STATISTICAL ANALYSIS PLAN

Final Version 4.0, dated 22-Feb 2024

Phase 3 Open-label, Multicenter, Randomized Study of ASP2215 versus Salvage Chemotherapy in Patients with Relapsed or Refractory Acute Myeloid Leukemia (AML) with FLT3 Mutation

ISN/Protocol: 2215-CL-0303

Sponsor:

Astellas Pharma Inc. (API)

Table of Contents

I. LIST OF ABBREVIATIONS AND KEY TERMS 5
1 INTRODUCTION 9
2 FLOW CHART AND VISIT SCHEDULE 10
3 STUDY OBJECTIVES AND DESIGN 11
3.1 Study Objectives 11
3.1.1 Primary Objectives 11
3.1.2 Secondary Objectives 11
3.1.3 Exploratory Objectives: 11
3.2 Study Design 12
3.2.1 Study Design 12
3.3 Randomization 14
4 SAMPLE SIZE 15
5 ANALYSIS SETS 15
5.1 Intention to Treatment Set (ITT) 15
5.2 Full Analysis Set (FAS) 16
5.3 Safety Analysis Set (SAF) 16
5.4 Pharmacokinetics Analysis Set (PKAS) 16
6 ANALYSIS VARIABLES 16
6.1 Efficacy Endpoints 16
6.1.1 Primary Efficacy Endpoint 16
6.1.2 Secondary Efficacy Endpoints 17
6.1.3 Exploratory Endpoints 23
6.1.4 Other Efficacy Variables 25
6.2 Safety Variables 25
6.3 Pharmacokinetics Variables 26
6.4 Other Variables 27
7 STATISTICAL METHODOLOGY 29
7.1 General Considerations 29
7.2 Study Population 30
7.2.1 Disposition of Subjects 30
7.2.2 Protocol Deviations 30

7.2.3 Demographic and Other Baseline Characteristics ··· 31
7.2.4 Previous and Concomitant Medications ··· 31
7.2.5 Previous and Concomitant Transfusions ··· 32
7.2.6 Prior AML Chemotherapy ··· 32
7.2.7 Non-Medication Therapy ··· 32
7.2.8 Clinical/Diagnostic Procedures ··· 32
7.2.9 Subsequent AML Therapy ··· 32
7.3 Study Drugs ··· 33
7.3.1 Exposure ··· 33
7.4 Analysis of Efficacy ··· 33
7.4.1 Analysis of Primary Endpoint of OS ··· 33
7.4.2 Analysis of Secondary Endpoints ··· 35
7.4.3 Analysis of Exploratory Endpoints ··· 38
7.4.4 Analysis of Other Variables ··· 38
7.5 Analysis of Safety ··· 39
7.5.1 Adverse Events ··· 39
7.5.2 Clinical Laboratory Evaluation ··· 41
7.5.3 Vital Signs ··· 42
7.5.4 Electrocardiograms (ECGs) ··· 43
7.5.5 Pregnancies ··· 44
7.5.6 Eastern Cooperative Oncology Group (ECOG) Performance Scores ··· 44
7.6 Analysis of PK ··· 44
7.7 Analysis of PD ··· 45
7.8 Subgroups of Interest ··· 45
7.9 Other Analyses ··· 47
7.9.1 PK-PD Analysis ··· 47
7.10 Interim Analysis (and Early Discontinuation of the Clinical Study) ··· 47
7.11 Handling of Missing Data, Outliers, Visit Windows, and Other Information ··· 48
7.11.1 Missing Data ··· 48
7.11.2 Outliers ··· 50
7.11.3 Visit Windows ··· 50
7.11.4 Pooling Strata ··· 50
7.11.5 Blinding ··· 51
8 DOCUMENT REVISION HISTORY ··· 52

9 REFERENCES ····· 57

10 APPENDICES ····· 58

10.1 Appendix 1: SIGNATURE ····· 58

10.2 Appendix 2: Blinding Process ····· 59

10.3 Appendix 3: Search Strategy for Adverse Events of Interest ····· 60

10.4 Appendix 4: Search Strategy for medication ····· 61

10.5 Appendix 5: China Specific Analysis····· 61

10.6 Appendix 6: FACIT-Dys SF ····· 62

10.7 Appendix 7: FACT-Leu Scoring ····· 65

I. LIST OF ABBREVIATIONS AND KEY TERMS

List of Abbreviations

Abbreviations	Description of abbreviations
ADaM	Analysis Data Model
AE	Adverse Event
AESI	Adverse Events of special Safety Interest
ALP	Alkaline Phosphatase
ALT	Alanine Transaminase
AML	Acute Myeloid Leukemia
ANC	Absolute Neutrophil Count
ANCOVA	Analysis of Covariance
ANOVA	Analysis of Variance
ASP2215	Astellas Compound code for 2215
AST	Aspartate Transaminase
ATC	Anatomical Therapeutic Chemical Classification
BFI	Brief Fatigue Inventory
BSA	Body Surface Area
CDE	Centre for Drug Evaluation
CMH	Cochran-Mantel-Haenszel
CR	Complete Remission
CRc	Composite Complete Remission
CRF	Case Report Form
CRh	Complete Remission with partial hematologic recovery
CRi	Complete Remission with incomplete hematological recovery
CRp	Complete Remission with incomplete platelet recovery
CSR	Clinical Study Report
CT	Computed tomography
CTCAE	Common Terminology Criteria for Adverse Events
DBL	Database Lock
DBP	Diastolic Blood Pressure
DCR	Duration of CR
DCRc	Duration of CRc
DCRCRh	Duration of CR/CRh
DR	Duration of Response
ECG	Electrocardiogram
ECHO	Echocardiogram
ECOG	Eastern Cooperative Oncology Group
eCRF	Electronic Case Report Form
EDC	Electronic Data Capture
EFS	Event-Free Survival
EQ-5D-5L	EuroQol Group-5 Dimension-5 Level Instrument
EWB	Emotional Well-Being
FAB	French-American-British

Abbreviations	Description of abbreviations
FACIT	Functional Assessment of Chronic Illness Therapy
FACIT-Dys-SF	Functional Assessment of Chronic Illness Therapy-Dyspnea-Short Forms
FACT-G	Functional Assessment of Cancer Therapy-General
FACT-Leu	Functional Assessment of Cancer Therapy-Leukemia
FACT-Leu TOI	FACT-Leu Trial Outcome Index
FAS	Full Analysis Set
FLAG	Fludarabine, cytarabine and granulocyte colony-stimulating factor
FLT3	FMS-like Tyrosine kinase
FSI	First Subject In
FWB	Functional Well-Being
GVHD	Graft-versus-host disease
HRQoL	Health-Related Quality of Life
HSCT	Hematopoietic Stem Cell Transplant
IAP	Interim Analysis Plan
ICF	Informed Consent Form
ICH	International Conference on Harmonization
IDAC	Independent Data Analysis Center
IDMC	Independent Data Monitoring Committee
INR	International Normalization Ratio
IRT	Interactive Response Technology
ISN	International Study Number
ITD	Internal Tandem Duplication
ITT	Intention to Treatment Set
IV	Intravenous
LeuS	Leukemia Subscale
LFS	Leukemia-Free Survival
LoDAC	Low-Dose cytarabine
MEC	Mitoxantrone, etoposide and intermediate-dose cytarabine
MMRM	Mixed-Effect Model Repeated Measure
MUGA	Multigated acquisition scan
NCI	National Cancer Institute
NE	Not evaluable
NR	No Response
OS	Overall Survival
PD	Protocol Deviation
PD	Pharmacodynamics
PGx	Pharmacogenetics
PK	Pharmacokinetic
PKAS	Pharmacokinetic Analysis Set
PPS	Per Protocol Set
PR	Partial Remission

Abbreviations	Description of abbreviations
PRO	Patient reported outcome
PT	Preferred Term
PWB	Physical Well-Being
QTc	QT interval corrected for heart rate
QTcF	QT interval corrected for heart rate using Fridericia's factor
RBC	Red Blood Cell
RR	Interval between 2 consecutive r waves on an ECG
SAE	Serious Adverse Event
SAF	Safety analysis set
SAP	Statistical Analysis Plan
SAS	Statistical Analysis System
SBP	Systolic Blood Pressure
SDTM	StudyData Tabulation Model
SMQ	Standardised MedDRA Queries
SOC	System Organ Class
SWB	Social/family Well-Being
TEAE	Treatment Emergent Adverse Event
TKD	Tyrosine kinase domain
TLF	Tables, Listings and Figures
TTCR	Time To CR
TTCRc	Time To CRc
TTCRCRh	Time To CR/CRh
TTR	Time To Response
ULN	Upper limit of normal
VAS	Visual analogue scale
WHO-DD	World Health Organization–Drug Dictionary

List of Key Terms

Terms	Definition of terms
Baseline	Observed values/findings which are regarded as the starting point for comparison.
Enroll	To register or enter into a clinical trial. NOTE: Once a subject has been enrolled, the clinical trial protocol applies to the subject.
Intervention	The drug, therapy or process under investigation in a clinical study that is believed to have an effect on outcomes of interest in a study (e.g., health-related quality of life, efficacy, safety, pharmacoeconomics).
Investigational period	Period of time where major interests of protocol objectives are observed and where the test drug or comparative drug (sometimes without randomization) is usually given to a subject and continues until the last assessment after completing administration of the test drug or comparative drug.
Post investigational period	Period of time after the last assessment of the protocol. Follow-up observations for sustained adverse events and/or survival are done in this period.
Randomization	The process of assigning trial subjects to treatment or control groups using an element of chance to determine assignments in order to reduce bias.
Screen failure	Potential subject who signed consent but did not meet 1 or more criteria required for participation in a trial and did not randomize to the trial.
Screening	A process of active consideration of potential subjects for enrollment in a trial.
Screening period	Period of time before entering the investigational period, usually from the time of starting a subject signing consent until just before the randomization.
Study period	Period of time from the first site initiation date to the last site completing the study.
Variable	Any quantity that varies; any attribute, phenomenon or event that can have different qualitative or quantitative values.

1 INTRODUCTION

This Statistical Analysis Plan (SAP) contains a more technical and detailed elaboration of the principal features of the analysis described in the protocol, and includes detailed procedures for executing the statistical analysis of the primary and secondary endpoints and other data.

The SAP is finalized and signed prior to any of the following: study unblinding, database hard lock, interim analysis, or accumulation of substantial amount of data in an open-label study to ensure lack of bias. For operational efficiency an earlier time is usually targeted and wherever possible, the SAP should be developed in parallel with protocol finalization. For phase 2-4 studies the SAP should be developed and approved before First Subject In (FSI). If the expected interval between FSI and soft-lock is less than 12 weeks, then the SAP should be approved by 12 weeks prior to the planned date of soft-lock. If needed, revisions to the approved SAP may be made prior to database hard lock. Revisions will be version controlled.

This statistical analysis is coordinated by the responsible biostatistician of API. Any changes from the analyses planned in the SAP will be justified in the Clinical Study Report (CSR).

2 FLOW CHART AND VISIT SCHEDULE

Refer Section V. FLOW CHART AND SCHEDULE OF ASSESSMENTS of the protocol.

3 STUDY OBJECTIVES AND DESIGN

3.1 Study Objectives

3.1.1 Primary Objectives

The primary objective is to:

- Determine the clinical benefit of ASP2215 therapy in subjects with FLT3-mutated AML who are refractory to or have relapsed after first-line AML therapy as shown with OS compared to salvage chemotherapy.

3.1.2 Secondary Objectives

The key secondary objectives are to:

- Determine the overall efficacy in event-free survival (EFS) of ASP2215 compared to salvage chemotherapy.
- Determine the overall efficacy in complete remission (CR) rate of ASP2215 compared to salvage chemotherapy.

The secondary objectives are to evaluate the safety and efficacy of ASP2215 therapy versus salvage chemotherapy in terms of:

- Leukemia-free survival (LFS)
- Duration of remission
- Composite complete remission (CRc = CR+CRp+CRi) rate
- Complete Remission and Complete Remission with Partial Hematologic Recovery (CR/CRh) Rate
- Transfusion conversion rate and transfusion maintenance rate
- Transplantation rate
- Patient reported fatigue (Brief Fatigue Inventory [BFI])
- Adverse events (AEs), safety labs, vital signs, electrocardiograms (ECGs) and Eastern Cooperative Oncology Group (ECOG) performance scores
- Evaluate the PK of ASP2215 therapy in the Chinese population

3.1.3 Exploratory Objectives:

Evaluate the safety and efficacy of ASP2215 therapy versus salvage chemotherapy in terms of:

- Pharmacogenomics (PGx)
- FLT3 gene mutation status
    - mutation types and frequency
    - relationship to efficacy and safety

- Exploratory (predictive) biomarkers of ASP2215 activity
- Resource utilization in this study population including hospitalization, blood transfusion, antibiotic iv infusions, medication for AEs and opioid usage
- Patient reported dyspnea (Functional Assessment of Chronic Illness Therapy-Dyspnea-Short Forms [FACIT-Dys-SF])
- Patient reported signs, symptoms and impacts of AML (Functional Assessment of Cancer Therapy-Leukemia [FACT-Leu], dizziness and mouth sore items)
- EuroQol Group-5 Dimension-5 Level Instrument (EQ-5D-5L)

3.2 Study Design

3.2.1 Study Design

This is phase 3, open-label, multicenter, randomized study to compare the efficacy and safety of ASP2215 therapy to salvage chemotherapy in FLT3-mutated AML subjects who are refractory to or have relapsed after first-line AML therapy. Approximately 50 centers in China, Russia, Singapore, Thailand, and Malaysia will participate in this study.

Three hundred eighteen subjects will be randomized. The randomization of the 318 subjects will be in a 1:1 ratio to receive ASP2215 or salvage chemotherapy. Subjects will enter the screening period up to 14 days prior to the start of treatment. Prior to randomization, the investigator will preselect a salvage chemotherapy regimen for each subject; options will include LoDAC, MEC or FLAG. The randomization will be stratified by response to first-line therapy and preselected salvage chemotherapy. Subjects will be administered treatment over continuous 28-day cycles and per institutional guidelines for chemotherapy product preparation and administration. The dose and duration of study treatments are outlined in Section 5.1.1 of the protocol.

Two (or more) study sites in China will be designated as a “pharmacokinetic analysis site” (PK analysis site). Approximately 20 Chinese subjects (10 male and 10 female subjects) (they will be included in 318 subjects) at PK analysis sites, who are randomized to the ASP2215 arm will be allocated to the “PK cohort”. Subjects in the PK cohort will be requested to be hospitalized on Day 1 and Day15 until the completion of all the assessments. Subjects in the PK cohort will be administered ASP2215 in the same manner and undergo the same efficacy and safety assessments as other subjects except for blood sampling for additional PK analysis. For subjects in the PK cohort, written informed consent must be obtained by the specific informed consent form (ICF) explaining that more frequent blood sampling will be done.

The first 20 subjects (10 male and 10 female) randomized into ASP2215 arm at PK cohort sites will participate in PK cohort.

For subjects taking ASP2215 or LoDAC, treatment should continue until the subject meets a treatment discontinuation criterion.

Subjects receiving MEC or FLAG will receive 1 cycle of therapy and will be assessed for response on or after day 15, per institutional guidelines. If the bone marrow cellularity is 20%

or greater with at least a 50% reduction in blasts, the subject may receive a second cycle of the same chemotherapy. If bone cellularity is between 5% and 20%, the investigator should make the decision whether the subject should receive another treatment cycle or be observed for recovery. If marrow cellularity is 5% or less, the subject will be observed for recovery. Subjects achieving CR, CRi or CRp may receive a second cycle of chemotherapy at the investigator's discretion. Subjects with no response (NR) or progressive disease following cycle 1 will discontinue study treatment.

Dose adjustments for ASP2215 are described in Section 5.1.2 of the protocol.

Subjects who have a donor identified and achieve a response allowing them to undergo HSCT per each institution's assessment can undergo HSCT without leaving the study. However, ASP2215 should be stopped and a pre-HSCT visit should be performed prior to starting the conditioning regimen for HSCT. ASP2215 can be resumed after stem cell transplantation if the following conditions are met:

- Subject is between 30 - 90 days post HSCT
- Subject has had successful engraftment as demonstrated by absolute neutrophil count (ANC) ≥ 500/mm3 and platelets ≥ 20000/mm3 without transfusions
- Subject does not have ≥ grade 2 acute graft-versus-host disease (GVHD)
- Subject is in CRc

For subjects resuming treatment, subjects will follow the procedures listed under subsequent cycles day 1 in the Schedule of Assessments. Subjects who do not resume ASP2215 will be followed for primary endpoint.

After treatment discontinuation, subjects will have an end of treatment visit within 7 days after treatment discontinuation, followed by a 30 day follow-up, in which a telephone contact with the subject is sufficient unless any assessment must be repeated for resolution of treatment-related AEs. After which the subjects will enter the long-term follow-up period for collection of patient reported outcome (PRO) using EQ-5D-5L, subsequent AML treatment, remission status and survival (cause of death and date of death). The long-term follow-up will be every 3 months, for up to 3 years from the subject's end of treatment visit until the implementation and reconsent of the current protocol version 7.0, at which time they will discontinue from the study.

A formal interim analysis by an Independent Data Monitoring Committee (IDMC) will be done when approximately 50% of deaths by any cause have occurred. This analysis will be utilized to determine whether the study should be terminated earlier than planned if ASP2215 has more favorable or harmful outcome than the salvage chemotherapy group. If the interim analysis demonstrates a more favorable outcome for ASP2215, enrollment to the study may be stopped in advance. If it demonstrates a harmful outcome, the enrollment will be stopped. However, any subject continuing to derive clinical benefit from ASP2215 as assessed by the investigator will be allowed to continue treatment until they meet a discontinuation criterion as outlined in Section 6 of the protocol or upon marketing authorization and commercial availability and applicable reimbursement of ASP2215 in the country of residence.

Subjects will be managed per the local institution's standard of care for safety and efficacy assessments while on study treatment after subjects reconsents under this protocol version 7.0. No data (including PK assessments) will be collected in the eCRFs after subjects' reconsent under this protocol version 7.0. Only SAEs, as defined in Section 5.5.2 of the protocol, will be collected and reported to Astellas Pharma Inc. Product Safety & Pharmacovigilance. SAE data will be reported in the safety database. SAE collection will continue until 30 days after the last dose of study treatment. Once subjects receiving study treatment meet the study discontinuation criteria or upon marketing authorization, commercial availability and applicable reimbursement of ASP2215 in the country of residence, subjects will be discontinued from the study.

Subjects in long-term follow-up who are no longer receiving study treatment will be followed every 3 months for up to 3 years until implementation and reconsent of the current protocol version 7.0, at which time they will discontinue from the study.

The detailed assessments and procedures are presented in Appendix 12.8 of the protocol.

The interim analysis of the study demonstrated a positive outcome and hence the Crossover Extension as outlined in Section 12.7 of the protocol was implemented.

3.3 Randomization

Randomization and study drug assignment will be performed via Interactive Response Technology (IRT). Prior to the initiation of the study treatment, the site staff will contact the IRT in order to determine the randomly assigned treatment. Specific procedures for randomization through the IRT are contained in the study procedures manual.

Subjects will be randomized in a 1:1 ratio to receive ASP2215 or salvage chemotherapy.

Randomization will be stratified by response to first-line AML therapy and preselected salvage chemotherapy:

Response to first-line therapy:

- Relapse within 6 months after allogeneic HSCT
- Relapse after 6 months after allogeneic HSCT
- Primary refractory without HSCT
- Relapse within 6 months after CRc and no HSCT
- Relapse after 6 months after CRc and no HSCT

Preselected chemotherapy:

- High intensity chemotherapy (FLAG, MEC)
- Low intensity chemotherapy (LoDAC)

4 SAMPLE SIZE

This is a group sequential design using the O'Brien-Fleming boundaries as implemented by Lan-DeMets alpha spending method (East®). One interim analysis and 1 final analysis are planned. The interim and final analyses will be performed after the prespecified number of death events. All statistical tests of treatment effects will be conducted at the 2-sided 0.05 level of significance.

Approximately 318 subjects (the planned sample size with an accrual rate of 12.7 subjects/ month, an accrual period of 25.0 months, and a dropout rate of 10%) will be randomized in a 1:1 ratio to receive ASP2215 or salvage chemotherapy (159 subjects in the ASP2215 treatment arm and 159 subjects in the salvage chemotherapy arm). The planned 230 events during the study will provide 90% power to detect a difference in OS between the ASP2215 arm with 7.7 months median survival time and salvage chemotherapy arm with 5 months median survival time (hazard ratio = 0.65) at the overall 2-sided 0.05 significance level.

The planned sample size with 250 EFS events will provide 90% power to detect the difference in EFS (6 months median EFS for ASP2215 arm and 3.9 months for salvage chemotherapy arm with hazard ratio = 0.65) and > 90% power to detect a difference in CR rate between ASP2215 with 25% CR rate and the salvage chemotherapy with 10% CR rate.

Approximately 20 Chinese subjects (10 male and 10 female subjects) will be allocated to the PK cohort. This sample size is calculated based on the following considerations:

- "The Technical Guidelines for Clinical Pharmacokinetics Research of Chemical Drugs" in China (March 2005) require 8 to 12 subjects of each gender in each dose group.
- To account for several subjects dropping out from the study and to ensure at least 16 evaluable subjects in the ASP2215 arm, approximately 20 subjects will be enrolled.

5 ANALYSIS SETS

In accordance with International Conference on Harmonization (ICH) recommendations in guidelines E3 and E9, the following analysis sets will be used for the analyses.

Intention to Treatment Set (ITT) and Full Analysis Set (FAS) will be used for efficacy analysis. Safety Analysis Set (SAF) will be used for the analyses of safety and biomarker variables. Pharmacokinetic Analysis Set (PKAS) will be used for pharmacokinetic analyses. The data from all randomized subjects will be included in the data listings.

5.1 Intention to Treatment Set (ITT)

The Intention to Treatment Set (ITT) consists of all subjects who are randomized. The subjects will be analyzed based on the randomized treatments.

The ITT will be used for primary analyses of efficacy data, as well as selected demographic and baseline characteristics, patient reported outcomes and resource utilization.

5.2 Full Analysis Set (FAS)

The Full Analysis Set (FAS) consists of all randomized subjects with FLT3 mutation based on central test. The subjects will be analyzed based on the randomized treatments.

The FAS will be used for sensitivity analyses of efficacy data, as well as selected demographic and baseline characteristics.

5.3 Safety Analysis Set (SAF)

The Safety Analysis Set (SAF) consists of all subjects who received at least one dose of study treatment (ASP2215 or salvage chemotherapy). The subjects will be analyzed based on the actual treatment received.

The SAF will be used for summaries of demographic and baseline characteristics and all safety variables.

5.4 Pharmacokinetics Analysis Set (PKAS)

The pharmacokinetic analysis set (PKAS) consists of all participants who receive at least 1 administration of study intervention for which at least 1 concentration data with time of dosing and sampling are available. Inclusion of participants in the PKAS with missing data or important protocol deviations will be considered by the pharmacokineticist on a case-by-case basis.

The PKAS will be used for summaries of plasma concentration of ASP2215, as well as selected demographic and baseline characteristics.

6 ANALYSIS VARIABLES

The cutoff date will be considered in the definitions of all analysis variables that no data after cutoff will be used, unless otherwise specified.

6.1 Efficacy Endpoints

6.1.1 Primary Efficacy Endpoint

Overall survival (OS) is the primary efficacy endpoint. OS is defined as the time from the date of randomization until the date of death from any cause (death date – randomization date + 1). For a subject who is not known to have died by the end of study follow-up, OS is censored at the date of last contact (date of last contact – randomization date + 1).

The date of last contact is the latest date the subject is known to be alive by the cutoff date. The last contact date will be derived for subjects alive at the analysis cutoff date. Subjects with last contact date beyond the analysis cutoff date will be censored at the analysis cutoff date.

As a sensitivity analysis, OS will be defined similarly as the above primary analysis, however, subjects who undergo an allogeneic HSCT will be censored at the time of HSCT (HSCT date – randomization date +1).

In addition, as another sensitivity analysis, OS will be defined similarly as the above primary analysis, however, subjects who initiate any new anti-leukemia therapy will be censored at the time of first new anti-leukemia therapy (first new anti-leukemia therapy date – randomization date +1).

During interim and final analysis, only deaths occurring on or prior to the cutoff date are counted as OS events. Subjects with death or last known alive date after the cutoff date will be censored at the cutoff date.

6.1.2 Secondary Efficacy Endpoints

6.1.2.1 Key Secondary Efficacy Endpoints

- Event-free survival (EFS)

  EFS is defined as the time from the date of randomization to the earliest date of the following events [earliest events date – randomization date + 1], including the long-term follow-up data.

  - Documented relapse (excluding relapse after PR)
  - Treatment failure
  - Death from any cause
  - Reported off-treatment relapse prior to new AML therapy in long-term follow up
  - New AML therapy (excluding subsequent HSCT) start without reported off-treatment relapse in long-term follow up

  If a subject experiences relapse, death or new AML therapy start, the subject is defined as having EFS event, and the event date is the earliest date of relapse, death, or new AML therapy start.

  If a subject ends the treatment and fails to achieve any of the response of CR, CRp, CRi during the treatment period (subject with best response of PR or NR [excluding NE]), the subject is defined as having EFS event related to treatment failure, and the event date is the randomization date.

  For a subject who is not known to have EFS events, the subject will be censored at the date of last relapse-free disease assessment or subsequent HSCT date (whichever is later). Subject is not censored at on-study HSCT. For subjects who are censored, last relapse-free disease assessment date refers to the subject's last disease assessment date. Subject without post-treatment disease assessment or without study treatment will be censored at randomization date.

  EFS will be tested at the interim and final analysis only when OS is rejected. During interim analysis and final analysis, only EFS events occurring on or prior to the cutoff date are counted. Subjects without these events before cutoff date will be censored at the last relapse-free disease assessment date or subsequent HSCT date on or before the cutoff date.

Disease assessment date refers to the date of local bone marrow aspiration or biopsy assessment. If no aspirate or biopsy is available and subject is evaluated based on blast count from peripheral blood, the date the peripheral blood sample is drawn will be used. If none of the above is available and subject is evaluated based on the presence of extramedullary leukemia, the assessment date of extramedullary leukemia will be used.

The following sensitivity analyses for EFS will be conducted:

EFS will be defined similarly as the above primary analysis, however the date of the first new anti-leukemia therapy after end of study treatment or the last treatment evaluation (when new anti-leukemia therapy date is not available), will be used as the event date of treatment failure.

Sensitivity analysis of EFS will be performed to evaluate the impact of loss to follow-up on EFS. EFS will be defined similarly as the above primary analysis, however, subjects who discontinued either study treatment or study follow up (including 30 day and long-term follow up) due to "Lost to follow up" will also be considered as an EFS event and the event date is withdrawal date or last evaluation date as collected on end of treatment, end of 30-day follow up, and end of long-term follow up CRF page whichever occurs first. "Lost to follow up" event includes subjects who discontinued either study treatment or study follow up (including 30 day and long-term follow up) due to either "LOST TO FOLLOW-UP" or "WITHDRAWAL BY SUBJECT". For subjects who discontinued the study treatment due to "WITHDRAWAL BY SUBJECT", only those who revoked the authorization to follow up will be included.

EFS will be defined similarly as the above primary analysis with the following differences:

- Subjects who end the treatment and achieve best response of CR, CRp, CRi or PR and experience relapse (including relapse after PR) will be defined as relapse and the event date is the date of first NR after CR, CRp, CRi or PR.
- Subjects who end the treatment and achieve best response of NR will be defined as treatment failure and the event date is the randomization date

EFS will be defined similarly as the above primary analysis, however, subjects who end the treatment with post-treatment disease assessment and achieve best response of PR, NR or NE will be defined as treatment failure and the event date is the randomization date.

Sensitivity analysis of EFS not using the long-term follow-up data, defined as the time from the date of randomization to the earliest date of documented relapse (excluding relapse after PR), treatment failure or death from any cause within 30 days after the last dose of study drug, will be performed. For a subject who is not known to have EFS events, the subject will be censored at the date of last relapse-free disease assessment.

- Complete remission (CR) rate

CR rate will be evaluated at the interim and final analysis when both OS and EFS are rejected.

CR rate is defined as the number of subjects who achieve the best response of CR divided by the number of subjects in the analysis population.

6.1.2.2 Other Secondary Efficacy Endpoints

- Leukemia-free survival (LFS)

  LFS is defined as the time from the date of first CRc until the date of first documented relapse (excluding relapse from PR) or death for subjects who achieve CRc (relapse date or death date – first CRc disease assessment date + 1). For a subject who is not known to have relapsed or died, LFS is censored on the date of last relapse-free disease assessment date (last relapse-free disease assessment date – first CRc disease assessment date + 1).

- Composite complete remission (CRc) rate

  CRc rate is defined as the number of subjects who achieve CRc (CR, CRp or CRi) divided by the number of subjects in the analysis population by the cutoff date.

- Complete Remission and Complete Remission with Partial Hematologic Recovery (CR/CRh) Rate

  CR/CRh rate is defined as the number of subjects who achieve CR or CRh at any of the postbaseline visits by the number of subjects in the analysis population.

- Duration of remission

  Duration of remission includes duration of CRc (DCRc), duration of CR (DCR), duration of CR/CRh (DCRCRh), and duration of response (DR) (i.e., CRc + PR).

  Duration of CRc is defined as the time from the date of first CRc until the date of first documented relapse for subjects who achieve CRc (relapse date – first CRc disease assessment date + 1). Subjects who die without report of relapse are considered non-events and censored at their last relapse-free disease assessment date (last relapse-free disease assessment date – first CRc disease assessment date + 1). Other subjects who do not relapse on study are considered non-events and censored at the last relapse-free disease assessment date.

  Duration of CR is defined similarly as duration of CRc for subjects who achieve CR.

  Duration of CR/CRh is defined similarly as duration of CRc for subjects who achieve CR/CRh.

  Duration of response is defined as the time from the date of either first CRc or PR until the date of documented relapse of any type (i.e., the date of first NR after CRc or PR) for subjects who achieve CRc or PR (relapse date – first CRc or PR disease assessment date + 1). Subjects who die without report of relapse are considered non-events and censored at their last relapse-free disease assessment date (last relapse-free disease assessment date – first CRc or PR disease assessment date + 1). Other subjects who do not relapse on

study are considered non-events and censored at the last response relapse-free disease assessment date.

- Transfusion conversion rate and transfusion maintenance rate

  Transfusion conversion rate and transfusion maintenance rate will only be defined for subjects in ASP2215 arm.

  For the purpose of defining transfusion conversion rate and transfusion maintenance rate, transfusion status (independent vs. dependent) at baseline period and post-baseline period are defined in the following for subjects who took at least one dose of study drug:

  Baseline transfusion status:

  - Baseline period is defined as the period from 28 days prior to the first dose to 28 days post first dose. For subjects who are on treatment <28 days, baseline period is from 28 days prior to the first dose until the end of treatment.
  - Subjects are classified as baseline transfusion independent if there is no RBC or platelet transfusions within the baseline period; otherwise, the subject is baseline transfusion dependent.

  Post-baseline transfusion status (only defined for ASP2215 arm):

  - Post-baseline period is defined as the period from 29 days post first dose until last dose.
  - For subjects who are on treatment >=84 days, subjects are classified post-baseline transfusion independent if there is one consecutive 56 days without any RBC or platelet transfusion within post-baseline period.
  - For subjects who are on treatment >28 days but <84 days, if there is no RBC or platelet transfusion within post-baseline period, post-baseline transfusion status is not evaluable.
  - For subjects who are on treatment <=28 days, post-baseline transfusion status is not evaluable.
  - Otherwise, the subject is considered post-baseline transfusion dependent.

  Both transfusion conversion rate and maintenance rate are defined for subjects who has evaluable post-baseline transfusion status.

  Transfusion conversion rate is defined as the number of subjects who were transfusion dependent at baseline period but become transfusion independent at post-baseline period divided by the total number of subjects who were transfusion dependent at baseline period.

  Transfusion maintenance rate is defined as the number of subjects who were transfusion independent at baseline period and still maintain transfusion independent at post-baseline period divided by the total number of subjects who were transfusion independent at baseline period.

As a sensitivity analysis, transfusion conversion rate and transfusion maintenance rate will be defined alternatively by considering all subjects who had not-evaluable post-baseline transfusion status as transfusion dependent.

- Transplantation rate

  Transplantation rate is defined as the percentage of subjects undergoing HSCT during the study period.

- Brief fatigue inventory (BFI)

  BFI was developed to assess the severity of fatigue and the impact of fatigue on daily functioning in patients with fatigue due to cancer and cancer treatment. The BFI short form has 9 items and a 24-hour recall.

  The BFI first asks whether the patient has been feeling unusually tired or fatigued (Y/N). Then three items ask patients to rate on a 0 (no fatigue) to 10 (as bad as you can imagine) scale the severity of their fatigue at its "worst", "usual", and "now" during normal waking. A composite fatigue severity score is calculated by averaging the ratings on these three items. All three severity items must have been completed for the mean to be calculated.

  The severity items are followed by six items that assess how much fatigue has interfered with different aspects of the patient's life during the past 24 hours. The interference items include general activity, mood, walking ability, normal work (includes both work outside the home and housework), relations with other people, and enjoyment of life. The interference items are measured on a 0 (does not interfere) to 10 (completely interferes) scale. BFI interference score is calculated as the mean of the six interference items. This mean can be used if more than 50%, or four of six, of the interference items have been completed on a given administration.

  A global BFI score is obtained by averaging all 9 items measured on the numeric rating scale. The global BFI score will be calculated only if at least 5 of the 9 items are answered.

  A higher score indicates a higher degree of fatigue.

6.1.2.3 Response Definition

Response to treatment will be defined per modified Cheson criteria [2003] as outlined below.

Response will be derived using myeloblast counts from locally evaluated bone marrow aspirate if it's adequate. In case of non-adequacy, myeloblast counts from locally evaluated bone marrow biopsy will be used. Centrally evaluated hematology results including ANC, platelet count and blast count in peripheral blood will be used in response derivation. Missing central hematology results will be imputed with local hematology results as collected on the eCRF.

Response will be derived for all post-baseline visits on or after 21 days from first dosing date.

- Complete Remission (CR)

  For subjects to be classified as being in CR at a post-baseline visit, they must have bone marrow regenerating normal hematopoietic cells and achieve a morphologic leukemia-free state and must have an absolute neutrophil count (ANC) ≥1 x 10^9/L and platelet count ≥ 100 x 10^9/L, and normal marrow differential with < 5% blasts, and they will be red blood cell (RBC) and platelet transfusion independent (defined as 1 weeks without RBC transfusion and 1 week without platelet transfusion). There must be no presence of Auer rods. There should be no evidence of extramedullary leukemia. The blast counts in peripheral blood must be ≤ 2%.

- Complete Remission with Incomplete Platelet Recovery (CRp)

  For subjects to be classified as being in CRp at a post-baseline visit, they must achieve CR except for incomplete platelet recovery (< 100 x 10^9/L).

- Complete Remission with Incomplete Hematological Recovery (CRi)

  For subjects to be classified as being in CRi at a post-baseline visit, they must fulfill all the criteria for CR except for incomplete hematological recovery with residual neutropenia < 1 x 10^9/L with or without complete platelet recovery. RBC and platelet transfusion independence is not required.

- Partial Remission (PR)

  For subjects to be classified as being in PR at a post-baseline visit, they must have bone marrow regenerating normal hematopoietic cells with evidence of peripheral recovery with no (or only a few regenerating) circulating blasts and with a decrease of at least 50% in the percentage of blasts in the bone marrow aspirate with the total marrow blasts between 5% and 25%. A value of less or equal than 5% blasts is also considered a PR if Auer rods are present. There should be no evidence of extramedullary leukemia.

- Not Evaluable (NE)/No Response (NR)

  In the situation where no bone marrow assessments are performed or myeloblast value is missing, blast value from peripheral blood is missing or <=2%, and extramedullary leukemia is missing or not present, the response will be classified as not evaluable (NE). In any case response cannot be categorized as CR, CRp, CRi, PR or NE, it will be categorized as NR

- Composite Complete Remission (CRc)

  For subjects to be classified as being in CRc at a post-baseline visit, they must either achieve CR, CRp or CRi at the visit.

- Complete Remission with Partial Hematologic Recovery (CRh)

  At a post baseline visit, subjects will be classified as CRh if they have marrow blasts <5%, partial hematologic recovery ANC >= 0.5 x 10^9/L and platelets >= 50 x 10^9/L, no evidence of extramedullary leukemia and cannot be classified as CR. The blast counts in peripheral blood must be ≤ 2%

- Complete Remission and Complete Remission with Partial Hematologic Recovery (CR/CRh)

  For subjects to be classified as being in CR/CRh at a post-baseline visit, they must either achieve CR or CRh at the visit.

- Relapse

  Relapse after CR, CRh, CRp or CRi is defined as a reappearance of leukemic blasts in the peripheral blood (> 2%) or ≥ 5% blasts in the bone marrow aspirate not attributable to any other cause or reappearance or new appearance of extramedullary leukemia.

  Relapse after PR is similarly defined with reappearance of significant numbers of peripheral blasts and an increase in the percentage of blasts in the bone marrow aspirate to > 25% not attributable to any other cause or reappearance or new appearance of extramedullary leukemia.

- Best Response

  Best response is defined as the best measured response to treatment for all visits (in the order of CR, CRp, CRi, PR, NR and NE) post-baseline. Subjects who achieve the best responses of CR, CRp, CRi, or PR will be classified as responders. Subjects who do not achieve at least a best response of PR will be classified as non-responders.

6.1.3 Exploratory Endpoints

- Pharmacogenomics
- FLT3 gene mutation status
  - mutation types (FLT3-ITD mutation and D835/I836 TKD mutation) and frequency
  - relationship to efficacy and safety
- Exploratory (predictive) biomarkers of ASP2215 activity
- Resource utilization, including hospitalization, blood transfusion, antibiotic intravenous infusions, medication for AEs and opioid usage (antibiotic intravenous infusions and opioid usage are specified in Appendix 4).
- Functional Assessment of Chronic Illness Therapy-Dyspnea-Short Forms (FACIT-Dys-SF)

  The FACIT-Dys-SF was developed to assess dyspnea severity and related functional limitations. The FACIT-Dys SF comprises of two parts. Part I asks patients about ten common daily tasks (e.g., dressing without help, walking up 20 stairs). Patients rate either the amount of dyspnea (0 “No shortness of breath” to 3 “Severely short of breath”), or that they did not complete the task, which is marked “due to shortness of breath” or other reason. Part II lists the same ten activities and inquiries about the difficulty completing them (0 “No difficulty” to 3 “Great difficulty”). It has a 7-day recall period and 20 items. Higher scores indicate more dyspnea.

  Part I item scores are summed and adjusted for missing values to produce a raw score, FACIT-Dys (higher scores indicate worse dyspnea). Part II is scored in a similar manner

to produce a raw "functional limitations" score, FACIT-Dys FL (higher scores indicate more limitation). The FACIT-Dys SF and the scoring guide is provided in Appendix 6.

If there are missing items, subscale scores can be prorated. This is done by multiplying the sum of the subscale by the number of items in the subscale, then dividing by the number of items actually answered. This can be done by using the formula below:

$$\text{Prorated subscale score} = \frac{[\text{Sum of item scores}] * [\text{N of items in subscale}]}{[\text{N of times answered}]}$$

When there are missing data, prorating by subscale in this way is acceptable as long as more than 50% of the items are answered (i.e., a minimum of 4 to 7 items, 4 of 6 items, etc).

- Functional Assessment of Cancer Therapy-Leukemia (FACT-Leu)

  The FACT-Leu is a 44-item questionnaire designed to measure health-related quality of life (HRQoL) and leukemia-specific symptoms. The questionnaire includes a global score and 5 subscales including physical well-being (PWB), social/family well-being (SWB), emotional well-being (EWB), functional well-being (FWB), and a leukemia subscale (LeuS). In addition to the 5 subscales, the FACT-Leu can be used to calculate the Functional Assessment of Cancer Therapy-General (FACT-G) score, the FACT-Leu Trial Outcome Index (FACT-Leu TOI), and the FACT-Leu total score which is comprised of all 44 items included in the questionnaire. The FACT-Leu contains some of the most common patient reported signs, symptoms, and impacts of AML. The FACT-Leu has a 7-day recall period. A higher score indicates a better quality of life.

  Each item is rated on a five-point Likert scale ranging from 0 = "not at all" to 4 = "very much". The subscale scores and the summary scores are calculated using the Manual of Functional Assessment of Chronic Illness Therapy (FACIT) Measurement System (Version 4) by David Cella. The scoring guide is presented Appendix 7.

  If there are missing items, subscale scores can be prorated. This is done by multiplying the sum of the subscale by the number of items in the subscale, then dividing by the number of items actually answered. This can be done by using the formula below:

$$\text{Prorated subscale score} = \frac{[\text{Sum of item scores}] * [\text{N of items in subscale}]}{[\text{N of times answered}]}$$

  When there are missing data, prorating by subscale in this way is acceptable as long as more than 50% of the items are answered (i.e., a minimum of 4 to 7 items, 4 of 6 items, etc).

  The FACT-Leu total score, FACT-G total score and FACT-Leukemia Trial Outcome Index (TOI) will be calculated only if all subscales that are summed up are not missing.

- Dizziness and mouth sore items which utilize a 5-point scale (0 to 4). A lower score indicates less experience of the symptom.

- EuroQol Group-5 Dimension-5 Level Instrument (EQ-5D-5L)

  The EQ-5D-5L is a common measure of patients' HRQoL. The EQ-5D-5L consists of the EuroQol Group-5 Dimension descriptive system and the EuroQol Group VAS.

  The EuroQol Group-5 Dimension descriptive system consists of 5 items assessing 5 key dimensions of health: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Each dimension has 5 levels: no problems, slight problems, moderate problems, severe problems and extreme problems.

  The VAS records the respondent's self-rated health status on a graduated (0 - 100) scale, where the endpoints are labeled 'worse imaginable health state' (0) and 'best imaginable health state' (100) with higher scores indicating higher HRQoL.

6.1.4 Other Efficacy Variables

- Best response rate – Defined as the number of subjects who achieve the best response of CR, CRp, CRi or PR each divided by the number of subjects in the analysis population.
- Response rate – Defined as the number of subjects who achieve the best response of CRc or PR divided by the number of subjects in the analysis population.
- CRh rate – Defined as the number of subjects who achieve CRh at any of the postbaseline visits and do not have best response of CR divided by the number of subjects in the analysis population.
- Time to remission

  Time to remission includes time to CRc (TTCRc), time to CR (TTCR), time to response (TTR), time to CR/CRh (TTCRCRh).

  TTCRc is defined as the time from the date of randomization until the date of first CRc (first CRc disease assessment date – randomization date +1). TTCRc will only be evaluated for subjects who achieved CRc.

  TTCR is defined similarly as time to CRc for subjects who achieve CR.

  TTR is defined as the time from the date of randomization until the date of either first response (CRc or PR), (first CRc or PR disease assessment date – randomization date +1). TTR will only be evaluated for subjects who achieved CRc or PR.

  TTCRCRh is defined similarly as time to CRc for subjects who achieve CR/CRh.

6.2 Safety Variables

Safety will be assessed by evaluation of the following variables:

- Treatment-emergent adverse events (TEAEs; frequency, severity, CTCAE grade, seriousness, and relationship to study drug)
  - TEAE is defined as an adverse event observed after starting administration of the study treatment (ASP2215 or salvage chemotherapy). If the adverse event occurs on Day 1 and the onset check box is marked "Onset after first dose of study drug" or the

onset check box is left blank, then the adverse event will be considered treatment emergent. If the adverse event occurs on Day 1 and the onset check box is marked "Onset before first dose of study drug", then the adverse event will not be considered treatment emergent. If a subject experiences an event both during the pre-investigational period and during the investigational period, the event will be considered as TEAE only if it has worsened in severity (i.e., it is reported with a new start date). All adverse events collected that begin within 30 days after taking the last dose of study drug will also be counted as TEAE, except for subjects that undergo HSCT without leaving the study and plan to resume ASP2215 treatment after HSCT. For these subjects, TEAE is defined as adverse events observed after starting administration of the study treatment until the last dose before pre-HSCT visit plus 30 days, and adverse events that begin after resumption of ASP2215 and within 30 days after the last dose of ASP2215 will also be counted as TEAE. Any AEs with onset dates completely missing will be considered TEAEs in summaries. AEs with partially missing onset dates will be assumed TEAEs unless the available portion of the date indicates that the onset was strictly before start of study medication or 30 days after the last study treatment. Missing or partial AE onset date will be imputed per Section 7.11.1.

  - A drug-related TEAE is defined as any TEAE with at least possible relationship (possibly or probably related) to study treatment as assessed by the investigator or with missing assessment of the causal relationship.

- Serious adverse events (SAEs) include adverse events that are flagged as serious by the investigator on eCRF, or upgraded by the Sponsor based on review of the Sponsor's list of Always Serious term.
- Adverse events of special safety interest (AESI) are defined in the Safety Review Plan for ASP2215 (as specified in Appendix 3).
- Clinical laboratory variables (hematology, biochemistry including liver function test and thyroid function test, coagulation, and urinalysis)
- Vital signs (systolic and diastolic blood pressure, pulse rate, and body temperature)
- 12-lead electrocardiogram (ECG)
- ECOG performance scores

6.3 Pharmacokinetics Variables

Plasma concentration data of ASP2215 will be used in pharmacokinetic analysis.

6.4 Other Variables

- Duration of exposure

  Duration of exposure to each component of a study drug will be calculated in days, using the following formula:

  Last date of exposure – First dose date + 1 – (on-study HSCT period for subjects undergo on-study HSCT)

  For ASP2215, last date of exposure = last dose date of ASP2215

  For chemotherapy, last date of exposure = (initial dose date of the last cycle + 28 – 1) or death date if death occur within last cycle

  When the start or stop date is missing, then the exposure will be treated as missing.

  Duration of exposure for a chemotherapy regimen, i.e., MEC and FLAG, will be the maximum (duration of exposure of all treatment components)

  When the last date of exposure is beyond cutoff date, the cutoff date will be used as the last date of exposure.

- Number of dosing days

  Number of days with non-zero dosing.

- Number of cycles

  Number of cycles for each component of study drug refers to total number of cycles with non-zero dosing in the cycle.

  Number of cycles for a chemotherapy regimen refers to total number of cycles with non-zero dosing of any treatment component in the cycle.

- Cumulative dose (mg or ug)

  Cumulative dose will be calculated using the following formula for each type of study drug:

  - ASP2215: Sum of ([stop date – start date + 1] * dose captured on dosing CRF) across all records
  - LoDAC: Sum of (number of dose taken * 20mg) across all records and all cycles. If number of dose take is “unknown”, it will be treated as missing.
  - Each component of MEC and FLAG: Sum of (actual delivered dose captured on dosing CRF) across the days when study drug was administered. If answer to “entire infusion administered” is No, actual delivered dose will be treated as missing.

- Average daily dose (mg/days or ug/days)

  $$\frac{\text{Cumulative dose (mg or ug)}}{\text{Number of dosing days (days)}}$$

- Dose intensity

Dose intensity will be calculated in mg/day for ASP2215 as:
Cumulative dose (mg)

Duration of exposure (days)

Dose intensity will be calculated in mg/cycle or ug/cycle for chemotherapy as:
Cumulative dose (mg or ug)
--- x 28 (days/cycle)
Duration of exposure (days)

- Relative dose intensity

  Relative dose intensity for ASP2215 will be calculated as:

  Dose intensity (mg/day)

  --- x 100%

  120 (mg/day)

  Relative dose intensity for each component of chemotherapy will be calculated as:
  Dose intensity (mg/cycle or ug/cycle)
  --- x 100%
  Planned dose intensity (mg/cycle or ug/cycle)

  Where planned dose intensity (mg/cycle) will be calculated based on protocol specified dose:

  - LoDAC: 40 mg/day * 10 days/cycle
  - MEC

    Mitoxantrone: 6 mg/m^2/day * BSA * 5 days/cycle

    Etoposide: 100 mg/m^2/day * BSA * 5 days/cycle

    Cytarabine: 1000 mg/m^2/day * BSA * 5 days/cycle
  - FLAG

    G-CSF: 300 ug/m^2/day * BSA * 5 days/cycle

    Fludarabine: 30 mg/m^2/day * BSA * 5 days/cycle

    Cytarabine: 2000 mg/m^2/day * BSA * 5 days/cycle

  Average of BSA for all dosed cycles will be used in the calculation of planned dose intensity.
- Relative dose intensity for a chemotherapy regimen will be the average of relative dose intensity of all treatment components.
- Duration of AML

  Duration of AML will be calculated in days using the following formula:

  (Randomization date – date of initial diagnosis of AML) + 1

  Partial date of initial diagnosis of AML will be imputed per Section 7.11.1.

- Previous and concomitant medication

  Previous medication is defined as medication with at least one dose taken before the date of the first dose of study drug (exclusive).

  Concomitant medication is defined as medication with at least one dose taken between the date of first dose (inclusive) and the date of last exposure (inclusive) of study drug.

  For subjects that undergo HSCT without leaving the study and plan to resume ASP2215 treatment after HSCT, concomitant medication is defined as medication with at least one dose taken between the date of first dose (inclusive) and the date of last exposure (inclusive) before pre-HSCT visit, or between resumption of ASP2215 (inclusive) and the date of last exposure (inclusive).
- Prior Use of FLT3 inhibitor

  Prior Use of FLT3 inhibitor is defined as "Yes" if subjects received prior AML therapy of Midostaurin, Sorafenib or Quizartinib.
- Previous and concomitant transfusion

  Previous transfusion is defined as transfusion received before the date of first dose of study drug, i.e., transfusion completed before the date of first dose (exclusive).

  Concomitant transfusion is defined as transfusion received between the date of first dose (inclusive) and the date of last exposure (inclusive) of study drug.

  For subjects that undergo HSCT without leaving the study and plan to resume ASP2215 treatment after HSCT, concomitant transfusion is defined as transfusion received between the date of first dose (inclusive) and the date of last exposure (inclusive) before pre-HSCT visit, or between resumption of ASP2215 (inclusive) and the date of last exposure (inclusive).

7 STATISTICAL METHODOLOGY

7.1 General Considerations

Since the study enrollment was terminated on 12Mar2021 due to favorable outcome of OS at the interim analysis by IDMC, and the formal interim analysis was conducted as final primary analysis based on data cutoff on 30Jun2020, the pre-specified hierarchal testing procedure to maintain the overall 2-sided Type I error rate at the 0.05 will not be applied for end-of-study analysis.

For continuous variables, descriptive statistics will include the number of subjects (n), mean, standard deviation, median, minimum and maximum. When needed, the use of other percentiles (e.g., 10%, 25%, 75% and 90%) will be specified in the relevant section. In addition, for plasma concentrations, the coefficient of variation (CV), the geometric mean, and the geometric CV will also be calculated. Frequencies and percentages will be displayed for categorical data. Percentages by categories will be based on the number of subjects with no missing data, i.e., will add up to 100%.

Summaries based on ITT and FAS (e.g., disposition, baseline and efficacy data) will be presented by randomized treatment group, unless specifically stated otherwise. Safety

analysis and other summaries based on SAF will be presented by actual treatment received. Pharmacokinetic summaries based on PKAS will be presented by actual treatment received. For subjects with dose increase/decrease, actual treatment refers to the initial dose received before dose change.

All statistical comparisons will be made using two-sided tests at the α=0.05 significance level unless specifically stated otherwise. All null hypotheses will be of no treatment difference, all alternative hypotheses will be two-sided, unless specifically stated otherwise.

All data processing, summarization, and analyses will be performed using SAS® Version 9.4 or higher on Red Hat Enterprise Linux. Specifications for table, figure, and data listing formats can be found in the TLF specifications for this study.

Baseline is defined as the last available measurement prior to the first dose of study drug. Unless otherwise specified, all summaries will be presented by treatment groups.

For the definition of subgroups of interest please refer to Section 7.8.

No patients entered the crossover extension. So, only the number of subjects with informed consent for crossover extension was added in Section 7.2.1. No additional analysis for crossover extension is planned.

7.2 Study Population

7.2.1 Disposition of Subjects

The following subject data will be presented:

[Subjects with informed consent obtained]

- Number and percentage of subjects with informed consent, re-screening, discontinued before randomization, randomized (overall only);

[All randomized subjects]

- Number and percentage of subjects randomized in each analysis set, by treatment group and overall;
- Number and percentage of subjects completed and discontinued treatment, by primary reason for treatment discontinuation and by treatment group;
- Number and percentage of subjects completed the 30-day follow-up evaluation, by 30-day follow-up status and by treatment group;
- Number and percentage of subjects completed the long-term follow-up evaluation, by long-term follow-up status and by treatment group;
- Number of subjects with informed consent for crossover extension.

7.2.2 Protocol Deviations

Protocol deviations as defined in the study protocol (Section 8.1.6 Protocol Deviations) will be assessed for all randomized subjects. The number and percentage of subjects meeting any criteria will be summarized for each criterion and overall, by treatment group and total as well as by study site. Subjects deviating from a criterion more than once will be counted once

for the corresponding criterion. Any subjects who have more than one protocol deviation will be counted once in the overall summary. A data listing will be provided by site and subject.

The protocol deviation criteria will be uniquely identified in the summary table and listing. The unique identifiers will be as follows:

PD1 - Entered into the study even though they did not satisfy entry criteria,

PD2 - Developed withdrawal criteria during the study and was not withdrawn,

PD3 - Received wrong treatment or incorrect dose,

PD4 - Received excluded concomitant treatment.

7.2.3 Demographic and Other Baseline Characteristics

Demographic and other baseline characteristics will be summarized by descriptive statistics.

Number and percentage of subjects allocated to treatment in each country and site will be presented by treatment group.

Descriptive statistics for age, weight, body surface area (BSA) and height at study entry will be presented. Frequency tabulations for sex, ethnicity, age group (defined in Section 7.8), race, region, baseline central FLT3 status, prior use of FLT3 inhibitor, cytogenetic risk status and baseline ECOG will be presented. This will be done for the SAF, ITT and FAS by treatment group.

Frequency tabulations for AML disease history including AML subtype as classified by World Health Organization (WHO) classification and French-American-British (FAB) classification, risk status, antecedent hematological disorder, central nervous system leukemia, FLT3 mutation status, FLT3-ITD mutation status, FLT3 point mutation status will be presented by treatment group for the SAF, ITT and FAS.

Medical history other than AML and conditions existing at Baseline will be coded in MedDRA and summarized by System Organ Class (SOC) and Preferred Term (PT) as well as by PT alone, by treatment group for the SAF, ITT and FAS. Baseline conditions are defined as those ongoing at the time of informed consent or arise following the time of informed consent and before the first dose of study drug. For ongoing medical conditions, Common Terminology Criteria for Adverse Events (CTCAE) grade will be provided in listing.

Frequency tabulations for prior transplant including number of prior transplant, graft type, donor relatedness, match type and outcome of transplant will be presented by treatment group for the SAF, ITT and FAS.

Results from lumbar puncture, baseline extramedullary leukemia and MUGA scan, if performed, will be provided in listing.

7.2.4 Previous and Concomitant Medications

Previous medications are coded with World Health Organization – Drug Dictionary (WHO-DD), and will be summarized by therapeutic subgroup (Anatomical Therapeutic

Chemical Classification [ATC] 2nd level) and chemical subgroup (ATC 4th level) and preferred WHO name by treatment group for the SAF, ITT and FAS.

As with previous medication, concomitant medication will be summarized for each treatment group by therapeutic subgroup (ATC 2nd level) and chemical subgroup (ATC 4th level) and preferred WHO name. Subjects taking the same medication multiple times will be counted once per medication and investigational period. A medication which can be classified into several chemical and/or therapeutic subgroups is presented in all chemical and therapeutic subgroups.

Concomitant medication will also be summarized by treatment group and preferred WHO name, and presented in decreasing order of frequency based on the total number of subjects took each medication.

7.2.5 Previous and Concomitant Transfusions

Frequency tabulations of subjects received transfusions and blood product will be presented for previous transfusion and concomitant transfusion by treatment group for SAF, ITT and FAS. Descriptive statistics will be presented for number of transfusion unit received per subject for each type of blood product.

7.2.6 Prior AML Chemotherapy

Frequency tabulations of subjects with prior AML chemotherapy, response to first line therapy, regimen, type of treatment, prior use of FLT3 inhibitor, and best response to prior AML therapy will be presented by treatment group for SAF, ITT and FAS. Descriptive statistics will be presented for duration of response to prior AML therapy.

7.2.7 Non-Medication Therapy

Frequency tabulations of subjects with non-medication therapy and reason for use will be presented by treatment group for SAF, ITT and FAS. Number of non-medication therapy received per subject will be summarized using descriptive statistics.

7.2.8 Clinical/Diagnostic Procedures

Frequency tabulations of subjects with clinical/diagnostic procedures and reason for use will be presented by treatment group for SAF, ITT and FAS. Number of clinical/diagnostic procedures received per subject will be summarized using descriptive statistics.

7.2.9 Subsequent AML Therapy

Frequency tabulations of subjects with subsequent AML therapy, regimens, relapse prior to subsequent AML therapy, reason of starting subsequent AML therapy, and response to subsequent AML treatment will be presented by treatment group for SAF, ITT and FAS. Descriptive statistics will be presented for duration of subsequent AML therapy.

7.3 Study Drugs

7.3.1 Exposure

The following information on drug exposure will be presented for each treatment regimen for the SAF:

- Descriptive statistics for cumulative amount of the drug subject was exposed to, number of dosing days, number of dosing cycles, average daily dose, dose intensity, and relative dose intensity; and
- Number and percent of subject with dose increases, decreases or interruptions.

Duration of exposure will be summarized in two ways.

- Descriptive statistics will be presented by treatment regimen.
- Exposure time will be categorized according to the following categories by treatment regimen:
  - less than or equal to 5 days
  - at least 6 days, less than 28 days
  - at least 28 days, less than 84 days
  - at least 84 days, less than 168 days
  - 168 days or more
  - Unknown.

Counts and percentages of subjects in each of these categories will be summarized for each treatment regimen for the SAF.

Listing of subjects with dose reduction and dose escalation will also be provided.

7.4 Analysis of Efficacy

7.4.1 Analysis of Primary Endpoint of OS

The primary efficacy endpoint of OS will be analyzed on ITT for primary analysis. In order to compare the OS between ASP2215 and the salvage chemotherapy, the null hypothesis will be constructed:

- H_{01}: OS is equal between ASP2215 and salvage chemotherapy

The accompanying alternative hypothesis is:

- H_{11}: OS is different between ASP2215 and salvage chemotherapy

The two-sided p-value for the above hypothesis test will be calculated using the stratified log-rank test (primary test) with strata (per IRT) to control for response to first-line AML therapy and preselected salvage chemotherapy.

The SAS code to implement stratified log-rank will be similar to that shown below:

```
PROC LIFETEST;

     TIME time * status (1);
```

```
      STRATA stratification variables/group=treatment;
RUN;
```

The hazard ratio of the treatment effect along with 95% confidence interval will be calculated by the stratified Cox proportional hazard model. The same stratification factors will be applied to both the stratified log-rank test and the stratified Cox proportional hazard model. The same Cox proportional hazard model will be included as a sensitivity analysis for testing H_{01}.

The SAS code to calculate the hazard ratio of the treatment effect along with 95% confidence interval will be similar to that shown below:

```
PROC PHREG;
      MODEL time * status (1) = treatment/RL;
      STRATA stratification variables;
RUN;
```

Stratification variables will include response to first-line AML therapy and preselected salvage chemotherapy.

Kaplan-Meier survival plots will be used to describe the OS in each treatment group. Median OS and 95% confidence interval, survival rates at 6, 12, 24 and 36 months and 95% confidence interval will be estimated from the Kaplan-Meier curve using the SAS code similar to that shown below:

```
PROC LIFETEST DATA=XX TIMELIST=(XX XX XX);
      TIME time * status (1);
      STRATA treatment;
RUN;
```

Where TIMELIST option contains timepoint of interest to estimate survival rate.

The sensitivity analyses for OS include:

- The same analysis as primary analysis but on FAS
- Stratified Cox proportional hazard model with strata to control for response to first-line AML therapy and preselected salvage chemotherapy on ITT (as described in usage of Cox proportional hazard model)
- The same analysis as primary analysis but with OS censoring at HSCT as defined in Section 6.1.1
- The same analysis as primary analysis but with OS censoring at first new anti-leukemia therapy as defined in Section 6.1.1
- To assess potential COVID-19 impact, below sensitivity analyses maybe conducted:
    - Same as OS primary analysis, except that subjects with COVID-19 death will be censored at the death date.

- Same as OS primary analysis, except excluding the subjects with COVID-19 death from the analysis.

7.4.2 Analysis of Secondary Endpoints

7.4.2.1 Key Secondary Efficacy Analysis

The key secondary efficacy endpoint of EFS will be analyzed in the same manner as OS on ITT. In order to compare the EFS between ASP2215 and the salvage chemotherapy, the null hypothesis will be constructed:

- H_{02}: EFS is equal between ASP2215 and salvage chemotherapy

The accompanying alternative hypothesis is:

- H_{12}: EFS is different between ASP2215 and salvage chemotherapy

To maintain the overall 2-sided Type I error rate at the 0.05, the key secondary efficacy endpoint of EFS will only be tested at the interim and final analyses. The pre-specified hierarchal testing procedure and significance level as given in Table 1.

The two-sided p-value for the above hypothesis test will be calculated using stratified

log-rank test with strata (per IRT) to control for response to first-line AML therapy and preselected salvage chemotherapy on ITT.

The hazard ratio of the treatment effect along with 95% confidence interval will be calculated using stratified Cox proportional hazard model. The same Cox proportional hazard model will be included as a sensitivity analysis for testing H_{02}.

Kaplan-Meier survival plots will be used to describe the EFS in each treatment group. Median EFS and 95% confidence interval, EFS rates at 6, 12, 24 and 36 months and 95% confidence interval will be estimated from the Kaplan-Meier curve. The number and percentage of EFS event (relapse, off-treatment relapse, new AML therapy, death or treatment failure) and censoring will be summarized for each treatment group and for each chemotherapy regimen.

The sensitivity analyses for the key secondary efficacy endpoints of EFS include:

- The same analysis as primary analysis but on FAS
- Stratified Cox proportional hazard model with strata to control for response to first-line AML therapy and preselected salvage chemotherapy
- The same analysis as primary analysis but with the date of the first new anti-leukemia therapy after the end of study treatment or the last treatment evaluation date (when the date of new anti-leukemia is not available) will be used as the event date of treatment failure as defined in Section 6.1.2.1
- The same analysis as primary analysis but with “Lost to follow up” also considered as an EFS event as defined in Section 6.1.2.1.

- The same analysis as primary but consider subjects who discontinue the treatment and achieve best response of CR, CRp, CRi or PR and experience relapse (including relapse after PR) will be defined as relapse, and consider the subjects who discontinued the treatment and only achieved best response of NR as treatment failure as defined in Section 6.1.2.1.
- The same analysis as primary but consider the subjects who discontinued the treatment with post-treatment disease assessment and achieved best response of PR, NR or NE as treatment failure as defined in Section 6.1.2.1.
- To assess potential COVID-19 impact, below sensitivity analyses maybe conducted:
  - Same as EFS primary analysis, except excluding COVID-19 death form EFS event.
  - Same as EFS primary analysis, except excluding the subjects with COVID-19 death from the analysis.
- EFS not using the long-term follow-up data as defined in Section 6.1.2.1.

The key secondary efficacy endpoint of CR rate will be analyzed using the Cochran-Mantel-Haenszel (CMH) test to control for response to first-line AML therapy and preselected salvage chemotherapy (Per IRT) on ITT. In order to compare the CR rate between ASP2215 and the salvage chemotherapy, the null hypothesis will be constructed:

- H_{03}: CR rate is equal between ASP2215 and salvage chemotherapy

The accompanying alternative hypothesis is:

- H_{13}: CR rate is different between ASP2215 and salvage chemotherapy

To maintain the overall 2-sided Type I error rate at the 0.05, the key secondary efficacy endpoint of CR rate will only be analyzed at the interim and final analyses follow the pre-specified hierarchal testing procedure and significance level as given in Table 1.

The treatment difference along with 95% confidence interval will be calculated using Mantel-Haenszel estimate.

The SAS code to implement the above test will be similar to that shown below:

```
PROC FREQ;
      TABLES stratification variables * response * treatment/cmh;
RUN;
```

CR rate along with the two-sided 95% exact confidence interval based on binomial distribution will be calculated for each treatment group using SAS code similar to that shown below:

```
PROC FREQ;
      TABLES outcome/BIONOMIAL (EXACT);
      BY treatment;
```

```
RUN;
```

The sensitivity analysis for CR rate will be performed as following:

- The same analysis as primary analysis but on FAS
- The same analysis as primary analysis but only including subjects who received at least one dose of study treatment
- The same analysis as primary analysis but only including subjects who had at least one post baseline bone marrow assessment
- Apply the same analysis method used for primary analysis to evaluate CR rate prior to HSCT (defined as the number of subjects who achieve CR prior to HSCT divided by the number of subjects in the analysis population)
- Un-stratified Fisher's exact test on ITT

7.4.2.2 Other Secondary Efficacy Analysis

LFS and duration of remission (DCRc, DCR, DCRCRh, and DR) will be analyzed in the same manner as OS for subjects who achieved remission on ITT using the stratified Cox proportional hazard model with strata to control for response to first-line AML therapy and preselected salvage chemotherapy. The hazard ratio of the treatment effect along with 95% confidence interval will be calculated. The survival curve and the median will be estimated using the Kaplan-Meier method and will be reported along with the corresponding 95% confidence interval.

CRc rate, CR/CRh rate and transplantation rate will be analyzed in the same manner as CR rate using CMH test to control for response to first-line AML therapy and preselected salvage chemotherapy on ITT. The number and percentage of subject with CRc, CR/CRh and transplantations will be summarized for each treatment group along with two-sided exact 95% confidence interval based on binomial distribution. The treatment difference along with 95% confidence interval will also be calculated. Similarly, CRc rate prior to HSCT and CR/CRh rate prior to HSCT will be defined and analyzed in the same manner as CR rate prior to HSCT.

Transfusion status (independent vs. dependent) at baseline period and post-baseline period will be summarized in a two-by-two contingency table.

Transfusion conversion rate and transfusion maintenance rate will be calculated based on the transfusion status. 95% confidence interval will be presented for both transfusion conversion and maintenance rates.

BFI global fatigue score will be summarized using mean, standard deviation, minimum, maximum and median by treatment group at each visit on ITT. Additionally, a within subject change will be calculated as the post-baseline measurement minus the baseline measurement and summarized in the same way. Change from baseline BFI score will be analyzed using Analysis of Covariance (ANCOVA) including treatment as a fixed factor, baseline BFI score, response to first-line AML therapy and preselected salvage chemotherapy as covariates.

7.4.3 Analysis of Exploratory Endpoints

Central FLT3 mutation status at screening, including subgroups of FLT3-ITD mutation and D835/I836 TKD mutations, will be summarized by the number and percentage of subjects in each category by treatment group. An exploratory analysis of FLT3 mutation type and selected clinical efficacy endpoints will be conducted (see Section 7.8). Incidence of resource utilization including hospitalization, blood transfusion, antibiotic intravenous infusions, medication for AEs and opioid medication will be summarized by treatment group on ITT. The difference between treatment groups will be tested using CMH test while controlling for response to first-line AML therapy and preselected salvage chemotherapy. Duration of hospital stays, blood transfusions, antibiotic intravenous infusions, medication for AEs and opioid medication will be summarized by treatment group using descriptive statistics (mean, standard deviation, minimum, maximum and median) on ITT. The difference between treatment groups will be tested with ANOVA while controlling for response to first-line AML therapy and preselected salvage chemotherapy.

FACIT-Dys-SF domain scores will be summarized by treatment group at each visit using descriptive statistics (mean, standard deviation, minimum, maximum and median) on ITT. Additionally, a within-subject change will be calculated as the post-baseline score minus the baseline score and summarized in the same way. ANCOVA model will be used to analyze the change in the FACIT-Dys-SF domain scores from baseline to post-baseline visits including treatment as a fixed factor, baseline score, response to first-line AML therapy and preselected salvage chemotherapy as covariates.

FACT-Leu global score and domain scores will be summarized by treatment group at each visit using descriptive statistics (mean, standard deviation, minimum, maximum and median) on ITT. Additionally, a within-subject change will be calculated as the post-baseline score minus the baseline score and summarized in the same way. ANCOVA model will be used to evaluate change from baseline to post-baseline visits for the global and domain scores of the FACT-Leu including treatment as a fixed factor, baseline score, response to first-line AML therapy and preselected salvage chemotherapy as covariates. The same analytic approach will be used for the dizziness and mouth sore items. EQ-5D-5L VAS will be summarized by treatment group at each visit using descriptive statistics (mean, standard deviation, minimum, maximum and median) on ITT. Additionally, a within-subject change will be calculated as the post-baseline score minus the baseline score and summarized in the same way. ANCOVA model will be used to evaluate change from baseline to post-baseline visits for the EQ-5D-5L VAS including treatment as a fixed factor, baseline score, response to first-line AML therapy and preselected salvage chemotherapy as covariates. Shift table showing shift in each dimension score from baseline to each post-baseline visit will be provided for the 5 dimensions (mobility, self-care, usual activities, pain/discomfort and anxiety/depression).

7.4.4 Analysis of Other Variables

Best Response and Best Response prior to HSCT will be summarized by treatment group and by chemotherapy regimen on ITT. The number and percentage of subjects in each category will be presented.

TTCRc, TTCR, TTCRCRh, TTR will be summarized by treatment group using descriptive statistics (mean, standard deviation, minimum, maximum and median) for subjects who achieved remission on ITT.

7.5 Analysis of Safety

All analysis of safety will be presented by treatment group for SAF, unless specified otherwise.

7.5.1 Adverse Events

All adverse event (AE) recorded on treatment including within 30 days from the last study treatment will be summarized.

Summaries and listings of SAEs and Serious TEAEs include SAEs upgraded by the sponsor based on review of the Sponsor's list of Always Serious terms if any upgrade was done.

The coding dictionary for this study will be MedDRA. It will be used to summarize AEs by SOC and PT. AEs will be graded using National Cancer Institute's Common Terminology Criteria for AEs (NCI-CTCAE).

An overview table to report the number and percentage of subjects and an overview table to report number of events and events adjusted by patient year from drug exposure for each treatment group will include the following details:

- TEAEs
- Drug related TEAEs
- Serious TEAEs
- Drug-related serious TEAEs
- TEAEs leading to death
- Drug-related TEAEs leading to death
- TEAEs leading to withdrawal of treatment
- Drug related TEAEs leading to withdrawal of treatment
- Grade 3 or higher TEAEs
- Drug related Grade 3 or higher TEAEs
- Any deaths
- TEAE occurred within the first 30 days of study drug
- Drug-related TEAE occurred within the first 30 days of study drug
- TEAE leading to study drug reduction
- Drug-related TEAE leading to study drug reduction
- TEAE leading to study drug interruption
- Drug-related TEAE leading to study drug interruption

In addition, subgroup analysis by preselected salvage chemotherapy (high intensity or low intensity) will be conducted.

The number and percentage of subjects with TEAEs and the number of events and events adjusted by patient year from drug exposure, as classified by SOC and PT will be summarized for each treatment group. Summaries will be provided for:

- TEAEs
- Drug related TEAEs
- Serious TEAEs
- Drug related serious TEAEs
- TEAEs leading to death
- Drug related TEAEs leading to death
- TEAEs leading to withdrawal of treatment
- Drug related TEAEs leading to withdrawal of treatment
- Grade 3 or higher TEAE
- Drug-related Grade 3 or higher TEAE
- TEAE occurred within the first 30 days of study drug
- Drug-related TEAE occurred within the first 30 days of study drug
- TEAE leading to study drug reduction
- Drug-related TEAE leading to study drug reduction
- TEAE leading to study drug interruption
- Drug-related TEAE leading to study drug interruption

In addition, subgroup analysis by preselected salvage chemotherapy (high intensity or low intensity) will be conducted.

The number and percentage of subjects with TEAEs, as classified by SOC and PT, will also be summarized for each treatment group for the following:

- TEAEs excluding serious adverse events that equal to or exceed a threshold of 5% in any treatment group
- Common TEAEs that equal to or exceed a threshold of 10% in any treatment group
- Drug-related common TEAEs that equal to or exceed a threshold of 10% in any treatment group

The number and percentage of subjects with TEAEs, as classified by PT only, will be summarized for each treatment group for the followings:

- TEAEs
- Drug related TEAEs

In addition, subgroup analysis by preselected salvage chemotherapy (high intensity or low intensity) will be conducted.

The number and percentage of subjects with TEAE of special safety interest (AESI), as classified by AESI category and PT will be summarized for each treatment group for the followings:

- TEAEs with special safety interest
- Grade 3 or higher TEAEs with special safety interest
- Drug-related Grade 3 or higher TEAEs with special safety interest

In addition, subgroup analysis by preselected salvage chemotherapy (high intensity or low intensity) will be conducted.

AE summary tables will include subject counts as opposed to AE counts. If a subject experiences more than one episode of a particular AE, that subject will be counted only once for that event. If a subject has more than one AE that code to the same preferred term, the subject will be counted only once for that preferred term. Similarly, if a subject has more than one AE within a body system, the subject will be counted only once in that body system.

The number and percentage of subjects with TEAEs, as classified by SOC and PT will also be summarized by NCI-CTCAE severity grade and by relationship to study drug. In the subject count, if a subject has multiple TEAEs with the same SOC or PT, but with differing severity grade or relationship, then the subject will be counted only once with the worst severity grade and highest degree of relationship, however, if any of the severity grade or relationship values are missing then the subject will be counted only once with missing severity grade or relationship. Drug related TEAEs will be presented in a similar way by severity grade only. Serious TEAE will be presented in a similar way by relationship to study drug.

All AEs, deaths, SAEs and withdrawals due to adverse events will be displayed in listings.

7.5.2 Clinical Laboratory Evaluation

The baseline visit is the last measurement taken prior to initial study drug administration.

Quantitative clinical laboratory variables, i.e., hematology, biochemistry, coagulation will be summarized using mean, standard deviation, minimum, maximum and median for each treatment group at each visit. Additionally, a within-subject change will be calculated as the post-baseline measurement minus the baseline measurement and summarized in the same way. Each laboratory result will be classified as low (L), normal (N), or high (H) at each visit according to the laboratory supplied reference ranges.

The number and percentage of subjects below and above reference range will be summarized for each treatment group at each visit.

For hematology and biochemistry two types of shift tables will be presented:

- Shift tables of reference range changes from baseline to each treatment visit as well as worst finding during the treatment period (low, normal, high), and
- Summary shifts of reference range changes from baseline to each treatment visit as well as worst finding during the treatment period (shift from normal or high to low, shift from normal or low to high, categorized increase [shift from low to normal, low to high or from normal to high], categorized no change [value stays in the same reference range], categorized decrease [shift from high to normal, high to low, or from normal to low]).

Laboratory results will also be graded using NCI-CTCAE, where possible. Parameters that have criteria available for both low and high values, i.e., hypo- and hyper-, will be summarized for both criteria. The same subject can be counted for both values if the subject has different laboratory values meeting each criterion. NCI-CTCAE grade of laboratory evaluations will be summarized by number and percentage of subjects for each visit. Shift tables of NCI-CTCAE grade change from baseline to worst post-baseline grade will also be

presented. The number and percentage of subjects with grade 3 or 4 laboratory test result will be summarized by treatment group and laboratory parameter (the name of the adverse event associated with the abnormal laboratory test result will be presented).

Laboratory results based on central assessment will be used for summaries as described above. Laboratory results based on local assessment and bone marrow results will be listed only.

7.5.2.1 Liver Enzymes and Total Bilirubin

The following potentially clinically significant criteria in liver function tests for Alkaline Phosphatase (ALP), Alanine Transaminase (ALT), total bilirubin, Aspartate Transaminase (AST) and their combination are defined. The subject's highest value during the investigational period will be used.

Parameter	Criteria
ALT	> 3xULN
	> 5xULN
	> 10xULN
	> 20xULN
AST	> 3xULN
	> 5xULN
	> 10xULN
	> 20xULN
ALT or AST	> 3xULN
Total Bilirubin	> 2xULN
ALP	> 1.5xULN
ALT and/or AST AND Total Bilirubin(*)	(ALT and/or AST > 3xULN) and total bilirubin > 2xULN

(*) Combination of values measured within same sample

The number and percentage of subjects with potentially clinically significant values in liver enzymes and total bilirubin during the investigational period will be presented by treatment group.

7.5.3 Vital Signs

The baseline visit is the last measurement taken prior to initial study drug administration.

Vital signs (systolic blood pressure, diastolic blood pressure, pulse rate, and body temperature) will be summarized using mean, standard deviation, minimum, maximum and median by treatment group and visit. Additionally, a within-subject change will be calculated per visit as the post-baseline measurement minus the baseline measurement and summarized by treatment group and visit.

Tables for potentially clinically significant vital signs will be generated using baseline value and highest value obtained during treatment for each subject for each treatment group.

The following potentially clinically significant criteria are defined for each parameter:

Vital Sign Variable	Criteria
SBP	≥180 mmHg AND ≥20 mmHg change from baseline
DBP	≥105 mmHg AND ≥15 mmHg change from baseline
Pulse Rate	≥120 bpm AND ≥15 bpm change from baseline

DBP: diastolic blood pressure; SBP: systolic blood pressure

7.5.4 Electrocardiograms (ECGs)

12-lead ECGs will be recorded in triplicate at the scheduled time points. Each ECG tracing will be taken 5 minutes apart. ECGs will be read at the site for clinical decision making and transmitted to a central reviewer. Data from the central reviewer will be used in summary presentations. The three values of each ECG parameter within a time point from the central reviewer will be averaged to determine time-specific parameter for a subject, and used in summaries.

ECG variables will be summarized using mean, standard deviation, minimum, maximum and median for each treatment group at each treatment visit and time point, including changes from baseline.

Number and percentage of subjects with normal and abnormal results as assessed by central review for the overall interpretation will be tabulated by treatment group at each treatment visit and time point. A shift analysis table showing shift in overall ECG interpretation from baseline to each time point will be provided. The worst of the three overall ECG interpretations will be used as the time-specific overall ECG interpretation for a subject.

The QT interval corrected for heart rate by Fridericia's formula, QTcF, is defined as: $QTc\ (F) = QT/(RR)^{0.33}$, where RR interval is inversely proportional to heart rate (approximately RR = 60/heart rate).

The QTcF interval will be summarized using frequency tables for each treatment group at each treatment visit and time point for values of clinical importance using the range criteria below.

	QTc Interval Criteria Value (msec)	
	Cumulative Category	**Interval Category**
Normal	≤ 450	≤ 450
Borderline	> 450	> 450 to ≤ 480
Prolonged	> 480	> 480 to ≤ 500
Clinically significant	> 500	> 500

QTc: QT interval corrected

The QTcF interval will also be summarized by the frequencies of subjects with a change from baseline of clinical importance using the criteria identified below. These summaries will be provided for each treatment group at each treatment visit and time point.

	Change from Baseline	
Variable	**Cumulative Category**	**Interval Category**
QTc Interval (msec)	<0 ≥ 0 > 30 > 60	<0 ≥ 0 to ≤ 30 > 30 to ≤ 60 > 60

QTc: QT interval corrected

Number and percent of subjects with 12 lead ECG abnormalities as well as number and percent of subjects whose 12 lead ECG reading changed from normal at baseline to abnormal will be tabulated by treatment group at each treatment visit and time point.

7.5.5 Pregnancies

A detailed listing of all pregnancies will be provided.

7.5.6 Eastern Cooperative Oncology Group (ECOG) Performance Scores

Number of percent of subjects for each category of the ECOG performance status at each assessment time will be provided by treatment group. Negative change scores indicate an improvement and positive scores indicate a decline in performance.

ECOG will also be summarized using shift table from baseline to post-baseline score for each treatment group by visit.

7.6 Analysis of PK

Plasma concentrations at pre-dose sampling will be summarized by visit. The summary statistics of demographic data (body weight, age, gender) of subjects will be also described.

Table creation will be done in Chinese population and overall subjects in PKAS, respectively, while all randomized subjects with ASP2215 treatment will be used for listing creation.

Dosing and Sampling Time

Actual elapsed times from dosing will be used for calculating presenting individual concentrations in listings and figures, while scheduled sampling times will be used for presenting the summary statistics of concentrations in tables and figures.

The acceptable time ranges for sample collection are shown below. If the actual sampling time deviates from the allowance, that point will not be included in the summary statistics calculation of concentrations by scheduled time; however, it will be presented in listings and plots of the individual concentration-time profiles.

Time windows for plasma sample collection of ASP2215

Sampling time point	Allowable time range
Predose	Scheduled date, Within 1 hour before dosing
Day1/Day 15: 0.5 hour post-dose	Scheduled date, scheduled time ± 5 minutes
Day 1/Day 15: 1, 2, 3, 4, 6, 10 hours post-dose	Scheduled date, scheduled time ± 10 minutes
Day 2/Day 16: 24 hours post-dose	Scheduled date, scheduled time ± 20 minutes

- Blood sampling for pharmacokinetic analysis in approximately 20 Chinese subjects (10 male and 10 female subjects) enrolled in the PK cohort will be done at the following time points;

Cycle 1 Day 1: predose, 0.5, 1, 2, 3, 4, 6, 10, 24 hours after dosing of study drug

Cycle 1 Day 8: predose

Cycle 1 Day 15: predose, 0.5, 1, 2, 3, 4, 6, 10, 24 hours after dosing of study drug

Day 1 of each subsequent Cycle: predose

- Blood sampling for pharmacokinetic analysis in subjects who are not allocated to PK cohort (Non-PK cohort) will be done at the following time points;

Cycle 1 Day 1: predose,

Cycle 1 Day 8: predose,

Cycle 1 Day 15: predose,

Day 1 of each subsequent Cycle: predose

7.7 Analysis of PD

Not applicable.

7.8 Subgroups of Interest

Primary efficacy endpoint (OS) and key secondary endpoints (EFS and CR rate) will be summarized by treatment groups for the subgroups defined on the basis of the categorized variables listed below:

Grouping variable	Subgroups
Age group	< 65 years
	>=65 years
Sex	Female
	Male
Race	White
	Black or African American
	Asian
	Other
Baseline ECOG	0-1
	>=2
Region	China
	Non-China
Central FLT3 Mutation Type	FLT3-ITD alone
	FLT3-TKD alone
	FLT3-ITD & TKD
	Others (Unknown/Missing/Negative)
Response to First-line Therapy	Relapse within 6 months after allogeneic HSCT
	Relapse after 6 months after allogeneic HSCT
	Primary refractory without HSCT
	Relapse within 6 months after CRc and no HSCT
	Relapse after 6 months after CRc and no HSCT
Salvage Chemotherapy	High intensity chemotherapy (FLAG, MEC)
	Low intensity chemotherapy (LoDAC)
Prior Use of FLT3 inhibitor	Yes
	No
Cytogenetic Risk Status	Favorable
	Intermediate
	Unfavorable
	Other

Pooled stratification factors, if used for stratified analysis, will also be included in subgroup analysis.

7.9 Other Analyses

7.9.1 PK-PD Analysis

Not applicable.

7.10 Interim Analysis (and Early Discontinuation of the Clinical Study)

To evaluate whether ASP2215 is particularly beneficial or harmful compared to the salvage chemotherapy while the study is ongoing, a formal interim analysis is planned when approximately 50% of the planned death events (or 115 OS events) have occurred in the study.

The interim analyses will be conducted by the IDMC. Production of TLFs will be performed by independent statistician and data analysis group. No member of the clinical trial team will have access to the TLFs created for the IDMC. For more details consult the IDMC Charter.

A group sequential design using the O'Brien-Fleming type alpha-spending function [Lan & DeMets, 1983] will be used to control the overall two-sided 0.05 significance level (East®). OS will be analyzed based on the method described in Section 7.4.1. The decision rule will be based on the significance level given by Lan and DeMets with O'Brien-Fleming boundary for the interim analysis. The IDMC may recommend terminating the trial for favorable or unfavorable results at the interim analysis. In the case of favorable results, the 2-sided significance level for superiority is 0.00305 for the interim analysis. If the observed hazard ratio is less than 1 and 2-sided p-value of the interim analysis is less than 0.00305, the IDMC may recommend terminating the trial for success. In the case of unfavorable results in the interim analysis, the 2-sided significance level for futility (nonbinding) is 0.84431. If the estimated hazard ratio is less than 1 and the 2-sided p-value is greater than 0.84431 or the estimated hazard ratio is greater than 1, the DMC may recommend terminating the trials for futility.

If the study is not stopped after the interim analysis, a final analysis will occur after 100% of events have been observed. The statistical 2-sided significance level for the final analysis is at 0.0490.

The key secondary efficacy endpoints of EFS, CR rate as well as other secondary efficacy endpoints will be analyzed based on the method described in Section 7.4.2. According to the sequential testing procedure on controlling overall type I error rate for multiple endpoints, EFS will be tested when the null hypothesis of OS is rejected at the interim or at the final analyses. CR rate will be tested when the null hypotheses of both OS and EFS are rejected hierarchically at the interim or at the final analyses. Their significance levels at the interim and final analyses will be based on Pocock alpha spending function.

The details about significance levels at the interim and final analyses for the primary and secondary key endpoints are specified as in Table 1 below.

Table 1 Summary of Timing, Sample Size and Decision Guidance at the Planned Analyses

Analysis	Criteria for conduct of analysis (Projected timing)	Endpoint/ Analysis Set	Efficacy Boundary [a]		Futility Boundary [a]	
			p-value (2-sided) at the Boundary	Approx. Observed HR at Boundary	p-value (2-sided) at the Boundary	Approx. Observed HR at Boundary
IA: OS; EFS when null hypothesis of OS is rejected; CR rate when null hypotheses of both EFS and OS are rejected	Approx. 115 OS events were observed	OS/ITT	0.00305	0.57	0.84431	0.96
		EFS/ITT	0.03101			
		CR rate/ ITT	0.03101			
Final: OS; EFS when null hypothesis of OS is rejected; CR rate when null hypotheses of both EFS and OS are rejected	Approx. 230 OS events were observed	OS/ITT	0.0490	0.77		
		EFS/ITT	0.02774			
		CR rate/ITT	0.02774			

CR: Complete Remission; EFS: Event-Free Survival; HR: Hazard Ratio; OS: Overall Survival

a: P-value at both efficacy and futility boundaries are based on 50% information fraction for OS, EFS and CR rate, and need update based on observed information fraction at the interim analysis.

Safety data including AE, lab, vital signs, ECG, and ECOG will also be reviewed by IDMC in the interim analysis. Other safety data reviews during the trials will be conducted by IDMC on a periodic basis. The statistical analysis plan will be described in interim analysis plan (IAP). The procedures for DMC safety review will be described in the DMC charter.

7.11 Handling of Missing Data, Outliers, Visit Windows, and Other Information

7.11.1 Missing Data

Every effort will be made to resolve incomplete dates for death and disease relapse. If a partial date cannot be resolved, the most conservative imputation methods will be used to complete the missing information.

For primary endpoint OS, missing or incomplete death date will be imputed as the earliest feasible date on or after the date of last contact as the examples shown in the table below. The date of last contact will be obtained as described in Section 6.1.1.

Incomplete Date of Death (YYYY MMM DD)	Date of Last Contact (YYYY MMM DD)	Imputed Date of Death (YYYY MMM DD)
2005 APR ??	2005 MAR 31	2005 APR 01
2005 ??? 13	2005 MAR 31	2005 APR 13
2005 ??? ??	2005 MAR 31	2005 MAR 31
???? APR ??	2005 MAR 31	2005 APR 01
???? APR 13	2005 MAR 31	2005 APR 13
???? ??? ??	2005 MAR 31	2005 MAR 31

Partial relapse dates will be imputed to the first day of the month of the missing parameter but not earlier than the last disease assessment date. A month and year must be present or the date will remain missing.

Non-responder imputation will be used for binary response variables.

Missing or partial start and stop dates of adverse events and concomitant medication will be imputed using the following algorithm:

- Imputation rules for partial or missing stop dates:
  - If the month and year are present, then impute as the last day of that month.
  - If only the year is present, impute as December 31 of that year.
  - If the stop date is entirely missing, assume the event or medication is ongoing.
- Imputation rules for partial or missing start dates:

Start Date		Stop Date						
		Complete: *yyyymmdd*		Partial: *yyyymm*		Partial: *yyyy*		missing
		< 1st dose	≥ 1st dose	< 1st dose *yyyymm*	≥ 1st dose *yyyymm*	< 1st dose *yyyy*	≥ 1st dose *yyyy*	
Partial: *yyyymm*	= 1st dose *yyyymm*	2	1	n/a	1	n/a	1	1
	≠ 1st dose *yyyymm*		2	2	2	2	2	2
Partial: *yyyy*	= 1st dose *yyyy*	3	1	3	1	n/a	1	1
	≠ 1st dose *yyyy*		3		3	3	3	3
Missing		4	1	4	1	4	1	1

1 = Impute as the date of first dose; 2 = Impute as the first of the month; 3 = Impute as January 1 of the year; 4 = Impute as January 1 of the stop year

The imputed dates will be used to determine whether an AE is/is not treatment emergent. Listings of AEs and concomitant medications will present the actual partial dates; imputed dates will not be shown.

In the case of partial date of initial diagnosis of AML, the date will be imputed to the first day of the month. A month and year must be present or the date will remain missing.

In the case of partial starting date of subsequent AML therapy, the date will be imputed to the first day of the month but not earlier than the last dosing date of the study drug. A month and year must be present or the date will remain missing.

In the case of partial stop date of study drug dosing, the date will be imputed to the first day of the month but not earlier than the start date of the study drug dosing. A month and year must be present or the date will remain missing.

7.11.2 Outliers

All values will be included in the analyses.

7.11.3 Visit Windows

Visit windows are allowed for certain visits per the schedule of assessments. Subject data will not be excluded from analyses due to the subject's failure to comply with the visit schedule, unless otherwise noted. CRF visit will be used for analysis. In the case of multiple observations at a specific visit, the observation which is closest to the target date will be used. If the observations have the same distance to the target visit, the latest one will be used. If more than one observation is made on the same day, an average value if continuous or the worst value if categorical will be included in the analysis.

7.11.4 Pooling Strata

In the time to event endpoint, e.g., OS and EFS analyses, if all events are from one treatment group in at least one stratum combination, or the Cox proportional hazard model does not converge due to small event size in some stratum combinations, the stratum combinations will be pooled per following steps until the issue is resolved or the normal (un-stratified) Cox proportional hazard model is applied.

Step1: Pooling within preselected chemotherapies

Within each preselected chemotherapy (high intensity or low intensity), if the criterions described above don't meet, pool the 5 levels of response to first-line therapy into 3 levels:

- Relapse after allogeneic HSCT
- Primary refractory without HSCT
- Relapse after CRc and no HSCT

After the above pooling, if the criterions still don't meet, pool the 3 levels into 2 levels:

- Relapse after allogeneic HSCT or after CRc without HSCT
- Primary refractory without HSCT

After the above pooling, if the criterions still don't meet, pool the 2 levels into 1 level:

- Relapse after allogeneic HSCT or after CRc without HSCT or Primary refractory without HSCT

Step2: Pooling across preselected chemotherapies

If after Step1 pooling, the criterions still don't meet, pool the preselected high and low dose chemotherapies into one level:

- High intensity chemotherapy (FLAG, MEC) or Low intensity chemotherapy (LoDAC)

After the above pooling, re-evaluate the pooling of response to first-line therapy after merging the subjects from both preselected high and low intensity chemo groups, i.e., start the pooling from the 5 levels and follow the similar order described in Step1 until the criterions meet.

In the binary endpoint, e.g., CR rate analysis using CMH test, similar pooling strategy will be applied when all the CR subjects are from one treatment group appears in at least one stratum combination.

All sensitivity analyses for the above primary and key secondary endpoints will apply the same pooled strata used for the primary analysis. In the case that the criterions don't meet, the un-stratified analysis will be used.

7.11.5 Blinding

Although the study is an open label study, to maintain trial integrity and increase the credibility of study results, the sponsor statistician's access to the randomized treatment assignment information will be limited. This will reduce potential bias due to the sponsor knowing the treatment effect due to unintentional efficacy and safety summary by treatment. On the other hand, the clinical data should be used appropriately for clinical operation, data cleaning and generating statistical programs. Thus, we will follow the procedures specified below (more details can be found in Section 10.2):

- The study statistician and programmers will have no access to the randomized treatment information before the database lock. The randomized treatment code will not be transferred to programmers and study statistician before the database lock. Instead, datasets with scrambled treatment code will be transferred to prepare analysis programs.
- Study manager and other study team members may have the access to the treatment information at the individual subject level.
- No by treatment summary and treatment difference will be generated during the study, except planned interim analysis conducted by IDMC, or early submission package preparation conducted by an external independent data analysis team.

8 DOCUMENT REVISION HISTORY

Version	Date	Changes	Comment/rationale for change
1.0	05-Oct-2017	NA	Document finalized
2.0 (Amendment 1)	25-Feb-2020	**Overall**	To correct errors, clarify the details of analyses, and delete the analyses out of scope of this SAP.
		Secondary Objectives Added the followings to the secondary objectives. -CR/CRh rate -Transfusion conversion rate and transfusion maintenance rate	To keep consistent with CL-0301 analyses.
		Analysis Sets Changed the analysis set for primary analyses of efficacy data from FAS to ITT.	To keep consistent with CL-0301 analyses.
		Analysis Sets Removed PPS.	To evaluate the effect of each major protocol deviation by sensitivity analyses instead of PPS analysis.
		Key Secondary Efficacy Endpoints Changed the primary definition of EFS to include long-term follow-up data.	To modify to the most appropriate definition of EFS referring to CL-0301 results.
		Sensitivity analysis on EFS Added the followings. - EFS with relapse after PR also considered as event - EFS with best response of NE also considered as treatment failure - EFS not using long-term follow up data	To keep consistent with CL-0301 analyses. To keep the previous primary EFS definition as sensitivity analysis.
		Sensitivity analysis on CR, CRc, CR/CRh and Best Response Added the followings. - CR rate prior to HSCT - CRc rate prior to HSCT - CR/CRh rate prior to HSCT - Best Response rate prior to HSCT	To keep consistent with CL-0301 analyses and add necessary sensitivity analyses.

Version	Date	Changes	Comment/rationale for change
		Other secondary efficacy endpoints Added the followings and the related analyses. - CR/CRh rate - DCRCRh - Transfusion conversion rate and transfusion maintenance rate Deleted the followings and the related analyses. - CRi, DCRp	To keep consistent with CL-0301 analyses.
		Other efficacy variables Added the followings and the related analyses. - Response rate, CRh rate - TTCRCRh Deleted the followings. - CRp rate, CRi rate, PR rate - TTCRi, TTCRp, TTBR	To keep consistent with CL-0301 analyses.
		Other variables Added the following. - Prior use of FLT3 inhibitor	To keep consistent with CL-0301 analyses.
		Protocol Deviation Added the section of "Protocol Deviations".	To keep consistent with CL-0301 analyses.
		Analysis of efficacy Changed the primary hypothesis test on OS, EFS, etc. from Wald test to log-rank test.	To keep consistent with CL-0301 analyses.
		Analysis of Primary Endpoint of OS Deleted the following from sensitivity analyses. - Stratified Cox proportional hazard model with time dependent binary covariate	To delete unnecessary analysis.
		Analysis of efficacy Added summary of OS, CR, CRc and CR/CRh by dose adjustment for ASP2215 group.	To investigate the impact of dose adjustment.
		Analysis of Exploratory Endpoints Deleted MMRM analysis on patient reported outcomes and health outcome.	To delete unnecessary analysis.

Version	**Date**	**Changes**	**Comment/rationale for change**
		Analysis of AE Added the number of events and events adjusted by patient year. Updated AE overview, SOC/PT analyses, PT only analyses, and AESI analyses. Added AE subgroup analysis by preselected salvage chemotherapy (high intensity or low intensity).	To keep consistent with CL-0301 analyses and ISS analyses.
		Subgroups of interest Added the followings to subgroup of interest. - Region - Prior Use of FLT3 inhibitor - Cytogenetic Risk Status - Pooled stratification factors Removed the followings - Country of the site	To keep consistent with CL-0301 analyses.
		Visit windows Changed to use CRF visit for analysis.	To keep consistent with CL-0301 analyses.
		Appendix Added the following appendix. - Appendix 3: Search Strategy for Adverse Events of Interest - Appendix 4: Search Strategy for medication - Appendix 5: China Specific Analysis	To clarify the details of analyses.
3.0 (Amendment 2)	31-May-2021	**Other Secondary Efficacy Endpoints** Updated the details of the endpoint - BFI	To clarify the details of analyses.
		Exploratory Endpoints **Appendix 6: FACIT-Dys SF** **Appendix 7: FACT-Leu Scoring** Updated the details of the endpoint - FACIT-Dys-SF - FACT-Leu	To clarify the details of endpoints.

Version	Date	Changes	Comment/rationale for change
		Appendix 3: Search Strategy for Adverse Events of Interest Updated the definition of adverse events of interests.	To update the definition of adverse events of interests.
		Sample Size Corrected the planned sample size with EFS event from 230 to 250.	To correct the error in the sample size calculation of EFS event.
		Missing Data Added the imputation in the case of partial stop date of study drug dosing.	To address partial stop date of study drug dosing.
		Analysis of Primary Endpoint of OS **Key Secondary Efficacy Analysis** Added sensitivity analyses for COVID-19 impact on OS and EFS.	To investigate the potential impact of COVID-19 impact.
4.0 (Amendment 3)	22-Feb-2024	**Study Design** Added texts for crossover extensions and eCRF data collection close.	To update the study design based on the protocol version 7.0.
		General Considerations Added texts for not applying the pre-specified hierarchal testing procedure for end-of-study.	Since the study enrollment was terminated on 12Mar2021 due to favorable outcome of OS at the interim analysis by IDMC, and the formal interim analysis was conducted based as final primary analysis on data cutoff on 30Jun2020, the pre-specified hierarchal testing procedure to maintain the overall 2-sided Type I error rate at the 0.05 will not be applied at the end-of-study.
		General Considerations **Disposition of Subjects** Added general texts and analysis for crossover extension.	No patients entered the crossover extension. So, only the number of subjects with informed consent for crossover extension was added in Section 7.2.1 Disposition of Subjects. No additional analysis for crossover extension is planned.

Version	**Date**	**Changes**	**Comment/rationale for change**
		FLOW CHART AND VISIT SCHEDULE Removed the flow chart and visit schedule and changed to refer the protocol. **General Considerations** Updated SAS version. **Appendix 1: SIGNATURE** Removed the list of key contributors.	To follow the contents of latest Statistical Analysis Plan format.
		ANALYSIS SETS **Pharmacokinetics Analysis Set (PKAS)** Added details of PKAS definition. **Analysis of PK Individual Data and Summary Statistics** Added details of PK analysis.	To incorporate PK analysis for end-of-study from PK-SAP.

9 REFERENCES

Cella D, Jensen SE, Webster K, Hongyan D, Lai JS, Rosen S, et al. Measuring health-related quality of life in leukemia: the Functional Assessment of Cancer Therapy--Leukemia (FACT-Leu) questionnaire. Value Health. 2012;15:1051-8.

Cheson BD, Bennett JM, Kopecky KJ, Büchner 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;21:4642-4649.

Choi SW, Victorson DE, Yount S, Anton S, Cella D. Development of a conceptual framework and calibrated item banks to measure patient-reported dyspnea severity and related functional limitations. Value Health. 2011;14:291-306.

ICH Harmonized Tripartite Guideline E 3. Structure and Content of Clinical Study Reports, November 1995. (www.ich.org; Guidelines; "Efficacy" Topics)

ICH Harmonized Tripartite Guideline E 9. Statistical Principles for Clinical Trials, February 1998. (www.ich.org; Guidelines; "Efficacy" Topics)

Karnofsky DA, Burchenal JH. “The Clinical Evaluation of Chemotherapeutic Agents in Cancer." In: MacLeod CM (Ed), *Evaluation of Chemotherapeutic Agents.* Columbia Univ Press. 1949;196.

Lan KKG, DeMets DL. Discrete sequential boundaries for clinical trials. Biometrika. 1983;70:659-63.

Mendoza TR, Wang XS, Cleeland CS, Morrissey M, Johnson BA, Wendt JK, et al. The rapid assessment of fatigue severity in cancer patients: use of the Brief Fatigue Inventory. Cancer. 1999;85:1186-96.

Sievers EL, Larson RA, Stadtmauer EA, et al. Efficacy and safety of Gemtuzumab Ozogamicin in Patients with CD33-Positive Acute Myeloid Leukemia in First Relapse. J Clin Oncol. 2001;19:3244-3254.

10 APPENDICES

10.1 Appendix 1: SIGNATURE

Author and Approver Signatories

(E-signatures are attached at end of document)

PPD, Statistical & Real World Data Science, Data Science, was the study statistician for this study and the primary author of this Statistical Analysis Plan

PPD, Statistical & Real World Data Science, Data Science, was the Statistical Lead for this study and the biostatistics peer reviewer of this Statistical Analysis Plan

This Statistical Analysis Plan was approved by:

PPD
, Oncology Development

10.2 Appendix 2: Blinding Process

As specified in SAP Section 7.11.5, although the study is an open label study, to increase the credibility of study results, the sponsor statistician's access to the randomized treatment assignment information in study database will be limited. This will reduce potential bias due to the sponsor knowing the treatment effect due to unintentional efficacy and safety summary by treatment.

For the purpose of controlling unintentional assessment of efficacy and safety summary by treatment, the study statisticians and programmers will have no access to treatment codes until the time of final database lock (DBL). For this study, treatment codes can be obtained from two sources: IRT vendor and EDC system. The following procedures should be applied when handling data extraction, data transferring and TLFs preparation to maintain the blinding during the entire study conducting period before the final DBL.

- The treatment codes collected in EDC system will not be extracted until the final DBL. The unblinded data including treatment variable will be blocked during any data extraction before the final DBL. The option for blocking treatment variable will be set up before the first data extraction.
- Dummy treatment codes will be created by programmers for preparing the SDTM, ADaM and TLFs for interim analysis validation and final data analysis.
- ADaM without treatment codes (or with dummy codes) will be provided by the study team to IDAC to conduct interim analysis or other IDMC requested analysis before final DBL.
- The Astellas independent statistician will transfer the treatment codes and the unblinded data to IDAC during the formal interim analysis or other IDMC requested analysis, and may transfer the treatment codes to an external independent data analysis team to prepare the early submission package before final DBL.
- Dosing or other data (including PK data) which may be related to treatment assignment will be extracted/transferred to an unblinding folder which is only accessible to the unblinding programmers or Astellas independent statistician.

Study manager and other study team members may have the access to the treatment assignment information at the individual subject level. Efficacy and safety information will not be summarized by treatment during the study, except at the planned interim analysis or other IDMC requested analysis conducted by IDAC, or early submission package preparation conducted by an external independent data analysis team.

10.3 Appendix 3: Search Strategy for Adverse Events of Interest

Risk	Search Strategy: MedDRA Version 23.0
Anaphylactic reaction	Anaphylactic reaction (SMQ Broad)
Cardiac failure	Cardiac failure (SMQ Narrow)
Creatine phosphokinase increased	Rhabdomyolysis/myopathy (SMQ Narrow)
	Blood creatine phosphokinase abnormal (PT=10005468 and Grade >=3)
	Blood creatine phosphokinase increased (PT=10005470 and Grade >=3)
	Blood creatine phosphokinase MM increased (PT=10005477 and Grade >=3)
	PT: Myalgia
	PT: Myositis
	PT: Muscular weakness
Diarrhea	Noninfectious diarrhoea (SMQ Broad)
Differentiation Syndrome [a]	PT: Acute interstitial pneumonitis, Acute kidney injury, Acute lung injury, Acute pulmonary oedema, Acute respiratory distress syndrome, Acute respiratory failure, Anuria, Atypical pneumonia, Blood creatinine increased, Blood pressure systolic decreased, Body temperature increased, Capillary leak syndrome, Cardiopulmonary failure, Cardiorenal syndrome, Cardiorespiratory distress, Cough, Differentiation syndrome, Dyspnoea, Febrile neutropenia, Fluid overload, Fluid retention, Generalised oedema, Hepatorenal failure, Hydraemia, Hypervolaemia, Hypotension, Lower respiratory tract infection, Lower respiratory tract inflammation, Lung infection, Lung infiltration, Multiple organ dysfunction syndrome, Noncardiogenic pulmonary oedema, Oedema, Oedema peripheral, Pericardial effusion, Pleural effusion, Pneumonia, Pneumonitis, Prerenal failure, Pulmonary congestion, Pulmonary oedema, Pulmonary toxicity, Pyrexia, Renal failure, Renal impairment, Renal injury, Respiratory arrest, Respiratory distress, Respiratory failure, Weight increased PTs
Gastrointestinal obstruction	Gastrointestinal obstruction (SMQ Narrow)
Gastrointestinal perforation	Gastrointestinal perforation (SMQ Narrow)
Liver transaminase increased	Liver related investigations, signs, and symptoms (SMQ Narrow)
Pancreatitis	Acute pancreatitis (SMQ Broad)
Pericarditis/Pericardial effusion	HLT Noninfectious pericarditis
	PT Pericardial effusion
PRES	Noninfectious encephalopathy/delirium (SMQ Narrow)
QT Prolongation	Torsade de pointes/QT prolongation (SMQ Narrow)
Teratogenicity and Embryo- Fetal Deaths	SMQ Broad-All Pregnancy

HLT: high level term; PRES: posterior reversible encephalopathy syndrome; PT: preferred term; SMQ: standard MedDRA queries

a: Only AEs occur within the first 90 days

10.4 Appendix 4: Search Strategy for medication

Medication	Search Strategy: WHODDE(B2) (V2016MAR)
Antibiotic iv infusion	The following medications with route = "INTRAVENOUS" WHO-DD ATC 4th level = "ANTIBIOTICS"
Opioid	The following medications. WHO-DD ATC 3rd level = "OPIOIDS"

ATC: Anatomical Therapeutic Chemical Classification; WHO-DD: World Health Organization – Drug Dictionary

10.5 Appendix 5: China Specific Analysis

Introduction

This appendix describes the planned regional specific analyses in addition to the analyses defined by the main body of this SAP. The regional analyses specified in this appendix are for the CDE submission only.

Region Specific Analyses of China

The following regional specific analyses of China will be performed as appropriate:

- All analyses as described in the main body of this SAP will be repeated for the subgroups of China vs Non-China
- All listings will be presented by sites ordered as China followed by Non-China sites

Below shows a couple of examples where the regional specific analyses of China will be handled differently from what are described in the main body of this SAP. The exceptions will be reflected in the corresponding TLF specifications, where the detailed Specifications for table, figure, and data listing formats can be found.

Example(s) of Region Specific Analyses of China with Difference

Section(s)	Change(s)	Comment/rationale for change
7.2.3	Geographical Region will not be included in the summary of demographic and other baseline characteristics	The summary of demographic and other baseline characteristics will be repeated for the subgroups of China vs Non-China
7.4	The efficacy endpoints will be evaluated using un- stratified analysis only.	Due to the small number of China patients, all these efficacy analyses will be performed using un-stratified analysis instead of stratified analysis.
7.8	Subgroup analyses won't be performed for Geographical Region	The subgroup analyses will be repeated for the subgroups of China vs Non-China

10.6 Appendix 6: FACIT-Dys SF

Each item on the Dyspnea subscale can be scored as 0 = 1 (no shortness of breath), 1 = 2 (mildly short of breath), 2 = 3 (moderately short of breath), and 3 = 4 (severely short of breath). If a subject indicated 0 (I did not do this in the past 7 days) for an item in the Dyspnea subscale:

- If 4 ("I have stopped trying, or knew I could not do this activity because of my shortness of breath") was selected for the reason, score = 3
- If 0 ("I did not do this activity for some other reason (including not having a chance to do it, other health issues etc).") was selected for the reason, score = missing

For the functional limitations scale, if participant also completed Dyspnea subscale and indicated 0 (I did not do this in the past 7 days) for an item, the same item on Functional Limitations subscale is not presented to participant, but scoring is as follows:

- If 4 ("I have stopped trying, or knew I could not do this activity because of my shortness of breath") was selected for the reason on Dyspnea subscale, score on Functional Limitations subscale = 3
- If 0 ("I did not do this activity for some other reason (including not having a chance to do it, other health issues etc).") was selected for the reason on Dyspnea subscale, score on Functional Limitations subscale = missing

FACIT-Dys SF scoring guide

Domain	Item codes	Score calculation	Score
Dyspnea	FDYSP01 to FDYSP10 with corresponding FDYSP01A to FDYSP10A	Mean(Score of FDYSPxx) * 10 Score of FDYSPxx is as below. 0: FDYSPxx = 1; 1: FDYSPxx = 2; 2: FDYSPxx = 3; 3: FDYSPxx = 4; 3: FDYSPxx = 0 and FDYSPxxA = 4; missing: FDYSPxx = 0 and FDYSPxxA = 0; If [# of non-missing score of FDYSPxx] is 5 or less, then Score = missing.	0-30
Functional limitations	FDYSP11 to FDYSP20 with corresponding FDYSP01A to FDYSP10A	Mean(Score of FDYSPxx) * 10 Score of FDYSPxx is as below. 0: FDYSPxx = 1; 1: FDYSPxx = 2; 2: FDYSPxx = 3; 3: FDYSPxx = 4; 3: FDYSPxx = missing and corresponding FDYSPxxA = 4; missing: FDYSPxx = missing and corresponding FDYSPxxA = 0; If [# of non-missing score of FDYSPxx] is 5 or less, then Score = missing.	0-30

Item codes are as follows.

FDYSP01: P1Q1DressingYourself [a]

FDYSP01A: P1Q1DressingYourselfWhyNot [b]

FDYSP02: P1Q2Walking50Steps [a]

FDYSP02A: P1Q2Walking50StepsWhyNot [b]

FDYSP03: P1Q3Walking20Stairs [a]

FDYSP03A: P1Q3Walking20StairsWhyNot [b]

FDYSP04: P1Q4PrepareMeals [a]

FDYSP04A: P1Q4PrepareMealsWhyNot [b]

FDYSP05: P1Q5WashingDishes [a]

FDYSP05A: P1Q5WashingDishesWhyNot [b]

FDYSP06: P1Q6Sweeping [a]

FDYSP06A: P1Q6SweepingWhyNot [b]

FDYSP07: P1Q7MakingBed [a]

FDYSP07A: P1Q7MakingBedWhyNot [b]

FDYSP08: P1Q8Lifting [a]

FDYSP08A: P1Q8LiftingWhyNot [b]

FDYSP09: P1Q9Carrying [a]

FDYSP09A: P1Q9CarryingWhyNot [b]

FDYSP10: P1Q10WalkingHalfMile [a]

FDYSP10A: P1Q10WalkingHalfMileWhyNot [b]

FDYSP11: P2Q1DifficultyDressing [c]

FDYSP12: P2Q2DifficultyWalking50Steps [c]

FDYSP13: P2Q3DifficultyWalking20Stairs [c]

FDYSP14: P2Q4DifficultyPreparingMeals [c]

FDYSP15: P2Q5DifficultyWashingDishes [c]

FDYSP16: P2Q6DifficultySweeping [c]

FDYSP17: P2Q7DifficultyMakingBed [c]

FDYSP18: P2Q8DifficultyLifting [c]

FDYSP19: P2Q9DifficultyCarrying [c]

FDYSP20: P2Q10DifficultyWalkingHalfMile [c]

Value codes of [a] are 1 (no shortness of breath), 2 (mildly short of breath), 3 (moderately short of breath), 4 (severely short of breath), and 0 (I did not do this in the past 7 days).

Value codes of [b] are 4 (“I have stopped trying, or knew I could not do this activity because of my shortness of breath”, 0 (“I did not do this activity for some other reason (including not having a chance to do it, other health issues etc).”).

Value codes of [c] are 1 (No difficulty), 2 (A little difficulty), 3 (Some difficulty), 4 (Much difficulty).

10.7 Appendix 7: FACT-Leu Scoring

FACT-Leu scoring guide

Domain	Item codes	Score calculation	Score range
Physical Well-Being (PWB)	FACLEU01 to FACLEU07	Mean(FACLEUxx) * 7 If [# of non-missing FACLEUxx] is 3 or less, then Score = missing.	0-28
Social/Family Well-Being (SWB)	FACLEU08 to FACLEU15 except FACLEU14	Mean(FACLEUxx) * 7 If [# of non-missing FACLEUxx] is 3 or less, then Score = missing.	0-28
Emotional Well-Being (EWB)	FACLEU16 to FACLEU21	Mean(FACLEUxx) * 6 If [# of non-missing FACLEUxx] is 3 or less, then Score = missing.	0-24
Functional Well-Being (FWB)	FACLEU22 toFACLEU28	Mean(FACLEUxx) * 7 If [# of non-missing FACLEUxx] is 3 or less then Score = missing.	0-28
Leukemia Subscale (LeuS)	FACLEU29 toFACLEU45	Mean(FACLEUxx) * 17 If [# of non-missing FACLEUxx] is 8 or less then Score = missing.	0-68
FACT-G total score	PWB, SWB, EWB, FWB	PWB + SWB + EWB + FWB If any of subscale score is missing, total score = missing.	0-108
FACT-Leukemia Trial Outcome Index (TOI)	PWB, FWB, LeuS	PWB + FWB + LeuS If any of subscale score is missing, total score = missing.	0-124
FACT-Leu total score	PWB, SWB, EWB, FWB, LeuS	PWB + SWB + EWB + FWB + LeuS If any of subscale score is missing, total score = missing.	0-176

Item codes are as follows.

FACLEU01: Q1LackEnergy [a]

FACLEU02: Q2Nausea [a]

FACLEU03: Q3FamilyNeeds [a]

FACLEU04: Q4Pain [a]

FACLEU05: Q5SideEffects [a]

FACLEU06: Q6Ill [a]

FACLEU07: Q7TimeInBed [a]

FACLEU08: Q8CloseToFriends [b]

FACLEU09: Q9EmotionalSupportFamily [b]

FACLEU10: Q10EmotionalSupportFriends [b]

FACLEU11: Q11FamilyAccept [b]

FACLEU12: Q12FamilyCommunication [b]

FACLEU13: Q13CloseToPartner [b]

FACLEU14: SexLife equal to Prefer not to answer
FACLEU15: Q14SexLife [b]
FACLEU16: Q15Sad [a]
FACLEU17: Q16Coping [b]
FACLEU18: Q17HopeLost [a]
FACLEU19: Q18Nervous [a]
FACLEU20: Q19Worried [a]
FACLEU21: Q20WorryWorse [a]
FACLEU22: Q21AbleWork [b]
FACLEU23: Q22WorkFulfilling [b]
FACLEU24: Q23EnjoyLife [b]
FACLEU25: Q24AcceptedIllness [b]
FACLEU26: Q25SleepWell [b]
FACLEU27: Q26HaveFun [b]
FACLEU28: Q27QualityOfLife [b]
FACLEU29: Q28HighBodyTemp [a]
FACLEU30: Q29CertainPain [a]
FACLEU31: Q30Chills [a]
FACLEU32: Q31NightSweats [a]
FACLEU33: Q32LumpsSwelling [a]
FACLEU34: Q33BleedEasily [a]
FACLEU35: Q34BruiseEasily [a]
FACLEU36: Q35WeakAllOver [a]
FACLEU37: Q36TireEasily [a]
FACLEU38: Q37LosingWeight [a]
FACLEU39: Q38GoodAppetite [b]
FACLEU40: Q39UsualActivities [b]
FACLEU41: Q40Infections [a]
FACLEU42: Q41FutureHealth [a]
FACLEU43: Q42WorryNewSymptoms [a]
FACLEU44: Q43EmotionUpDown [a]
FACLEU45: Q44FeelIsolated [a]
Value codes of [a] are 4 (Not at all), 3 (A little bit), 2 (Somewhat), 1 (Quite a bit), 0 (Very much).
Value codes of [b] are 0 (Not at all), 1 (A little bit), 2 (Somewhat), 3 (Quite a bit), 4 (Very much).