Official Title: A Phase 3, Multicenter, Double-Blind, Randomized, Placebo-Controlled Study of AG-120 in Combination with Azacitidine in Subjects ≥18 Years of Age with Previously Untreated Acute Myeloid Leukemia with an IDH1 Mutation

NCT Number: NCT03173248

Document Dates: SAP Version 1.0: 23-June-2020

STATISTICAL ANALYSIS PLAN

A Phase 3, Multicenter, Double-Blind, Randomized, Placebo-Controlled Study of AG-120 in Combination with Azacitidine in Subjects ≥18 Years of Age with Previously Untreated Acute Myeloid Leukemia with an IDH1 Mutation

AG120-C-009

Version: v1.0

Date: 22-Jun-2020

Prepared by:

[redacted], PhD
Study Statistician

DocuSigned by:
[redacted]
Signer Name: [redacted]
Signing Reason: I am the author of this document
Signing Time: 22-Jun-2020 | 4:42 PM EDT
040E544D84204815B0793151FB226AE5

22-Jun-2020 | 4:43 PM EDT

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

Approved by:

[redacted], MD
Study Medical Monitor

DocuSigned by:
[redacted]
Signer Name: [redacted]
Signing Reason: I approve this document
Signing Time: 22-Jun-2020 | 9:22 PM EDT
9749FA1E99B34303A41D7451A6872A04

22-Jun-2020 | 9:22 PM EDT

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

[redacted], PhD
[redacted]

DocuSigned by:
[redacted]
Signer Name: [redacted]
Signing Reason: I approve this document
Signing Time: 23-Jun-2020 | 9:25 AM EDT
1733C50FC95F44F39DEF81FD0E9BEA2A

23-Jun-2020 | 9:25 AM EDT

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

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND DEFINITIONS OF TERMS

Abbreviation	Definition
2-HG	2-hydroxyglutarate
AE	Adverse event
AESI	Adverse event of special interest
ALP	Alkaline phosphatase
ALT	Alanine aminotransferase
AML	Acute Myeloid Leukemia
ANC	Absolute neutrophil count
AST	Aspartate aminotransferase
ATC	Anatomical Therapeutic Chemical
BAS	Biomarker Analysis Set
BMI	Body mass index
BMMC	Bone marrow mononuclear cell
BSA	Body surface area
CI	Confidence interval
CMH	Cochran-Mantel-Haenszel
CR	Complete remission
CRh	CR with partial hematologic recovery
CRi	CR with incomplete hematologic recovery
CRp	CR with incomplete platelet recovery
CSR	Clinical study report
CTCAE	Common Terminology Criteria for Adverse Events
DBL	Database lock
DI	Dose intensity
DOCR	Duration of CR
DOCRh	Duration of CR+CRh
DOCRi	Duration of CR+CRi (including CRp)
DOR	Duration of response
ECG	Electrocardiogram
ECHO	Echocardiogram
ECOG PS	Eastern Cooperative Oncology Group performance status
eCRF	Electronic case report form

EFS	Event-free survival
EORTC	European Organisation for Research and Treatment of Cancer
EORTC-QLQ-C30	European Organization of Research and Treatment of Cancer – Quality of Life Questionnaire – Core Questionnaire
EOS	End of study
EOT	End of treatment
FAS	Full Analysis Set
HR	Hazard ratio
HSCT	Hematopoietic stem cell transplant
IA	Interim analysis
IC	Intensive chemotherapy
IDH1m	Isocitrate dehydrogenase 1 mutation-positive
I-DMC	Independent Data Monitoring Committee
IRT	Interactive response technology
ITT	Intention-to-treat
IV	Intravenous
IWG	International Working Group
KM	Kaplan-Meier
LDH	Lactate dehydrogenase
LLN	Lower limit of normal
LVEF	Left ventricular ejection fraction
MC	Mutation clearance
MDS	Myelodysplastic syndrome
MedDRA	Medical Dictionary for Regulatory Activities
MLFS	Morphologic leukemia-free state
MPD	Myeloproliferative disorders
MUGA	Multi-gated acquisition
NCCN	National Comprehensive Cancer Network
NCI	National Cancer Institute
NE	Not evaluable
OR	Objective response
ORR	Objective response rate
OS	Overall survival
PBMC	Peripheral blood mononuclear cell

PD	Pharmacodynamic
PK	Pharmacokinetic
PH	Proportional hazard
PO	Oral
PPS	Per-Protocol Set
PR	Partial remission
PS	Performance status
PT	Preferred Term
Q	Every
QD	Once daily
QoL	Quality of life
QTc	Heart rate-corrected QT interval
QTcB	Heart rate-corrected QT interval using the Bazett's formula
QTcF	Heart rate-corrected QT interval using the Fridericia's formula
RBC	Red blood cell
RDI	Relative dose intensity
RMST	Restricted mean survival time
ROW	Rest of world
SAE	Serious adverse event
SAP	Statistical analysis plan
SC	Subcutaneous
SD	Standard deviation
SD	Stable disease
SE	Standard error
SOC	System Organ Class
TEAE	Treatment-emergent adverse event
TF	Treatment failure
TTCR	Time to CR
TTCRh	Time to CR+CRh
TTCRi	Time to CR+CRi (including CRp)
TTR	Time to first response
ULN	Upper limit of normal
VAF	Variant allele frequency

WBC	White blood cell
WHO	World Health Organization

1. VERSION HISTORY

This statistical analysis plan (SAP) describes the analysis associated with protocol AG120-C-009, Version 7.0 (dated 04-Mar-2020).

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

Version	Version Date	Rationale and Summary of Changes
1.0	22-Jun-2020	Original version.

2. INTRODUCTION

This SAP provides the detailed methodology for summary and statistical analyses of the data collected in Study AG120-C-009, except for pharmacokinetic (PK)/pharmacodynamic (PD) data, which will be described in a separate SAP. This document may modify the plans outlined in the protocol; however, any major modifications of the primary endpoint definition or its analysis will also be reflected in a protocol amendment.

The clinical study report (CSR) will include all data up to a data cutoff date that is determined based on the number of events required for the final analysis of the primary endpoint [event-free survival (EFS)] and a minimum follow-up of 24 weeks for all subjects randomized.

The data cutoff date will be defined prospectively once a data extract [before database lock (DBL)] is available that indicates that all subjects have been randomized and have the minimum duration of follow-up and that the required number of events for EFS is expected to occur by the cutoff date. The final number of events might deviate from the planned number; the data cutoff date will not be adjusted retrospectively in this case.

End of Study (EOS) is defined as the time at which all subjects have died, discontinued the study, are lost to follow-up, or have withdrawn consent; or when the Sponsor ends the study.

3. TRIAL OBJECTIVES AND ENDPOINTS

3.1. Objectives

3.1.1. Primary Objective

The primary objective of the study is to compare EFS between AG-120+azacitidine and placebo+azacitidine.

3.1.2. Secondary Objectives

The key secondary objectives of the study are:

- To compare the complete remission (CR) rate between AG-120+azacitidine and placebo+azacitidine
- To compare overall survival (OS) between AG-120+azacitidine and placebo+azacitidine

- To compare the CR + CR with partial hematologic recovery (CRh) rate between AG-120+azacitidine and placebo+azacitidine; CRh will be derived by the Sponsor
- To compare the objective response rate (ORR) between AG-120+azacitidine and placebo+azacitidine

Additional secondary objectives are:

- To compare the CR+CR with incomplete hematologic recovery (CRi) (including CR with incomplete platelet recovery [CRp]) rate between AG-120+azacitidine and placebo+azacitidine
- To compare duration of CR (DOCR), duration of CR+CRh (DOCRh), duration of response (DOR), and duration of CR+CRi (including CRp) (DOCRi) between AG-120+azacitidine and placebo+azacitidine
- To compare time to CR (TTCR), time to CR+CRh (TTCRh), time to first response (TTR), and time to CR+CRi (including CRp) (TTCRi) between AG-120+azacitidine and placebo+azacitidine
- To assess the safety and tolerability of treatment with AG-120+azacitidine compared with placebo+azacitidine
- To compare transfusion requirements (platelet and red blood cell [RBC]; number of units transfused), infection rates, days spent hospitalized, and other efficacy and safety measures that are potentially indicative of clinical benefit between AG-120+azacitidine and placebo+azacitidine
- To assess the impact of treatment on quality of life (QoL) using the European Organisation for Research and Treatment of Cancer (EORTC) QLQ-C30 and EQ-5D-5L
- To evaluate the PK of AG-120 as administered in combination with azacitidine
- To evaluate the PK/PD relationship of AG-120 and 2-HG in blood samples in comparison with placebo
- To compare rates of CR with IDH1 mutation clearance (MC) between AG-120+azacitidine and placebo+azacitidine

3.2. Endpoints

Investigator disease assessments per modified IWG response criteria will be used for all response related efficacy endpoints, with the exception of CRh, which will be derived by the Sponsor.

3.2.1. Primary Endpoint

The primary endpoint of the study is EFS, which is defined as the time from randomization until treatment failure (TF), relapse from remission, or death from any cause, whichever occurs first. TF is defined as failure to achieve CR by Week 24.

3.2.2. Secondary Endpoints

The key secondary endpoints are:

- CR rate (CR defined as bone marrow blasts <5% and no Auer rods, absence of extramedullary disease, absolute neutrophil count (ANC) $\geq 1.0\times10^9$/L [1000/μL], platelet count $\geq 100\times10^9$/L [100,000/μL], and independence of RBC transfusions)
- OS, defined as the time from date of randomization to the date of death due to any cause
- CR+CRh rate (CRh is defined as a CR with partial recovery of peripheral blood counts where ANC is $>0.5\times10^9$/L [500/μL], and platelet count is $>50\times10^9$/L [50,000/μL]; CRh will be derived by the Sponsor)
- ORR, defined as the rate of CR, CRi (including CRp), Partial remission (PR), and Morphologic leukemia-free state (MLFS)

Additional secondary endpoints are:

- CR+CRi (including CRp) rate (CRi [including CRp] is defined as all CR criteria except for residual neutropenia where ANC is $<1.0\times10^9$/L [1000/μL] or thrombocytopenia where platelet count is $<100\times10^9$/L [100,000/μL]; without platelet transfusion for at least 1 week prior to disease assessment)
- DOCR, among subjects who achieved CR; DOCRh, among subjects who achieved CR or CRh; DOR, among subjects who achieved CR, CRi (including CRp), PR or MLFS; and DOCRi, among subjects who achieved CR or CRi (including CRp)
- TTCR, among subjects who achieved CR; TTCRh, among subjects who achieved CR or CRh; TTR, among subjects who achieved CR, CRi (including CRp), PR or MLFS; and TTCRi, among subjects who achieved CR or CRi (including CRp)
- Vital signs, and results of Eastern Cooperative Oncology Group performance status (ECOG PS), electrocardiogram (ECG), and echocardiogram (ECHO) or multi-gated acquisition (MUGA) for left ventricular ejection fraction (LVEF) as clinically indicated (method per institutional standard of care, with the same method used for an individual throughout the study; sites in Germany may only use ECHO.)

- Clinical laboratory assessments (hematology, chemistry, and coagulation)
- Adverse Events (AEs), AEs of special interest (AESIs), SAEs, and AEs leading to discontinuation or death
- Concomitant medication use
- Transfusion requirements (platelet and RBC; number of units transfused), rates of infection, days spent hospitalized, and other efficacy and safety measures that are potentially indicative of clinical benefit
- Changes from baseline in QoL assessments (EORTC QLC-C30 and EQ-5D-5L)
- Rates of CR with IDH1 MC
- AG-120/placebo and azacitidine drug exposure, including dose modifications and dose intensities
- AG-120 and 2-HG concentrations in circulating plasma

4. STUDY DESIGN

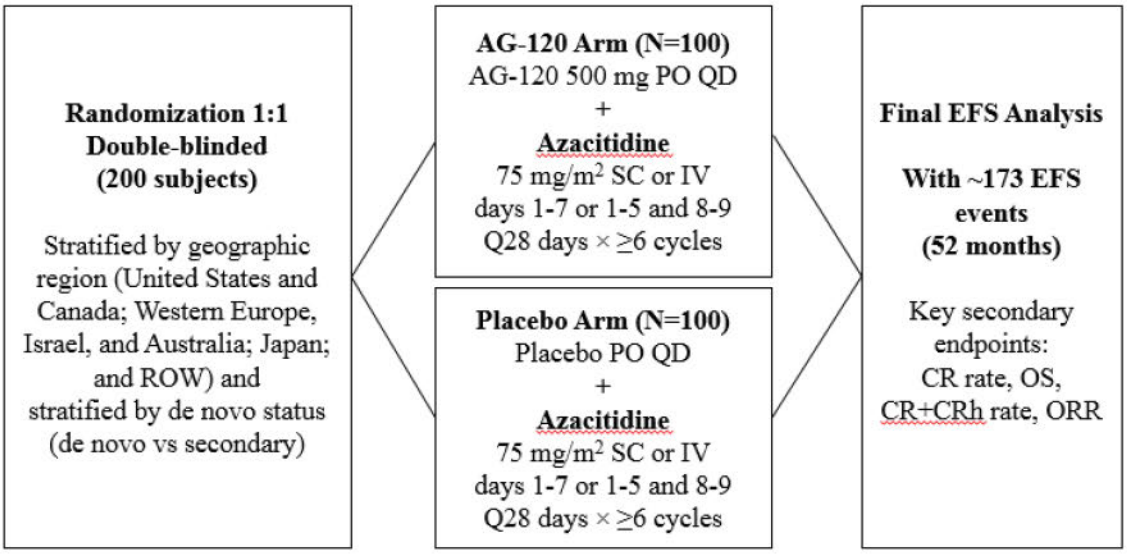
Study AG120-C-009 is a global, Phase 3, multicenter, double-blind, randomized, placebo-controlled clinical trial to evaluate the efficacy and safety of AG-120+azacitidine vs placebo+azacitidine in adult subjects with previously untreated isocitrate dehydrogenase 1 mutation-positive (IDH1m) AML and who are considered appropriate candidates for non-intensive therapy.

Following provision of informed consent, all subjects will undergo Screening procedures within 4 weeks (28 days) prior to randomization to determine eligibility.

- Gene mutation analysis for confirmation of IDH1m disease from a bone marrow and germ-line mutation analysis from a buccal swab will be conducted for all subjects, and can be conducted prior to the 28-day Screening window.
- Central laboratory confirmation of IDH1m status is required for study eligibility.
- With Medical Monitor approval, subjects may be eligible and randomized with local IDH1m testing results; however, bone marrow aspirate for central laboratory testing must have been sent with proof of shipment to the central laboratory prior to randomization. In cases where bone marrow aspirate is not available (ie, dry tap), peripheral blood samples may be used for IDH1m confirmation with Medical Monitor approval.

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

Figure 1: Study Schema



Abbreviations: CR = complete remission; CRh = CR with partial hematologic recovery; EFS = event-free survival; IV = intravenous; ORR = objective response rate; OS = overall survival; PO = oral; Q = every; QD = once daily; ROW = Rest of World; SC = subcutaneous.

Subjects should be treated for a minimum of 6 cycles of combination therapy unless they experience relapse after achieving a CR, CRi (including CRp), or MLFS; disease progression after having previously attained PR or stable disease (SD); unacceptable toxicity

(AE); confirmed pregnancy; withdrawal by subject; protocol violation; death; or End of Study in which cases the subjects may be treated for less than 6 cycles.

Treatment will be administered as follows:

- All subjects will receive azacitidine 75 mg/m^2/day SC or IV for the first week (7 days) (or on a 5-2-2 schedule) of each 4-week (28-day) cycle in combination with AG-120 or placebo QD on each day of the 4-week cycle. The same schedule should be used for each subject throughout the duration of treatment, when possible.
- Subjects should continue to receive therapy with AG-120 or placebo+azacitidine until death, disease relapse, disease progression, development of unacceptable toxicity (AE), confirmed pregnancy, withdrawal by subject, protocol violation, or End of Study.
    - Disease progression (defined only for subjects who have previously attained PR or SD) is defined as evidence for an increase in bone marrow blast percentage and/or increase of absolute blast counts in the blood: 1) >50% increase in bone marrow blast count over baseline (a minimum 15% point increase is required in cases with <30% blasts at baseline); or persistent marrow blast percentage of >70% over at least 3 months; without at least a 100% improvement in ANC to an absolute level (>500/µL, and/or platelet count to >50,000/µL non-transfused); 2) >50% increase in peripheral blasts (white blood cells [WBCs] × % blasts) to >25,000/µL in the absence of treatment-related differentiation syndrome; or 3) new extramedullary disease
    - Subjects with a response less than CR at 24 weeks or beyond can continue on treatment if demonstrating treatment benefit, defined as any of the following: 1) transfusion-independence while on study treatment; 2) ANC >500/µL; or 3) platelets >50,000/µL

5. ANALYSIS DATA SETS

Only subjects who sign informed consent and are screened will be included in the analysis sets below.

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

- The Full Analysis Set (FAS) will include all subjects who are randomized. Subjects will be classified according to the randomized treatment arm. (Note: this data set is referred to as the Intent-to-Treat Analysis Set in the protocol.)
- The Safety Analysis Set will include all subjects who receive at least 1 dose of the study treatment. Subjects will be classified according to the treatment received, where treatment received is defined as:
  - The randomized treatment if it is received at least once, or
  - The first treatment received if the randomized treatment is never received
- The Per-Protocol Set (PPS) is a subset of the FAS. Subjects who meet any of the following criteria will be excluded from the PPS:
  - Do not receive at least 1 dose of the randomized treatment
  - Eligible for intensive chemotherapy (IC) (ie, do not meet Inclusion Criterion 1)
  - Do not have an IDH1 mutation as determined by central laboratory testing (ie, do not meet Inclusion Criterion 3)
  - Have an ECOG PS score >2 (ie, do not meet Inclusion Criterion 4)
  - Have received any prior treatment for AML with the exception of nononcolytic treatments to stabilize disease such as hydroxyurea or leukapheresis (ie, do not meet Exclusion Criterion 2)
  - Have received any prior hypomethylating agent (ie, do not meet Exclusion Criterion 3)
  - Have received any prior IDH1 inhibitor (ie, do not meet Exclusion Criterion 5)
- The Biomarker Analysis Set (BAS) is a subset of the FAS and will include all subjects who have at least 1 on-treatment biomarker sample providing valid IDH1m variant allele frequency (VAF) data.

Table 2 summarizes the use of the analysis sets.

Table 2: Analysis Sets for Each Endpoint

Endpoints	Full Analysis Set (FAS)	Per-Protocol Set (PPS)	Safety Analysis Set	Biomarker Analysis Set (BAS)
Demographic and other baseline characteristics	✓		✓	
Disposition	✓			
Major protocol deviations	✓			
Subsequent therapies	✓			
Exposure and concomitant therapies			✓	
Efficacy	✓	✓ (primary and key secondary only*)		
Safety			✓	
Biomarkers				✓

* Key secondary endpoints are CR, OS, CR+CRh and objective response (OR).

6. GENERAL STATISTICAL CONSIDERATIONS

6.1. Randomization, Blinding, Unblinding, and Crossover

Subjects will be randomized in a 1:1 ratio to one of the following treatment arms:

- AG-120+azacitidine=AG-120 500 mg QD orally plus azacitidine 75 mg/m^2/day SC or IV
- Placebo+azacitidine=AG-120-matched placebo QD orally plus azacitidine 75 mg/m^2/day SC or IV

Randomization assignment will be implemented by interactive response technology (IRT) and stratified by:

- AML status (de novo AML; secondary AML)
- Geographic region (United States and Canada; Western Europe, Israel, and Australia; Japan; ROW)

This is a double-blind study therefore the subjects, Investigators, Sponsor, and the clinical research unit staff who work directly with subjects will all be blinded to study treatment assignment.

At the time of the DBL for the final analysis of the primary endpoint, the study will be unblinded. Subjects on treatment at that time may continue receiving study treatment.

After unblinding, and if the benefit-risk profile favors AG-120+azacitidine, subjects randomized to placebo+azacitidine may be offered the choice to receive AG-120.

6.2. Sample Size Determination and Decision Rules

6.2.1. Sample Size Determination

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

$H_{01}: S_{EFS\_T}(t) \leq S_{EFS\_C}(t)$ vs. $H_{11}: S_{EFS\_T}(t) > S_{EFS\_C}(t), t \geq 0$

where $S_{EFS\_T}(t)$ is the survival distribution function of EFS in the AG-120+azacitidine arm (referred as treatment arm) and $S_{EFS\_C}(t)$ is the survival distribution function of EFS in the placebo+azacitidine arm (referred as control arm).

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

$H_{02}: \text{OddsRatio}_{(CR)}=1$ vs $H_{12}: \text{OddsRatio}_{(CR)} > 1$

where $\text{OddsRatio}_{(CR)}$ is the ratio of the odds of CR between the treatment arm and the control arm.

$H_{03}: S_{OS\_T}(t) \leq S_{OS\_C}(t)$ vs $H_{13}: S_{OS\_T}(t) > S_{OS\_C}(t), t \geq 0$

where $S_{OS\_T}(t)$ is the survival distribution function of OS in the treatment arm and $S_{OS\_C}(t)$ is the survival distribution function of OS in the control arm.

H_{04}: $OddsRatio_{(CR+CRh)}=1$ vs H_{14}: $OddsRatio_{(CR+CRh)}>1$

where $OddsRatio_{(CR+CRh)}$ is the ratio of the odds of CR+CRh between the treatment arm and the control arm.

H_{05}: $OddsRatio_{(OR)}=1$ vs H_{15}: $OddsRatio_{(OR)}>1$

where $OddsRatio_{(OR)}$ is the ratio of the odds of OR between the treatment arm and the control arm.

To control the overall Type 1 error rate at the 1-sided 2.5% level, the fixed sequence testing procedure (Westfall and Krishen, 2001) will be used to adjust for multiple statistical testing of the primary and key secondary efficacy endpoints. These endpoints will be tested in the following order:

- EFS
- CR
- OS
- CR+CRh
- OR

No interim analyses for efficacy are planned.

A total of approximately 200 subjects with previously untreated IDH1m AML will be randomized in this study. The calculations and simulations described below were performed using East Version 6.05 and R Version 3.4.4.

Primary Endpoint - EFS

Based on the definition of EFS, subjects who do not achieve CR by 24 weeks [treatment failure (TF)] will be considered to have had an EFS event on the day of randomization; for subjects who achieve CR by 24 weeks (responders), EFS will be the time from randomization to the earliest of relapse from remission or death from any cause. Whereas it is reasonable to assume that EFS for responders in each treatment arm will follow an exponential distribution, this assumption will be clearly violated for the EFS endpoint which includes both TF and responders components as described below.

Assumptions for the placebo+azacitidine arm are based on results from Study AZA-AML-001 in newly diagnosed AML subjects who are ineligible for IC receiving azacitidine. Data from Study AZA-AML-001 were obtained under a data exchange agreement with Celgene (now Bristol-Myers Squibb), and based on results from a retrospective analysis of these data, the following are assumed in this study:

- CR rate at 24 weeks=20%
- Median EFS=14.6 months, for subjects who achieve CR by 24 weeks

Assumptions for the AG-120+azacitidine arm are based on results from Study AG-221-AML-005 in newly diagnosed AML subjects who are ineligible for IC receiving AG-120 in combination with azacitidine:

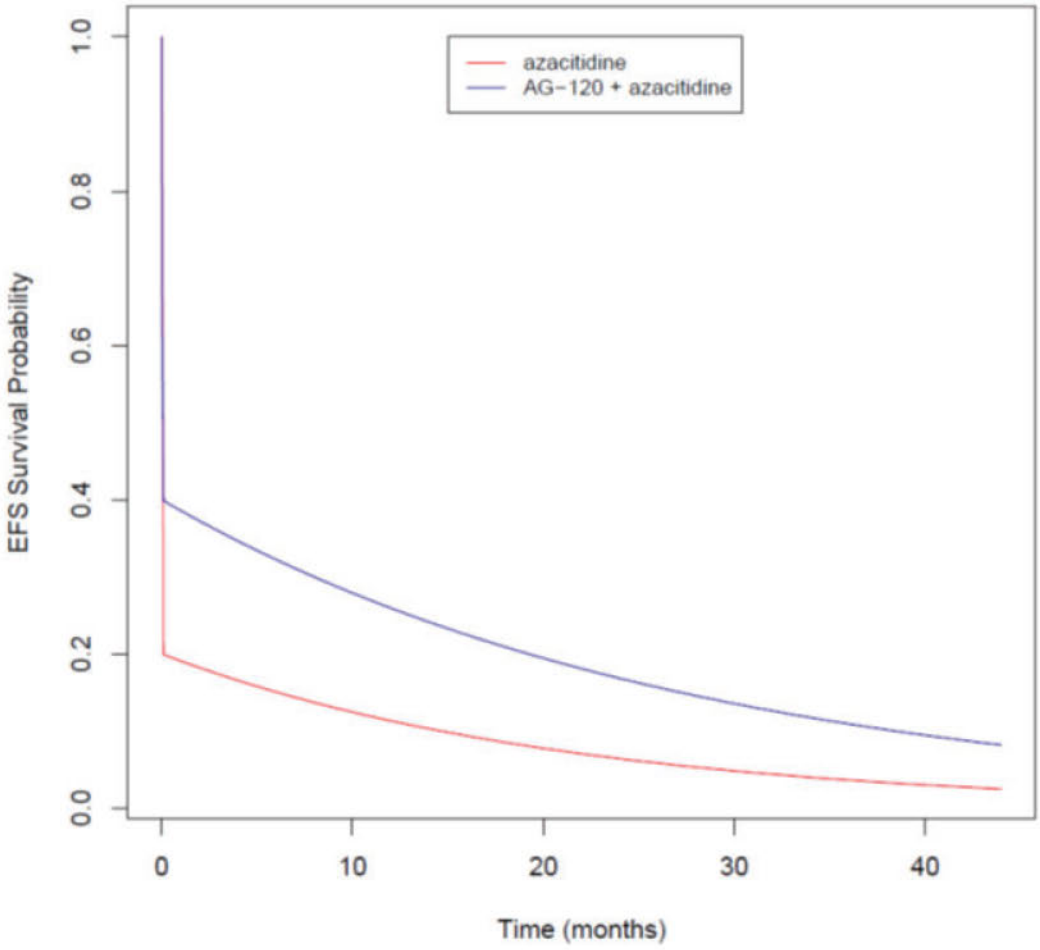
- CR rate at 24 weeks=40%

- Median EFS=19.2 months, for subjects who achieve CR by 24 weeks; assuming an exponential distribution, this corresponds to a hazard ratio (HR) of 0.76 for the estimated effect of treatment compared to control.

The EFS curves for both the placebo+azacitidine and AG-120+azacitidine arms based on the above assumptions are displayed in Figure 2. Censoring and dropout considerations are not included in the diagram for simplicity. The EFS curve has 2 distinct components with the first component driven by the subjects who have TF and the second component driven by responders.

The treatment effects for the 2 components are clearly distinct and the assumption of proportional hazards will be violated. Therefore, instead of the overall HR, the assumptions for EFS are presented for the 2 components separately. Based on simulation results, the average overall HR over 10,000 simulations for the entire population is 0.641. However, the overall HR for the entire population was not part of the study design assumptions. Additionally, based on the current assumptions, for the overall population, the median EFS will be 1 day for both the placebo+azacitidine and AG-120+azacitidine arms.

Figure 2: Event-Free Survival under Protocol Assumptions



Abbreviations: EFS = event-free survival.

Under these assumptions, a total of 173 EFS events in approximately 200 randomized subjects with previously untreated IDH1m AML are required to provide 80% power at a 1-sided 2.5% level of significance to reject H_{01} using a stratified log-rank test (Table 3).

Table 3: Study Design Parameters for EFS

Design Parameters and Outcomes	
Power	80%
Significance level (1-sided)	2.5%
CR rate by 24 weeks (treatment vs control)	40% vs 20%
HR (treatment vs control) that can be detected for subjects who achieve CR by 24 weeks	0.76
Median EFS for subjects who achieve CR by 24 weeks (treatment vs control)	19.2 vs 14.6 months
Number of subjects randomized (1:1)	200
Target number of EFS events	173 EFS events
Accrual rate	3 subjects/month for the first 10 months, 5 subjects/month thereafter
Accrual time	44 months
Dropout rate	5%
Time to final analysis under H_{11}	52 months after first subject randomized

Abbreviations: CR = complete remission; EFS = event-free survival; HR = hazard ratio.

Overall Survival

Based on available data (Dombret et al, 2015), the median OS in the placebo+azacitidine arm is assumed to be 10.4 months. A target HR of 0.71 for OS (equivalent to a median OS of 10.4 months in the placebo+azacitidine arm vs 14.6 months in the AG-120+azacitidine arm, assuming an exponential distribution) is assumed. Based on 200 randomized subjects and the above assumptions, it is estimated that at the time of the final analysis for EFS, approximately 145 deaths will be observed. With 145 deaths, the study will have 54% power to detect a HR for OS of 0.71 at a 1-sided 2.5% level of significance using a stratified log-rank test.

CR, CR+CRh, and OR

Based on a retrospective analysis of Study AZA-AML-001, the response rates in the placebo+azacitidine arm are assumed to be 20%, 25%, and 32% for CR, CR+CRh, and OR, respectively.

Based on 200 randomized subjects, the power for the analysis of each of these endpoints at a 1-sided 2.5% level of significance to reject the corresponding null hypotheses using a Cochran-Mantel-Haenszel (CMH) test are presented in Table 4.

Table 4: Power for Analyses of CR, CR+CRh, and OR

Endpoints	Number of Subjects	Assumed Response Rate under the Alternative Hypothesis (Treatment vs Control)	Power
CR	200	40% vs 20%	86%
CR+CRh	200	45% vs 25%	84%
OR	200	52% vs 32%	81%

Abbreviations: CR = complete remission; CRh = complete remission with partial hematologic recovery; OR = objective response; vs = versus.

6.2.2. Decision Rules

The final analysis for the primary and key secondary efficacy endpoints will be performed after all subjects have been randomized in the study and followed for a minimum of 24 weeks and the target number of events for EFS has occurred (see Section 2).

The basis for a claim of efficacy will be the statistical significance of EFS in favor of the AG-120+azacitidine arm when the 1-sided p-value <0.025. To protect the integrity of the study and to preserve the Type 1 error at or below 1-sided α=0.025, a hierarchical testing strategy will be used as described in Section 6.2.1 to test each of the key secondary efficacy endpoints and statistical significance will be achieved with 1-sided p-value <0.025 for each endpoint.

6.3. Definitions

6.3.1. Study Drug and Study Treatment

Study drug is defined as AG-120 or matched placebo or azacitidine.

Study treatment is defined as AG-120+azacitidine or matched placebo+azacitidine.

6.3.2. Start and End Dates of Study Drug and Study Treatment

The **start of study drug** (AG-120 or matched placebo or azacitidine) is the earliest date of administration of a non-zero dose of the study drug.

The **end of study drug** (AG-120 or matched placebo or azacitidine) is the latest date of administration of a non-zero dose of the study drug on or before the data cutoff date.

The **start of study treatment** is the earliest of the non-missing start of any study drug.

The **end of study treatment** is the latest of the non-missing end of any study drug.

6.3.3. Study Day

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

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

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

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

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

6.3.4. Baseline

For efficacy evaluations, the last adequate assessment on or before the date of randomization will be used as the baseline. Per protocol, the first QoL assessment is planned to occur on Cycle 1 Day 1 before the start of study treatment. Therefore, for QoL assessments only, if there is no value available on or before the date of randomization, then the last measurement on or before the start of study treatment will be used as the baseline.

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

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

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

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

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

6.3.5. On-Treatment Period

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

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

6.3.6. Start of Subsequent Anticancer Therapy

The start of subsequent anticancer therapy is used in censoring for efficacy analyses. The earliest start date, on or after randomization, captured in the Anticancer Therapy eCRF page will be used in the analyses as the start of subsequent anticancer therapy. If the option "HSCT Conditioning" is selected, the corresponding date will not be used in the derivation of start of subsequent anticancer therapy.

6.3.7. Last Contact Date

The last contact date will be derived using the last complete date in the eCRF on or before the data cutoff date, from among the following:

- All assessment dates (eg, vital signs assessment, ECG assessment)
- Dates of administration of study drug, concomitant medications, and subsequent anticancer therapies
- Start and end dates of AEs
- Last contact date collected on the Survival Follow-up eCRF when the subject status is alive
- Randomization date
- Withdrawal of consent date
- Date of discontinuation on disposition eCRF pages. If the option "Lost to Follow-up" is selected, the corresponding date will not be used in the derivation of last contact date

Notes:

- Only dates associated with actual examinations of the subject will be used in the derivation. Dates associated with a technical operation unrelated to subject status, such as the date a blood sample is processed, will not be used.
- Assessment dates after the data cutoff date will not be used to derive the last contact date.

6.4. General Methods

6.4.1. Data Handling After Cutoff Date

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

6.4.2. Standard Derivations and Reporting Conventions

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

The following derivations will be implemented.

- Age (years):
  - (date of informed consent−date of birth+1)/365.25
  - If only day of birth is missing: Age (years): (year/month of informed consent−year/month of birth)
  - If day and month of birth are missing: Age (years): (year of informed consent−year of birth)

  The integer part of the calculated age will be used for reporting purposes.

- BMI (kg/m^2)=weight (kg)/height $(m)^2$
- BSA (m^2) as collected in the eCRF
- Duration (in days) from a reference date (eg, randomization date, start date of study treatment)=
  - date of event−reference date+1, if the date of the event is on or after the reference date
  - date of event−reference date, if the date of the event is before the reference date

Reporting conventions will be as follows:

- Mean and median will be displayed to one more decimal place than the raw data
- Standard deviation (SD) will be displayed to two more decimal places than the raw data
- Percentages will be displayed to 1 decimal place (however, percentages corresponding to 0 counts will be reported as 0 rather than 0.0 and 100 percent will be reported as 100 rather than 100.0)
- p-values will be reported with 4 decimal places; all p-values should be specified to be 1-sided or 2-sided
- Unless otherwise specified, rounding will be performed to the closest integer/first decimal using the common mid-point between the two consecutive values, eg, 5.11 to 5.14 will be rounded to 5.1, and 5.15 to 5.19 will be rounded to 5.2
  - Non-zero percentages that are <0.1 before rounding will be displayed as "<0.1", eg, 0.09 will be reported as <0.1 rather than as 0.1
  - p-values <0.0001 before rounding will be displayed as "<0.0001", eg, a p-value of 0.00009 will be displayed as <0.0001 rather than as 0.0001

6.4.3. Pooling of Data Across Sites

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

6.4.4. Continuous and Categorical Variables

Continuous variables will be summarized using descriptive statistics, ie, number of non-missing values, mean, standard deviation (SD), median, quartiles, minimum, and maximum. Time-to-event endpoints in the presence of censoring will be estimated using Kaplan-Meier (KM) methodology.

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

counts of missing observations will be included in the denominator and presented as a separate category. For summaries by visit, percentages will be based on the number of subjects with data available for that visit, unless otherwise specified.

6.4.5. Unscheduled Visits

Generally, data collected at unscheduled visits will be included and summarized for both safety and efficacy analyses in the same manner as the data collected at scheduled visits. Descriptive statistics (mean, SD, median, quartiles, minimum, and maximum) by nominal visit for safety endpoints such as laboratory measurements, ECG parameters and vital signs will include only data from scheduled visits. Summaries of outliers [eg, worst value, worst change from baseline, worst Common Terminology Criteria for Adverse Events (CTCAE) grade] during the on-treatment period for safety endpoints such as AEs, laboratory measurements and ECG parameters will include data from both scheduled and unscheduled visits.

6.5. Methods for Handling Missing Data

6.5.1. Adverse Event and Concomitant Medication Start Dates

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

(1) Missing day only

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

(2) Missing day and month

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

(3) Missing day, month, and year

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

6.5.2. Adverse Event and Concomitant Medication End Dates

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

the death date, the death date will be used as the end date. If an imputation for an AE end date results in an AE end date that is after the data cutoff date, the AE will be considered as ongoing at the data cutoff date.

(1) Missing day only

- The last day of the month will be used.

(2) Missing day and month

- 31 December will be used.

(3) Missing day, month, and year

- The event will be regarded as ongoing.

6.5.3. Exposure

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

If the date of the last dose of study drug (AG-120 or matched placebo) is missing or partially missing, it will be imputed as follows:

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

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

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

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

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

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

  =max(the latest complete dosing date on or before the data cutoff date, 01 January of the year), if only the year is available

=max(the latest complete dosing date on or before the data cutoff date, first day of the month), if both the year and month are available

6.5.4. Death Date

Missing or partial death dates will be imputed based on the last contact date (as derived in Section 6.3.7), as follows:

- If the death date is missing it will be imputed as the day after the date of last contact
- If the day is missing or both the day and month are missing, the death date will be imputed as follows:
  - Missing day only: max(1st day of the month and year of death, last contact date+1)
  - Missing day and month: max(01 January of the year of death, last contact date+1)

If the imputed death date is after the data cutoff date, the subject will be considered to be alive at the time of the data cutoff date.

6.5.5. Date of Start of New Anticancer Therapy

Incomplete dates for start of new anticancer therapy will be imputed and the imputed full date will be used for determining the censoring date for efficacy analyses as described in Section 7. If the imputation results in an end date prior to the imputed start date, then the imputed start date will be set to the end date.

- The end date of new anticancer therapy will be included in the imputations for start date of new anticancer therapy. If the end date of new anticancer therapy is
  - Completely missing then it will be ignored in the imputations below
  - Partially missing with only year (YYYY) available then the imputations below will consider 31DECYYYY as the end date of the new anticancer therapy
  - Partially missing with only month and year available then the imputations below will consider the last day of the month for MMMYYYY as the end date of the new anticancer therapy
- For subjects who have not discontinued study treatment at the data cutoff date, last dose of study treatment is set to the data cutoff date in the imputations below.
- If the start date of new anticancer therapy is completely or partially missing, then the imputed start date of new anticancer therapy is derived as follows:
  - Start date of new anticancer therapy is completely missing

    Imputed start date=min[max(relapse date+1, last dose of study treatment+1), end date of new anticancer therapy]
  - Only year (YYYY) for start of anticancer therapy is available

IF YYYY <Year of min[max(relapse date+1, last dose of study treatment+1), end date of new anticancer therapy]

THEN imputed start date=31DECYYYY

ELSE IF YYYY=year of min[max(relapse date+1, last dose of study treatment+1), end date of new anticancer therapy]

THEN imputed start date=min[max(relapse date+1, last dose of study treatment+1), end date of new anticancer therapy]

ELSE IF YYYY >Year of min[max(relapse date+1, last dose of study treatment+1), end date of new anticancer therapy]

THEN imputed start date=01JANYYYY

- Both year (YYYY) and month (MMM) for start of anticancer therapy are available

IF

YYYY=Year of min[max(relapse date+1, last dose of study treatment+1), end date of new anticancer therapy], AND

MMM <Month of min[max(relapse date+1 day, last dose of study treatment+1 day), end date of new anticancer therapy]

THEN

imputed start date=DAY (Last day of MMM) MMM YYYY ;

ELSE IF

YYYY=Year of min[max(relapse date+1, last dose of study treatment+1), end date of new anticancer therapy], AND

MMM=Month of min[max(relapse date+1 day, last dose of study treatment+1 day), end date of new anticancer therapy]

THEN

imputed start date=min[max(relapse date+1 day, last dose of study treatment+1 day), end date of new anticancer therapy];

ELSE IF

YYYY=Year of min[max(relapse date+1, last dose of study treatment+1), end date of new anticancer therapy], AND

MMM >Month of min[max(relapse date+1 day, last dose of study treatment+1 day), end date of new anticancer therapy]

THEN

imputed start date=01 MMM YYYY;

ELSE IF

YYYY <Year of min[max(relapse date+1, last dose of study treatment+1), end date of new anticancer therapy]

THEN

imputed start date=DAY (Last day of MMM) MMM YYYY

ELSE IF

YYYY >Year of min[max(relapse date+1, last dose of study treatment+1), end date of new anticancer therapy]

THEN

imputed start date=01 MMM YYYY.

7. STATISTICAL ANALYSES

7.1. Subject Disposition

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

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

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

The following summaries will be presented by treatment arm based on the FAS:

- Frequency of subjects in each randomization strata and combination of randomization strata (per IRT)
- Frequency of subjects in each randomization strata and combination of randomization strata (as derived from data in the eCRF)
    - De novo AML from the Underlying AML Diagnosis eCRF, the "Specify Nature of AML per Investigator" field with the option "de novo" selected
    - Secondary AML from the Underlying AML Diagnosis eCRF, the "Specify Nature of AML per Investigator" field with any of these options selected, "treatment-related AML", "history of MDS", "history of MPD", or "other"
    - Geographic region will be based on the country code from the Demographics eCRF
- Frequency of subjects randomized in each geographic region, country, and site
- Frequency of subjects randomized and not treated, overall and by reason for discontinuation
- Frequency of subjects with study drug ongoing (separately for each study drug)
- Frequency of subjects who discontinue study drug, overall and by the reason for discontinuation of study drug (separately for each study drug)
- Frequency of subjects ongoing in the study
- Frequency of subjects who discontinue the study, overall and by the reason for study discontinuation

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

In addition, the following cross-tabulations will be performed:

- Cross-tabulation of randomization strata by IRT vs randomization strata as derived from data in the eCRF

- Cross-tabulation of subjects randomized (AG-120+azacitidine, placebo+azacitidine, none) vs subjects who have received at least 1 dose of study drug (AG-120+azacitidine, placebo+azacitidine, none)
- Cross-tabulation of subjects who have discontinued/are ongoing treatment with AG-120 or matched placebo vs subjects who have discontinued/are ongoing treatment with azacitidine

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

7.2. Protocol Deviations

All major protocol deviations that impact the safety of the subjects, the conduct of the study, or the evaluation of the study results will be reported by treatment arm based on the FAS. These will include:

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

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

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

7.3. Demographic and Other Baseline Characteristics

The following summaries will be presented by treatment arm and overall based on the FAS, unless otherwise specified.

7.3.1. Demographics and Physical Measurements

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

- Demographic characteristics
  - Sex: male, female
  - Race: American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or other Pacific Islander, White, other, unknown
  - Ethnic origin: Hispanic or Latino, not Hispanic or Latino, not reported
  - Age (years): summary statistics
  - Age categories:

  - <65 years, ≥65 years
  - <75 years, ≥75 years
- Physical measurements
  - Height (cm)
  - Weight (kg)
  - BMI (kg/m^2)
  - BSA (m^2)

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

7.3.2. Disease Characteristics

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

- Disease type
  - Nature of AML per Investigator [de novo; secondary (treatment-related AML; history of MDS; history of MPD; other)]
  - WHO classification of AML (AML with genetic abnormalities; AML with myelodysplasia-related changes; therapy-related myeloid neoplasms; AML not otherwise specified)
- ECOG PS at baseline (0; 1; 2; 3; 4)
- IDH1 mutation type at baseline based on central testing (R132C; R132G; R132H; R132L; R132S; wild type). Results from bone marrow will be used as the primary source. If results from bone marrow are not available, the results from peripheral blood will be used instead.
- IDH1 mutation status at baseline based on local testing (positive; negative)
- Cytogenetic results (normal karyotype; abnormal karyotype) and cytogenetic risk status according to the Investigator based on NCI NCCN guidelines (favorable; intermediate; poor; other)
- Baseline bone marrow aspirate blasts (%)
- Baseline bone marrow biopsy blasts (%)
- Baseline peripheral blood blasts (%)
- Baseline values of the following laboratory parameters. These parameters will be summarized as continuous variables as well as categorical variables using the specified category:
  - WBC (<15×10^9/L; 15 to <30×10^9/L; ≥30×10^9/L)
  - ANC (<0.5×10^9/L; 0.5 to <1×10^9/L; ≥1.0×10^9/L)
  - Hemoglobin (<80 g/L; ≥80 g/L)

- Platelet count (<50×10^9/L; 50 to <100×10^9/L; ≥100×10^9/L)
- Lactate dehydrogenase (LDH)
- Creatinine clearance (<15 mL/min; 15 to <40 mL/min; 40 to <60 mL/min; 60 to <90 mL/min; ≥90 mL/min)

- Extramedullary disease at baseline (yes; no; unknown; not assessed)

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

7.3.3. Medical History

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

Medical history will be provided in by-subject listings.

7.3.4. Prior Therapies

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

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

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

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

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

7.4. Exposure to Study Drug and Compliance

The following summaries will be presented by treatment arm based on the safety analysis set. The derivations are provided for the following study drugs:

- AG-120 or matched placebo administered orally at a dose of 500 mg QD
- Azacitidine administered SC or IV at a dose of 75 mg/m^2/day for the first week (days 1 through 7 days) or on a 5-2-2 schedule (days 1 through 5 and days 8 and 9) of each 4-week (28-day) cycle

Exposure and compliance will be summarized for each study drug by treatment arm.

7.4.1. Treatment Duration and Exposure

Exposure will be summarized as dose received [cumulative dose, actual dose intensity (DI)] and as dose received relative to planned dose [relative dose intensity (RDI)]. Duration of

exposure to each study drug will be summarized as a continuous variable as well as in categories (>0-≤4, >4-≤8, >8-≤12, >12-≤16, >16-≤20, >20-≤24, and >24 weeks).

Exposure to AG-120 or matched placebo

The dose level for AG-120 or matched placebo is calculated as actual dose administered (mg/day). Only non-zero doses of AG-120 or matched placebo are considered in the derivations.

- Duration of exposure (days)=last dose date–first dose date+1
- Cumulative dose (mg)=sum of the actual doses
- Planned DI (mg/day)=500
- Actual DI (mg/day)=cumulative dose (mg)/duration of exposure (days)
- RDI (%)=100×Actual DI (mg/day)/Planned DI (mg/day)

Exposure to Azacitidine

The dose level for azacitidine is calculated as actual dose administered (mg/m^2/day). In what follows, dose, last dose date and first dose date refer to azacitidine. Only non-zero doses are considered in the derivations.

- Planned duration of exposure (4-week cycle)=(start date of azacitidine last cycle–start date of azacitidine first cycle+28)/28
- Duration of exposure (4-week cycle)=(last dose date–first dose date+1)/28
- Cumulative dose (mg/m^2)=sum of the actual doses
- Planned DI (mg/m^2/4-week cycle)=525
- Actual DI (mg/m^2/4-week cycle)=cumulative dose (mg/m^2)/[planned duration of exposure (4-week cycle)]
- RDI (%)=100×Actual DI (mg/m^2/4-week cycle)/Planned DI (mg/m^2/4-week cycle)

7.4.2. Dose Modifications

The summary of dose modifications will include:

- The frequency of subjects with at least 1 dose reduction
- Summary of the number of days with dose reductions
- The frequency of subjects with at least 1 interruption of study drug
- Summary of the number of days with interruptions of study drug

Dose reduction is defined as an administered non-zero dose that is lower than the planned dose (AG-120 or matched placebo: 500 mg; azacitidine: 75 mg/m^2). Data from the AG-120/Placebo Dosing or Azacitidine Dosing eCRF when the option "Dose Reduced" is chosen will be used in the derivation.

An interruption of study drug is defined as:

- A 0 mg dose given on one or more days for AG-120 or matched placebo, or
- A 0 mg dose given on one or more days within the dosing period (where a subject is not on the "off" part of a treatment cycle) for azacitidine, after which >0 mg dose resumes. Azacitidine will be given 7 days (days 1-7 or days 1-5, 8-9) for each 28-day treatment cycle. The 21 days when azacitidine is not administered following protocol specified dosing regimen will not be considered as a dose interruption. An interruption is not considered a dose change.

Will use the data from the AG-120/Placebo Dosing or Azacitidine Dosing eCRF when the option "Drug Interrupted" is chosen.

7.5. Concomitant Therapies

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

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

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

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

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

7.6. Subsequent Therapies

The following summaries will be presented by treatment arm based on the FAS. Subsequent therapies are defined as therapies that are started after the last dose of study treatment.

7.6.1. Subsequent Stem Cell Transplants for AML

For subsequent stem cell transplants for AML, type of hematopoietic stem cell transplant (HSCT), and disease status at the time of HSCT will be summarized in frequency tabulations.

HSCT data will be presented in by-subject listings for subjects who receive HSCT.

7.6.2. Subsequent Anticancer Therapies

Subsequent anticancer therapies will be coded according to the ATC code and PT using the latest version of the WHO Drug Dictionary. All subsequent anticancer therapies will be

summarized in frequency tabulations according to the WHO ATC third level and PT. Line of therapy and type of therapy will also be summarized in frequency tabulations.

Anticancer therapies will be presented in by-subject listings for subjects with at least 1 anticancer therapy.

7.7. Efficacy Analyses

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

7.7.1. Primary Endpoint

7.7.1.1. Primary Analyses for EFS

EFS is defined as the time from randomization until TF, relapse from remission, or death from any cause, whichever occurs first. TF is defined as failure to achieve CR by Week 24. Subjects who do not achieve CR by Week 24 will be considered to have had an EFS event at Day 1 of randomization. For subjects who achieve CR by Week 24 (responders), the EFS time will be the time from randomization to relapse or death, whichever occurs first.

EFS will be tested using the log-rank test stratified by the randomization stratification factors. The basis for a claim of efficacy will be the statistical significance of EFS in favor of the AG-120+azacitidine arm when the 1-sided p-value <0.025.

Kaplan-Meier estimates (product-limit estimates) will be presented by treatment arm together with a summary of associated statistics. In particular, the EFS rate at 1 day, and 3, 6, 9, 12, 18, 24 and 36 months will be estimated with corresponding 2-sided 95% CIs. The CIs for the survival function estimates at the time points defined above will be derived using the log-log transformation (conftype=loglog default option in SAS PROC LIFETEST) with back transformation to a CI on the untransformed scale. The estimate of the standard error (SE) will be computed using Greenwood's formula.

The HR will be estimated using a Cox's proportional hazards (PH) model stratified by the randomization strata. Each stratum will define a separate baseline hazard function (using the "STRATA" statement in SAS PROC PHREG), ie for the i-th stratum the hazard function is expressed as: $h(i;t)=h(i,0;t)\exp(x\beta)$, where $h(i,0;t)$ defines the baseline hazard function for the i-th stratum and x defines the treatment arm (0= control arm, 1=treatment arm) and β is the unknown regression parameter. Ties will be handled by replacing the proportional hazards model by the discrete logistic model (Ties=Discrete option in SAS PROC PHREG). Given that the PH assumption is not met based on the EFS definition, the overall HR may not be meaningful.

As EFS is a composite endpoint, the estimates for each component will be summarized:

- CR rate by 24 weeks (following the methodology outlined in Section 7.7.2.1 without presenting p-values)
- EFS among subjects who achieve CR by 24 weeks (following the methodology outlined above restricted only to subjects who achieve CR by 24 weeks without

presenting p-values). In addition, the median EFS time with 2-sided 95% CI will be calculated based on Brookmeyer and Crowley.

Restricted Mean Survival Time (RMST)

The HR estimate from the Cox PH model is routinely used to empirically quantify the between-arm difference under the assumption that the ratio of the two hazard functions is constant over time. When this assumption is plausible, such a ratio estimate captures the relative difference between two survival curves. However, the clinical meaning of such a ratio estimate is difficult, if not impossible, to interpret when the underlying PH assumption is violated (ie, the HR is not constant over time).

The RMST is a robust and clinically interpretable summary measure of the survival time distribution. Unlike median survival time, it is estimable even under heavy censoring. There is a considerable body of methodological research (Royston and Parmar, 2011; Uno, et al., 2014; Zhang, 2013) about the use of RMST to estimate treatment effects as an alternative to the HR approach.

The RMST methodology is applicable independently of the PH assumption and can be used, at a minimum, as a sensitivity analysis to explore the robustness of the primary analysis results. However, when large departures from the PH assumption are observed, as is expected in this study for EFS, the log-rank test is underpowered to detect differences between the survival distributions for the treatment arms, and a test of the difference between the RMST for the treatment arm and the control arm may be more appropriate to determine superiority of the treatment arm compared to the control arm with respect to the time-to-event endpoint.

As it pertains to the **cut-off point (τ)** to evaluate the RMST, the cut-off point should not exceed the minimum of the largest observed time for both treatment arms so that the RMST of all treatment arms being evaluated can be adequately estimated and comparison between treatments is feasible; τ should be clinically meaningful and closer to the end of the study follow-up so that the majority of survival outcomes will be covered by the time interval.

τ=min(largest observed survival time for the treatment arm, largest observed survival time for the control arm).

The RMST up to time τ can then be interpreted as the expected survival time restricted to the common follow-up time τ among all subjects.

In this section, "survival" is meant to denote EFS.

The treatment effect between the treatment arm and the control arm will also be assessed based on the difference in RMST. The associated 95% CI for the difference in RMST and 1-sided p-value will be generated.

7.7.1.1.1. Determination of Event or Censoring for EFS

Evaluation of disease status, including evaluation of bone marrow and/or peripheral blood, is planned to be conducted at Screening (or as part of Pre-screening if within 28 days prior to randomization); Day 1 (±7 days) of Weeks 9, 17, 25, 33, 41, 53, and every 24 weeks thereafter; at End of Treatment (EOT); during EFS follow-up on the same schedule; as clinically indicated; and/or any time that disease progression is suspected.

Determination of relapse date

Only disease assessments performed before the start date of subsequent anticancer therapies will be considered in the determination of relapse.

Confirmation is required for relapse. Assessments which are not done or not evaluable are ignored in the derivation of relapse confirmation. A subject will be considered to have relapsed if either of the following criteria are met:

- Relapse in 2 consecutive assessments that are at least 4 weeks apart
- Relapse with no further evaluable disease assessments before discontinuation from study or initiation of subsequent anticancer therapy

The date of relapse considered in the analyses will be the date when the first relapse, that is subsequently confirmed, is observed.

Determination of CR by 24 weeks

CR will be assessed until the date of relapse (that is subsequently confirmed). Only assessments performed on or before the start date of subsequent anticancer therapies will be considered in the determination of CR.

The protocol allows a 1-week window for disease assessments. Therefore, a subject will be considered to have achieved "CR by 24 weeks" if the date of first CR is within 25 weeks (24 weeks target+1 week window) after the date of randomization:

CR by 24 weeks if date of first CR–date of randomization$\leq$24×7+7.

Determination of two or more missing assessments

In the primary analysis of EFS, if a subject has an EFS event (relapse or death) after two or more missing disease assessments, the subject will be censored at the last adequate disease assessment documenting no relapse before the missing assessments. The followings are the rules for determining whether two or more disease assessments are missed.

- If the event date is prior to day 1 of week 53 (relative day since randomization $\leq$52×7=364), and the difference of event date and the last adequate disease assessment >(8×2+1)×7=119, then the event is considered as occurring after two or more missing disease assessments
- If the event date is between day 1 of week 53 and day 1 of week 77 (relative day since randomization >364 and $\leq$532) and the difference of event date and the last adequate disease assessment >(8+12+1)×7=147, then the event is considered as occurring after two or more missing disease assessments
- If the event date is between day 1 of week 77 and day 1 of week 101 (relative day since randomization >532 and $\leq$700) and the difference event date and the last adequate disease assessment is >(12+24+1)×7=259, then the event is considered as occurring after two or more missing disease assessments
- If the event date is after day 1 of week 101 (relative day since randomization >700) and the difference of event date and the last adequate disease assessment is

>(24+24+1)×7=343, then the event is considered as occurring after two or more missing disease assessments

The outcome (event or censor), type of event and the date of event or censoring to be considered for the primary analysis of EFS are presented in Table 5.

Table 5: Outcome and Event or Censoring Dates for EFS

Scenario	Date of event/censoring	Event (event type)/ Censored
CR by 24 weeks[a] then relapse	Date of relapse	Event (relapse)
CR by 24 weeks[a] then death (no relapse)	Date of death	Event (death)
CR by 24 weeks[a] then start subsequent anticancer therapy (prior to relapse or no relapse)	Date of the last adequate disease assessment documenting no relapse prior to start of subsequent anticancer therapy or missed response assessments	Censored
CR by 24 weeks[a] then relapse or death after two or more missing or inadequate disease assessments		Censored
CR by 24 weeks[a] and neither relapse nor death		Censored
On treatment ≥24 weeks without CR by 24 weeks[a]	Date of randomization	Event (TF, on treatment ≥24 weeks without CR)
Treatment discontinuation prior to 24 weeks, without CR by 24 weeks[a]		Event (TF, treatment discontinuation prior to 24 weeks without CR)

Abbreviations: CR = complete response; EFS = event-free survival; TF = treatment failure.
[a] Based on the protocol schedule of assessments, "+1 week window" is allowed.

Frequency of subject with each event type and censoring reasons will be presented by treatment arm. Reasons for censoring will be summarized according to the categories in Table 6 following the hierarchy shown.

Table 6: EFS Censoring Reasons and Hierarchy

Hierarchy	Condition	Censoring Reason
1	CR by 24 weeks then start subsequent anticancer therapy	CR by 24 weeks, start subsequent anticancer therapy
2	CR by 24 weeks then relapse or death after two or more missing or inadequate disease assessments	CR by 24 weeks, relapse/death documented after two or more missing disease assessments
4	CR by 24 weeks, no relapse or death, lost to follow-up	CR by 24 weeks, lost to follow-up
5	CR by 24 weeks, no relapse or death, withdrawal by subject from the study	CR by 24 weeks, withdrawal by subject
6	CR by 24 weeks, no relapse or death, ongoing in study	CR by 24 weeks, ongoing without relapse or death

Abbreviations: CR = complete response; EFS = event-free survival.

The EFS event or censoring time, event type and the reason for censoring will also be presented in a by-subject listing.

7.7.1.2. Sensitivity Analyses for EFS

The following sensitivity analyses will be performed:

1. EFS will be tested using the log-rank test stratified by the IRT randomization stratification factors and based on the FAS. The time of relapse or death is determined using the actual date of relapse or death, even in situations where relapse or death is observed after two or more missing disease assessment or start of subsequent anticancer therapy.
2. EFS will be tested using the unstratified log-rank test and based on the FAS.
3. EFS will be tested using the log-rank test stratified by the IRT randomization stratification factors and based on the PPS.
4. EFS will be tested using the log-rank test stratified by the randomization stratification factors derived based on data provided by the Investigator in the eCRF and based on the FAS.
5. EFS will be tested using the log-rank test stratified by the IRT randomization stratification factors and based on the FAS. For the subjects who do not achieve CR by Week 24, instead of being considered to have had an EFS event at Day 1 of randomization, the event time will be either 24 weeks or EOT, whichever is earlier.

7.7.2. Key Secondary Endpoints

The key secondary efficacy endpoints are CR, OS, CR+CRh, and OR. CR, CR+CRh and OR will be assessed until the date of relapse (that is subsequently confirmed). Only assessments performed on or before the start date of subsequent anticancer therapies will be considered in the determination of these response endpoints.

7.7.2.1. Complete Remission

Complete remission rate is defined as the proportion of subjects who achieve a CR. If the primary analysis of EFS is significant, a CMH test stratified by the randomization stratification factors will be used to compare CR rate between the 2 treatment arms at 1-sided 2.5% level of significance. The odds ratio and its associated 95% CI will be presented.

7.7.2.2. Overall Survival

Overall survival is defined as the time from date of randomization to the date of death due to any cause. If a subject is not known to have died by the data cutoff date, then OS will be censored at the date of last contact (see Section 6.3.7).

If the primary analyses of EFS and CR rate are significant, OS will be tested using the log-rank test stratified by the randomization stratification factors at 1-sided 2.5% level of significance.

The HR will be estimated using a Cox's PH model stratified by the randomization strata. Each stratum will define a separate baseline hazard function (using the "STRATA" statement in SAS PROC PHREG), ie for the i-th stratum the hazard function is expressed as: $h(i;t)=h(i,0;t)\exp(x\beta)$, where $h(i,0;t)$ defines the baseline hazard function for the i-th stratum

and x defines the treatment arm (0=control arm, 1=treatment arm) and β is the unknown regression parameter. Ties will be handled by replacing the PH model by the discrete logistic model (Ties=Discrete option in SAS PROC PHREG).

Kaplan-Meier estimates (product-limit estimates) will be presented by treatment arm together with a summary of associated statistics including the median OS time with 2-sided 95% CIs. In particular, the OS rate at 3, 6, 9, 12, 18, 24 and 36 months will be estimated with corresponding 2-sided 95% CIs. The CIs for the median will be calculated according to Brookmeyer and Crowley and the CIs for the survival function estimates at the time points defined above will be derived using the log-log transformation (conftype=loglog default option in SAS PROC LIFETEST) with back transformation to a CI on the untransformed scale. The estimate of the SE will be computed using Greenwood's formula.

Frequency of subjects with an event (death) and censoring reasons will be presented by treatment arm. Reasons for censoring will be summarized according to the categories in Table 7 following the hierarchy shown.

Table 7: OS Censoring Reasons and Hierarchy

Hierarchy	Condition	Censoring Reason
1	No event and (EOS date ≥date of randomization when reason for EOS=Withdrawal by Subject)	Withdrawal of consent
2	No event and [(lost to follow-up in any disposition page or survival follow-up page) or (data cutoff date–last contact date >8 weeks)]	Lost to follow-up
3	No event and none of the conditions in the prior hierarchy are met	Alive

Abbreviations: OS=overall survival; EOS=end of study.

The OS event or censoring time and the reasons for censoring will also be presented in a by-subject listing.

7.7.2.3. Complete Remission Plus Complete Remission with Partial Hematologic Recovery

The CR+CRh rate is defined as the proportion of subjects who achieved a CR or CRh. CRh is defined as a CR with partial recovery of peripheral blood counts [<5% bone marrow blasts, ANC >0.5×10^9/L (500/μL), and platelets >50×10^9/L (50,000/μL)].

Since CRh is not part of modified IWG criteria, it will be derived by the Sponsor based on the following rules, for each individual assessment which is not CR.

- Bone marrow blast <5%, and
- Closest ANC within±8 days >0.5×10^9/L, and
- Closest platelet counts within±8 days >50×10^9/L, and
- Assume negative if Auer rod is not reported, and
- Assume negative if extramedullary disease is not reported

For bone marrow blasts, bone marrow aspirate will be used as the primary source. If a bone marrow aspirate assessment is not available, a bone marrow biopsy assessment will be used.

For ANC or platelet count, if there are 2 values which have the same distance to the bone marrow assessment date (eg, one is 1 day prior and the other is 1 day after), the value from the earlier assessment date will be used.

If the primary analyses of EFS, CR and OS are statistically significant, a CMH test stratified by the randomization stratification factors will be used to compare the CR+CRh rate between the 2 treatment arms at 1-sided 2.5% level of significance. The odds ratio and its associated 95% CI will be presented.

7.7.2.4. Objective Response

Objective response rate is defined as the rate of CR, CRi (including CRp), PR, and MLFS.

The best response is calculated using the following hierarchy: 1) CR; 2) CRi (including CRp); 3) PR; and 4) MLFS.

Frequency of subjects with best response of CR, CRi (including CRp), PR or MLFS will be tabulated by treatment arm.

If the primary analysis of EFS, CR, OS and CR+CRh are statistically significant, a CMH test stratified by the randomization stratification factors will be used to compare ORR between the 2 treatment arms at 1-sided 2.5% level of significance. The odds ratio and its associated 95% CI will be presented.

7.7.3. Additional Secondary Efficacy Endpoints

Only disease assessments performed on or before the start date of any further anticancer therapies will be considered in the analyses of these endpoints.

7.7.3.1. Complete Remission and CRi (Including CRp)

The CR+CRi (including CRp) rate is defined as the proportion of subjects who achieved a CR or CRi (including CRp). A CMH test stratified by the randomization stratification factors will be used to compare CR+CRi (including CRp) rate between the 2 treatment arms and the nominal 1-sided p-value will be presented. The odds ratio and its associated 95% CI will be presented as well.

7.7.3.2. DOCR, DOCRh, DOR, and DOCRi

DOCR is defined, for subjects who achieved CR, as the time from the first occurrence of CR to confirmed relapse or death due to any cause.

DOCR (months)=(date of event or censoring–first date of CR+1)/30.4375

DOCRh is defined, for subjects who achieved CR or CRh, as the time from the first occurrence of CR or CRh to confirmed relapse or death due to any cause.

DOCRh (months)=(date of event or censoring–first date of CRh+1)/30.4375

DOR is defined, for subjects who achieved CR, CRi (including CRp), PR, or MLFS, as the time from the first response to confirmed relapse, disease progression, or death due to any cause.

DOR (months)=(date of event or censoring– first date of response+1)/30.4375

DOCRi is defined, for subjects who achieved CR or CRi (including CRp), as the time from the first occurrence of CR or CRi (including CRp) to confirmed relapse or death due to any cause.

DOCRi (months)=(date of event or censoring–first date of CR or CRi+1)/30.4375

The outcome (event or censor), type of event and the date of event or censoring to be considered for the analysis of DOCR, DOCRh, DOR, and DOCRi are presented in Table 8.

Table 8: Outcome and Event or Censoring Dates for DOCR, DOCRh, DOR, and DOCRi

Scenario	Date of event/censoring	Event (event type)/ Censored
Relapse/progression[a]	Date of relapse/progression	Event (relapse/progression)
Death (no relapse/progression)	Date of death	Event (death)
Start subsequent anticancer therapy (prior to relapse/progression or no relapse/progression)	Date of the last adequate disease assessment documenting no relapse/progression prior to start of subsequent anticancer therapy or missed disease assessments	Censored
Relapse/progression or death after two or more missing or inadequate disease assessments		Censored
Neither relapse/progression nor death		Censored

[a]Progression only applies to the calculation of DOR, for which subjects with a BOR of PR may experience disease progression afterwards.

Frequency of subject with each event type and censoring reasons will be presented by treatment arm. Reasons for censoring will be summarized according to the categories in Table 9 following the hierarchy shown.

Table 9: Censoring Reasons and Hierarchy for DOCR, DOCRh, DOR, and DOCRi

Hierarchy	Condition	Censoring Reason
1	Start subsequent anticancer therapy	Start subsequent anticancer therapy
2	Experience relapse/progression or die after two or more missing or inadequate disease assessments	Relapse/progression/death documented after two or more missing disease assessments
3	No relapse/progression or death, lost to follow-up	Lost to follow-up
4	No relapse/progression or death, withdrawal by subject from the study	Withdrawal by subject
5	No relapse/progression or death, ongoing in study	Ongoing without relapse/progression or death

[a]Progression only applies to the calculation of DOR, for which subjects with a BOR of PR may experience disease progression afterwards.

Kaplan-Meier estimates (product-limit estimates) will be presented by treatment arm together with a summary of associated statistics including the median DOCR, DOCRh, DOR, and DOCRi time with 2-sided 95% CIs. In particular, the DOCR, DOCRh, DOR, and DOCRi rates at 3, 6, 9, 12, 18, 24 and 36 months will be estimated with corresponding

2-sided 95% CIs. The CIs for the median will be calculated according to Brookmeyer and Crowley and the CIs for the survival function estimates at the time points defined above will be derived using the log-log transformation (conftype=loglog default option in SAS PROC LIFETEST) with back transformation to a CI on the untransformed scale. The estimate of the SE will be computed using Greenwood's formula.

DOCR, DOCRh, DOR, and DOCRi will be displayed graphically and analyzed using Kaplan-Meier methodology. If the number of subjects with responses is small, the Kaplan-Meier method may not provide reliable estimates. In this case, only descriptive statistics or listings will be provided.

7.7.3.3. TTCR, TTCRh, TTR, and TTCRi

TTCR is defined, for subjects who achieve CR, as the time from randomization to the first occurrence of CR.

TTCR (months)=(first date of CR–date of randomization+1)/30.4375

TTCRh is defined, for subjects who achieve CR or CRh, as the time from randomization to the first occurrence of CR or CRh.

TTCRh (months)=(first date of CR or CRh–date of randomization+1)/30.4375

TTR is defined, for subjects who achieve CR, CRi (including CRp), PR, or MLFS, as the time from randomization to the first response.

TTR (months)=(first date of response–date of randomization+1)/30.4375

TTCRi is defined, for subjects who achieve CR or CRi (including CRp), as the time from randomization to the first occurrence of CR or CRi (including CRp).

TTCRi (months)=(first date of CR or CRi–date of randomization+1)/30.4375

TTCR, TTCRh, TTR, and TTCRi will be summarized by treatment arm using descriptive statistics.

7.7.3.4. Quality of Life Assessments

Summaries of QoL data will be presented by treatment arm based on the FAS.

7.7.3.4.1. Descriptions of Questionnaires

7.7.3.4.1.1. European Organization of Research and Treatment of Cancer – Quality of Life Questionnaire – Core Questionnaire (EORTC-QLQ-C30)

The EORTC QLQ-C30 is a validated and reliable self-reported measure of QoL for subjects with cancer who are receiving cancer treatment. The QLQ-C30 contains 30 items in total and each item is a 4-point or 7-point Likert scale. These 30 items can be categorized into 1 global health status/HRQOL scale, 5 functional scales, 3 symptom scales, and 6 single item scales. Each of the multi-item scales includes a different set of items. No item occurs in more than one scale:

- 1 global health status scale (2 items)

- 5 functional scales: physical function (5 items), role function (2 items), emotional function (4 items), cognitive function (2 items), social function (2 items)
- 3 symptom scales: fatigue (3 items), nausea and vomiting (2 items), pain (2 items)
- 6 single-item scales relating to dyspnea, insomnia, appetite loss, constipation, diarrhea, and financial difficulties

7.7.3.4.1.2. EuroQoL EQ-5D-5L (EQ-5D-5L)

The EuroQoL EQ-5D-5L is a widely used standardized measure of health status that provides a descriptive profile and single index value to appraise respondents' health status. The measure was designed for both clinical and economic appraisal and has been validated in numerous clinical populations. The EQ-5D-5L is comprised of five single-item dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. These dimensions are measured on a 5-level scale: "No problems", "Slight problems", "Moderate problems", "Severe problems", and "Extreme problems."

EQ-5D-5L responses can also be converted into summary index values, using established population-based value sets. Index values reflect general population valuations of a given health state, as constituted by the EQ-5D-5L responses, where a score of 0 is equivalent to being dead and 1 is equivalent to perfect health.

The EQ-5D-5L also contains a visual analog scale (VAS) to assess the subject's overall health on a scale of 0 ("The worse health you can imagine") to 100 ("The best health you can imagine").

7.7.3.4.2. Statistical Analyses

Scores for the EORTC QLQ-C30 and EQ-5D-5L will be summarized at each visit where data are available.

Completion rates will be calculated for each instrument at each assessment visit as the proportion of subjects alive at the assessment visit with a completed instrument.

7.7.3.4.2.1. EORTC-QLQ-C30

Each scale will be transformed into a 0-100 score following EORTC guidelines:

- Estimate the average of the items that contribute to the scale; this is the raw scale score
- Use a linear transformation to standardize the raw scale score, so that it ranges from 0 to 100. Scale score=(raw score–1)/range*100, where range is the difference between the maximum and minimum possible value of the raw score from each item

Transformed scores for each scale, and the absolute and percent changes from baseline will be summarized by treatment arm at each visit.

Mixed models will also be applied in the analysis of the EORTC QLQ-C30. All available data for each subject will be used in the analysis. The mixed model on the change from

baseline across visits for all scales will be performed with baseline score, treatment arm, time, randomization stratification factors and an interaction between treatment arm and time as fixed effect, and subject as random effects. All parameter estimates will be obtained using restricted maximum likelihood. The unstructured covariance structure will be used to define covariance between random effects (using option "Type=UN" as a part of the RANDOM statement in PROC MIXED). For the degrees-of-freedom calculations the Kenward and Roger algorithm will be used (using option "ddfm=kr" as a part of the MODEL statement in PROC MIXED). The least square means and 95% CI will be summarized and graphically presented by treatment arm over time.

7.7.3.4.2.2. EQ-5D-5L

The first response is coded as a 1 (indicating no problems), the second response is coded as a 2 (indicating slight problems), the third response is coded as a 3 (indicating moderate problems), the forth response is coded as a 4 (indicating severe problems), the fifth response is coded as a 5 (indicating extreme problems). Ambiguous responses (eg, more than 1 response in a dimension) are treated as missing values.

For each dimension, the number and percentage of subjects with no problems, slight problems, moderate problems, severe problems and extreme problems will be summarized by treatment arm at each visit.

EQ-5D-5L responses will also be converted into index values using a US value set. Summary statistics for index scores and the absolute and percent change from baseline will be reported by treatment arm at each visit.

The EQ-5D-5L also contains a utility scale of health state (VAS) ranging between 0 (worst health) and 100 (best health). Summary statistics for VAS scores and the absolute and percent change from baseline will be reported by treatment arm at each visit.

[redacted]

7.7.4. Subgroup Analyses

Subgroup analyses to be performed for EFS are presented in Table 10.

Table 10: Subgroup Analyses to be Performed for EFS

Subgroup	Categories
De novo status based on IRT	Yes; No
De novo status based on Investigator from eCRF	Yes; No
Region	United States and Canada; Western Europe, Israel or Australia; Japan; ROW
Age	<75; ≥75 years
Baseline ECOG PS	0 or 1; ≥2
Sex	Female; Male
Race	White; Asian; Black or African American; Other (Other includes American Indian or Alaska Native, Native Hawaiian or other Pacific Islander, and not reported)
Baseline cytogenetic risk status	Favorable risk; Intermediate risk; Poor risk
WHO classification of AML	AML with genetic abnormalities; AML with myelodysplasia-related changes; Therapy-related myeloid neoplasms; AML not otherwise specified
Baseline WBC count	$\leq 5\times10^9$/L; $>5\times10^9$/L
Baseline percent bone marrow blasts*	≤50%; >50%

Abbreviations: AML = acute myelogenous leukemia; ECOG PS = Eastern Cooperative Oncology Group performance status; ROW = rest of world; WBC = white blood cell; WHO = World Health Organization.
*For bone marrow blasts, bone marrow aspirate will be used as the primary source. If a bone marrow aspirate assessment is not available, a bone marrow biopsy assessment will be used.

Treatment arms will be compared for EFS using a 2-sided unstratified log-rank test for each category and the unstratified HR and its corresponding 95% CI will be computed for each category and depicted in a forest plot.

If there is a low number of subjects within a category (<5% of the subjects in the FAS), the categories will be pooled (if 3 or more categories are pre-specified for the subgroup) or the subgroup will not be analyzed (if only 2 pre-specified categories in the subgroup). [redacted]

7.8. Safety Analyses

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

7.8.1. Adverse Events

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

All AEs will be listed by subject and AEs with onset outside of the on-treatment period will be flagged in the listings. Unless otherwise specified, TEAEs will be summarized according

to the latest version of MedDRA by SOC and/or PT, severity (based on CTCAE v4.03 grading), seriousness, and relation to study treatment in decreasing frequency based on the frequencies observed for the AG-120+azacitidine arm.

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

The following will be summarized:

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

Treatment-related TEAEs (including TEAEs, Grade ≥3 TEAEs, serious TEAEs, and TEAEs leading to death) will be summarized separately for

- TEAEs related to AG-120/matched placebo only
- TEAEs related to azacitidine only
- TEAEs related to both AG-120/matched placebo and azacitidine

TEAEs leading to discontinuation of study drug, TEAEs leading to interruption of study drug, and TEAEs leading to dose reduction, will be summarized separately for

- AG-120/matched placebo only

- Azacitidine only
- Both AG-120/matched placebo and azacitidine

7.8.1.1. Adverse Events of Special Interest

The following are considered AESIs:

- QT prolongation (see "AG-120-specified Safety Search Criteria" for the criteria used to identify the relevant AEs)
- Leukocytosis (see "AG-120-specified Safety Search Criteria" for the criteria used to identify the relevant AEs)
- IDH differentiation syndrome (see "AG-120-specified Safety Search Criteria" for the criteria used to identify the relevant AEs)

Additional TEAEs of interest indicative of clinical benefit are as follows (see "AG-120-specified Safety Search Criteria" for the criteria used to identify the relevant AEs):

- Infection
- Bleeding

The following will be summarized for each AESI category and additional TEAEs of interest:

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

AESIs/TEAEs of interest leading to discontinuation of study drug will be summarized separately for:

- AG-120/matched placebo only
- Azacitidine only
- Both AG-120/matched placebo and azacitidine

7.8.1.2. Adverse Events Associated with COVID-19

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

- TEAEs associated with COVID-19, by SOC and PT
- Grade ≥3 TEAEs associated with COVID-19, by SOC and PT
- Serious TEAEs associated with COVID-19, by SOC and PT

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

TEAEs associated with COVID-19 leading to discontinuation of study drug, leading to interruption of study drug, and leading to dose reduction will be summarized separately for

- AG-120/matched placebo only
- Azacitidine only
- Both AG-120/matched placebo and azacitidine

7.8.2. Death

The frequency of subjects in the safety analysis set who died, along with the cause of death, will be tabulated based on information from the Death Report eCRF. Cause of death will be summarized for the following categories:

- On-treatment death: Deaths within 28 days after the last dose of study treatment (ie, deaths during the on-treatment period)
- Post-treatment death: Deaths more than 28 days after the last dose of study treatment (ie, deaths after the end of the on-treatment period)
- Overall: All deaths

In addition, for each cause of death reported in the eCRF, those related to COVID-19 will be summarized.

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

7.8.3. Clinical Laboratory Data

Clinical laboratory test results will be expressed in SI units.

For all laboratory tests (chemistry, hematology, coagulation), the actual values and the changes from baseline will be summarized by study visit.

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

Parameters with CTCAE grades available:

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

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

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

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

Parameters with CTCAE grades not available:

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

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

7.8.3.1. Hematology

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

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

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

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

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

7.8.3.2. Chemistry

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

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

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

Concurrent measurements are those occurring on the same date.

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

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

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

In addition, a listing of all total bilirubin, ALT, AST, and ALP values for subjects with a postbaseline total bilirubin >2×ULN, ALT >3×ULN, or AST >3×ULN will be provided.

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

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

7.8.3.3. Pregnancy Tests

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

7.8.4. Vital Signs and Physical Measurements

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

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

7.8.5. Electrocardiograms

ECG summaries will include all ECG assessments from the on-treatment period. QTcB and QTcF interval will be derived based on RR and QT interval (see below), if not collected in the eCRF.

Selecting Primary QT Interval Correction for Heart Rate

The analysis of QT interval data is complicated by the fact that the QT interval is highly correlated with heart rate. Because of this correlation, formulas are routinely used to obtain a corrected QT interval, denoted QTc, which is independent of heart rate. This QTc is intended to represent the QT interval at a standardized heart rate. Several correction formulas have been proposed in the literature. For this analysis several of those methods of correction will be used, as described below. The QT interval corrected for heart rate by Bazett's formula, QTcB, is defined as

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

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

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

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

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

Linear regression method:

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

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

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

ECG Summaries

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

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

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

7.8.6. Left Ventricular Ejection Fraction

The LVEF data will be presented in a by-subject listing.

7.8.7. ECOG Performance Status

The ECOG PS shift from baseline to highest score during the on-treatment period will be summarized.

ECOG PS will be presented in a by-subject listing.

7.8.8. Safety Measures Indicative of Clinical Benefit

Transfusion requirements (platelet and RBC; number of units), days spent hospitalized, and other safety measures that are potentially indicative of clinical benefit will be summarized by treatment arm using descriptive statistics.

7.8.8.1.1. On-treatment Transfusions

On-treatment transfusions are defined as transfusions that are started during the on-treatment period.

The summary of on-treatment transfusion data will include the frequency of subjects with any on-treatment transfusions, the frequency of subjects with each type of on-treatment

transfusion (whole blood, packed RBCs, platelet, plasma, and other), the total number of units per subject with each type of on-treatment transfusion, and the reasons for which each type of transfusion is administered.

Transfusion type, number of units, reasons for transfusion administration, and dates of transfusions will be presented in a by-subject listing.

7.8.8.1.2. Hospitalizations for Adverse Event

Hospitalizations for AE will be summarized, including number of events, total number of days hospitalized, the rate of events and days hospitalized per person-year.

Hospitalizations will be provided in a by-subject listing.

7.9. Biomarker Analyses

7.9.1. Methods

Genomic DNA from baseline and on-treatment bone marrow mononuclear cells (BMMCs) and/or peripheral blood mononuclear cells (PBMCs) will be used for molecular studies. mIDH1 VAF will be quantified through BEAMing digital PCR (dPCR) technology (OncoBEAM™; Sysmex Inostics, Baltimore, MD, USA) for assessment of R132 (C/G/L/S/H) alleles. A valid mIDH1 VAF result is defined as a VAF value other than "not evaluable". IDH1 MC is defined as mIDH1 VAF reduction to below the lower limit of detection (0.02%-0.04%) for at least one on-treatment time point.

7.9.2. Statistical Analyses

Complete remission with IDH1 MC is defined as a response of CR where there is no evidence of the IDH1 mutation by molecular techniques to below the lower limit of detection (0.02%-0.04%) for ≥1 on-treatment time point. A CMH test stratified by the randomization stratification factors will be used to compare the rate of CR with IDH1 MC between the 2 treatment arms and the nominal 1-sided p-value will be presented. The odds ratio and its associated 95% CI will be presented.

IDH1 MC by sample type (BMMCs and PBMCs) and treatment arm will be summarized in a table by best objective response. The difference in percent of subjects who achieved IDH1 MC in each sample type and treatment arm will be compared between subjects achieving CR to subjects not achieving CR as best objective response using a Fisher's exact test.

[redacted]

7.10. Interim Analyses

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

Safety data will be reviewed regularly by an I-DMC to ensure the safety of the combination therapy. These reviews will occur after the first 6, 12, 24, and 36 subjects have completed 1 cycle of therapy or discontinued, whichever should occur first. Thereafter, safety reviews

will be conducted approximately every 6 months until the study is unblinded for the analysis of the primary endpoint.

8. REFERENCES

Dombret H, Seymour JF, Butrym A, Wierzbowska A, Selleslag D, Jang JH, 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.

Royston P, Parmar MK. The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt. *Statistics in Medicine.* 2011;30(19):2409-2421.

Uno H, et al.. Moving Beyond the Hazard Ratio in Quantifying the Between-Group Difference in Survival Analysis. *J Clin Oncol.* 2014;32.

Westfall PH, and Krishen A. Optimally weighted, fixed sequence, and gatekeeping multiple testing procedures. *Journal of Statistical Planning and Inference.* 2001;99:25-40.

Zhang X. Comparison of restricted mean survival times between treatments based on a stratified Cox model. *Bio-Algorithms and Med-Systems*. 2013;9(4):183–9.