Adverse Events	Search Criteria
Tumor Lysis Syndrome (AE) occurring within 7 days from the first dose study drug	SMQ – "Tumor Lysis Syndrome" (narrow)
Grade ≥ 3 neutropenia	PT terms – "neutropenia," "neutrophil count decreased," "febrile neutropenia," "agranulocytosis," "neutropenic infection," and "neutropenic sepsis"
Grade ≥ 3 infection, including opportunistic infections	SOC of "infections and infestations"
Haemorrhages	SMQ – "Haemorrhages" (narrow)
Anemia	PT terms – "Anaemia" and "Haemoglobin decreased"
Thrombocytopenia	PT terms – "Thrombocytopenia" and "Platelet count decreased"

There will be no statistical comparison for TEAEs between two treatment arms.

11.2 Deaths

The number of subject deaths will be summarized (1) for all deaths in this study regardless of the number of days after the last dose of study drug, (2) for deaths occurring within 30 days of the first dose of study drug, (3) for deaths occurring within 60 days of the first dose of study drug, (4) for deaths occurring within 30 days of the last dose of study drug, and (5) for deaths occurring more than 30 days of the last dose of study drug. There will be no statistical comparison for above analyses.

11.3 Analysis of Laboratory and Vital Signs Data

The value for baseline used in laboratory analyses is defined in Section 6.0. Post baseline visits windows are specified in Table 3. There will be no statistical comparison for laboratory data between two treatment arms.

11.3.1 Analysis of Mean Changes from Baseline in Clinical Laboratory and Vital Signs Data

Changes from baseline will be analyzed for each chemistry, hematology, and vital sign parameters at scheduled post-baseline visits and the final observation (Table 3). The mean and 95% confidence interval of change from baseline values are depicted for each visit by treatment arms. Laboratory tests to be summarized are included in Table 12 (if data is available).

Table 12. Clinical Laboratory Tests

Hematology	Clinical Chemistry
Hematocrit	Blood urea nitrogen (BUN)
Hemoglobin	Creatinine
Red blood cell (RBC) count	Calculated or Measured creatinine clearance
White blood cell (WBC) count	Total bilirubin
Neutrophils	Serum glutamic-pyruvic transaminase (SGPT/ALT)
Bands (if detected)	Serum glutamic-oxaloacetic transaminase (SGOT/AST)
Lymphocytes	Alkaline phosphatase
Monocytes	Sodium
Basophils (if detected)	Potassium
Eosinophils (if detected)	Calcium
Platelet count (estimate not acceptable)	Inorganic phosphorus
Blast count	Uric acid
Coagulation	Total protein
Prothrombin time (PT)	Glucose
Activated partial thromboplastin time (aPTT)	Albumin
	Lactate dehydrogenase (LDH)
	Chloride
	Bicarbonate

11.3.2 Analyses of Shift from Baseline in Clinical Laboratory Data

For shifts relative to National Cancer Institute Common Toxicity Criteria for Adverse Events (NCI CTCAE version 4.03), baseline and post-baseline laboratory observations will be categorized as Grade 0, Grade 1, Grade 2, Grade 3, or Grade 4 according to NCI CTCAE grade version 4.03.

The baseline and final grades will be defined respectively as the grade of the last measurement collected on or prior to the first dose of study drug (or randomization for non-treated subjects), and as the last post-baseline measurement collected no more than 30 days after the last dose of study drug.

The maximum NCI toxicity grade value is the value with highest NCI toxicity grade collected after the first dose of study drug (or randomization for non-treated subjects) and within 30 days following the last dose of study drug. 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 each variable, cross tables will be generated for the number of subjects with baseline values of Grade 0, Grade 1, Grade 2, Grade 3, Grade 4, or missing grade, versus maximum post-baseline/final observations of Grade 0, Grade 1, Grade 2, Grade 3, Grade 4, missing grade. All subjects in the safety analysis set 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).

For above shift tables, baseline grade of 0 (normal) was imputed for all subjects with at least one post-baseline but missing a baseline value for each lab test.

Detailed listings of data for subjects experiencing NCI CTCAE Grade 3 to 4 blood chemistry and hematology values will be provided. All measurements collected, regardless of the number of days after the last dose of study drug, will be included in these listings.

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 TBILI > 2 × ULN within 72 hours of each other) will be presented.

Number and percentage of subjects meeting the Howard criteria for TLS will be presented. The Howard Criteria for TLS is defined as two or more of the following treatment-emergent lab changes within a 24-hour period. The evaluation period for TLS is after 1st dose of study drug to 7 days from the 1st of dose study drug:

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

11.4 Safety Subgroup Analyses

Selected Safety analysis as described in Section 11.1, Section 11.2, and Section 11.3 may also be summarized for the subgroups in the safety analysis set defined below (but not limited to):

- Region (China, Japan, EU)
- Gender (Male, Female)
- Age (18 – < 65 years, 65 – < 75 years, ≥ 75 years)
- Hepatic impairment at baseline (Yes, No)
- Renal impairment at baseline (Yes, No)

12.0 Pharmacokinetic (PK) Analyses

Pharmacokinetic (PK) Analyses Plasma concentrations of venetoclax and azacitidine will be listed for each subject by arm and scheduled visit. Summary statistics will be computed for each arm and dose level by scheduled visit. Samples with significant sampling time deviations will be excluded from summary statistics calculations.

For the China subjects in the AML open label safety cohort, pharmacokinetic parameter values for venetoclax will be tabulated for each subject by visit, and summary statistics will be computed for each parameter.

13.0 References

1. Cheson BD, Bennett JM, Kopecky KJ, et al. Revised recommendations of the International Working Group for Diagnosis, Standardization of Response Criteria, Treatment Outcomes, and Reporting Standards for Therapeutic Trials in Acute Myeloid Leukemia. J Clin Oncol. 2003;21(24):4642-9.
2. Döhner H, Estey E, Grimwade D, et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. Blood. 2017;129(4):424-47.
3. Lan KKG, DeMets DL. Discrete sequential boundaries for clinical trials. Biometrika. 1983;70(3):659-63.
4. Bretz F, Maurer W, Brannath W, et al. A graphical approach to sequentially rejective multiple test procedures. Stat Med. 2009;28(4):586-604.
5. Howard SC, Jones DP, Pui CH. The tumor lysis syndrome. N Engl J Med. 2011;364(19):1844-54.
6. 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.
7. Xu T, Qin Q, Wang X. Defining information fraction in group sequential clinical trials with multiple endpoints. Contemp Clin Trials Commun. 2018;10:77-9.

14.0 Appendix

List of Abbreviations

AML	acute myeloid leukemia
BCL	B cell lymphocyte
BTD	Breakthrough Therapy Designation
CI	confidence interval
CR	complete remission
CRi	complete remission with incomplete blood count recovery
CRh	complete remission with partial hematologic recovery
CTCAE	Common Terminology Criteria for Adverse Events
ECOG	Eastern Cooperative Oncology Group
EFS	event-free survival
EORTC-QLQ-C30	European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Core
FLT3	FMS-like tyrosine kinase 3
GFR	glomerular filtration rate
HMA	hypomethylating agent
IDMC	Independent data monitoring committee (also called DMC)
IRC	independent review committee
ITD	internal tandem duplication FLT3
IWG	International Working Group
MDS	myelodysplastic syndrome
MLFS	morphologic leukemia-free state
MR	morphologic relapse
MRD	minimal residual disease
OS	overall survival
PD	progressive disease
PR	partial remission
PROMIS	Patient Reported Outcomes Measurement Information System
QD	once daily
TLS	tumor lysis syndrome
US	United States