Official Title of Study:

A Phase 3, Multicenter, Open-label, Randomized Study Comparing the Efficacy and Safety of AG-221 (CC-90007) Versus Conventional Care Regimens in Older Subjects with Late Stage Acute Myeloid Leukemia Harboring an Isocitrate Dehydrogenase 2 Mutation

PROTOCOL(S) AG-221-AML-004

NCT Number: NCT02577406

Document Date (Date in which document was last revised): May 29, 2020

# STATISTICAL ANALYSIS PLAN

A Phase 3, Multicenter, Open-label, Randomized Study Comparing the Efficacy and Safety of AG-221 (CC-90007) Versus Conventional Care Regimens in Older Subjects with Late Stage Acute Myeloid Leukemia Harboring an Isocitrate Dehydrogenase 2 Mutation

**STUDY DRUG:** **AG-221 (CC-90007)**

**PROTOCOL NUMBER:** **AG-221-AML-004**

**DATE FINAL:** 29 May 2020

Prepared by:

[redacted]

Celgene Corporation

86 Morris Avenue

Summit, NJ 07901

[redacted]

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# SIGNATURE PAGE

| STATISTICAL ANALYSIS PLAN (SAP) AND SAP AMENDMENT APPROVAL SIGNATURE PAGE | | |
|---|---|---|
| SAP TITLE | AG-221-AML-004 Statistical Analysis Plan | |
| SAP VERSION, DATE | Version 2, 29 May 2020 | |
| SAP AUTHOR | [redacted] | *{See appended electronic signature page}* |
| | Printed Name and Title | Signature and Date |
| PROTOCOL TITLE | A PHASE 3, MULTICENTER, OPEN-LABEL, RANDOMIZED STUDY COMPARING THE EFFICACY AND SAFETY OF AG-221 (CC-90007) VERSUS CONVENTIONAL CARE REGIMENS IN OLDER SUBJECTS WITH LATE STAGE ACUTE MYELOID LEUKEMIA HARBORING AN ISOCITRATE DEHYDROGENASE 2 MUTATION | |
| INVESTIGATIONAL PRODUCT | AG-221 (CC-90007/ENASIDENIB) | |
| PROTOCOL NUMBER | AG-221-AML-004 | |
| PROTOCOL VERSION, DATE | Amendment No 2.0, 30 Nov 2017 | |
| SIGNATURE STATEMENT | By my signature, I indicate I have reviewed this SAP and find its contents to be acceptable. | |

**Statistical Therapeutic Area Head**

Signature _____

Printed Name [redacted] Date _____

**Lead Clinical Research Physician / Clinical Research Physician**

Signature _____

Printed Name [redacted] Date _____

**Lead Product Safety Physician**

Signature [redacted]

Printed Name [redacted] Date [redacted]

# 1. LIST OF ABBREVIATIONS

**Table 1: Abbreviations and Specialist Terms**

| Abbreviation or Specialist Term | Explanation |
|---|---|
| ADaM | Analysis Data Model |
| ADT | Analysis Date |
| AE | Adverse event |
| ALT | Alanine aminotransferase (SGPT) |
| AML | Acute myeloid leukemia |
| ANC | Absolute Neutrophil Count |
| AST | Aspartate aminotransferase (SGOT) |
| ATC | Anatomical therapeutic chemical |
| BID | Twice a day |
| BMI | Body mass index |
| BSA | Body surface area |
| BSC | Best supportive care |
| BPM | Beats per minute |
| CCR | Conventional care regimen |
| CI | Confidence interval |
| CR | Morphologic complete remission |
| CRc | Cytogenetic Complete Remission |
| CRF | Case report form |
| CRh | Morphologic complete remission with partial hematologic recovery |
| CRi | Morphologic complete remission with incomplete neutrophil recovery |
| CRp | Morphologic complete remission with incomplete platelet recovery |
| CTCAE | Common Terminology Criteria for Adverse Events |
| DMC | Data Monitoring Committee |
| ECG | Electrocardiogram |
| ECOG | Eastern Cooperative Oncology Group |
| eCRF | Electronic case report form |
| EFS | Event-free survival |
| EORTC | European Organization for Research and Treatment of Cancer |

| Abbreviation or Specialist Term | Explanation |
|---|---|
| EOT | End of treatment |
| FCBP | Females of childbearing potential |
| HI | Hematologic improvement |
| HI-E | Hematologic improvement erythroid response |
| HI-N | Hematologic improvement neutrophil response |
| HI-P | Hematologic improvement platelet response |
| HLT | High-Level Term |
| HSCT | Hematopoietic stem cell transplantation |
| IDAC | Intermediate-dose cytarabine |
| IDH2 | Isocitrate dehydrogenase isoform 2 |
| IRAC | Independent Response Assessment Committee |
| ITT | Intent-to-treat |
| IV | Intravenous |
| IVRS | Interactive voice response system |
| IWG | International Working Group |
| KM | Kaplan-Meier |
| [redacted] | |
| MedDRA | Medical dictionary for regulatory activities |
| MLFS | Morphologic leukemia-free state |
| NCI | National Cancer Institute |
| ORR | Overall response rate |
| OS | Overall survival |
| [redacted] | |
| PR | Partial remission; same as partial response |
| PD | Progressive disease |
| QTcB | Heart-rate corrected QT with Bazett's correction |
| QTcF | Heart-rate corrected QT with Fridericia's correction |
| SAE | Serious adverse event |
| SD | Stable disease |
| SGPT | Serum glutamic pyruvic transaminase |
| SMQ | Standardized MedDRA Query |

| Abbreviation or Specialist Term | Explanation |
|---|---|
| SOC | System organ class |
| TEAE | Treatment emergent adverse event |
| ULN | Upper limit of normal |

# 2. INTRODUCTION

This statistical analysis plan (SAP) describes the analyses and data presentations for Celgene's protocol AG-221-AML-004 "A Phase 3, Multicenter, Open-label, Randomized Study Comparing the Efficacy and Safety of AG-221 (CC-90007) Versus Conventional Care Regimens (CCR) in Older Subjects with Late Stage Acute Myeloid Leukemia Harboring an Isocitrate Dehydrogenase 2 Mutation," which was issued on 13 Aug 2015 and amended to Amendment 2.0 on 30 Nov 2017. It contains definitions of analysis populations, derived key efficacy variables, and statistical methods for the analysis of efficacy and safety data.

The efficacy analyses include one interim analysis and one final analysis. Throughout this SAP, the treatment arms will be referred to as AG-221 and CCRs (Conventional Care Regimens). Treatment arm AG-221 refers to AG-221 Orally (PO) plus best supportive care (BSC). The CCR treatment options include BSC only, azacitidine subcutaneously (SC) plus BSC, low-dose cytarabine (LDAC) SC plus BSC, or intermediate-dose cytarabine (IDAC) intravenously (IV) plus BSC. Following review of eligibility, subjects will be assigned by the investigator to one of the CCR treatment options based on the investigator's assessment of subjects' eligibility. Following the selection of a CCR treatment option, subjects will be randomized centrally in a 1:1 ratio to receive either the AG-221 treatment or the CCR treatment option pre-selected by the investigator.

A planned interim analysis was completed in March 2019, following the reporting of 65% of the expected number of deaths (approximately 163 deaths). Following this the study was continued to final analysis. This SAP provides a comprehensive and detailed description of the strategy, rationale, and statistical techniques to evaluate the efficacy and safety endpoints. The purpose of the SAP is to ensure the credibility of the study findings by pre-specifying the statistical approaches to the primary efficacy analysis of study data prior to database lock for the final analysis. A data cut-off date for the locked database for the final analysis will be determined and the database will be locked after medical/scientific review of the data has been completed, protocol violations/deviations have been identified, and the data have been declared to be clean. All statistical analyses detailed in this SAP will be conducted using SAS® Version 9.2 or higher.

The analysis methods and timing of analysis for the patient reported outcomes/quality of life [redacted] will be described in a separate pharmacoeconomics SAP. [redacted] Operational details for the Data Monitoring Committee (DMC) during the course of the study will be discussed in a separate DMC charter, and will not be included in this SAP either.

# 3. STUDY OBJECTIVES

## 3.1. Primary Objective

- To determine the primary efficacy, measured as overall survival (OS), of AG-221 compared with CCRs in subjects 60 years or older with acute myeloid leukemia (AML) refractory to or relapsed after second- or third-line AML therapy and positive for an isocitrate dehydrogenase isoform 2 (IDH2) mutation

## 3.2. Secondary Objectives

- To determine the supporting efficacy of AG-221 compared with CCRs
- To determine the safety and tolerability of AG-221 compared with CCRs
- To determine the effect of AG-221 compared with CCRs on Health-related Quality-of-Life (HRQoL)

[redacted]

# 4. INVESTIGATIONAL PLAN

## 4.1. Overall Study Design and Plan

This is an international, multicenter, open-label, randomized, Phase 3 study comparing the efficacy and safety of AG-221 versus CCRs in subjects 60 years or older with AML refractory to or relapsed after second- or third-line AML therapy and positive for an IDH2 mutation.

It is planned that approximately 316 subjects will be randomized in a 1:1 ratio to receive either the AG-221 treatment or the CCR treatment option pre-selected by the Investigator prior to randomization. Randomization will be stratified by prior intensive therapy for AML (yes versus no), primary refractory (i.e., morphologic complete remission [CR], CR with incomplete neutrophil recovery [CRi], or CR with incomplete platelet recovery [CRp] has never been attained) (yes versus no), and prior allogeneic hematopoietic stem cell transplantation (HSCT) for AML (yes versus no).

The treatment options in the CCR treatment arm include:

- BSC only: continuous 28-day cycles of BSC. Best supportive care includes, but is not limited to, hydroxyurea for leukocytosis and/or differentiation-like syndrome, anti-infectives, analgesics, antiemetics, antipyretics, transfusions and nutritional support (refer to protocol Section 8 for details);
- Azacitidine SC plus BSC: continuous 28-day cycles of azacitidine 75 mg/m$^2$/day SC for 7 days, plus BSC;
- LDAC SC plus BSC: continuous 28-day cycles of cytarabine 20 mg SC twice a day (BID) for 10 days, plus BSC;
- IDAC IV plus BSC: 28-day cycles of cytarabine 0.5 to 1.5 g/m$^2$/day IV for 3 to 6 days, per standard institutional practice, plus BSC; only BSC given after IDAC therapy concludes per standard institutional practice.

Subjects randomized to the AG-221 treatment arm will receive AG-221 100 mg PO once a day on Days 1 to 28 of each 28-day cycle, plus BSC. Dose escalation to 200 mg may occur for augmenting treatment response.

Study treatment starts within 3 days after randomization. No crossover between any of the treatment options will be permitted during the course of study treatment per protocol.

Dosing interruptions, delays, or dose modifications may occur for managing toxicities and/or augmenting treatment response during study treatment.

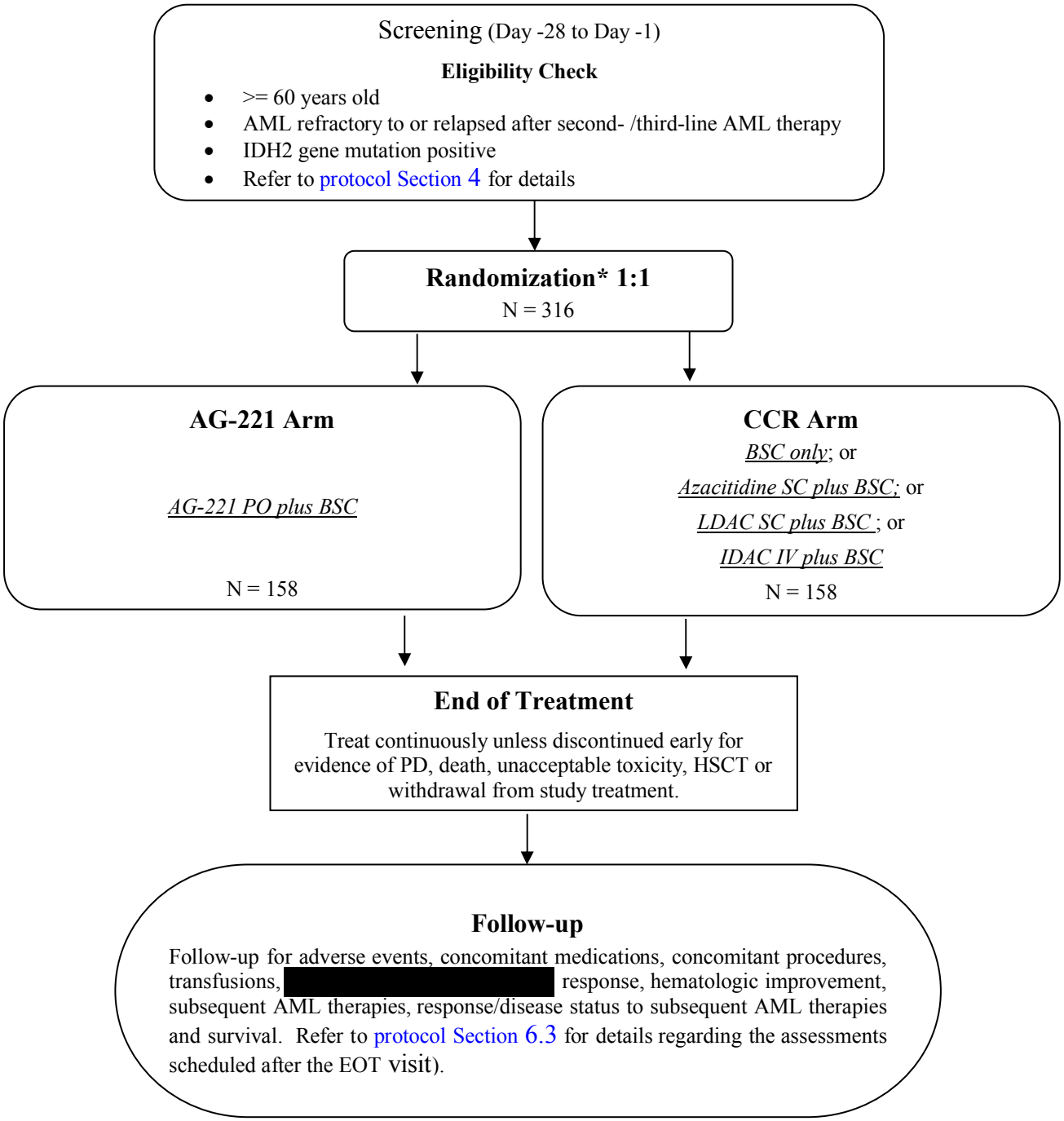
All subjects who have received at least one dose of study treatment should undergo end-of-treatment (EOT) evaluations when study treatment is discontinued. The reason for discontinuation will be recorded in the electronic case report form (eCRF) pages and in the source document.

All subjects discontinued from study treatment for any reason other than withdrawal of consent for follow-up will continue to be assessed for adverse events (AEs), concomitant medications,

concomitant procedures, transfusions, [redacted] response, hematologic improvement (HI), subsequent AML therapies, and survival.

The study schematic is presented in Figure 1 below.

**Figure 1: Overall Study Design**




Key: AML = acute myeloid leukemia; BSC = best supportive care; CCR = conventional care regimen; HSCT = hematopoietic stem cell transplantation; IDAC = Intermediate-dose cytarabine; IDH2 = isocitrate dehydrogenase isoform 2; IV = intravenously; LDAC = low-dose cytarabine; PD = progressive disease; PO = orally; SC = subcutaneously.

* Stratification factors: prior intensive therapy for AML (yes versus no), primary refractory (yes versus no), and prior allogeneic HSCT for AML (yes versus no).

## 4.2. Study Endpoints

### 4.2.1. Primary Endpoint

The primary endpoint is overall survival (OS), defined as time from randomization to death due to any cause.

### 4.2.2. Secondary Endpoints

The secondary endpoints are listed in Table 2 below.

**Table 2: Secondary Endpoints**

| Name | Description |
|---|---|
| Overall response rate | Rate of Best Response in CR + CRi + CRp + PR + MLFS according to modified IWG AML response criteria |
| Event-free survival | Time from randomization to documented morphologic relapse, PD according to modified IWG AML response criteria or death from any cause, whichever occurs first |
| Duration of response | Time from the first documented CR/CRi/CRp/PR/MLFS to documented morphologic relapse, PD according to modified IWG AML response criteria or death due to any cause, whichever occurs first |
| Time to response | Time from randomization to first documented CR/CRi/CRp/PR/MLFS according to modified IWG AML response criteria |
| Time to best response | Time from randomization to the documented best response of CR/CRi/CRp/PR/MLFS according to modified IWG AML response criteria |
| Treatment mortality at 30- and 60-day | Rate of death from any cause within 30 and 60 days of initiation of study treatment |
| One-year survival | The probability of survival at 1 year from randomization |
| Overall remission rate | Rate of CR + CRi + CRp according to modified IWG AML response criteria |

| Name | Description |
|---|---|
| Complete remission rate | Rate of CR according to modified IWG AML response criteria |
| Hematologic improvement rate | Rate of HI-N + HI-P + HI-E according to IWG MDS HI criteria |
| Rate of HSCT | Rate of bridge-to-HSCT through study treatment |
| Time to treatment failure | Time from randomization to discontinuation of study treatment due to any cause |
| Rate of CR/CRh | Rate of CR + CRh (Appendix 15.5) |
| Duration of CR/CRh | Time from the first documented CR/CRh to relapse according to modified IWG AML response criteria or death due to any cause, whichever occurs first |
| Time to CR/CRh | Time from randomization to first documented CR/CRh |
| Safety and tolerability | Type, frequency, severity, seriousness and relationship of adverse events to study treatments; vital signs; ECG; clinical laboratory evaluations |
| HRQoL | European Organization for Research and Treatment of Cancer Quality-of-Life questionnaire (EORTC QLQ-C30) and EuroQoL Group EQ-5D-5L instrument |

Abbreviations: AML = acute myeloid leukemia; CR = morphologic complete remission; CRi = morphologic complete remission with incomplete neutrophil recovery; CRp = morphologic complete remission with incomplete platelet recovery; CRh=morphologic complete remission with partial hematologic recovery; ECG = electrocardiogram; EORTC= European Organization for Research and Treatment of Cancer; HI = hematologic improvement; HI-E = hematologic improvement erythroid response; HI-N = hematologic improvement neutrophil response; HI-P = hematologic improvement platelet response; HRQoL = Health-related Quality of Life; HSCT = hematopoietic stem cell transplantation; IWG = International Working Group; MDS = myelodysplastic syndromes; MLFS = morphologic leukemia-free state; PR = partial remission; PD = progressive disease.

[redacted]

[redacted]

### 4.2.4. Safety Endpoints

Safety assessments include physical examination, vital signs, electrocardiogram (ECG), hematology, serum chemistry, cardiac markers, fasting lipid panel, pregnancy testing (for females of childbearing potential [FCBPs] only), AEs, concomitant medications, concomitant procedures and transfusions. After screening, echocardiogram (ECHO), urinalysis and coagulation will be repeated as clinically indicated.

## 4.3. Stratification, Randomization, and Blinding

Randomization will be carried out using an Interactive Voice Response System (IVRS). Following the selection of a CCR treatment option, subjects will be randomized centrally in a 1:1 ratio to receive either AG-221 treatment or one of the CCR treatment options pre-selected by the Investigator based on standard institutional practice and on evaluation of the subject's underlying disease condition. To avoid potential bias, the intended treatment selection from the CCR treatment arm by the Investigator will be recorded for all subjects prior to randomization; treatment assignment is not to be subsequently changed after randomization.

To further minimize the heterogeneity of treatment modalities and clinical outcomes in AML, randomization in this study is stratified by the following three factors that are believed to have the most impact to the study's primary endpoint, OS: prior intensive therapy for AML (yes versus no), primary refractory (i.e., CR/CRi/CRp has never been attained) (yes versus no), and prior allogeneic HSCT for AML (yes versus no).

Although this is an open-label study, the interim efficacy analysis will be performed by an independent third party and results will be kept strictly confidential in order to maintain the integrity of the study. The study team will be blinded at summary-level information using actual treatment, i.e. no summary tables and figures will be produced or reviewed by the study team using actual treatment information. An independent unblinded team will produce the unblinded

results and send to an independent Data Monitoring Committee (DMC) who will review and give advice to the Sponsor regarding the study conduct.

## 4.4. Sample Size Determination

The equality of OS curves will be compared between the AG-221 and combined CCR treatment arms using a stratified log-rank test. Subjects will be randomized to receive AG-221 treatment or a CCR treatment option pre-selected by the Investigator in a 1:1 ratio. Assuming a median OS of 5.6 months in the CCR treatment arm, a median OS of 8 months in the AG 221 treatment arm (42.9% improvement), and a drop-out rate of approximately 9%, this design requires 316 subjects (158 per treatment arm) to be randomized, and 250 deaths to be observed in order to achieve 80% power to detect a hazard ratio (HR) of 0.7 and demonstrate a statistically significant difference in OS at type I error rate of 0.05 (two-sided). It is assumed that the OS distribution is exponential with a constant failure rate and that accrual rate is 3 subjects per month for the first 12 months and 9 subjects per month afterwards. The interim analysis for superiority will be conducted at 65% information (i.e., approximately 163 deaths) based on Lan-DeMets version of the O'Brien-Fleming alpha spending function (DeMets and Lan, 2008)

[redacted]

# 5. GENERAL STATISTICAL CONSIDERATIONS

## 5.1. Reporting Conventions

The following reporting conventions apply generally to tables, listings, and figures:

- Data from all study centers will be combined for analysis;
- Efficacy and Safety analyses will be done by treatment groups/options (AG-221 versus CCR overall and furthermore AG-221 versus, BSC Only, Azacitidine + BSC, LDAC + BSC, and IDAC + BSC by pre-selected CCR group). Exceptions will be noted in corresponding sections;
- All stratified efficacy analyses will use the randomization stratification factors taken from CRF (primary) including prior intensive therapy for AML (yes versus no), primary refractory (yes versus no), and prior allogeneic HSCT for AML (yes versus no) and repeated for stratification factors taken from IVRS.
- In stratified analysis, if stratified model does not converge due to small sample size, the stratification factor leads to least subjects will be dropped from the model. If a stratification factor has less than 40 subjects in one of its two levels across treatment groups, then this factor will not be used in stratified analysis;
- All statistical tests of the treatment effect will preserve a significance level of 0.05 for two-sided tests, unless otherwise specified;
- P-values will be rounded to 4 decimal places. P-values that round to 0.0000 will be presented as '<0.0001' and p-values that round to 1.000 will be presented as '>0.9999';
- Confidence intervals (CIs) will be presented as two-sided 95% CIs unless specified differently in specific analysis;
- Summary statistics will consist of the number and percentage of subjects (or cycles, if appropriate) in each category for discrete variables, and the sample size, mean, median, standard deviation, minimum, and maximum for continuous variables;
- All mean and median values will be formatted to one more decimal place than the measured value. Standard deviation values will be formatted to two more decimal places than the measured value;
- All percentages will be rounded to one decimal place. The number and percentage of responses will be presented in the form XX (XX.X%), where the percentage is in the parentheses;
- All listings will be sorted for presentation in order of treatment group, study center, subject, and date/time of procedure or event. Subjects who are screened but not randomized will be listed separately as a non-randomized group;
- All analysis and summary tables will have the analysis population sample size (i.e., number of subjects);

- The day of the first dose of any study drug will be defined as Day 1; for BSC only subjects, Day 1 is the randomization day.
- Unless noted otherwise, baseline value will be defined as the last non-missing value prior to or on the start of study treatment on Day 1 of Cycle 1. For subjects in the BSC only group, the baseline value will be defined as the last non-missing value on or prior to date of randomization. For subjects who are randomized but not treated, the baseline will be the assessment value on or prior to randomization date.
- To summarize efficacy/safety data by visit, the post-baseline results will be summarized by the scheduled visit or the derived visit as appropriate.
- In general, the derived visit is the same as the scheduled visit in the protocol which is defined by study day; data that is closest to derived visit (before and after) will be used as the value for that visit. In case that two data collected at the equal distance to the derived visit, the later one will be used.
- Calculation of Cycles:

The start date of each treatment cycle will be calculated based on study drug exposure records for each subject, except for BSC only subjects. The start date of the first cycle will be the earliest date the subject receives any study drug or the date of randomization for subjects receiving BSC only. Once the start dates, e.g., $S_1$, $S_2$, $S_3$… are calculated, the end date of each cycle is calculated as the day before the start date of the following cycle, i.e., $E_i = S_{i+1} - 1$. For subjects receiving BSC only, each cycle has a defined length of 28 days; for the last cycle, the end date will be calculated as the date of discontinuation from the treatment period. The cycle number for each date of interest, e.g., adverse event (AE) will be calculated based on event start date and the cycle window start and end dates. If an event date is on or after $S_i$ and before $S_{i+1}$, the corresponding cycle number will be i.

## 5.2. Analysis Populations

### 5.2.1. Intent-to-Treat Population

The intent-to-treat (ITT) population includes all subjects who are randomized to treatment, regardless of whether they received treatment or not.

### 5.2.2. Modified Intent-to-Treat Population

The modified intent-to-treat (mITT) population includes all subjects who have met all eligibility criteria and experienced no major protocol deviations during the study, received at least one dose of study treatment (except for BSC subjects) and had at least one post-randomization treatment response assessment performed. Major protocol deviations leading to exclusion from the mITT population are defined as any of the following:

- Had at least one inclusion/exclusion criteria violation;
- Disease status at screening not confirmed by retrospective central review;
- Initial study treatment given is not the one assigned in randomization;

- Did not have at least one post-randomization efficacy assessment performed;

### 5.2.3. Safety Population

The safety population includes all randomized subjects who received at least 1 dose of study treatment. Because BSC only treatment may consist of blood products or antibiotics, etc. administered as needed, subjects who are randomized to the BSC arm option will be included in the safety population if they have had at least 1 safety assessment after randomization. The safety population will be used for all safety analyses and drug exposure. Subjects will be analyzed according to the initial treatment actually received.

[redacted]

# 6. SUBJECT DISPOSITION

Subject disposition (analysis population allocation, entered, discontinued, along with primary reason for discontinuation) will be summarized using frequency and percent for both treatment and follow-up phases. Enrollment by country and site will also be summarized.

The disposition of subjects will be summarized with counts and percentages. A summary of subject disposition will be presented by treatment group for the following analysis populations:

- ITT population
- mITT population
- Safety population

A listing of reasons for exclusion from the mITT population will be provided. A separate listing will be provided for subjects who were not randomized (screen failures) with reasons for screening failure.

Reasons for treatment discontinuation will be collected on the case report form (CRF) and will be summarized for all randomized subjects with the following categories:

- Death
- Adverse event
- Pregnancy
- Progressive disease
- Withdrawal by subject
- Non-compliance with study drug
- Lost to follow-up
- Study terminated by Sponsor
- Transition to commercially available treatment
- Physician decision
- Disease relapse
- Protocol violation (to be specified on the eCRF)
- Other (to be specified on the eCRF)
- Other: Hematopoietic Stem Cell Transplantation

Reasons for study discontinuation will be collected on the CRF and will be summarized for all randomized subjects and all subjects who enter the follow-up phase with the following categories:

- Screen failure
- Death
- Adverse event
- Pregnancy
- Progressive disease
- Withdrawal by subject
- Lost to follow up
- Study terminated by sponsor
- Transition to commercially available treatment
- Physician decision
- Disease relapse
- Protocol violation (to be specified on the eCRF)
- Other (to be specified on the eCRF)

Additional summaries will be provided for number of subjects per randomization stratum by randomized treatment group, as well as number of subjects by study site.

# 7. PROTOCOL DEVIATIONS/VIOLATIONS

The protocol deviations/violations will be identified and assessed by clinical research physician or designee following company standard operational procedure. These protocol deviations/violations will be summarized by treatment group for the ITT population.

Protocol deviations and violations will be reviewed prior to database hard lock. Major protocol violations will be used to determine the mITT population (defined in Section 5.2.2).

Major protocol deviations and violations will be summarized using frequency tabulations. A by-subject listing of subjects with all protocol deviations/violations in the ITT population will be provided.

# 8. DEMOGRAPHICS AND BASELINE CHARACTERISTICS

Demographic and baseline disease characteristics will be summarized by treatment group and overall for the ITT, mITT and safety populations, and for Bridge to HSCT subjects in safety population.

Individual subject listings will be provided to support the summary tables.

## 8.1. Demographics

Age (year), height (cm), weight (kg), body mass index (kg/m$^2$), body surface area (m$^2$), and other continuous baseline characteristics will be summarized using descriptive statistics (N, mean, standard deviation, median, minimum, maximum), while age group ≥60 years and <70 years, ≥70 years and <80 years, ≥80 years), gender, ethnicity, race, geographic region and other categorical variables will be provided using frequency tabulations (count, percent) by treatment group and overall.

Age or year of birth will be recorded on CRF. Where age is not recorded, age will be calculated as follows: Age = year of informed consent – year of birth.

Body mass index (BMI) will be calculated as follows: BMI (kg/m$^2$) = weight in kg / (height in m)$^2$.

If not collected in CRF, body surface area (BSA) will be calculated per the Dubois & Dubois formula: BSA (m$^2$) = weight (kg)$^{0.425}$ x height (cm)$^{0.725}$/139.2.

## 8.2. Baseline Disease Characteristics and AML Diagnosis History

The number and percentage of subjects in each of the following categories will be summarized by treatment group and overall. Continuous variable will be summarized by mean, standard deviation, median, minimum, and maximum.

Depending on the way in which data is collected, information on the same CRF page will be listed in one listing. Multiple listings will be provided for the information above.

### 8.2.1. Baseline Disease Characteristics

1. IDH2 gene mutate type (R140 or R172);
2. Eastern Cooperative Oncology Group (ECOG) performance status;
3. Prior intensive therapy for AML (yes versus no);
4. Prior refractory (yes versus no);
5. Prior stem cell transplants for AML (yes versus no; if yes, type);
6. Prior systemic anti-cancer therapies for AML;
7. Time (days) from last prior HSCT, calculated from starting date of the last prior HSCT to randomization date (≤ 100 days; > 100 days to ≤ 365 days; > 365 days);

8. Bone marrow blasts (%) and category (< 20%; 20% to < 30%;30% to < 50%; ≥50%) from the screening bone marrow aspirate sample (local, central) and bone marrow biopsy sample (local, central);
9. Peripheral blood blast (%) (local, central);
10. Cytogenetic Risk Status (better-risk, intermediate-risk, poor-risk, failure, normal, not done) (local);
11. Select Chromosomal Abnormalities at Baseline;
12. Co-occurring mutations (FLT3-ITD, NPM, TET2, IDH1) if data is available;
13. Baseline values of the following lab parameters: Hemoglobin (< 80 g/L , ≥ 80 g/L), platelet count (< 50 x10^9/L, ≥ 50 x10^9/L), absolute neutrophil count (ANC) (< 0.5 x10^9/L, 0.5 to < 1 x10^9/L, ≥ 1.0 x10^9/L), white blood cells (WBC) (<15 x10^9/L, 15 to < 30 x10^9/L, ≥ 30 x10^9/L), creatinine clearance (< 45 mL/min, 45 to < 60 mL/min, ≥ 60 mL/min);
14. Number of transfusions and number of units transfused of red blood cell (RBC) and platelets within 56 days prior to randomization, and baseline transfusion dependence for both RBC and platelets.
15. Left Ventricular Ejection Fraction (LVEF) from Echocardiogram and Multigated Acquisition (MUGA)
16. Prior lines of AML Treatment and category (0 lines, 1lines, 2 lines, 3 lines, 4 lines, >4 lines), Prior refractory status, and prior relapse status
17. Best response to Prior lines of AML Treatment, split by Low-Intensity AML Therapy and Intensive Chemotherapy for First line, Second Line, Third Lines, Fourth Line, and Last line of prior AML treatment

### 8.2.2. AML Diagnosis History

1. WHO Classification
2. AML initial diagnosis: primary (i.e., de novo) or secondary (MDS progression or myeloproliferative neoplasms (MPN), or therapy-related) according to WHO classification. If AML is secondary, number and percentage of subjects who has an antecedent hematologic disorder will be summarized, and type of disorder will be tabulated too;
3. IPSS Classification for subjects with Higher Risk MDS
4. Time (month) from the initial diagnosis of AML, calculated from time of initial diagnosis to randomization date.

## 8.3. Medical History and Concomitant Disease

The medical and surgical history, as well as concomitant disease will be summarized by the Medical Dictionary for Regulatory Activities (MedDRA) latest version by system organ class (SOC) and preferred term (PT) for the ITT population and safety population. Active concomitant disease will be summarized in separate tables.

## 8.4. Prior/Concomitant/Follow-up Medications/Procedures/Therapies

Prior medications/procedures/transfusions are defined as medications/procedures/transfusions that were started prior to start of study treatment (start of study treatment refers to randomization date for BSC only group and first dose of study drug for other treatment groups) or have start dates missing. If start date is missing and end date is before start of study treatment medication will be classed as prior only.

Concomitant medications/procedures/transfusions are defined as non-study medications/procedures/transfusions that are started on or after the start of study treatment but before the end of the study treatment period or started before the start of the study treatment and ended or remain ongoing during the study treatment. For non-BSC groups, the end of the study treatment period is defined as the earlier date of last dose of study drug + 28 days and date of death. For BSC groups, the end of the study treatment period is defined as the earlier date of discontinuation from the treatment period and date of death.

Medication/procedure/transfusion with missing start date but end date on or after start of the study treatment (randomization date for BSC only group) and before the end of the study treatment period or with both missing start date and end date is also considered as concomitant medication/procedure/transfusion.

Follow-up medications/transfusions are defined as non-study medications/transfusions that are started after the date of the end of the study treatment period. All non-study medications and therapies will be coded by Anatomical Therapeutic Chemical (ATC) code and PT using the World Health Organization Drug Dictionary latest version. All procedures and surgeries will be coded by the Medical Dictionary for Regulatory Activities (MedDRA) latest version by system organ class (SOC) and preferred term (PT).

### 8.4.1. Prior/Concomitant/Follow-up Medications

All prior, concomitant and follow-up medications will be summarized in frequency tabulations (subject counts and percentages) and by World Health Organization ATC first level and PT. Listings will also be provided for general medications and concomitant QT prolonging medications. All tables described in this section will be produced on safety population.

### 8.4.2. Prior/Concomitant/Follow-up Procedures

All prior, concomitant and follow-up procedures will be summarized in frequency tabulations (subject counts and percentages) and by SOC and PT. Listings will also be provided for general medications/procedures and concomitant QT prolonging medications. All tables described in this section will be produced on safety population.

### 8.4.3. Prior/Follow-up Anti-Cancer Therapies

This section includes both AML therapies and non- AML therapies. AML therapies include surgery, radiation, systemic or any other therapy for the subject's AML disease, particularly intensive AML chemotherapies as induction/re-induction/salvage chemotherapies, consolidation or maintenance therapies, HSCT, low-intensity AML therapies such as azacitidine, decitabine or LDAC, or other medications considered supportive care for AML, regardless of discontinuation

date of treatment. Non-AML therapies may consist of the same therapies like AML therapies but for other hematologic or non-hematologic malignancies.

All tables described in this section will be produced on both ITT and safety population.

Prior and follow-up therapies will be summarized separately in tables.

#### 8.4.3.1. Anti-Cancer Procedures/Surgeries

A listing will be provided for procedure/surgery including indication and date of procedures/surgeries. All prior, concomitant and follow-up procedures will be summarized in frequency tabulations (subject counts and percentages) and by SOC and PT.

#### 8.4.3.2. Radiations

The number and percentage of subjects who had any prior/follow-up radiation therapy will be presented by treatment group for the safety and ITT populations. Summaries will include total frequency of subjects receiving any prior/follow-up radiation therapy, as well as the frequency by type of radiation therapy received (i.e., external beam, radio-immuno therapy, brachytherapy, etc.) will be presented by treatment group and overall. For subjects with radiation therapy, treatment site, type of radiation therapy, location of external beam, start/stop date, dose, number of fraction, intent of therapy, and settings of therapy will be presented in a listing.

#### 8.4.3.3. Systemic Anti-Cancer Medication for AML

The number and percentage of subjects with any prior/follow-up systemic anti-cancer medication for AML will be presented by treatment group and overall for ITT and safety populations. A similar summary will also be presented for medication for AML or Higher Risk MDS. A summary of the Follow-up AML therapies, and their associated responses will also be presented. For subjects with systemic anti-cancer therapy for AML, detailed information will be presented in a listing.

#### 8.4.3.4. Systemic Anti-Cancer Medication Other than AML

The number and percentage of subjects with any prior/follow-up systemic anti-cancer therapy for diseases other than AML will be presented by treatment group and overall for ITT and safety populations. For subjects with systemic anti-cancer therapy, detailed information will be presented in a listing.

#### 8.4.3.5. Stem Cell Transplants for AML

The number and percentage of subjects with any prior/follow-up stem cell transplants for AML will be presented by treatment group and overall for ITT and safety populations. For subjects with stem cell transplants, type, date of procedure, intent of therapy, and disease status at the time of HSCT will be presented in a listing.

### 8.4.4. Prior/Concomitant/Follow-up Transfusions

The number of subjects receiving any transfusion product, and the number of subjects receiving any transfusion of particular type (red blood cells, whole blood, platelets, plasma, other) will be

summarized by treatment groups. Prior /concomitant/follow-up transfusions will be summarized in separate tables for the safety population.

The number of subjects who are RBC transfusion dependent versus independent at baseline, the median transfusion burden at baseline, the changes from dependence to independence, independence remaining independent achieved at least 56 days RBC transfusion independence during treatment period together with the time to transfusion independence and duration of transfusion independence will be summarized by treatment groups.

The number of subjects who are platelet transfusion dependent versus independent at baseline, the median transfusion burden at baseline, the changes from dependence to independence, independence remaining independent achieved at least 56 days platelet transfusion independence during treatment period together with the time to transfusion independence and duration of transfusion independence will be summarized by treatment groups.

Transfusion type, number of units, reasons for taking transfusion, and dates of transfusions will be listed.

# 9. STUDY TREATMENTS AND EXTENT OF EXPOSURE

All analyses of treatment exposure will be conducted using the safety population. Only exposure to the AG-221, azacitidine, LDAC, and IDAC will be included in the treatment exposure summaries. Treatment exposure is not applicable to the BSC only treatment group. Duration of treatment and number of cycles will be calculated for all treatment groups, including the BSC only group.

Descriptive statistics for duration of treatment exposure, duration of treatment, average prescribed daily dose, average calculated daily dose, average daily dose, cumulative dose, dose intensity and relative dose intensity, will be presented by treatment groups. Individual subject listings will be provided to support the tables.

## 9.1. Duration of Treatment Exposure and Duration of Treatment

Duration of treatment exposure and duration of treatment will be summarized by treatment group.

Duration of treatment exposure (days) is defined as: Treatment end date – treatment start date + 28 days. The treatment start date is the date of the first dose of study drug, and the treatment end date is the last dose date. If death date or data cut-off is earlier than 28 days after treatment end date, duration of treatment exposure is defined as: the earlier of death date and data cut-off – treatment start date + 1. Duration of treatment exposure will be summarized by treatment group of AG-221, azacitidine, LDAC, and IDAC. For BSC only group, duration of treatment exposure is the same as duration of treatment defined below.

For AG-221, azacitidine, LDAC, and IDAC groups, duration of treatment (days) is defined as: Treatment end date – treatment start date + 1. For subjects who are still on treatment at a data cut-off date, treatment end date will be the earlier date of data cut-off and the last dose date.

For BSC only group, duration of treatment is defined as the period from date of randomization through the date of discontinuation from the treatment period, date of death, or clinical data cut-off date, whichever is earliest (calculated as the earliest of death, clinical data cut-off, or discontinuation date – randomization date + 1). Duration of treatment will be summarized for all treatment groups including the BSC only group.

See Section 5 for cycle length and cycle number calculations. Average cycle length (days), average number of days dosed per cycle, and number of treatment cycles will be summarized by treatment group. Average cycle length and average number of days dosed per cycle will not be summarized for BSC only group. The number of days dosed per cycle is defined as the number of days with a non–missing dosing date within that cycle. For average cycle length and average number of days dosed per cycle, a single average value will be computed for each subject first and then the descriptive statistics will be computed for each treatment group. Additionally, the count and percent of subjects by number of cycles of treatment will be provided for each treatment group.

Person-years of exposure will be calculated as duration of treatment exposure (days)/365.25 for all treatment groups except BSC only group. Person-years of exposure will be calculated as duration of treatment (days)/365.25 for the BSC only group.

## 9.2. Cumulative Dose

Cumulative dose will be calculated for AG-221, azacitidine, LDAC, and IDAC treatment groups. Cumulative dose is defined as the sum of all doses taken across the treatment period in mg.

Cumulative dose for AG-221 and LDAC = Sum of (administered dose of each visit in mg);

Cumulative dose for IDAC = Sum of (administered dose in $g/m^2$ of each visit).

Cumulative dose for azacitidine = Sum of (administered dose in $mg/m^2$ of each visit).

The dosage will be counted as zero for days when the study drug is not taken.

## 9.3. Dose Intensity

Dose intensity during the treatment is defined as the cumulative dose divided by the duration of treatment. Dose intensity will be calculated in the following way:

Dose intensity for AG-221 and LDAC = [cumulative dose in mg]/[duration of treatment];

Dose intensity for IDAC = [cumulative dose in $g/m^2$]/[duration of treatment].

Dose intensity for azacitidine = [cumulative dose in $mg/m^2$]/[duration of treatment].

## 9.4. Relative Dose Intensity

Relative dose intensity is the dose intensity divided by the planned dose intensity.

For AG-221, the planned dose intensity is based on the assigned dose each time.

For LDAC, subjects randomized to LDAC treatment option will receive cytarabine 20 mg SC BID for the first 10 days of each every 28-day treatment cycle. Therefore, the planned dose intensity is 14.3 mg/day throughout the study;

For azacitidine, subjects in this group will receive 75 $mg/m^2$/day for 7 days of each 28-day treatment cycle, so the planned dose intensity is 18.75 $mg/m^2$/day;

For IDAC, subjects randomized to IDAC treatment option will receive cytarabine 0.5 to 1.5 $g/m^2$ IV for the first 3 to 6 days of each 28-day treatment cycle. The dosing schedule is flexible with each site, so for calculation purpose, prior to the final analysis, the planned dose intensity for one specific cycle will be calculated in the following way:

Planned dose intensity = [dose administered ($g/m^2$) x actual duration (Days)]/28, which equals to X.X $g/m^2$/day.

Relative dose intensity for all four drugs mentioned above will be categorized into < 75%, 75% to < 90%, and $\geq$ 90%, and frequency counts will be provided by treatment group.

## 9.5. Average Daily Dose

Average daily dose during treatment is defined as the cumulative dose divided by number of days dosed. A single average value will be computed for each subject, overall and within each cycle, and then the descriptive statistics will be computed for each treatment group. Average daily dose will be summarized for the prescribed daily dose (mg/day for AG221 and LDAC; $g/m^2$/day for IDAC; and $mg/m^2$/day for azacitidine), and the total daily dose.

## 9.6. Dose Modification

This part does not apply to BSC only group.

The number of subjects who have at least one dose modification, number of dose modifications per subject, and reasons for dose modifications will be summarized by treatment group. Details of dose adjustments will be provided in a by subject listing.

According to protocol, under specific circumstances, AG-221 dose may increase for augmenting treatment responses. Number of subjects who undergo dose escalation, number of subjects who have dose modification after dose escalation will be summarized. Reasons for dose modifications after dose escalations will also be summarized.

## 9.7. Compliance

Overall and by cycle study medication compliance, which is measured by relative dose intensity (see Section 9.4 for details), will be summarised by treatment.

# 10. EFFICACY ANALYSIS

All efficacy analysis will be performed on the ITT population unless otherwise specified. Key efficacy analyses will also be performed on the mITT population as supportive evidence and to assess the robustness of the efficacy findings. Subjects will be analyzed according to randomized treatment group. Statistical comparisons will be made between AG-221 and CCR overall and by pre-selected CCR therapy whenever applicable.

All statistical tests will be two-sided at the significance level of 0.05, and the corresponding two-sided p-values and two-sided 95% confidence interval (CI) for intended point estimates will be reported, unless specified otherwise.

## 10.1. Multiplicity

In order to perform hypothesis testing on multiple endpoints while controlling the overall Type I error rate, a closed, sequential testing approach is proposed to test endpoints in a pre-specified order. The primary efficacy endpoint of OS will be tested first at a two-sided significance level of 0.05. If superiority of AG-221 is demonstrated in OS, then the first key secondary endpoints of overall response rate (ORR) will be tested next. The second key secondary endpoint of the event-free survival (EFS) will be tested only if the test results for both OS and ORR are significant each at a two-sided 0.05 significance level. The interim superiority analysis adopts Lan-DeMets version of the O'Brien-Fleming alpha spending function to control the overall Type I error rate. The alpha spending completed in the interim analysis is further detailed in Section 12.2 of this document. The multiplicity strategy is applied to hypothesis tests between AG-221 and CCR overall group only.

## 10.2. Analysis of Primary Efficacy Endpoint

The primary efficacy endpoint of OS is defined as the time between randomization and death from any cause. Subjects who are alive or whose death status is unknown will be censored at the last date known alive on or before data cutoff date. The OS period in months is calculated as (the earlier of death date/the last date known alive and cutoff date – date of randomization + 1)/30.4375. For subjects who were known to be alive and in the study at data cut-off (were not lost to follow-up, and had not withdrawn consent), the last known-to-be-alive date is set to be the data cut-off date.

The analysis of the primary endpoint will be based on the ITT population. Subjects will be analyzed as they were randomized, regardless of the actual treatment received. The survival distribution of OS will be estimated using Kaplan-Meier (KM) method. The 25$^{th}$ percentile, median and 75$^{th}$ percentile of OS including two-sided 95% CI for each treatment arm will be provided, and their differences with 95% CIs between two arms will be calculated using Kosorok's method. KM curves will be provided by treatment arm in figures.

A stratified log-rank test will provide the pivotal p-value for the comparison of the two survival curves of AG-221 vs. CCR arm overall. Hazard ratio and its two-sided 95% CI will be estimated from a stratified Cox proportional hazards regression model. The stratification factors are:

- Prior intensive therapy for AML: yes versus no
- Primary refractory: yes versus no
- Prior allogeneic HSCT for AML: yes versus no

They are assessed by investigators and captured both in the IVRS system at the time of screening phase and in the CRF. In case the site made a stratification error at the time of screening/randomization, stratification errors are corrected in the CRF. Stratified analysis will use stratification factors taken from the CRF. If data does not converge in the stratified analysis due to small sample size in stratum or if there are less than 40 subjects in a stratum, the stratification factor leads to least subjects will be dropped from the model.

Reasons for censoring (e.g., lost to follow-up, withdrew consent, alive at data cut-off, Discontinuation due to other reasons) will be summarized (n, percent) for the OS endpoint.

An unstratified log-rank analysis and Cox regression will be considered as supporting analyses. Stratified analyses based on IVRS stratification factors will be included as a sensitivity analysis.

The analyses on the primary endpoint will be repeated for AG-221 versus BSC Only, Azacitidine + BSC, LDAC + BSC, IDAC + BSC individually by pre-selected CCR therapy.

In addition, sensitivity analyses will be done in the following ways:

1. Analysis of OS using mITT population;
2. Censoring subjects at the time of transplantation after study treatments;
3. Censoring subjects at the start date of any subsequent AML therapy;
4. Censoring subjects at the start date of any subsequent therapy for diseases other than AML excluding the use of hydroxyurea;
5. Censoring subjects for subsequent treatment of AG-221 after initial CCRs;
6. Inverse Probability of Censoring Weighted (IPCW) method censoring subjects at off-protocol AG-221;
7. IPCW method censoring subjects at any subsequent AML therapies if applicable;
8. Rank Preserving Structural Failure Time (RPSFT) model adjusting the confounding effect of off-protocol AG-221 if applicable;
9. Regression based imputation method adjusting subsequent AML therapies if applicable.

At the final analysis, the assumption of proportional hazards will be tested using a time-dependent Cox model with categorical variables for survival time (e.g., 0 – 3 months, 3 – 6 months, 6 – 9 months, etc.) and with following covariates:

- Treatment
- Prior intensive therapy for AML (yes, no) from CRF;
- Prior refractory (yes, no) from CRF;
- Prior stem cell transplants for AML (yes, no) from CRF;

If the model does not converge due to small sample size, the covariate leads to least subjects will be dropped from the model. If a stratification factor covariate has less than 40 subjects in one of its levels across treatment groups, then this factor will not be used in stratified analysis;
A statistically significant estimation of any of these survival time categories indicates that the proportional hazard assumption is violated. A figure of smoothed hazard function over time will give an instinct idea of hazard trend over time.

If proportional hazard assumption is violated, the following sensitivity analyses will be conducted using ITT population:

1. Restricted mean survival time will be calculated between AG-221 and CCR arms;
2. Piecewise cox regression will provide hazard estimates for every 3 months;
3. Piecewise cox regression will provide hazard estimates for every 6 months;
4. In the presence of non-proportional and early OS separation, a generalized Wilcoxon test will be utilized to compare survival curves.

A number of subgroups analyses will also be carried out. For details see Section 10.4.

Additionally, the primary endpoint will be repeated for a dose level group analysis of AG-221 subjects who had at least one dose > 100mg during the study versus those who received only dose ≤100mg.

## 10.3. Analyses of Secondary Efficacy Endpoints

The time-to-event secondary efficacy variables will be analyzed similarly to the primary efficacy variable with and without stratification when appropriate. KM methods will be used to estimate time-to-event curves, and hazard ratio and its two-sided 95% confidence interval (CI) will be estimated from Cox proportional hazards regression model, unless otherwise specified. Counts and percentages will be used to describe categorical secondary variables.

All secondary efficacy endpoints will be analyzed based on ITT population and key efficacy analysis will be performed on the mITT population as supportive evidence and to assess the robustness of the efficacy findings. Subjects will be analyzed according to randomized treatment group.

Throughout, summaries of Complete Remission will include both CR and CRc responses, and will be displayed as CR. CRc will be summarized separately once, and otherwise will be displayed in combined categories as CR.

### 10.3.1. Key Secondary Endpoints

The key secondary efficacy endpoints include ORR and EFS assessed by an Independent Response Assessment Committee (IRAC) and by local investigators based on modified International Working Group (IWG) AML response criteria. Details of the procedure used to proceed IRAC assessed IWG and CR/CRh responses, and the IRAC assessed HI response can be found in the separate IRAC Charter.

#### 10.3.1.1. Overall Response Rate

Overall response rate (ORR) is defined as the rate of Best Response in CR + CRi + CRp + partial remission (PR) + morphologic leukemia-free state (MLFS) according to modified IWG AML response criteria.

The ORR will be summarized by treatment group and be compared between arms by the Cochran-Mantel-Haenszel (CMH) test with stratification factors:

- Prior intensive therapy for AML: yes versus no
- Primary refractory: yes versus no
- Prior allogeneic HSCT for AML: yes versus no

If a stratification factor covariate has less than 40 subjects in one of its levels across treatment groups, then this factor will not be used in stratified analysis. This analysis based on IRAC assessment in the ITT population will be used as the main analysis of ORR.

As sensitivity analyses, a CMH test with the above strata will be performed in the mITT population. In addition, a Fisher's exact test without stratification factors will be performed in both ITT and mITT population. An Odds Ratio and its 95% C.I. of responders versus non-responders between treatment arms will also be calculated. This analysis will be repeated for the local assessment of response as a supportive analysis of ORR.

Additionally, overall response rate will be repeated for a Dose group analysis of AG-221 subjects who had at least one dose > 100mg during the study versus those who received only dose ≤100mg.

#### 10.3.1.2. Event-Free Survival

Event-free survival (EFS) is defined as the period from the date of randomization to the date of documented morphologic relapse (hereafter referred as relapse) after CR/CRi/CRp, PD according to modified IWG AML response criteria or death from any cause, whichever occurs first. Subjects who are alive without any of these events will be censored at the date of their last valid response assessment. Subjects who do not have post-randomization response assessment will be censored at date of randomization.

If a subject begins a new anti-cancer therapy prior to documented disease progression (or death), the subject will be censored at the date of last assessment before starting the new anti-cancer therapy. Subjects with two or more consecutive missing response assessments prior to documented progression (or death) will be censored at the last date of response assessment before the missing assessments.

Analysis of EFS will be similar to the analysis of OS. The stratified log-rank test based on the ITT population will be considered the primary analysis for EFS, with unstratified analyses and the analyses based on mITT populations considered supportive analyses. The analysis based on local assessment of response will be conducted as a supportive analysis of EFS.

Additionally, EFS will be repeated for a Dose group analysis of AG-221 subjects who had at least one dose > 100mg during the study versus those who received only dose ≤100mg.

**Table 4: Censoring Rules for EFS**

| Value of Event-free Survival Date (ADT) | Censored (Y,N) | Derivation |
|---|---|---|
| **ADT = min(death date, relapse date, progression date)** | **N** | **If a subject died, relapsed, or had disease progression by independent review, and the subject has less than two consecutive missing response assessments prior to a visit documented death, relapse, or progression which is before any subsequent anti-cancer therapies** |
| **ADT = the last event-free assessment date before missing assessments/ randomization date** | **Y** | **If a subject died, relapsed, or had disease progression by independent review, and the subject has two or more consecutive missing response assessments prior to a visit documented death, relapse, or progression** |
| **ADT = the last assessment date/randomization date** | **Y** | **If a subject did not die, or did not have relapse or progression, and did not receive any subsequent anti-cancer therapies.**<br>**If there was no post-randomization response assessment, then ADT = the randomization date.** |
| **ADT = the last assessment date on or prior to start of subsequent anti-cancer therapies/randomization date** | **Y** | **If a subject did not die, or did not have relapse or progression on or prior to start of subsequent anti-cancer therapies.**<br>**If there was no post-randomization response assessment on or prior to subsequent therapies, then ADT = the randomization date.** |
| **ADT = the last assessment date on or prior to start of subsequent anti-cancer therapies/randomization date** | **Y** | **If a subject died, relapsed, or had disease progression by independent review after receiving subsequent anti-cancer therapies.**<br>**If there was no post-randomization response assessment on or prior to subsequent therapies, then ADT = the randomization date.** |
| **ADT = analysis date; N = no; Y = yes.**<br>**Note: Event-free response refers to a response that was neither morphologic relapse, progressive disease (PD) nor death.** | | |

Sensitivity analyses for EFS will be done with the following censoring rules:

- Subjects with two or more consecutive missing response assessments prior to an event date (death, relapse or disease progression) will not be censored at the last date of last response assessment before the missing assessment;
- Subjects with subsequent anti-cancer therapies or with two or more consecutive missing assessments prior to an event date will not be censored;
- Subjects with subsequent transplantation after treatment will not be censored;
- Subjects with off-protocol AG221 prior to an event will not be censored;
- Subjects with any anti-cancer therapies (including AML therapy, any therapy for diseases other than AML excluding hydroxyurea) prior to an event date will not be censored.

Other sensitivity rules may be added when data is applicable.

### 10.3.2. Other Secondary Endpoints

The other secondary endpoints include duration of response, time to response, rates of overall remission, CR, complete remission rate and HI rates assessed by an IRAC and by local investigators according to modified IWG AML response criteria and IWG myelodysplastic syndromes (MDS) HI criteria, treatment mortality at 30- and 60-day, one-year survival, rate of HSCT and time to treatment failure. CR/CRh rate, duration of CR/CRh, and time to CR/CRh assessed by IRAC only and using a sponsor derived assessment will also be presented.

#### 10.3.2.1. Duration of Response

Duration of response is defined as time from first documented CR/CRi/CRp/PR/MLFS to documented morphologic relapse, PD, or death due to any cause, whichever occurs first. Subjects without morphologic relapse, PD, or death due to any cause will be censored at the date of the last valid response assessment.

The duration of response will be analyzed using the KM method. The 25$^{th}$ percentile, median and 75$^{th}$ percentile time (including two-sided 95% CI) will be summarized for each treatment arm. Only subjects with documented CR/CRi/CRp/PR/MLFS response as determined by the IRAC will be included in the analysis. Investigator assessed results will be summarized as a supportive analysis.

#### 10.3.2.2. Time to Response

Time to response is defined as time from randomization to first documented CR/CRi/CRp/PR/MLFS.

Time to response will be summarized with descriptive statistics in subjects with documented CR/CRi/CRp/PR/MLFS response as determined by the IRAC. Number of subjects who attain first response in a given cycle will also be summarized. Investigator assessed results will be summarized as a supportive analysis.

#### 10.3.2.3. Time to Best Response

Time to best response is defined as time from randomization to the documented best response of CR/CRi/CRp/PR/MLFS.

Time to best response will be summarized with descriptive statistics in subjects with documented CR/CRi/CRp/PR/MLFS response as determined by the IRAC. Number of subjects who attain best response in a given cycle will also be summarized. Investigator assessed results will be summarized as a supportive analysis.

#### 10.3.2.4. Treatment Mortality at 30- and 60-Day

Treatment mortality at 30- and 60-day is defined as rate of death from any cause within 30 or 60 days of initiation of study treatment.

Number and percentage of deaths at 30- and 60-day will be summarized for each treatment arm. Treatment mortality at 30- and 60-Day will also be evaluated by KM estimate for both AG-221 and CCR treatment groups.

#### 10.3.2.5. One-Year Survival

One-year survival is defined as the probability of survival at 1 year from randomization.

One-year survival will be evaluated by KM estimate with 95%CI for each treatment arm. Difference between the two treatment groups and corresponding two-sided 95% CI will also be displayed. The CI for the difference in the 1-year survival probabilities will be derived using Greenwood's variance estimate.

#### 10.3.2.6. Overall Remission Rate

Overall remission rate is defined as rate of CR + CRi + CRp according to modified IWG AML response criteria per IRAC.

Number and percentage of subjects with overall remission will be summarized by treatment groups. Comparisons between AG-221 and CCR will be analyzed by Fisher’s exact test. Investigator assessed results will be summarized as a supportive analysis

#### 10.3.2.7. Complete Remission Rate

Complete remission rate is defined as rate of CR according to modified IWG AML response criteria per IRAC.

Number and percentage of subjects with complete remission will be summarized by treatment groups. Comparisons between AG-221 and CCR will be analyzed by Fisher’s exact test. Investigator assessed results will be summarized as a supportive analysis

#### 10.3.2.8. Hematologic Improvement Rate

Hematologic improvement rate is defined as rate of neutrophil response (hematologic improvement neutrophil response [HI-N]) + platelet response (hematologic improvement platelet response [HI-P]) + erythroid response (hematologic improvement erythroid response [HI-E]) according to IWG MDS HI criteria (Section 15.6).

Number and percentage of subjects with hematologic improvement will be summarized by treatment groups. Comparisons between AG-221 and CCR will be analyzed by Fisher's exact test.

Subjects who achieved post-baseline 56-day RBC transfusion independence, i.e., without RBC transfusion for at least 56 consecutive days post baseline during treatment exposure period, will be summarized by baseline RBC transfusion dependence status (Yes vs. No), within each category of best response (e.g., CR, CRi/CRp, PR/MLFS, SD, PD, etc.) and overall. Subjects with at least one transfusion during the baseline period (defined as within 56 days before the first dose date) are considered transfusion dependent at baseline. Time to transfusion independence will be summarized for the subject's first transfusion independence, and duration of the independence will be summarized using the transfusion independence with the longest duration experienced by a subject.

The post-baseline 56-day Platelet transfusion independence will be analyzed similarly.

#### 10.3.2.9. Rate of HSCT

Rate of HSCT is defined as rate of bridge-to-HSCT through study treatment. This analysis is for the subjects who proceeded to a subsequent HSCT immediately after discontinuation of the study drugs. Subjects who received systemic anti-cancer page for AML after discontinuation of the study drugs and later underwent a subsequent HSCT are not included.

Number and percentage of subjects with HSCT will be summarized by treatment groups. Comparisons between AG-221 and CCR will be analyzed by Fisher's exact test. In addition, Numbers and percentages of subjects who with HSCT after reaching CR/CRi/CRp or without CR/CRi/CRp will be summarized.

#### 10.3.2.10. Time to Treatment Failure

Time to treatment failure is defined as time from randomization to discontinuation of study treatment due to any cause. For BSC subjects, the duration is from randomization to the date of treatment discontinuation or the last visit date if still ongoing. For subjects received treatment other than BSC, the duration is from randomization to the last dose date or day before start of non-BSC treatment whichever occurs first; subjects who have completed the study treatment or are still active will be censored at last dose date.

The time to treatment failure will be analyzed using the KM method. The $25^{th}$ percentile, median and $75^{th}$ percentile time (including two-sided 95% CI) will be summarized for each treatment arm; the associated HR with two-sided 95% CI will be estimated using Cox proportional hazard models with and without stratification factors. A stratified and an unstratified log-rank tests will provide the pivotal p-value for the comparison between AG-221 and CCR arm.

#### 10.3.2.11. CR/CRh Rate

CR/CRh rate is the rate of CR or CRh where CRh is defined as partial hematologic recovery that achieves all criteria of CR except ANC and platelet criteria. For CRh, those criteria are:

- ANC > 0.5 x $10^9$/L; and

- Platelet > 50 x $10^9$/L

Number and percentage of subjects with CR/CRh will be summarized by treatment groups. Comparisons between AG-221 and CCR will be analyzed by Fisher's exact test.

CR/CRh assessed by IRAC only and using a sponsor derived assessment.

#### 10.3.2.12. Duration of CR/CRh

Duration of CR/CRh is defined as time from first documented CR or CRh to documented morphologic relapse, PD, or death due to any cause, whichever occurs first. Subjects without morphologic relapse, PD, or death due to any cause will be censored at the date of the last response assessment.

The duration of CR/CRh will be analyzed using the KM method. The 25$^{th}$ percentile, median and 75$^{th}$ percentile time (including two-sided 95% CI) will be summarized for each treatment arm. Only subjects with CR/CRh response will be included and will be summarized for both IRAC and sponsor assessed responses.

#### 10.3.2.13. Time to CR/CRh

Time to CR/CRh is defined as time from randomization to first documented CR/CRh.

Time to CR/CRh will be summarized with descriptive statistics in subjects with documented CR/CRh as determined by IRAC and sponsor assessment. It will also be summarized by cycle.

## 10.4. Subgroup Analysis

Appropriate subgroup analyses for primary and key secondary efficacy endpoints (ORR and EFS) by the following subgroups:

1. Age groups (≥60 years and <70 years, ≥70 years and <80 years, ≥80 years);
2. Gender (male and Female);
3. Region (North America, South America, Europe, Asia, Australia; US, non-US);
4. Race;
5. Prior systemic anti-cancer therapies for AML (2$^{nd}$ versus 3$^{rd}$ lines);
6. Baseline cytogenetic risk status (better-risk, intermediate-risk, poor-risk, failure);
7. IDH2 gene mutate type (R140 vs. R172);
8. AML initial diagnosis (primary or secondary);
9. WHO classification of AML;
10. ECOG performance status at baseline (Grade 0, 1, 2, 3);
11. Prior intensive therapy for AML (yes versus no) (IVRS and CRF source);
12. Primary refractory (yes versus no) (IVRS and CRF source);
13. Prior allogeneic HSCT for AML (yes versus no) (IVRS and CRF source);
14. ELN risk classification

15. Non-Protocol Definition Primary Refractory, defined as having at least 2 lines prior AML or Higher MDS treatment and never attained CR, Cri, or CRP ( yes versus no).

A forest plot will be provided based on the HRs for each subgroup.

# 11. SAFETY ANALYSIS

All analyses of safety data will be conducted using the safety population unless otherwise specified. Safety summaries based on the safety population will be provided by treatment and pre-specified CCR option (Overall, BSC only, Azacitidine + BSC, LDAC + BSC, IDAC + BSC). Dosing summaries will be carried out for each study drug separately.

For summaries of safety data, the treatment period is defined as follows:
• AG-221, azacitidine, LDAC and IDAC AZA: date of first dose through date of last dose + 28 days, or until the EOT visit, whichever is longer;
• BSC only: date of randomization through date of treatment discontinuation.

## 11.1. Adverse Events

Adverse events will be analyzed in terms of treatment-emergent adverse events (TEAEs) which are defined as any AEs that begin or worsen on or after the start of study drug through 28 days after the last study treatment or until the EOT visit, whichever is longer, as well as those serious adverse events (SAEs) made known to the Investigator at any time thereafter that are suspected of being related to investigational product. All AEs will be coded using the Medical Dictionary for Regulatory Affairs® (MedDRA) version 22.0. The severity will be graded by the study personnel based on NCI National Cancer Institute Common Terminology Criteria for Adverse Events (CTCAE) Version 4.03.

A treatment-related TEAE is defined as TEAE that is suspected by the Investigator to be related to study treatment.

If a subject experiences the same AE more than once with different toxicity grade, then the event with the highest grade will be tabulated in "by grade" tables. If a subject experiences multiple AEs under the same preferred term (system organ class), then the subject will be counted only once for that preferred term (system organ class).

Summarizations of TEAEs per person-year of exposure are carried out in exposure-adjusted event rate, which is defined as number of events occurred in treatment exposure duration.

The number of subjects with at least one TEAE will be summarized. The incidence of TEAEs will be summarized by alphabetic order of SOC and descending order of PT for AG-221 group. Tables summarizing the incidence of TEAEs will be generated for each of the following:

- Summary of TEAEs;
- All TEAEs by SOC and PT;
- All TEAEs by decreasing frequency of PT for AG-221;
- All TEAEs per person-year of exposure, person-years of exposure is defined as (Sum of subjects' overall treatment exposure (days) from first dose date (or randomization date for BSC only group) to the end of the TEAE period [c])/365.25;
- TEAEs related to study treatment;
- TEAEs by maximum severity;

- TEAEs with CTCAE grade ≥ 3;
- Related TEAEs with CTCAE grade ≥ 3;
- TEAEs with CTCAE grade ≥ 3 per person-years exposure;
- Related TEAEs with CTCAE grade ≥ 3 per person-years exposure;
- All AEs resulting in death;
- Serious TEAEs;
- Serious TEAEs per person-years exposure;
- All serious TEAEs by decreasing frequency of PT for AG-221;
- Serious TEAEs related to study treatment;
- Serious TEAEs related to study treatment per person-years exposure;
- TEAEs leading to study drug withdrawal;
- TEAEs leading to study drug withdrawal per person-years exposure;
- TEAEs related to study drug and leading to study drug withdrawn;
- TEAEs with leading to dose reduction;
- TEAEs with leading to dose reduction per person-years exposure;
- TEAEs related to study drug and leading to study drug dose reduction;
- TEAEs leading to study drug dose interruption;
- TEAEs leading to study drug dose interruption per person-years exposure;
- TEAEs related to study drug and leading to study drug dose interruption;
- TEAEs leading to study drug dose reduction and dose interruption;
- TEAEs related to study drug and leading to dose reduction and study drug dose interruption;
- TEAEs for the following baseline subgroups (provided when the number of subjects is sufficient):
  - Gender (male and female);
  - Age (≥60 years and <70 years, ≥70 years and <80 years, ≥80 years);
  - Region (North America, South America, Europe, Australia, Asia; US vs. Non-US);
  - Baseline ECOG performance status (Grade 0, 1, 2);
  - Prior intensive therapy for AML (yes versus no) (IVRS and CRF source);
  - Primary refractory: yes versus no
  - Prior allogeneic HSCT for AML (yes versus no) (IVRS and CRF source);

- Baseline ANC value ($\leq 0.5x10^9$/L, $0.5<1x10^9$/L, $\geq 1x10^9$/L)
- Baseline creatinine clearance value (<45, 45-<60, ≥60 mL/min);
- Best clinical response (CR, PR, SD, PD) adjusted by person years on treatment.

In addition, the impact of dose escalation to AEs will be explored. Sensitivity analyses will be carried out for the following two tables:

- All TEAEs by AG-221 ≤100 mg and AG-221 > 100 mg cohorts:

  AEs which occurred in subjects which only received doses of AG-221 ≤100mg, and AEs which occurred prior to a dose switching to > 100 mg for subjects which had an escalated dose, or AEs which occurred following a reduction to <=100mg following an escalation are to be counted in ≤100 mg cohort. Only AEs which occurred after a dose > 100 mg was administered and occurred within 28 days following this dose administration, and where there were no subsequent dose reductions between escalated dose date and AE start date will be counted in the > 100 mg cohort.

- TEAEs with CTCAE grade ≥3 by AG-221≤ 100 mg and AG-221 > 100 mg cohorts.

Listings for the corresponding summary tables will be presented separately. Non-treatment-emergent AEs will also be included in one of the listings.

## 11.2. Adverse Events of Special Interest

The following TEAEs of special interest will be summarized:

1. QT prolongation;

   Standardized MedDRA Query (SMQ) Torsade de pointes/QT prolongation.

2. IDH inhibitor associated Differentiation syndrome (IDH-DS);
   a. TEAE AE differentiation syndrome as reported by the investigators.
   b. Time to IDH-DS: Summary statistics including mean, standard deviation, median, minimum, maximum will be provided.
      i. From start of treatment until an event of IDH-DS
      ii. From onset until end date of an event of IDH-DS
   c. Signs and symptoms of PT IDH-DS (% of subject with specific symptom with denominator of N of subjects with completed CRF page for the differentiation syndrome).
      - Symptoms of the IDH-DS (from the CRF AE page)
      - Treatment for IDH-DS: systemic steroids
      - Mean duration of steroid treatment for the IDH-DS
      - Study drug interruption, discontinuation, dose adjustment due to IDH-DS
   d. Concurrency of the IDH-DS and tumor lysis syndrome PT.
3. Bilirubinemia and liver safety;
      a. SMQ Biliary system related investigations, signs and symptoms (narrow terms)
      b. SMQ Drug related hepatic disorders (broad terms).
      c. Time to onset and time to resolution
4. Hepatic Disorders;
5. Acute Renal Failure;
6. Pulmonary Events;
7. Peripheral Neuropathy
8. Hepatobiliary Disorders;

The following summaries will be provided for TEAEs of special interest:

- All TEAEs by PT;
- Treatment related TEAEs by PT;
- Serious TEAEs by PT;
- Serious treatment related TEAEs by PT;
- TEAEs with CTCAE grade of Grade ≥ 3;

- Treatment related TEAEs with CTCAE grade of Grade ≥ 3;
- Time to first onset of TEAE;

## 11.3. Death

Cause of death will be collected in the CRF and will be summarized for all treated subjects with the following categories during on-treatment period and post-treatment period:

- Death from malignant disease under study, or complication due to malignant disease under study
- Death from adverse event (not otherwise specified)
- Death from other cause
- Death from unknown cause (not assessable or insufficient data)

A death listing will include deaths for all screened subjects during pre-treatment, on-treatment, and post-treatment period.

## 11.4. Clinical Laboratory Evaluations

Clinical laboratory results will be summarized descriptively by treatment group, which will also include a display of change from baseline. Laboratory values outside of the normal ranges will be identified. Normal ranges will be used to determine the categories of High, Low, and Normal for laboratory tests that have no severity grade.

Where multiple laboratory records exist for the same scheduled visit, the record closest to the scheduled clinical visit date will be summarized. If a record exists on the same day of a scheduled visit date, this record will be selected. If no such record exists then the closest record will be determined by first checking for a record with date one day before clinical record, then for a record one day after clinical record. If neither of these exist then check two days before clinical record, and then two days after clinical record. Follow a similar logic until a closest record can be found. If more than one record is selected using above process, then the record with sample type not equal to "U" will used. If a unique record still cannot be determined then an average of records fitting above criteria will be taken, and this will be used in summaries.

Clinically significant hematologic and non-hematologic laboratory abnormalities that meet Grade 3 or Grade 4 criteria according to the NCI CTCAE 4.03 will be summarized. Where applicable, shift to high values will be presented separately from shift to low value (e.g. leukocytosis and leukopenia). The worst grade during the treatment period will be summarized by treatment group and by cycle. Frequency distributions of CTC grade that shift from baseline to the worst grade during treatment period will be presented by treatment arm when data is applicable. Graphical display of select laboratory parameters over the course of the study will be provided where useful to assist in the interpretation of results.

Listings of all clinical laboratory data from central laboratory with abnormal flags will be provided by subjects and tests. Listings will also be provided for the local laboratory data.

### 11.4.1. Hematology

For specific haematological parameters (Hemoglobin, platelet counts, ANC, WBC, and RBC), the change from baseline to nadir value will be summarized by treatment group.

Shifts in CTCAE grades from baseline to worst grade will be summarized by treatment and by cycle.

### 11.4.2. Serum Chemistry

For serum chemistry parameters, the change from baseline will be summarized by treatment group and study visit. Change in total, direct and indirect bilirubin by UGT1A1 mutation will be analyzed.

Shifts in CTC grades from baseline to worst grade will be summarized by treatment and by cycle.

The proportion of subjects with clinically significant post-baseline change in hepatic, renal function and electrolytes will be assessed based on the criteria presented below.

**Table 5 Serum Chemistry Changes of Interest**

| Laboratory Test | Parameters for any value and the last value |
|---|---|
| Alanine Aminotransferase (ALT) | CTCAE Grades ≥3 |
| Aspartate Aminotransferase (AST) | CTCAE Grades ≥3 |
| Total Bilirubin | CTCAE Grades ≥3 |
| Total Bilirubin ≥ Grade 2 and ALT ≥ Grade 2 | Composite of Total Bilirubin ≥ Grade 2 and ALT ≥ Grade 2 within 1 cycle of Total Bilirubin elevation |
| Total Bilirubin ≥ Grade 2 and AST ≥ Grade 2 | Composite of Total Bilirubin ≥ Grade 2 and AST ≥ Grade 2 within 1 cycle of Total Bilirubin elevation |
| Potassium | Composite:<br>Hypokalemia Grade ≥ 2 and >15% decrease from baseline<br>Hyperkalemia Grade ≥ 3 and >15% increase from baseline |
| Phosphorus | Composite: Hyperphosphatemia > ULN and >25% increase from baseline |
| Calcium | Composite: Hypocalcemia Grade ≥ 3 and > 15 % decrease from baseline |
| Magnesium | Composite: Hypomagnesemia Grade ≥ 2 and >25% decrease from baseline |
| Uric Acid | Composite: Hyperuricemia >ULN and >25% increase from baseline |

### 11.4.3. Urinalysis

Urinalysis includes the following tests: specific gravity, glucose, ketones, blood, pH, and protein, and microscopic analysis. Since urinalysis will only be performed when clinically indicated after screening, the results will only be listed in a listing.

### 11.4.4. Coagulation

Coagulation includes the following tests: prothrombin time with international normalized ratio and partial thromboplastin time. Since coagulation will only be performed when clinically indicated after screening, the results will only be listed in a listing.

### 11.4.5. Fasting Lipid Panel

Fasting lipid panel includes total cholesterol, low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), and triglycerides. Lipid profiles will be described by concentration of the lipids at scheduled visits. The proportion of subjects with clinically meaningful post-baseline increase in lipids will be summarized based on criteria presented in Table 6 below.

**Table 6: Fasting Lipid Changes of Interest**

| Laboratory Test | Parameters for any value and the last value |
|---|---|
| LDL cholesterol | >ULN and >10% increase from BL |
| LDL-C/HDL-C ratio | >3.5 and >10% increase from BL |

### 11.4.6. Cardiac Markers

Cardiac markers include creatine kinase and cardiac troponins I. Incidence of the test values above upper limit of normal will be summarized at scheduled visits.

### 11.4.7. Females of Childbearing Potential

Pregnancy incidences will be reported as adverse events and listed out at subject level.

## 11.5. Vital Sign Measurements

Evaluation of vital signs includes recording height (only at screening), weight, temperature, systolic blood pressure (SBP), diastolic blood pressure (DBP), pulse rate, and respiratory rate at scheduled visits.

Vital sign measurements will be listed for each subject and by scheduled visit. Descriptive statistics for vital signs (except height), both observed values and changes from baseline, will be summarized by treatment group.

## 11.6. Physical Examination

Abnormal physical examination findings will be captured as medical history (screening) and will be summarized as medical history or be captured as adverse events and will be summarized as adverse events.

## 11.7. Electrocardiograms

Electrocardiogram (ECG) parameters include heart rate, PR interval, QRS duration, RR interval, QT, QTc Fridericia value (QTcF) and QTc Bazett value (QTcB). QTcB is not recorded in the CRF page and will be calculated in the following way: QTcB=QT/sqrt (RR).

Recorded values and of ECG parameters and change from baseline values will be summarized at each time point by treatment group.

In addition, at baseline and maximum post-baseline visits, the proportion of subjects having absolute QTcF and QTcB intervals of the following categories will be presented:

- $\leq 450$ ms
- $> 450$ to $\leq 500$ ms
- $> 500$ ms

At maximum post-baseline visits, the proportion of subjects who have an increase from baseline in QTcF and QTcB intervals of the following categories will be presented:

- $\leq 30$ ms
- $> 30$ to $\leq 60$ ms
- $> 60$ ms

The overall ECG interpretation will be summarized by presenting the shift from baseline to worst by treatment group and by cycle. The overall ECG interpretation will be displayed in cross–tabulations for each treatment.

# 12. INTERIM ANALYSIS

## 12.1. General Information

The interim analysis for superiority was performed when approximately 65% of the expected numbers of deaths (i.e., approximately 163 deaths) have been reported. The interim efficacy analysis was performed by an unblinded study team and the results will be kept strictly confidential in order to maintain the integrity of the study. The interim analysis results were provided to the DMC to review and give advice to the Sponsor regarding the study conduct. The analysis results were not disseminated among investigators and those directly involved with the study conduct. The analysis was shared with regulators as needed to support regulatory submissions.

**Data Monitoring Committee (DMC)**

An external DMC was convened. The DMC included a group of physicians with experience in treating subjects with AML and a statistician, all of whom are not otherwise involved in the study conduct. The DMC members reviewed the efficacy and safety data from the interim analysis and recommended whether the trial should continue according to the protocol. During the course of the study, the DMC members will also periodically review the safety data to ensure the safety of the study subjects. The detailed roles and responsibilities of DMC can be found in the DMC charter.

**Independent statistician who conducted the unblinded analyses**

[redacted] an independent CRO, was used to conduct unblinded statistical analyses. [redacted] unblinded statisticians/programmers received the unblinded randomization codes directly from the IVRS system, conducted the statistical analyses, and provided the unblinded summary results to the DMC members. Only DMC members and [redacted] statisticians/programmers who performed the unblinded analyses could access to the unblinded summary results.

## 12.2. Statistical Approaches for Control of Alpha

The type I error rate was controlled at the overall two-sided 0.05 level by using an O'Brien-Fleming monitoring boundary with a Lan-DeMets alpha spending function. The decision rules were as follows:

- 65% of expected events

Stop for superiority if one-sided p-value ≤0.0055 (HR ≤0.6715)

- 100% of expected events

Claim success if one-sided p-value ≤0.0233 (HR ≤0.7775)

The actual levels used depended on the actual number of events observed at the interims and the final analyses.

At the time of the interim analysis, the death rate was evaluated relative to the study assumptions, and enrollment into the study may be increased to ensure that the required number of 250 deaths can be reached in a reasonable time period.

# 13. CHANGES TO THE STATISTICAL ANALYSES SECTION OF THE PROTOCOL

[redacted]

Additional analysis of the Transfusion Independence of subjects will be conducted, which is not mentioned in the Protocol. For details see Section 10.3.2.8 of this SAP.

# 14. REFERENCES

David L. DeMets, K. K. Gordon Lan. Lan-DeMets Alpha-Spending Function. John Wiley & Sons, Inc. 2008

Cheson BD, Bennett JM, Kopecky KJ, Buchner T, Willman CL, Estey EH, et al. Revised recommendations of the International Working Group for diagnosis, standardization of response criteria, treatment outcomes, and reporting standards for therapeutic trials in Acute Myeloid Leukemia. J Clin Oncol. 2003 Dec; 21(24):4642-4649.

Cheson BD, Greenberg PL, Bennett JM, Lowenberg B, Wijermans PW, Nimer SD, et al. Clinical application and proposal for modification of the International Working Group (IWG) response criteria in myelodysplasia. Blood 2006; 108 (2):419-25.

Kosorok, MR. Two-sample quantile tests under general conditions. Biometrika 1999;86:909-921.

Robins JM, Finkelstein DM. Correcting for noncompliance and dependent censoring in a AIDS clinical trial with inverse probability of censoring weighted (IPCW) log rank tests. Biometrics 2000; 56:779-788.

Robins JM, Tsiatis AA. Correcting for non-compliance in randomized trials using rank preserving structural failure time models Commun. Statist.-Therory Meth. 1991; 20 (8): 2609-2631.

Luo X, Li M, Wu C, Xu Q, Chen G, Dornseif BE, Renschler MF, Koch G. A proposed approach for analyzing post-study therapy effect in survival analysis. J Biopharmaceutical Statistics 2016; 26 (4): 790-800

# 15. APPENDICES

## 15.1. Handling of Dates

Dates will be stored as numeric variables in the SAS analysis files and reported in DDMMMYYYY format (i.e., the Date9. datetime format in SAS). Dates in the clinical database are classified into the categories of procedure dates, log dates, milestone dates, outcome dates, and special dates.

- **Procedure Dates** are the dates on which given protocol-specified procedure are performed. They include the dates of laboratory testing, physical examinations, etc. They should be present whenever data for a protocol-specified procedure are present and should only be missing when a procedure are marked as NOT DONE in the database. Procedure dates will not be imputed.
- **Log Dates** are dates recorded in CRF data logs. Specifically, they are the start and end dates for AE and concomitant medications/procedures. They should not be missing unless an event or medication is marked as *ongoing* in the database. Otherwise, incomplete log dates will be imputed according to the rules in Section 15.2 (e.g., for duration or cycle assignment, etc.). However, in listings, log dates will be shown as recorded without imputation.
- **Milestone Dates** are dates of protocol milestones such as randomization, study drug start date, study drug termination date, study closure date, etc. They should not be missing if the milestone occurs for a subject. They will not be imputed.
- **Outcome Dates** are dates corresponding to study endpoints such as survival, progression, etc. In most cases they are derived either from a milestone (e.g., the survival date is derived from the death date), or a procedure date (e.g., the progression date is derived from the date of the hematology lab that was used to determine progression). They may be subject to endpoint-specific censoring rules if the outcome did not occur but are not otherwise subject to imputation.
- **Special Dates** cannot be classified in any of the above categories and they include the date of birth. They may be subject to variable-specific censoring and imputation rules.

Dates recorded in comment fields will not be imputed or reported in any specific format.

### 15.1.1. Calculation Using Dates

Calculations using dates (e.g., subject's age or relative day after the first dose of study drug) will adhere to the following conventions:

- Study days after the start day of study drug will be calculated as the difference between the date of interest and the first date of dosing of study drug (e.g., lenalidomide) plus 1 day. The generalized calculation algorithm for relative day is the following:

- If TARGET DATE ≥ DSTART then STUDY DAY = (TARGET DATE – DSTART) + 1;
- Else use STUDY DAY = TARGET DATE – DSTART.

Note that Study Day 1 is the first day of treatment of study drug. Negative study days are reflective of observations obtained during the baseline/screening period. Note: Partial dates for the first study drug are not imputed in general. All effort should be made to avoid incomplete study drug start dates. Some analyses may exclude subjects without study drug start date. The SAP may be amended on a case-by-case basis, including guideline and suggestion such as imputing Day 1 by consent date of treated subject.

- Age (expressed in days) is calculated: AGE = CONSENT – DATE of BIRTH + 1. In practice, age will be transformed to years by dividing the difference by 365.25 days, then truncating.
  - Preference is for using calculated age from clinical database. When not available, calculated age from CRF or IVRS may be used
  - Partial birth date: impute missing day as 15$^{th}$ of the month; impute missing month as July; set missing age for missing year
- Intervals that are presented in weeks will be transformed from days to weeks by using (without truncation) the following conversion formula:

  WEEKS = DAYS /7
- Intervals that are presented in months will be transformed from days to months by using (without truncation) the following conversion formula:

  MONTHS = DAYS /30.4375

### 15.1.2. Calculation of Cycles

In this study, subjects meeting inclusion/exclusion criteria will be randomized 1:1 via IVRS into 1 of 2 treatment arms to receive either AG-221 or CCR (one of the four options). Dose modification is permitted according to the protocol. The start date of each cycle will be calculated as follows:

The start date of a new cycle is the scheduled interval start date since there is no prescribed treatment break. One cycle is considered as 28 days. All qualified cycles will be sorted as the calculated cycle. An algorithm will be implemented to ensure that there are no gaps between Cycle X end date and Cycle X+1 start date. Details of the applied algorithm can be found in the Analysis Dataset specifications for the study.

## 15.2. Date Imputation Guideline

### 15.2.1. Impute Missing Start/Stop Dates for Adverse Events/ Prior or Concomitant Medications

Partially missing AE start dates will be imputed in the Analysis Data Model (ADaM) dataset for AEs, but partially missing AE end dates will not be imputed in the same dataset. If the AE end date is complete with no missing year, month, or day, and the partially missing start date imputed by the rules below is after the AE end date, then the start date will be imputed by the AE end date.

Partially missing start/stop dates for prior/concomitant medications and partially missing start dates for prior/concomitant procedures will be imputed in the ADaM dataset for prior/concomitant medications/procedures. For prior/concomitant medications, if the stop date is complete with no missing year, month, or day, and the partially missing start date imputed by the rules below is after the stop date, then the start date will be imputed by the stop date.

**Incomplete Start Date:** If the stop date is not missing, and the imputed start date is after the stop date, the start date will be imputed by the stop date.

*Missing day and month*

- If the year is the **same** as the year of the first dosing date, then the day and month of the first dosing date will be assigned to the missing fields.
- If the year is **prior to** the year of first dosing date, then December 31 will be assigned to the missing fields.
- If the year is **after** the year of first dosing, then January 1 will be assigned to the missing fields.

*Missing day only*

- If the month and year are the **same** as the year and month of first dosing date, then the first doing date will be assigned to the missing day.
- If either the year of the partial date is **before** the year of the first dosing date or the years of the partial date and the first dosing date are the same but the month of partial date is **before** the month of the first dosing date, then the last day of the month will be assigned to the missing day.
- If either the year of the partial date is **after** the year of the first dosing date or the years of the partial date and the first dose date are the same but the month of partial date is **after** the month of the first dosing date, then the first day of the month will be assigned to the missing day.
- If the stop date is not missing, and the imputed start date is after the stop date, the start date will be imputed by the stop date.

*Missing day, month, and year*

- No imputation is needed, the corresponding AE will be included as TEAE if end date of AE is after the first dose date or the end date is also missing.

**Incomplete Stop Date:** If the imputed stop date is before the start date, then the imputed stop date will be equal to the start date.

*Missing day and month*

- If the year of the incomplete stop date is the **same** as the year of the last dosing date, then the day and month of the last dosing date will be assigned to the missing fields.
- If the year of the incomplete stop date is **prior to** the year of the last dosing date or prior to the year of the first dosing date, then December 31 will be assigned to the missing fields.
- If the year of the incomplete stop date is **prior to** the year of the last dosing date but is the same as the year of the first dosing date, then the first dosing date will be assigned to the missing date.
- If the year of the incomplete stop date is **after** the year of the last dosing date, then January 1 will be assigned to the missing fields.

*Missing day only*

- If the month and year of the incomplete stop date are the **same** as the month and year of the last dosing date, then the day of the last dosing date will be assigned to the missing day.
- If either the year of the partial date is **not equal to** the year of the last dosing date or the years of the partial date and the last dosing date are the same but the month of partial date is **not equal to** the month of the last dosing date, then the last day of the month will be assigned to the missing day.

### 15.2.2. Medical History

Partially missing medical history start dates will be imputed in the ADaM dataset for medical history. The 16$^{th}$ of the month will be used to impute a partially missing start date that has only the day missing, and July 1$^{st}$ will be used to impute a partially missing start date that has both the month and day missing.

## 15.3. Independent Response Assessment Committee

Response to treatment is assessed locally by the Investigators and centrally by an IRAC according to modified IWG AML response criteria through collection of hematology laboratory parameters, peripheral blood smear, bone marrow aspirates and/or biopsies, and cytogenetics.

## 15.4. Table of Events

| Events | Screening Phase | Treatment Phase [a] | | | | | | | | Follow-up Phase [a] |
|---|---|---|---|---|---|---|---|---|---|---|
| | Screening | 28-day cycles | | | | | | | EOT [b] | Follow-up |
| | | Cycles 1 to 2 | | | | Cycles 3 to 4 | | Cycle 5 and beyond | | |
| | Days -28 to -1 | Day 1 | Day 8 | Day 15 | Day 22 | Day 1 | Day 15 | Day 1 | | |
| Informed Consent | × [c] | | | | | | | | | |
| IVRS Registration and Calls | × | × [d] | | | | × | | × | × | |
| Inclusion & Exclusion Criteria | × | × [e, f] | | | | | | | | |
| Demographics | × | | | | | | | | | |
| Initial AML Diagnosis | × | | | | | | | | | |
| Prior AML Therapies | × | | | | | | | | | |
| AML Disease Status | × [g] | | | | | | | | | |
| Central Testing of IDH2 Gene Mutations on BMA and PB | × [g] | | | | | | | | | |
| Medical History | × | | | | | | | | | |
| Prior Medications and Procedures | × [h] | | | | | | | | | |
| Eligibility of CCR Options | × [i] | | | | | | | | | |
| EORTC QLQ-C30 and EQ-ED-5L | | × | | | | × | | × | × | |
| ECOG Performance Status | × | × | | | | × | | × | × | |
| Physical Exam [j] | × | × | | | | × | | × | × | |
| Vital Signs | × | × | × | × | × | × | × | × | × | |

| Events | Screening Phase | Treatment Phase [a] | | | | | | | | Follow-up Phase [a] |
|---|---|---|---|---|---|---|---|---|---|---|
| | Screening | 28-day cycles | | | | | | | EOT [b] | Follow-up |
| | | Cycles 1 to 2 | | | | Cycles 3 to 4 | | Cycle 5 and beyond | | |
| | Day -28 to -1 | Day 1 | Day 8 | Day 15 | Day 22 | Day 1 | Day 15 | Day 1 | | |
| **Height** | × | | | | | | | | | |
| **Body Weight** | × | × | | | | × | | × | × | |
| **BSA Calculation [k]** | × | × | | | | × | | × | | |
| **ECHO / MUGA Scan [l]** | × [m] | As clinically indicated | | | | | | | | |
| **12-Lead ECG [l]** | × | × | | × | | × | | × | × | |
| **Pregnancy Test (FCBP only) [n]** | × [o] | × [l, p] | | | | × [l] | | × [l] | × [l] | |
| **Urinalysis [o]** | × | As clinically indicated | | | | | | | | |
| **Coagulation Laboratory [o]** | × | | | | | | | | | |
| **Hematology Laboratory [o]** | × | × | × | × | × | × | × | × | × | × [q] |
| **Chemistry Laboratory [o]** | × | × | | × | | × | | × | × | |
| **UGT1A1 Gene Mutation Test [o]** (for Diagnosis of Gilbert's syndrome; refer to Section 4.2) | × | | | | | | | | | |
| **Cardiac Markers [o]** | | × [f] | × [r] | | | | | | | |
| **Fasting Lipid Laboratory [o]** | | × [f] | × [s] | | | | | | | |
| **Adverse Event** | Continuous starting after signing ICF until 28 days after the last study treatment or until the EOT visit, whichever is longer | | | | | | | | | |
| [redacted] | | | | | | | | | | |
| **Concomitant Medications and Procedures** | | Continuous until 28 days after the last study treatment or until the EOT visit, whichever is longer | | | | | | | | |
| **Transfusions** | × [u] | | | | | | | | | |

[redacted]

| Events | Screening Phase | Treatment Phase [a] | | | | | | | | Follow-up Phase [a] |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 28-day cycles | | | | | | | | |
| | Screening | Cycles 1 to 2 | | | | Cycles 3 to 4 | | Cycle 5 and beyond | EOT [b] | Follow-up |
| | Day -28 to -1 | Day 1 | Day 8 | Day 15 | Day 22 | Day 1 | Day 15 | Day 1 | | |
| **BMA for Assessing Disease Status /Response to Treatment [v]** | × [g] | × [w] | | | | × [x] | | × [y] | × [z] | × [q] |
| [redacted] | | | | | | | | | | |
| **BMB [dd]** | × [g] | × [w] | | | | × [x] | | × [y] | × [z] | × [q] |
| **PB Smear** | × [g] | × [w] | | | | × [x] | | × [y] | × [z] | × [q] |
| **Cytogenetics Testing** | × [g] | × [w, ee] | | | | × [x, ee] | | × [y, ee] | × [z, ee] | × [q, ee] |
| **Response and HI** | | × [w] | | | | × [x] | | × [y] | × [z] | × [q] |
| [redacted] | | | | | | | | | | |

[redacted]

| Events | Screening Phase | Treatment Phase [a] | | | | | | | | Follow-up Phase [a] |
|---|---|---|---|---|---|---|---|---|---|---|
| | Screening | 28-day cycles | | | | | | | EOT [b] | Follow-up |
| | | Cycles 1 to 2 | | | | Cycles 3 to 4 | | Cycle 5 and beyond | | |
| | Day -28 to -1 | Day 1 | Day 8 | Day 15 | Day 22 | Day 1 | Day 15 | Day 1 | | |
| **Treatment Accountability** [ii] | | × [jj] | | | | × | | × | × | |
| **IP Dispensation** | | × | | | | × | | × | | |
| **Treatment Administration** | | See protocol Section 7.2 for details | | | | | | | | |
| **Survival Follow-up** | | | | | | | | | | × [kk] |
| **Subsequent AML Therapies** | | | | | | | | | | × [kk] |

Abbreviations: AML = acute myeloid leukemia; β- hCG = β-subunit of human chorionic gonadotropin; BMA = bone marrow aspirate; BSA = body surface area; CCR = conventional care regimen; CR = morphologic complete remission; ECHO = echocardiogram; eCRF = electronic case report form; ECG = electrocardiogram; ECOG = Eastern Cooperative Oncology Group; EORTC= European Organization for Research and Treatment of Cancer; FCBP = female of childbearing potential; EOT = End of Treatment; HI = hematologic improvement; HRQoL = Health-related Quality of Life; HSCT = hematopoietic stem cell transplantation; ICF = informed consent form; IDH2 = isocitrate dehydrogenase isoform 2; IP = investigational product; IVRS = Interactive Voice Response System; IWG = International Working Group; MUGA = multi-gated acquisition; [redacted] PB = peripheral blood; SOC = standard of care; UGT1A1 = uridine diphosphate-glucuronosyltransferase 1 family, polypeptide A1.

[a] Study Treatment is to be initiated on Day 1 of each treatment cycle. One treatment cycle (one month) is considered as 28 days (i.e., 4 weeks). Unless noted otherwise, an administrative window of ± 3 days is permitted for all subsequent visits after the start of study treatment in the Treatment Phase. Day 1 of Cycles 2 and beyond may be delayed from Day 28 of the prior cycle in order for subjects to recover from toxicity and meet criteria for initiating each treatment cycle (protocol Section 7.2). The study visit window for EOT or monthly-scheduled survival follow-up is ± 7 days.

[b] See protocol Section 6.2.3 for details.

[c] Including informed consent for mandatory genetic testing [redacted].

[d] The subject should start study treatment (i.e., Day 1 of Cycle 1) within 3 days after randomization.

[e] Subject should continue to be eligible for study entry prior to the start of study treatment on Day 1 of Cycle 1.

[redacted]

[g] See protocol Section 6.1 for details regarding collecting BMA, BMB, PBS and cytogenetics at screening for assessing AML disease status, collecting BMA and PB at screening for central testing IDH gene mutations [redacted].

[redacted]

[h] All prior medications (prescription and non-prescription) taken and treatment procedures received from the 4-week period (i.e., 28 days) prior to starting study treatment (including those prior to the start of study treatment on Day 1 of Cycle 1), and all prior anticancer therapies, regardless of discontinuation date of treatment.
[i] Documentation of CCR treatment option selected by the investigator must occur prior to randomization.
[j] Source documented only.
[k] The BSA calculation is per the Dubois & Dubois formula: BSA (m2) = weight (kg)0.425 x height (cm)0.725/139.2. Where applicable, the dose should be calculated on Day 1 of each treatment cycle.
[l] These assessments will be done locally.
[m] If an assessment has been performed within 28 days prior to the start of study treatment, it does not need to be repeated.
[n] Pregnancy test is required for all FCBPs (see protocol Section 4.2 for the definition). Serum β-hCG pregnancy test (sensitivity of at least 25 mIU/mL) will be performed centrally at screening. For FCBP subjects, a local serum or urine (investigator's discretion under local regulations) β- hCG pregnancy test (sensitivity of at least 25 mIU/mL) is to be done within 72 hours prior to study treatment administration on Day 1 of every treatment cycle in the Treatment Phase and at the EOT visit. Negative results are required for study treatment administration.
[o] These assessments will be done centrally.
[p] A local serum or urine (investigator's discretion under local regulations) pregnancy test (sensitivity of at least 25 mIU/mL) is to be done within 72 hours prior to the start of study treatment in the Treatment Phase for FCBP only (note that the screening central serum pregnancy test can be used as the test prior to the start of study treatment in the Treatment Phase if it is performed within the 72-hour timeframe).
[q] Whenever response/disease status is assessed in the Follow-up Phase (i.e., every 8 weeks [± 14 days]). See protocol Section 6.3 for details.
[r] Day 1 (± 14 days) of every 3rd cycle (e.g., Cycles 3, 6, 9, etc) or more frequently per standard institutional practice. Not necessary for the EOT visit if last performed within 14 days.
[s] Day 1 (± 28 days) of every 6th cycle (e.g., Cycles 6, 12, 18, etc) or more frequently per standard institutional practice. Not necessary for the EOT visit if last performed within 28 days.

[redacted]

[u] Including type, number of units, reasons and date of transfusions taken ≤ 8 weeks prior to the start of study treatment through 28 days after the last study treatment or until the EOT visit, whichever is longer. Thereafter, transfusions will continue to be collected until the next AML therapy after discontinuation from study treatment, death, lost to follow-up, withdrawal of consent for further data collection, or the End of Trial, whichever occurs first.
[v] In addition to the frequency specified in the table, samples will also be collected if clinically indicated (e.g., confirmation of CR/CRi/CRp, morphologic relapse after CR/CRi/CRp or PD by a repeated bone marrow assessment at least 1 month later) or required for toxicity assessment.
[w] Within 7 days prior to Day 1 of Cycle 2.
[x] Within 7 days prior to Day 1 of Cycle 3.
[y] Within 7 days prior to Day 1 of Cycle 5 and Day 1 of every 2nd cycle thereafter (e.g., Cycles 7, 9, etc).
[z] Not necessary for the EOT visit if last performed within 28 days.

[redacted]

[redacted]

[dd] A bone marrow biopsy can be collected in conjunction with an aspirate if it is standard institutional practice. A bone marrow biopsy must be collected if adequate aspirate is not attainable.

[ee] If the bone marrow aspirate is obtained for confirming CR/CRi/CRp, morphologic relapse after CR/CRi/CRp or PD, a sample of the bone marrow aspirate is to be sent to the local or central cytogenetics laboratory (for sites without local analysis capability) for standard metaphase preparation and analysis of a minimum of 20 metaphase cells. A bone marrow biopsy can be used for cytogenetics testing if adequate aspirate is not attainable (note that specific handling of the biopsy is required for cytogenetics testing).

[redacted]

[ii] Including diary cards, if utilized. See protocol Section 7.6 for details.

[jj] Cycle 2 only.

[kk] Every 4 weeks (± 7 days) for survival follow-up until death, lost to follow-up, withdrawal of consent for further data collection or the End of Trial, whichever occurs first. Subsequent AML therapies should be collected at the same time schedule. See protocol Section 6.3 for details.

## 15.5. International Working Group Acute Myeloid Leukemia Response Criteria

| **Hematologic Response According to modified IWG Criteria for AML** | |
|---|---|
| **Category** | **Definition** |
| **Morphologic Complete Remission (CR)** | Defined as less than 5% blasts in a BM aspirate sample with marrow spicules and with a count of at least 200 nucleated cells. There should be no blasts with Auer rods and absence of extramedullary disease. Plus, all the following conditions should be met:<br>• ANC ≥ 1 x $10^9$/L (1,000/μL)<br>• Platelet count ≥ 100 x $10^9$/L (100,000/μL)<br>• Independent of red cell transfusions for ≥ 1 week immediately before each response assessment |
| **Morphologic Complete Remission with Incomplete Neutrophil Recovery (CRi)** [a] | Defined as all criteria of morphologic CR except the following:<br>• ANC < 1 x $10^9$/L (1,000/μL) |
| **Morphologic Complete Remission with Incomplete Platelet Recovery (CRp)** [a] | Defined as all criteria of morphologic CR except the following:<br>• Platelet count < 100 x $10^9$/L (100,000/μL) |
| **Morphologic Complete Remission with Partial Hematologic Recovery (CRh)** [a] | Defined as all criteria of morphologic CR except the following:<br>• ANC > 0.5 x $10^9$/L (500/μL), and<br>• Platelet count > 50 x $10^9$/L (50,000/μL) |
| **Morphologic Leukemia-free State (MLFS)** | Defined as less than 5% blasts in a BM aspirate sample with marrow spicules and with a count of at least 200 nucleated cells. There should be no blasts with Auer rods and absence of extramedullary disease |
| **Partial Remission (PR)** | Defined as all hematologic criteria of morphologic CR with a > 50% decrease in the percentage of BM blasts to 5% to 25% (a blast count value of < 5% may also be considered a partial remission if Auer rods are present) [b] |
| **Cytogenetic Complete Remission (CRc)** | Defined as CR/CRi/CRp with a reversion to a normal karyotype in cases with an abnormal karyotype at baseline, based on evaluating ≥ 20 metaphase cells from BM |
| **Morphologic Relapse after CR/CRi/CRp** [a] | Defined as one of the following conditions:<br>• Reappearance of ≥ 5% blasts in the BM not attributable to any other cause (e.g., BM regeneration after consolidation therapy); or<br>• Development of extramedullary disease |
| **Not evaluable (NE)** [a] | Defined as without a post-treatment response assessment |
| **Stable Disease (SD)** [a] | Defined as failure to meet any of the above criteria and not meeting the criteria of progressive disease (see below) |

| | |
|---|---|
| **Progressive Disease (PD)** [a] | Defined as one of the following conditions:<br>• For subjects with 5 to 70% BM blasts at baseline: a > 50% increase of **BM blast count percentage** from baseline to ≥ 20%; or<br>• For subjects with > 70% BM blasts at baseline: a doubling of **absolute blast count in peripheral blood** from baseline to ≥ 10 x $10^9$/L (10,000/µL); or<br>• Development of new extramedullary disease since last response assessment<br>Progressive disease is to be confirmed by 2 consecutive response assessments separated by at least 1 month. The date of progressive disease is defined as the first date that one of three conditions listed above was met. |

AML = Acute Myeloid Leukemia; ANC= Absolute Neutrophil Count; BM= Bone Marrow; IWG= International Working Group.
[a] Modification to IWG response criteria.
[b] If the pre-treatment bone marrow blast percentage was 50% to 100%, the percentage of blasts must decrease to a value between 5% and 25%; if the pre-treatment blast percentage was 20% to less than 49%, they must decrease by at least half to a value of more than 5%.
Notes: Deletions to the IWG response criteria are not shown.
Source: protocol Appendix F.

## 15.6. Hematologic Improvement According to the International Working Group for Myelodysplastic Syndromes

| Hematologic Improvement According to IWG Criteria | |
|---|---|
| **Hematologic improvement** [a] | **Response criteria (responses must last at least 8 week)** [b] |
| Erythroid Response (HI-E)<br>(pre-treatment, < 11 g/dL) | • Hemoglobin increase by ≥ 1.5 g/dL<br>• Relevant Reduction in units of RBC transfusions by an absolute number of at least 4 RBC transfusions/8 week compared with the pretreatment transfusion number in the previous 8 weeks<br>Note: Only RBC transfusions given for a hemoglobin of ≤ 9.0 g/dL on treatment will count in the RBC transfusion response evaluation [b] |
| Platelet Response (HI-P)<br>(pre-treatment, < 100 X $10^9$/L) | • Absolute increase of ≥ 30 X $10^9$/L for subjects starting with > 20 X $10^9$/L platelets<br>• Increase from <= 20 X $10^9$/L to > 20 X $10^9$/L and by at least 100% [b] |
| Neutrophil Response (HI-N)<br>(pre-treatment, < 1.0 X $10^9$/L) | • At least 100% increase and an absolute increase > =0.5 X $10^9$/L [b] |
| Progression or Relapse After HI [c] | At least 1 of the following:<br>• At least 50% decrease from maximum response levels in granulocytes or platelets<br>• Reduction in hemoglobin by ≥ 1.5 g/dL<br>• Transfusion dependence |

HI-E = hematologic improvement erythroid response; HI-N = hematologic improvement neutrophil response; HI-P = hematologic improvement platelet response; IWG = International Working Group; RBC = red blood cell.

Pre-treatment counts averages of at least 2 measurements (not influenced by transfusions, i.e., no RBC transfusions for 2 weeks and no platelet transfusions for 1 week) ≥ 1 week apart (modification).
Modification to IWG (2000) response criteria.

[c] In the absence of another explanation, such as acute infection, repeated courses of chemotherapy (modification), gastrointestinal bleeding, hemolysis, and so forth. It is recommended that the 2 kinds of erythroid and platelet responses be reported overall as well as by the individual response pattern.

Note: Deletions to the IWG response criteria are not shown. To convert hemoglobin levels from grams per deciliter to grams per liter, multiply grams per deciliter by 10.

Source: Cheson, 2006.

# Celgene Signing Page

**This is a representation of an electronic record that was signed electronically in Livelink. This page is the manifestation of the electronic signature(s) used in compliance with the organizations electronic signature policies and procedures.**

UserName: [redacted]
Title: [redacted]
Date: Tuesday, 02 June 2020, 08:50 AM Eastern Daylight Time
Meaning: Approved, no changes necessary.
==========

UserName: [redacted]
Title: [redacted]
Date: Tuesday, 02 June 2020, 02:35 PM Eastern Daylight Time
Meaning: Approved, no changes necessary.
==========

UserName: [redacted]
Title: [redacted]
Date: Tuesday, 02 June 2020, 02:37 PM Eastern Daylight Time
Meaning: Approved, no changes necessary.
==========