STATISTICAL ANALYSIS PLAN

Final Version 3.0, dated 12-Aug-2021

A Phase 2 Multi-Center, Randomized, Double-Blind, Placebo-Controlled Trial of the FLT3 Inhibitor Gilteritinib (ASP2215) Administered as Maintenance Therapy Following Induction/Consolidation Therapy for Subjects with FLT3/ITD AML in First Complete Remission

ISN/Protocol: 2215-CL-0302

IND number: 117,548

NCT02927262

Sponsor:

Astellas Pharma Global Development, Inc. (APGD)
1 Astellas Way
Northbrook, IL 60062

Table of Contents

I. LIST of ABBREVIATIONS AND KEY TERMS

List of Abbreviations

Abbreviations	Description of abbreviations
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
APGD	Astellas Pharma Global Development
ASCM	Analysis Set Classification Meeting
ASP2215	Astellas Compound code for 2215
AST	Aspartate Transaminase
ATC	Anatomical Therapeutic Chemical
BMI	Body Mass Index
CMH	Cochran-Mantel-Haenszel
CR	Complete remission
CR1	First complete remission
CRc	Composite complete remission
CRi	Complete remission with incomplete hematologic recovery
CRp	Complete remission with incomplete platelet recovery
CS	Classification Specifications
CSR	Clinical Study Report
CT	Computed tomography
CTCAE	Common Terminology Criteria for Adverse Events
CV	Coefficient of variation
CYP	Cytochrome P450
DBP	Diastolic blood pressure
ECG	Electrocardiogram
ECHO	Echocardiogram
ECOG	Eastern Cooperative Oncology Group
eCRF	Electronic case report form
EDTA	Ethylenediaminetetraacetic acid
EFS	Event-free survival

Abbreviations	Description of abbreviations
EQ-5D-5L	EuroQol Group-5 dimension-5 level
ER	Emergency room
FAB	French-American-British
FACT-An	Functional Assessment of Cancer Therapy-Anemia
FACT-Leu	Functional Assessment of Cancer Therapy-Leukemia
FAS	Full analysis set
FLT3	FMS-like tyrosine kinase
FSI	First subject in
H	High
HEOR	Health Economics and Outcomes Research
HSCT	Hematopoietic stem cell transplant
ICF	Informed consent form
ICH	International Conference on Harmonization
ICU	Intensive care unit
IDAC	Independent data analysis center
IDMC	Independent Data Monitoring Committee
IND	Investigational new drug
IRC	Independent Review Committee
IRT	Interactive Response Technology
ISN	International Study Number
ITD	Internal tandem duplication
IU/L	International units/liter
IWG	International Working Group
L	Low
LFT	Liver function test
LLN	Lower limit of normal
LLOQ	Lower limit of quantification
MedDRA	Medical Dictionary for Regulatory Activities
mg	Milligrams
Min	Minute
mL	Milliliter
mmHg	millimeters of mercury
MRD	Minimal residual disease
msec	milliseconds
MUGA	Multigated acquisition scan
N	Number
NCI	National Cancer Institute

Abbreviations	Description of abbreviations
NE	Not Evaluable
NGS	Next-generation sequencing
NR	Non-Response
OS	Overall Survival
PD	Protocol Deviation
PGx	Pharmacogenomics
PK	Pharmacokinetic
PKAS	Pharmacokinetic analysis set
PRO	Patient reported outcome
PT	Preferred Term
QD	quaque die, a Latin phrase meaning “every day”
QTc	QT interval corrected for heart rate
QTcF	Fridericia-corrected QT interval
RBC	Red blood cell
RFS	Relapse-free survival
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
SOC	System Organ Class
TEAE	Treatment Emergent Adverse Event
TLF	Tables, Listings and Figures
TKD	Tyrosine kinase domain
ULN	Upper limit of normal
US	United State
VAS	Visual analogue scale
WHO	World Health Organization
WHO-DD	World Health Organization – Drug Dictionary

List of Key Terms

Terms	Definition of terms
Baseline	Observed values/findings which are regarded as the observed starting point(s) 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.
Screening	A process of active consideration of potential subjects for enrollment in a trial.
Screen failure	Potential subject who did not meet 1 or more criteria required for participation 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 test drug or comparative drug (sometimes without randomization) is given to a subject.
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.

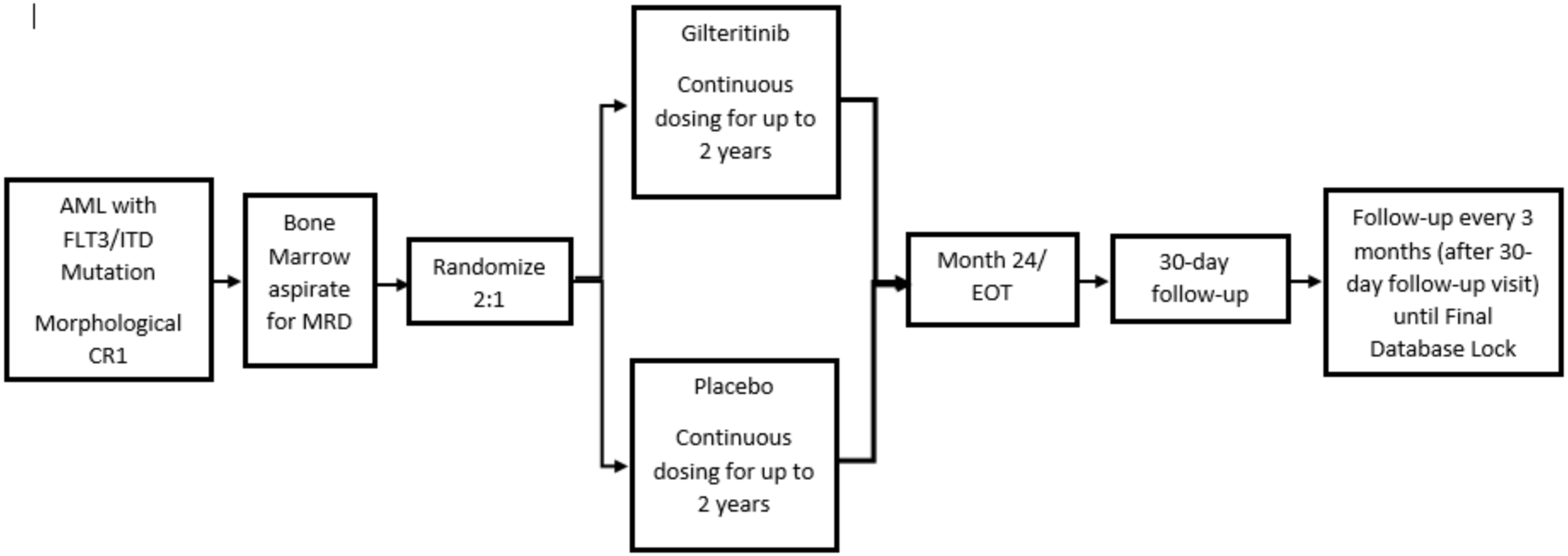
1 INTRODUCTION

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 final SAP will be approved prior to the primary database lock. Any changes from the analyses planned in the SAP will be justified in the Clinical Study Report (CSR).

2 FLOW CHART AND VISIT SCHEDULE

Flow Chart



AML: acute myeloid leukemia; CR1: first complete remission; EOT: end of treatment; FLT3: FMS-like tyrosine kinase 3; ITD: internal tandem duplication; MRD: minimal residual disease.

Note that French patients would be followed up every 3 months for 3 years after 30-day follow-up visit.

Table 1 Schedule of Assessments

Assessments	Screening	D1	D8	D9	D15	D29	Months 2-6[a]	Subsequent Visits[a,b] (every 2 months, i.e., Month 8, Month 10, etc.)	Month 24/ EOT[c]	30-Day Follow-up	Long-term Follow-up[d]
Windows	**Day -14 to -1**		**+/- 1d**		**+/-1d**	**+/- 1d**	**+/- 5d**	**+/- 10d**	**+/- 7d**	**+7d**	**+/-7d**
Signed ICF	X										
Medical and Disease History	X										
Randomization		X									
Physical Examination[e]	X	X	X		X	X	X	X	X		
Vital Signs	X	X	X		X	X	X	X	X		
ECOG Performance Status	X						X	X	X		
12-lead ECG[f] (all subjects)	X	X	X	X[g]	X	X	X	X	X		
12-lead ECG – ECG/PK sampling subset – additional time points[h]					X	X					
Chest X-ray (or CT of chest)	X										
Pregnancy Test for WOCBP[i]	X	X					X	X	X		
MUGA or ECHO[j]	X										
Clinical Laboratory Tests (chemistry, hematology, coagulation, urinalysis)[k]	X[k,l]	X	X		X	X	X[k]	X[k]	X[k]		
Thyroid Function Test[m]	X						X[m]	X[m]	X[m]		
Patient reported outcome tools[n]		X					X	X	X		
Resource utilization		X					X	X	X		
FLT3 Mutation Status (local results)[o]	X[o]										
Bone Marrow Biopsy and/ or Aspiration for disease assessment and MRD[p]	X[p]						X[p]	X[p]	X[p]		
PK Sample Collection (all subjects)[q]		X	X		X	X	X	X	X		
PK Sample Collection – ECG/PK sampling subset – additional time points					X[r]	X[r]					
PGx (whole blood and buccal swab)[s]		X									
AE/SAE Assessment	X	X	X		X	X	X	X	X	X[t]	X[u]
Prior and Concomitant Medications[v]	X[v]	X	X		X	X	X	X	X		
Survival and subsequent anti-leukemic treatments and their outcomes										X[t]	X
Gilteritinib or Placebo Dosing at the Clinic[w]		X	X		X	X	X	X			

Footnotes appear on next page

AE: adverse event; CT: computed tomography; D: day; ECG: electrocardiogram; ECHO: echocardiogram; ECOG: Eastern Cooperative Oncology Group; EOT: end of treatment; FLT3: FMS-like tyrosine kinase 3; ICF: informed consent form; MRD: minimal residual disease; MUGA: multigated acquisition scan; PGx: pharmacogenomics; PK: pharmacokinetic; QTcF: Fridericia-corrected QT interval; SAE: serious adverse event; WOCBP: women of childbearing potential.

[a] Visits should be scheduled based on day 1.

[b] After month 6, subjects will be seen every 2 months for the duration of study treatment, up to 2 years (month 8, month 10, month 12, etc.).

[c] If subject relapses and/or permanently discontinues treatment, an end of treatment visit should be conducted within 7 days of last dose.

[d] Telephone contact every 3 months after the 30-day follow-up visit. Additional contacts may be made to support key analyses. Follow-up will continue until the final database lock, which is estimated to occur after the last subject enrolled reaches the 30-day follow-up visit.

[e] Height measurement performed only at screening. Weight measurement should be performed at screening and each monthly visit.

[f] Screening ECG is required. ECG assessment will be evaluated in all subjects before dosing on day 1, day 8, day 15, day 29 and each subsequent visit. Predose assessments should be taken within 1 hour before drug administration. The 12-lead ECGs will be recorded in triplicate (3 separate ECGs with 10 minutes resting prior to first ECG and at least 5 minutes apart per time point) and transmitted electronically for central reading. The mean QTcF of the triplicate ECG tracings based on central reading will be used for all treatment decisions. If the mean triplicate QTcF is > 500 ms at any time point, the ECG will be repeated (within 2 hours if identified on machine read or as soon as possible if identified by central read).

[g] If the mean QTcF from day 1 to day 8 has increased > 30 ms with no other known etiology, a confirmatory ECG should be performed on day 9. If the day 8 and 9 ECGs confirm the > 30 ms increase in QTcF, then the investigator should assess if a dose modification should occur as per the dose interruption or reduction guideline in the protocol.

[h] For subjects participating in the ECG/PK sampling subset, additional ECGs will be performed 4 hours (+/-1 hour) postdose on day 15 and day 29. The 12-lead ECGs will be recorded in triplicate (3 separate ECGs with 10 minutes resting prior to first ECG and at least 5 minutes apart per time point) and transmitted electronically for central reading. Triplicate ECGs are to be performed prior to obtaining the time-matched PK sample (at day 15 and day 29), therefore must be started at least 10 to 15 minutes before the PK draw.

[i] Woman of childbearing potential must have a negative serum or urine pregnancy test (minimum sensitivity 25 IU/L or equivalent units of human chorionic gonadotropin).

[j] MUGA scans or ECHO are to be performed at screening for subjects with history of congestive heart failure New York Heart Association Class 3 or 4 (unless MUGA scan or ECHO performed within 1 month prior to screening revealed left ventricular ejection fraction ≥ 45%).

[k] Urinalysis is only required at screening. Additional laboratory tests may be performed according to institutional standard of care.

[l] All subjects must be randomized off of central results. Labs can be repeated during the screening period for eligibility.

[m] Thyroid function tests will be repeated after every 2 months of therapy beginning at month 2 (month 2, month 4, month 6, etc.).

[n] Includes EuroQol Group 5-dimension 5-level instrument, Functional Assessment of Cancer Therapy-Anemia and Functional Assessment of Cancer Therapy-Leukemia. If possible, subject reported outcome measures should be performed prior to any other assessments on that visit day.

[o] The documentation of a FLT3-ITD mutation in the past will be used to confirm eligibility criteria at screening. For randomized subjects, if bone marrow aspirate or peripheral blood sample and/or DNA derived from the sample at the time of diagnosis are available, additional testing of this sample may be performed.

Footnotes continued on next page

[p] Bone marrow aspirate is required and bone marrow biopsy in addition is preferred. In case of inadequate aspirate, bone marrow biopsy is required.

Bone marrow samples will be collected at screening and months 3, 6, 12, 18 and 24 and assessed by the local lab. If a subject relapses, a bone marrow sample should also be performed at the time of relapse. Only bone marrow samples showing relapse will be sent to a central lab. If bone marrow aspirate is unobtainable at relapse (e.g., dry tap), a peripheral blood smear should be collected along with bone marrow biopsy and sent to the central lab.

MRD assessment will be performed at screening and months 3, 6, 12, and 24/EOT. MRD performed at EOT should be performed only if the subject discontinued treatment for a reason other than relapse. For MRD, the first 0.25-0.75 mL of bone marrow aspirate will be collected and sent by overnight carrier to a central laboratory. A peripheral blood sample is not acceptable for MRD assessment.

If a subject relapses, FLT3 mutations will be assessed at the time of relapse by a central lab. If bone marrow aspirate is unavailable at relapse, a peripheral blood sample is acceptable to be collected for assessment of FLT3 mutations.

[q] Trough PK samples will be collected in all subjects predose (within 1 hour of dose administration) on day 1, day 8, day 15, day 29, at months 2-6 and every subsequent visit (month 8, month 10, etc.) and at EOT.

[r] For subjects participating in the ECG/PK sampling subset, additional PK sample will be collected 4 hours (+/-1 hour) postdose on day 15 and day 29.

[s] Whole blood and buccal swab collected predose on day 1 for subjects who consent to participate in the PGx study.

[t] Telephone contact with the subject is sufficient unless any assessment must be repeated for resolution of treatment-related AEs.

[u] Only SAE data that is possibly or probably related to study drug will be collected.

[v] Includes medications taken within 28 days prior to day 1.

[w] Gilteritinib or placebo is taken daily at home except for clinic days when it will be taken at the clinic.

3 STUDY OBJECTIVES AND DESIGN

3.1 Study Objectives

3.1.1 Primary Objective

The primary objective is to compare relapse-free survival (RFS) between subjects with FLT3/ITD AML in first complete remission (CR1) without transplant and who are randomized to receive gilteritinib or placebo beginning after completion of induction/consolidation chemotherapy for a two-year period.

3.1.2 Secondary Objectives

The key secondary objective is to compare overall survival (OS) in subjects treated with gilteritinib as maintenance therapy after induction/consolidation with those treated with placebo.

Additional secondary objectives are to evaluate the safety and efficacy of gilteritinib versus placebo in terms of:

- Event-free survival (EFS)
- Relationship of minimal residual disease (MRD), as determined using a next-generation sequencing (NGS) platform specific to FLT3/ITD mutations, with RFS and OS
- AEs
- Clinical laboratory
- Vital signs
- Electrocardiograms (ECGs)
- Eastern Cooperative Oncology Group (ECOG) performance scores

3.1.3 Exploratory Objectives:

The exploratory objectives are to:

- Assess relationship between gilteritinib exposure and QTcF for subjects participating in the ECG/PK sampling subset
- Explore the pharmacokinetics of gilteritinib (and metabolite, if applicable) in study population using a population pharmacokinetics approach
- Determine FLT3 mutation status at relapse
- Evaluate the safety and efficacy of gilteritinib under in terms of:
    - Patient reported signs, symptoms and impacts of AML (Functional Assessment of Cancer Therapy-Leukemia [FACT-Leu],
    - Functional Assessment of Cancer Therapy-Anemia [FACT-An])
    - EuroQol Group-5 Dimension-5 Level Instrument (EQ-5D-5L)
- Evaluate the healthcare resource utilization including hospitalization, intensive care unit (ICU) visits, emergency room (ER) visits, transfusion and use of antibiotics.

3.2 Study Design

This is a phase 2, randomized, placebo-controlled, double-blind, 2-arm study to compare the effect of gilteritinib as maintenance therapy versus placebo after induction/consolidation in subjects with FLT3/ITD AML in CR1 (including CRp and CRi). The trial will be conducted at approximately 200 centers in North America, Europe, South America, Central America, Asia/Pacific and rest of world.

Subjects in CR1 will be approached for this study after induction/consolidation therapy is complete and a decision not to proceed with transplantation is made or a suitable donor could not be identified.

Subjects will enter the screening period up to 14 days prior to the start of treatment. Subjects will be administered treatment up to 2 years, or until a discontinuation criterion is met.

After treatment completion/discontinuation, subjects will have a 30-day follow-up visit for safety, after which the subjects will enter the long-term follow-up period for collection of subsequent AML treatment including hematopoietic stem cell transplant (HSCT), remission status, and survival (cause of death and date of death).

3.3 Randomization

Enrollment, 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.

Approximately 85 subjects will be randomized in a 2:1 ratio to receive gilteritinib or placebo.

Randomization will be stratified based on:

- Age (< 60 or ≥ 60 years)
- Geographic region (North America or Europe or Asia/Pacific/Central America/South America/Rest of World)
- Presence of MRD at screening (yes or no)
- Use of FLT3 inhibiting agents during induction/consolidation (yes or no).

4 SAMPLE SIZE

The target number of randomized subjects is 85; 57 in gilteritinib arm and 28 in placebo arm.

The median RFS time in the placebo arm is assumed to be 15 months, based on RATIFY control arm data on subjects with FLT3 mutation achieving CR1 after induction and consolidation. A total of 54 relapse or death events will provide 83.2% power to detect a hazard ratio of 0.5 (corresponding to 24% difference in 2-year RFS rates) with 1-sided significance level of 0.075. The sample size estimation of 85 subjects assumes approximately 2 years of accrual.

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.

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 tables, listings, and figures. Detailed criteria for the FAS and SAF will be laid out in Classification Specifications (CS) and the allocation of subjects to FAS and SAF will be determined prior to the database hard lock.

5.1 Full Analysis Set (FAS)

The full analysis set (FAS) will consist of all subjects who are randomized and will be used for efficacy analyses. Subjects will be analyzed based on the randomized treatment.

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

5.2 Safety Analysis Set (SAF)

The safety analysis set (SAF) consists of all randomized subjects who took at least one dose of study drug (gilteritinib or placebo). 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.

5.3 Pharmacokinetics Analysis Set (PKAS)

The pharmacokinetic analysis set (PKAS) consists of the randomized subjects in SAF who were administered at least 1 dose of study drug, and for whom at least one measurable plasma concentration datum is available and for whom both the date and time of dosing on the day of PK sampling and the date and time of PK sampling is known. Additional subjects may be excluded from the PKAS at the discretion of the pharmacokineticist. Any formal definitions for exclusion of subjects or time points from the PKAS will be documented.

The PKAS will be used for all tables, listings and graphical summaries of the PK data.

At a subset of sites, subjects participating in the ECG/PK sampling subset (PKAS ECG/PK subset) will have additional PK sample collected 4 hours (± 1 hour) postdose on day 15 and day 29.

6 ANALYSIS VARIABLES

6.1 Efficacy Endpoints

Relapse after CR (including CRp and CRi), is defined as bone marrow blasts 5% or higher (not attributable to regenerating bone marrow), any circulating blasts, any extra-medullary blast foci as per Revised International Working Group (IWG) criteria. Relapse events will be

adjudicated by an independent review committee (IRC) as defined in IRC charter and will be used in the definitions of efficacy endpoints, unless specifically stated otherwise.

6.1.1 Primary Efficacy Endpoint

The primary efficacy endpoint is Relapse-free survival (RFS) per IRC adjudication, defined as the time from the date of randomization until the date of documented relapse, or death from any cause, whichever occurs first (relapse date or death date – randomization date + 1).

If a subject experiences relapse or death, the subject is defined as having RFS event related to either "relapse" or "death," and the event date is the date of relapse or death.

For a subject who is not known to have relapse or death, RFS is censored at the date of last evaluable relapse-free disease assessment (last relapse-free disease assessment date – randomization date +1). For subjects who are censored, last relapse-free disease assessment date refers to the subject's last disease assessment date. A subject who has no post-treatment disease assessment or has no study treatment will be censored at randomization date.

At each visit during treatment, disease assessment date refers to the date of bone marrow aspiration or biopsy assessment. In the event that central bone marrow assessment is not performed or bone marrow is not adequate, local bone marrow assessment date will be used. If no aspirate or biopsy is available and subject is evaluated based on blast count from peripheral blood at a visit that bone marrow is expected to be collected, the date when the peripheral blood sample is drawn will be used.

A disease assessment will be considered as not evaluable (NE), if only extramedullary leukemia is absent but neither bone marrow nor peripheral blood sample is available.

If a subject goes off study treatment without relapse but relapse is reported in eCRF during follow-up period, the IRC adjudication will still be performed. In such a case, if there is no an adjudicated relapse, the subject is defined as having RFS event related to "relapse" and the event date is the date of relapse reported in eCRF. Otherwise, the following date after end of treatment will be considered as disease assessment date and used for censoring of RFS for the subject who goes off study treatment without relapse:

- For subjects with subsequent AML therapy or HSCT: The start date of subsequent AML therapy or the date of initiation of HSCT (as reported in subsequent AML therapy eCRF), whichever is earlier.
- For subjects without subsequent AML therapy and HSCT: The last date known alive collected in survival status follow-up.

During the final analyses, only relapses or deaths occurring on or prior to the cutoff date are counted as RFS events. Subjects with relapses or deaths after the cutoff date will be censored at the last evaluable relapse-free disease assessment date on or before cutoff date.

RFS will also be defined in alternative ways to fulfill the sensitivity analyses:

- RFS will be defined similarly as above except that a subject is censored at the end of treatment.

- RFS will be defined similarly as above except that a subject is censored at the last evaluable disease assessment without relapse when relapse or death is documented after more than 1 missed disease assessment.
- RFS will be defined similarly as above by using investigator assessed relapse.

	Primary	**Sensitivity 1 (Censored at EOT)**	**Sensitivity 2 (Missing >1 assessment)**
Scenarios	**Indicator [Event Type]; RFS date**	**Indicator [Event Type]; RFS date**	**Indicator [Event Type]; RFS date**
Relapsed confirmed by adjudication*	Event [type=relapse]; Date of relapse	Event [type=relapse]; Date of relapse; If relapse after EOT, Non-Event [type=censor]; Date of EOT	Event [type=relapse]; Date of relapse; If relapse after ≥2 missed assessments, Non-Event [type=censor]; Date of last assessment before missing assessment
Relapsed after EOT	Event [type=relapse]; Date of relapse	Non-Event [type=censor]; Date of EOT	Non-Event [type=censor]; Date of last assessment
Death without relapse	Event [type=death]; Date of death;	Event [type=death]; Date of death; If death after EOT or no treatment, Non-Event [type=censor]; Date of EOT or randomization, respectively	Event [type=death]; Date of death; If death after EOT or no treatment or death after 2 missed assessments, Non-Event [type=censor]; Date of last assessment or randomization or date of last assessment before missing assessment, respectively
Not treated, no relapse, no death	Non-Event [type=censor]; Date of randomization	Non-Event [type=censor]; Date of randomization	Non-Event [type=censor]; Date of randomization
Treated, no post-baseline disease assessment, no relapse, no death	Non-Event [type=censor]; Date of randomization	Non-Event [type=censor]; Date of randomization	Non-Event [type=censor]; Date of randomization
Treated, ≥1 post-baseline disease assessment, no relapse, no death, and ended study treatment: Initiation of subsequent AML therapy including HSCT	Non-Event [type=censor]; Date at first initiation of subsequent AML therapy	Non-Event [type=censor]; Date of EOT	Non-Event [type=censor]; Date of last disease assessment
Treated, ≥1 post-baseline disease assessment, no relapse, no death, and ended study treatment: No initiation of subsequent AML therapy including HSCT	Non-Event [type=censor]; Date of last known alive date	Non-Event [type=censor]; Date of EOT	Non-Event [type=censor]; Date of last disease assessment
Treated, ≥1 post-baseline disease assessment, no relapse and ongoing with treatment	Non-Event [type=censor]; Date of last disease assessment	Non-Event [type=censor]; Date of last disease assessment	Non-Event [type=censor]; Date of last disease assessment

* RFS defined by using investigator assessed relapse will be the same as "Primary" except for replacing "Relapsed confirmed by adjudication" with "Relapsed assessed by Investigator".

EOT: end of treatment

6.1.2 Secondary Efficacy Endpoints

6.1.2.1 Key Secondary Efficacy Endpoints

- Overall survival (OS)

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, OS is censored at the date of last known alive (date of last known alive – randomization date + 1). The date of last known alive is the latest date that the subject is confirmed to be alive across the entire database.

During the final analyses, 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.

As a sensitivity analysis, OS will be defined similarly as above, however, subjects initiate other anti-leukemic treatment or undergo HSCT will be censored at the first HSCT and first subsequent anti-leukemic treatment, whichever occurs first.

6.1.2.2 Other Secondary Efficacy Endpoints

- Event-free Survival (EFS)

EFS is defined as the time from the date of randomization until the date of documented relapse (adjudicated), or premature discontinuation of the treatment before 2 years, or initiation of other anti-leukemic treatment, or death from any cause, whichever occurs first [earliest of (relapse date, treatment discontinuation date, start date of other anti-leukemic treatment, death date) – randomization date + 1].

If a subject experiences relapse or death, the subject is defined as having EFS event related to either “relapse” or “death”, and the event date is the date of relapse or death.

If a subject discontinues the treatment or initiates other anti-leukemic treatment, the subject is defined as having EFS event related to either “treatment discontinuation” or “initiation of other anti-leukemic treatment”, and the event date is the date of study drug discontinuation or start date of other anti-leukemic treatment, respectively. HSCT won’t be considered as an EFS event.

If multiple events occur on the same date, take the order of “relapse”, “death”, “treatment discontinuation”, and “initiation of other anti-leukemic treatment”. For example, if a subject has the same date for EOT and death, then the event should be “death”.

For a subject who is not known to have relapse or death or treatment discontinuation or initiation of other anti-leukemic treatment, EFS is censored at the last evaluable relapse-free disease assessment date. If a subject undergoes an HSCT without these events, EFS will be censored at the last evaluable relapse-free assessment or the first HSCT, whichever occurs later.

During the final analyses, only death, relapse, treatment discontinuation or initiation of other anti-leukemic treatment occurring on or prior to the cutoff date are counted as EFS events.

Subjects without these events before cutoff date will be censored at the last relapse-free disease assessment date or the first HSCT on or before the cutoff date, whichever occurs first.

- Minimal Residual Disease (MRD)

MRD will be analyzed from bone marrow samples by a Sponsor-designated central laboratory.

FLT3/ITD mutation ratio will be measured in relation to total FLT3. For a patient with multiple ITD mutations, the overall FLT3/ITD mutation ratio will be calculated from the sum of all ITD mutations. Changes in FLT3/ITD mutation ratio from baseline will be compared.

At each visit, the presence of MRD will be "Yes" if $\log_{10}$-transformed overall FLT3/ITD mutation ratio is greater than -4; otherwise, the presence of MRD will be "No".

6.1.3 Exploratory Endpoints

- FLT3/ITD mutation status at the time of AML diagnosis

FLT3/ITD mutation status of the diagnostic bone marrow or blood sample will be defined for local and central results, respectively.

- FLT3 mutation status at relapse

FLT3 mutation status at relapse will be assessed.

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

The EQ-5D-5L is being used as a measure of respondents' health-related quality of life. The EQ-5D-5L consists of the EuroQol Group-5 Dimension descriptive system and the EuroQol Group Visual Analogue Scale (VAS).

The EuroQol Group-5 Dimension descriptive system comprises of 5 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 'best imaginable health state' and 'worst imaginable health state' with higher scores for higher health related quality of life.

- Functional Assessment of Cancer Therapy-Leukemia (FACT-Leu)
- The FACT-Leu [Cella et al, 2012] 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. Functional Assessment of Cancer Therapy-Anemia (FACT-An)

- The FACT-An [Cella D., 1997] is to assess specific quality of life concerns related to anemia and fatigue in patients with cancer. It consists of 13 items measuring fatigue associated with anemia and 7 items measuring other symptoms and impacts associated with anemia, such as shortness of breath and dizziness. It also includes the same generic physical well-being, social/family well-being, emotional well-being, functional well-being domains as the FACT-Leu. These redundant scales will not be administered as part of the FACT-An; only the 20 anemia-specific items will be administered.
- Healthcare resource utilization including hospitalization, ICU visits, ER visits, transfusion and use of antibiotics

6.2 Safety Variables

Safety will be assessed by evaluation of the following variables:

- Treatment-emergent adverse events (TEAEs; frequency, CTCAE grade, seriousness, and relationship to study drug)
  - TEAE is defined as an AE observed after starting administration of the study treatment (gilteritinib or placebo) until 30 days from the last study treatment. If the AE 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 AE will be considered treatment emergent. If the AE occurs on Day 1 and the onset check box is marked "Onset before first dose of study drug", then the AE 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 a TEAE only if it has worsened in severity (i.e. it is reported with a new start date). 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 treatment 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 AEs 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 gilteritinib (as specified in Section 10.2, Appendix 2).
- Clinical laboratory variables (hematology, biochemistry, 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 gilteritinib will be used in pharmacokinetic analysis.

Sparse pharmacokinetic samples will be collected in all subjects as outlined in the Schedule of Assessments [Table 1].

At a subset of sites, approximately 40 subjects (targeting approximately 27 subjects in the gilteritinib arm and 13 subjects in the placebo arm) participating in the ECG/PK sampling subset will have additional PK sample collected 4 hours (± 1 hour) postdose on day 15 and day 29.

6.4 Other Variables

- Body Mass Index (BMI)

  BMI = weight (kg) / [height (m)]^2
- Duration of exposure (days)

  Duration of exposure = date of last dose – date of first dose + 1
- Number of dosing days (days)

  Number of dosing days = Date of last dose – Date of first dose + 1 – number of days without drug administration in between
- Cumulative dose (mg)

  Cumulative dose = total dose of study drug taken during the study
- Average daily dose (mg/day)

  Average daily dose = cumulative dose (mg)/number of dosing days (days)
- Dose intensity (mg/day)

  Dose intensity = cumulative dose (mg)/duration of exposure (days)
- Planned dose intensity (mg/day)

  Planned daily dose = 120 mg/day
- Relative dose intensity (%)

  Relative dose intensity = dose intensity (mg/day)/planned daily dose (mg/day)*100%
- Time from initial diagnosis of AML to time of randomization

  It 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 treatment (medication and non-medication therapy)

  Previous treatment is defined as treatment administered before the date of first dosing (exclusive).

  Concomitant treatment is defined as treatment administered between the date of first dose (inclusive) and the date of last dose (inclusive) of study drug.
- 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.

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

7 STATISTICAL METHODOLOGY

7.1 General Considerations

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 mentioned in the relevant section. In addition, for plasma concentrations, the coefficient of variation and the geometric mean 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 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 decrease, actual treatment refers to the first dose received before dose change.

All statistical comparisons will be made using 1-sided tests at α=0.075 significance level unless specifically stated otherwise.

All data processing, summarization, and analyses will be performed using SAS® Version 9.3 or higher on Unix. 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 before or at the first dose date of study drug, or before or at the randomization (for subjects who receive no study drug). Unless otherwise specified, all summaries will be presented by treatment group.

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

7.2 Study Population

7.2.1 Disposition of Subjects

The following subject data will be presented:

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

For all randomized subjects, the following subject data will be presented by treatment group:

- Number and percentage of subjects in each analysis set;
- Number and percentage of subjects who completed or discontinued treatment by primary reason for treatment discontinuation;

- Number and percentage of subjects who completed or discontinued the 30-day follow-up evaluation by status;
- Number and percentage of subjects who completed or discontinued the long term follow-up evaluation by status

7.2.2 Protocol Deviations

Protocol deviations (PDs) 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 mass index (BMI), and height at study entry will be presented. Frequency tabulations for sex, ethnicity, race, geographic region, age group, presence of MRD at screening, Use of FLT3 inhibiting agents during induction/consolidation, cytogenetic risk status and baseline ECOG (defined in Section 7.8) will be presented 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-ITD mutation status and FLT3 point mutation status will be presented by treatment group.

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 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.

FLT3/ITD mutation status of the diagnostic bone marrow or blood sample will be summarized for local results and central results.

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 [ATC] 2nd level) and chemical subgroup (ATC 4th level) and preferred WHO name (PT) by treatment group.

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 PT by treatment group and presented in decreasing order of frequency based on the total number of subjects who took each medication.

7.2.5 Previous and Concomitant Transfusions

Frequency tabulations of subjects who received transfusions and blood product will be presented for previous transfusion and concomitant transfusion by treatment group. 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. Descriptive statistics will be presented for duration of response to prior AML therapy before randomization.

7.2.7 Previous and Concomitant Non-Medication Therapy

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

7.2.8 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. 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 by treatment group for the SAF:

- Descriptive statistics for cumulative dose, average daily dose, dose intensity, relative dose intensity, and
- Number and percentage of subjects with dose decrease or interruption.

Duration of exposure and number of dosing days will be summarized by treatment group in two ways:

- Descriptive statistics will be presented.
- Exposure time will be categorized according to the following categories:
  - less than 3 months
  - at least 3 months, less than 6 months
  - at least 6 months, less than 12 months
  - at least 12 months, less than 18 months
  - 18 months or more
  - Unknown.

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

Listing of subjects with dose decrease or interruption will also be provided for SAF.

7.3.2 Treatment Compliance

Overall compliance (including relative dose intensity) with the dosing schedule will be examined for subjects in the SAF.

7.4 Analysis of Efficacy

7.4.1 Analysis of Primary Endpoint

7.4.1.1 Primary Analysis of the Primary Endpoint

The primary efficacy analysis will be performed on the FAS to compare the RFS between gilteritinib and placebo. The null and alternative hypotheses for this comparison are:

- H_{01}: RFS in the gilteritinib arm is worse than or equal to RFS in the placebo arm
- H_{11}: RFS in the gilteritinib arm is better than RFS in the placebo arm.

The analysis will be conducted using the stratified log-rank test with randomization strata as described in Section 3.3.

The hazard ratio of the treatment effect along with 95% confidence interval will be presented.

The assumption of proportionality of the hazard rates will be assessed.

The SAS codes to perform the log-rank test with strata will be similar to that shown below:

```
PROC LIFETEST;
        TIME time * status (1);
        STRATA stratification_variables/ GROUP = treatment;
RUN;
```

Kaplan-Meier survival plots will be used to describe the RFS in each treatment group. Median RFS and RFS rates at 1, 2 and 3 years with 95% confidence interval will be estimated from the Kaplan-Meier plots using the SAS code similar to that shown below:

```
PROC LIFETEST;
        TIME time * status (1);
        STRATA treatment;
RUN;
```

Median follow-up of RFS estimated from reverse Kaplan-Meier curve will be provided.

7.4.1.2 Sensitivity analyses of the Primary Endpoint

The sensitivity analyses for the primary efficacy endpoint of RFS will be performed as described below:

- Unstratified log-rank test on the FAS
- Stratified Cox proportional hazards model with the same strata as in primary analysis on the FAS
- Same analysis as primary analysis on the FAS, but RFS is censored at the last evaluable disease assessment without relapse when relapse or death is documented after more than 1 missed disease assessment.
- Same analysis as primary analysis on the FAS, but RFS is censored at end of treatment.
- Same analysis as primary analysis on the FAS, but RFS is defined by using investigator assessed relapse.

7.4.2 Analysis of Secondary Endpoints

7.4.2.1 Key Secondary Efficacy Analysis

The key secondary efficacy endpoint of OS will be analyzed on the FAS in the same manner as RFS using the stratified log-rank test with the same strata as described in Section 7.4.1.1. The null and alternative hypotheses for this comparison are:

- H_{02}: OS in the gilteritinib arm is worse than or equal to OS in the placebo arm
- H_{12}: OS in the gilteritinib arm is better than OS in the placebo arm.

With the sequential multiple test procedure to control for overall type I error, formal significance testing of OS will be conducted only if the RFS comparison is statistically significant. Otherwise, OS analysis will be considered exploratory.

Kaplan-Meier curves will be used to describe the OS in each arm. Median OS time and OS rates at 1, 2, 3 and 4 years along with 95% confidence interval will be estimated from the Kaplan-Meier curves.

For the key secondary endpoint of OS, sensitivity analyses will be performed as below:

- Unstratified log-rank test on the FAS
- Stratified Cox proportional hazards model with the same strata as in primary analysis on the FAS
- Same analysis as primary analysis on the FAS, but OS is censored at the first HSCT and first subsequent anti-leukemic treatment, whichever occurs first.

7.4.2.2 Other Secondary Efficacy Analysis

- Analysis of EFS

EFS will be analyzed in the same manner as RFS in the FAS using the stratified log-rank test with the same strata as described in Section 7.4.1.1. The hazard ratio of the treatment effect along with 95% confidence interval will be calculated. Kaplan-Meier curve will be created for each treatment group. Median time and 95% confidence interval, EFS rates at 1, 2 and 3 years if appropriate and 95% confidence interval will be estimated from the Kaplan-Meier curve.

- Analysis of MRD

MRD will be analyzed in two ways measured as $\log_{10}$-transformed overall FLT3/ITD mutation ratio and presence/absence of MRD as defined in Section 6.1.2.2.

Analysis of covariance (ANCOVA) model will be used to analyze the change from baseline of MRD measured as $\log_{10}$-transformed overall FLT3/ITD mutation ratio for post-baseline visits including treatment, age group, geographic region and use of FLT3 inhibiting therapy during induction/consolidation as fixed factors, and $\log_{10}$-transformed overall FLT3/ITD mutation ratio at baseline as covariate.

The Cochran-Mantel-Haenszel (CMH) method will be used to measure the difference between treatment groups for the proportion of subjects with the absence of MRD while controlling for age group, geographic region and use of FLT3 inhibiting therapy during induction/consolidation.

7.4.3 Analysis of Exploratory Endpoints

FLT3/ITD mutation status at the time of AML diagnosis will be summarized by the number and percentage of subjects by treatment group on FAS for local and central results, respectively.

FLT3 mutation status at relapse will be summarized by the number and percentage of subjects by treatment group on FAS.

Incidence of resource utilization including hospitalization, blood transfusion, antibiotic intravenous infusions, ICU visit and ER visit will be summarized by treatment group on FAS.

The difference between treatment groups will be tested using CMH test while controlling of stratification factors.

Resource utilization counts including duration of hospital stays, duration of medications, blood transfusions, antibiotic intravenous infusions, ICU visit and ER visit will be summarized by treatment group using descriptive statistics (mean, standard deviation, minimum, maximum and median) on FAS. The difference between treatment groups will be tested with ANOVA while controlling of stratification factors.

EQ-5D-5L VAS will be summarized by treatment group using descriptive statistics (mean, standard deviation, minimum, maximum and median) on FAS. Additionally, change from baseline 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 for the EQ-5D-5L VAS including treatment and stratification factors as fixed factors, and baseline score as covariate. 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).

FACT-Leu global score and domain scores will be summarized by treatment group using descriptive statistics (mean, standard deviation, minimum, maximum and median) on FAS. Additionally, change from baseline will be calculated as the post-baseline value minus the baseline value and summarized in the same way. ANCOVA model will be used to evaluate change from baseline to post-baseline for the global and domain scores, individual items and item clusters of the FACT-Leu treatment and stratification factors as fixed factor, and baseline score as covariate.

FACT-An score will be summarized by treatment group using descriptive statistics (mean, standard deviation, minimum, maximum and median) on FAS. Additionally, change from baseline will be calculated as the post-baseline value minus the baseline value and summarized in the same way. ANCOVA model will be used to evaluate change from baseline to post-baseline scores including treatment and stratification factors as fixed factor, and baseline score as covariate.

Additional exploratory analyses of patient reported outcomes (FACT-An, FACT-Leu) and heath outcome (EQ-5D-5L) may be performed and the details may be included in a separate Patient Reported Outcome Statistical Analysis Plan.

7.5 Analysis of Safety

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

7.5.1 Adverse Events

Any adverse event (AE) recorded on treatment including within 30 days from the last study treatment will be classified as treatment-emergent AE (TEAE) and will be summarized.

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) version 4.03.

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 will include the following details:

- TEAEs,
- Drug-related TEAEs,
- TEAEs before investigator assessed relapse,
- Drug-related TEAEs before investigator assessed relapse,
- 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,
- TEAEs leading to dose reduction,
- Drug-related TEAEs leading to dose reduction,
- TEAEs leading to dose interruption,
- Drug-related TEAEs leading to dose interruption,
- Grade 3 or higher TEAEs,
- Drug-related Grade 3 or higher TEAEs,
- Any death.

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 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,
- TEAEs before investigator assessed relapse,
- Drug-related TEAEs before investigator assessed relapse,
- 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,
- TEAEs leading to dose reduction,
- Drug-related TEAEs leading to dose reduction,
- TEAEs leading to dose interruption,
- Drug-related TEAEs leading to dose interruption,
- Grade 3 or higher TEAEs,
- Drug-related Grade 3 or higher TEAEs,
- Death

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

- TEAEs
- Drug-related TEAEs,

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 following:

- TEAEs with special safety interest,
- Serious TEAEs with special safety interest,
- TEAEs with special safety interest leading to withdrawal of treatment,
- Drug related-TEAEs with special safety interest leading to withdrawal of treatment,
- TEAEs with special safety interest leading to dose reduction,
- Drug-related TEAEs with special safety interest leading to dose reduction,
- TEAEs with special safety interest leading to dose interruption,
- Drug-related TEAEs with special safety interest leading to dose interruption,
- Grade 3 or higher TEAEs with special safety interest,
- Drug-related Grade 3 or higher TEAEs with special safety interest

AE summary tables will include subject counts as opposed to AE counts. If a subject experiences more than one episode of a particular AE, the subject will be counted only once for that AE. If a subject has more than one AE that codes to the same PT, the subject will be counted only once for that PT. Similarly, if a subject has more than one AE within a SOC, the subject will be counted only once in that SOC.

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 value is the last measurement taken prior to the first study drug administration.

Quantitative clinical laboratory variables, i.e., hematology, biochemistry, coagulation and urinalysis will be summarized using mean, standard deviation, minimum, maximum and median for each treatment group at each visit. Additionally, change from baseline will be calculated as the post-baseline value minus the baseline value 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 central laboratory 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 changes based on reference range defined category (low, normal, high) from baseline to each visit as well as worst finding during the treatment period, and
- Summary shifts of changes based on reference range defined category from baseline to each visit as well as worst finding during the treatment period (shift from high or normal to low, shift from low or normal to high, categorized increase [shift from low to normal, from 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, from high to low or from normal to low]).

Laboratory results will also be graded using NCI-CTCAE v4.03, 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 at 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. 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 post-baseline 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 will be presented by treatment group.

7.5.3 Vital Signs

The baseline value is the last measurement taken prior to the first study drug administration.

Vital signs (systolic blood pressure [SBP], diastolic blood pressure [DBP], pulse rate, and body temperature) will be summarized using mean, standard deviation, minimum, maximum and median by treatment group and visit. Additionally, change from baseline will be calculated as the post-baseline value minus the baseline value and summarized by treatment group and visit.

Tables for potentially clinically significant vital signs will be generated using baseline value and highest post-baseline value 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

7.5.4 Electrocardiograms (ECGs)

12-lead ECGs will be recorded pre-dose (within 1 hour before study drug) at the scheduled visits for all subjects.

Additional ECGs will be collected on day 15 and day 29 at 4 hours post-dose (+/- 1 hour) for subjects participating in the ECG/PK sampling subset.

Each ECG will be recorded in triplicate and transmitted electronically to central reading.

The mean of the triplicate ECGs from central read should be used for all final treatment decisions, AE reporting and in the summary for analysis at each visit.

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

Number and percentage of subjects with normal and abnormal results as assessed by central read for the overall interpretation will be tabulated by treatment group at each visit. A shift analysis table showing shift in overall ECG interpretation from baseline to each visit 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 visit for values of clinical importance using the range criteria below.

	QTcF 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

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 visit.

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

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.

All ECG analyses will also be performed for subjects participating in the ECG/PK sampling subset in addition to SAF.

7.5.5 Pregnancies

A detailed listing of all pregnancies will be provided.

7.5.6 Eastern Cooperative Oncology Group (ECOG) Performance Scores

Number and percent of subjects for each category of the ECOG performance scores 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 performance scores will also be summarized using shift table from baseline to post-baseline score for each treatment group.

7.6 Analysis of PK

PK analysis will be conducted on the PKAS.

7.6.1 Analysis of Sparse Pharmacokinetic Samples

The plasma concentrations of sparse pharmacokinetic samples will be summarized for gilteritinib by visit and nominal time points using descriptive statistics, including number of subjects, mean, standard deviation, minimum, median, maximum, geometric mean, and coefficient of variation of the mean and geometric mean.

Population pharmacokinetic analysis will be performed. Data from this study may be pooled with other studies for analysis. Details of this analysis will be specified in a separate population pharmacokinetic analysis plan.

7.6.2 Analysis of Pharmacokinetic for ECG/PK sampling subset

For subjects participating in the ECG/PK sampling subset with additional pharmacokinetic samples collected, plasma concentrations will be summarized for gilteritinib by visit and nominal time points using descriptive statistics as described in Section 7.6.1. Time-course of drug concentrations will be plotted as appropriate.

Population pharmacokinetic/pharmacodynamic analysis will be performed to assess the relationship between gilteritinib exposure and QTcF. Data from this study may be pooled with other studies for analysis. Details of this analysis will be specified in a separate population pharmacokinetic/pharmacodynamic analysis report.

7.7 Analysis of Pharmacodynamics

Not applicable.

7.8 Subgroups of Interest

Primary efficacy endpoint (RFS) and key secondary endpoint (OS) will be summarized by treatment group for the subgroups defined on the basis of the categorized variables listed below:

Grouping Variables	Subgroups
Age	< 60 years ≥ 60 years
Sex	Female Male
Race	White Black or African American Asian Other
Baseline ECOG	0-1 ≥ 2
Geographic region	North America Europe Asia/Pacific/South and Central America/Rest of World
Presence of screening MRD	Yes No
Use of FLT3 inhibiting therapy during induction/consolidation	Yes No
Cytogenetic Risk Status	Favorable Intermediate Unfavorable Other

7.9 Other Analyses

Not applicable.

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

Not applicable.

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 RFS and key secondary endpoint OS, missing or incomplete death date will be imputed as the earliest feasible date on or after the date of last known alive as the examples shown in the table below. The date of last known alive will be obtained as described in Section 6.1.1.

Incomplete Date of Death (YYYY MMM DD)	**Date of Last –Known Alive (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.

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	2	1	n/a	1	1
	≠ 1st dose *yyyymm*		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

*For Subsequent AML therapy, the partial date with year only will be imputed as EOT date.

The imputed dates will be used to determine whether an AE is 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. Concentrations below the lower limit of quantification (LLOQ) in PK should be assigned to be zero in the analysis of pharmacokinetic.

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. 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 RFS and OS analysis, if there is no event in at least one stratum combinations, or the Cox proportional hazard model does not converge due to small event size, the stratum combinations will be pooled in the order of geographic region, age group, use of FLT3 inhibiting agents during induction/consolidation, and screening MRD status until the issue is resolved or the normal (un-stratified) Cox proportional hazard model is applied.

All sensitivity analyses for the RFS and OS endpoints will apply the pooled strata. In the case that the criterions don't meet, the un-stratified analysis will be used.

7.11.5 Blinding

Although the unblinding of the study drug could be considered for participant safety and/or evidence of documented relapse contingent upon knowing the blinded study drug assignment (more details can be found in protocol Section 4.4.2), to maintain trial integrity and increase the credibility of study results, the study statistician, supporting statisticians and statistical programmers will have no access to the randomized treatment information before the database lock.

8 DOCUMENT REVISION HISTORY

Version	**Date**	**Changes**	**Comment/rationale for change**
1.0	10-Oct-2016	NA	Document finalized
2.0	29-Apr-2019	Updated Flow Chart, Schedule of Assessment and footnote	Updated to keep consistent with protocol version 3.0 (amendment 2) dated 25-Apr-2019.
		Updated Section 3 Study objectives and design with the following changes: • Reduced sample sizes from 354 to 85 and number of RFS events needed from 182 to 54 • Decreased Hazards ratio from 0.6 to 0.5 • Increased 1-sided alpha from 0.025 to 0.075 • Updated Power as 83.2% • Removed the interim analysis for sample size re-estimation	Updated due to Protocol version 3.0 (amendment 2) dated 25-Apr-2019.
		Updated Section 5.4 Pharmacokinetics Analysis Set (PKAS) to clarify both dosing date/time and sampling date/time must be available.	To clarify the detailed PKAS criteria.
		Updated Section 6.1.1 to clarify the definition of RFS: • Clarified the handling strategy while a subject goes off treatment without relapse that the relapse reported during long-term follow-up will be used for primary analysis of RFS • Clarified the RFS censoring rules to include the assessment date after end of treatment to censor RFS	Previous definition of RFS was not clear while a subject goes off treatment without relapse.
		Updated Section 6.1.2.1: • Clarified the definition of last date known alive and move it under OS endpoint	Previous definition of last date known alive didn't cover all possible dates to support subject alive status
		Updated Section 6.2 and added Section 10.2 to clarify Adverse events of special safety interest (AESI) searching strategy	Include AESI for more comprehensive analysis on AEs.

Version	Date	Changes	Comment/rationale for change
		Updated Section 6.3 for numbers of PK subset	Updated due to protocol version 3.0 (amendment 2) dated 25-Apr-2019.
		Updated Section 7.2 Study Population to add the following sections: • Added the section of prior AML chemotherapy • Added the section of subsequent AML therapy	Add these sections to keep consistent with CRF data collected to summarize the prior and subsequent AML therapies.
		Updated Section 7.2.3 and 7.8 to add subgroup of cytogenetic risk status for demographics and subgroup analysis.	Add the new subgroup per knowledge learned in ASP2215 project
		Updated Section 7.4.1, for all time-to-event analysis (RFS, OS and EFS): • Updated the primary analysis method as stratified log-rank test instead of stratified Cox proportional hazards model • Used the original primary analysis based on stratified Cox proportional hazards model as sensitivity	Updated due to protocol version 3.0 (amendment 2) dated 25-Apr-2019.
		Updated Section 7.5.1 to include additional summaries of AEs and AE summaries by patient-year.	Updated per knowledge learned in ASP2215 project
		Updated Section 7.5.2 to add the grade 3 and 4 summary for selected lab test parameters	Updated per knowledge learned in ASP2215 project
		Updated Section 7.5.4 to add ECG abnormalities	Updated per knowledge learned in ASP2215 project
		Updated Section 7.10 Interim Analysis as Not Applicable	Updated due to protocol version 3.0 (amendment 2) dated 25-Apr-2019.

Version	**Date**	**Changes**	**Comment/rationale for change**
		Updated Section 7.11: • Added how to handle partial date in initial diagnosis of AML and subsequent AML therapy; • Modified the pooling strata section to clarify how strata pooling will be performed as appropriate. • Added the blinding section	For clarification purpose
		Updated Section 7.11: • Removed the visit window table	It is redundant
3.0	12-Aug-2021	Removed PPS and corresponding sections and add necessary sensitivity analysis to section 7.4 as appropriate	The company did not endorse this analysis.
		Added the following summaries for overview and by SOC/PT: • TEAEs before investigator assessed relapse • Drug-related TEAEs before investigator assessed relapse,	Was requested by IDMC.
		Added a table to show the definition of RFS and RFS variables for sensitivity analyses	Easy for programming team to follow
		Added order of Event for EFS if multiple events on the same day	Multiple events might have occurred on the same day
		Removed treatment compliance	The calculation was based on unreliable drug accountability data
		Removed subgroup analysis per pooled strata	Due to small sample size, the pooled strata may end up with at most one factor which has ready have subgroup analysis

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.

Cella D. The Functional Assessment of Cancer Therapy-Anemia (FACT-An) Scale: a new tool for the assessment of outcomes in cancer anemia and fatigue. Semin Hematol. 1997 Jul;34(3 Suppl 2):13-9.

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)

10 APPENDICES

10.1 Appendix 1: Key Contributors and Approvers

List of Key Contributors and Approvers

Key Contributors

The following contributed to or reviewed this Statistical Analysis Plan as relevant to their indicated discipline or role.

Primary author (s)

PPD, MS PPD	Oncology Biostatistics Data Science Astellas Pharma Global Development, Inc.

Contributors and Reviewers

PPD, PhD PPD	Oncology Biostatistics Data Science Astellas Pharma Global Development, Inc.
PPD	Statistics Labcorp
PPD MD PPD	Medical Science Astellas Pharma Global Development, Inc.
PPD, MD PPD	Medical Science Astellas Pharma Global Development, Inc.
PPD	Clinical Sciences Astellas Pharma Global Development, Inc.
PPD PhD PPD	Clinical Pharmacology & Exploratory Development Astellas Pharma Global Development, Inc.
PPD, PhD PPD	Clinical Pharmacology & Exploratory Development Astellas Pharma Global Development, Inc.
PPD, PhD PPD	Clinical Pharmacology & Exploratory Development Astellas Pharma Global Development, Inc.

PPD, MPH PPD	Global Health Economics & Outcomes Research (HEOR) – Oncology Global Medical Affairs Astellas Pharma Global Development, Inc.
PPD, PhD PPD	Medical Writing Astellas Pharma Global Development, Inc

Author and Approver Signatories

(E-signatures are attached at end of document)

PPD /MS, PPD, Oncology Biostatistics/Global Data Science was the study statistician for this study and the primary author of this Statistical Analysis Plan

PPD PhD, PPD, Oncology Biostatistics/Global Data Science was the global statistical lead (STATL) for this project and biostatistics peer reviewer of this Statistical Analysis Plan

This Statistical Analysis Plan was approved by:

PPD, MD
PPD, Global Medical Science

10.2 Appendix 2: 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*	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

*Only AEs occur within the first 90 days