STATISTICAL ANALYSIS PLAN

Final Version 2.0, dated 21 Feb 2023

A Multi-center, Randomized, Double-blind, Placebo-controlled Phase III Trial of the FLT3 Inhibitor Gilteritinib Administered as Maintenance Therapy Following Allogeneic Transplant for Patients with FLT3/ITD AML

ISN/Protocol: 2215-CL-0304

IND number: 117,548

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 Gilteritinib
AST	Aspartate Transaminase
ATC	Anatomical Therapeutic Chemical
BERC	Blinded endpoint review committee
BMI	Body Mass Index
BM	Bone marrow
BMT	Bone marrow transplant
BMT-CTN	Bone Marrow Transplant Clinical Trial Network
BSA	Body surface area
CIBMTR	Center for International Blood and Marrow Transplant Research
CMH	Cochran-Mantel-Haenszel
CMV	Cytomegalovirus
CR	Complete remission
CR1	First complete remission
CS	Classification Specifications
CSR	Clinical Study Report
CTCAE	Common Terminology Criteria for Adverse Events
CV	Coefficient of variation
DBP	Diastolic blood pressure
DLI	Donor lymphocyte infusion
DSMB	Data and Safety Monitoring Board
ECG	Electrocardiogram
ECHO	Echocardiogram
eCRF	Electronic case report form
EDC	Electronic Data Capture
EFS	Event-free survival
EQ-5D-5L	EuroQol Group-5 dimension-5 level
FACT-BMT	Functional Assessment of Cancer Therapy-Bone Marrow Transplant
FACT-G	Functional Assessment of Cancer Therapy-General

Abbreviations	Description of abbreviations
FACT-Leu	Functional Assessment of Cancer Therapy-Leukemia
FLT3	FMS-like tyrosine kinase
FPI	First participant in
GI	Gastrointestinal
GRFS	GVHD-free relapse-free survival
GVHD	graft-versus-host disease
H	High
HSCT	Hematopoietic stem cell transplant
HLA	Human Leucocyte Antigen
HQL	Health-related quality of life
IAP	Interim Analysis Plan
ICH	International Conference on Harmonization
ICU	Intensive care unit
IPCW	Inverse probability of censoring weights
IRT	Interactive Response Technology
ISN	International Study Number
ITD	Internal tandem duplication
ITT	Intent-to-treat analysis set
IU/L	International units/liter
L	Low
MedDRA	Medical Dictionary for Regulatory Activities
mg	Milligrams
Min	Minute
mL	Milliliter
mmHg	millimeters of mercury
MOP	Manual of procedures
MRD	Minimal residual disease
msec	Milliseconds
MUGA	Multigated acquisition scan
N	Number
NCI	National Cancer Institute
NE	Not Evaluable
NGS	Next-generation sequencing
NHLBI	National Heart, Lung, and Blood Institute
NR	Non-Response
NRM	Non-relapse mortality
OS	Overall Survival
PB	Peripheral blood
PCR	Polymerase chain reaction
PD	Protocol Deviation
PK	Pharmacokinetic
PKAS	Pharmacokinetic analysis set

Abbreviations	Description of abbreviations
PT	Preferred Term
QTc	QT interval corrected for heart rate
QTcF	Fridericia-corrected QT interval
RBC	Red blood cell
R-IWG	Revised International Working Group
RDI	Relative Dose Intensity
RFS	Relapse-free survival
RMST	Restricted mean survival time
RPFST	Rank-preserving structural failure time method
RR	Interval between 2 consecutive r waves on an ECG
SAE	Serious adverse event
SAF	Safety analysis set
SAP	Statistical analysis plan
SBP	Systolic blood pressure
SOC	System Organ Class
SUSAR	Suspected unexpected serious adverse reaction
TEAE	Treatment Emergent Adverse Event
TLF	Tables, Listings and Figures
ULN	Upper limit of normal
US	United States
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 participant has been enrolled, the clinical trial protocol applies to the participant.
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 participant 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.
Screening period	Period of time before entering the investigational period, usually from the time of starting a participant signing consent until just before the test drug or comparative drug (sometimes without randomization) is given to a participant.
Randomization	The process of assigning trial participants 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 participants for enrollment in a trial.
Screen failure	Potential participant who did not meet 1 or more criteria required for participation in a trial.
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

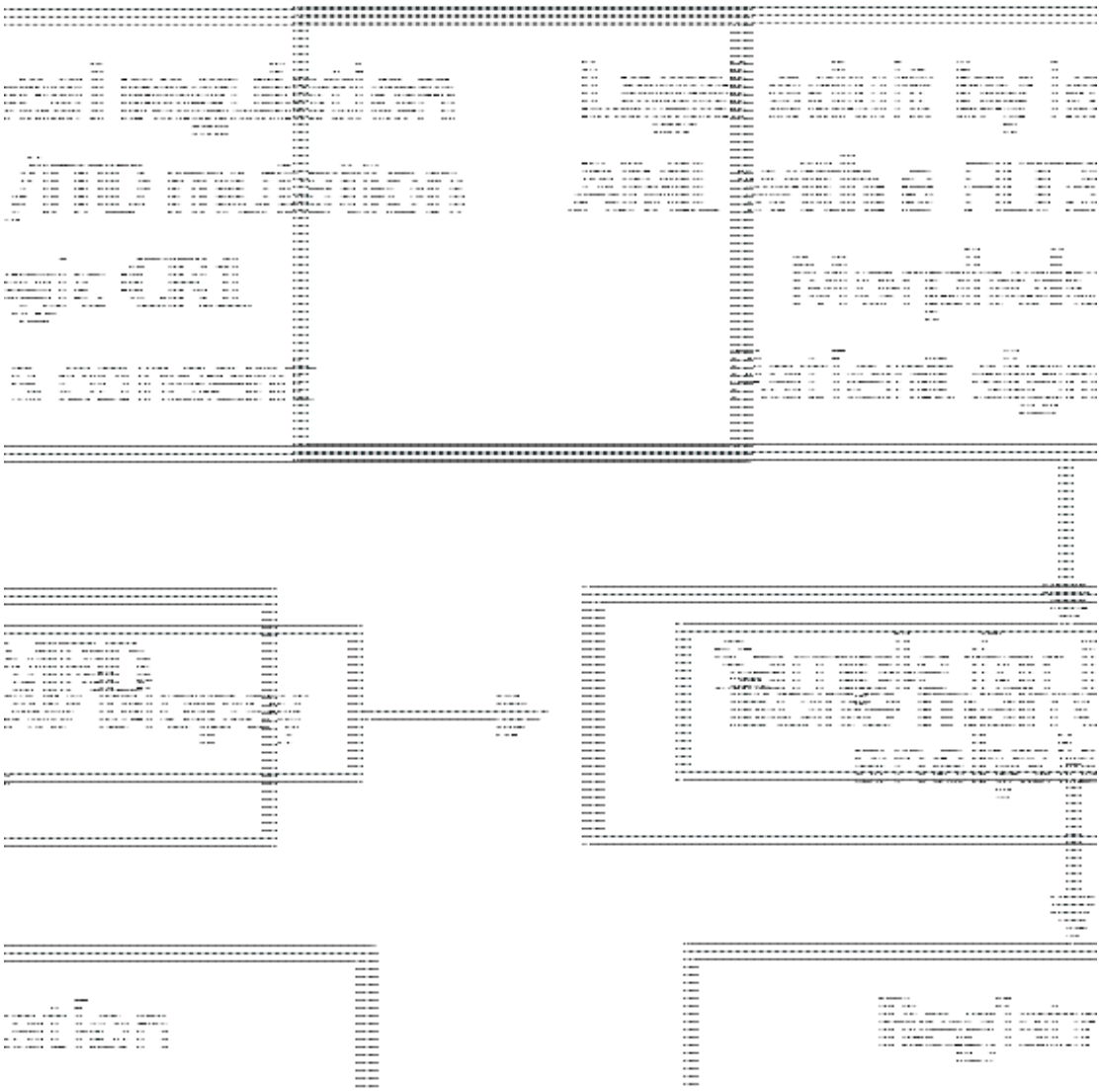
This Statistical Analysis Plan (SAP) contains a more technical and detailed elaboration of the principal features of the analysis described in the protocol, and includes detailed procedures for executing the statistical analysis of the primary and secondary endpoints and other data.

The final SAP will be approved prior to the primary database lock.

If there are any changes from the planned analyses in the final version of the SAP that impact the statistical analyses, then they will be documented in the Clinical Study Report (CSR).

2 FLOW CHART AND VISIT SCHEDULE

Flow Chart



AML: acute myeloid leukemia; ANC: absolute neutrophil count; BM: bone marrow; CR1: first morphologic complete remission; FLT3: FMS-like tyrosine kinase 3; GVHD: graft-versus-host disease; HSCT: hematopoietic stem cell transplant; ITD: internal tandem duplication; MRD: minimal residual disease.

Refer to Table 4 of the protocol for the schedule of assessments.

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 participants with FLT3/ITD AML in CR1 who undergo HSCT and are randomized to receive gilteritinib or placebo beginning after the time of engraftment for a two year period.

3.1.2 Secondary Objectives

The key secondary objective is to compare overall survival (OS) in participants treated with gilteritinib as maintenance therapy after HSCT compared to those treated with placebo.

Additional secondary objectives are the following:

- To determine the safety and tolerability of gilteritinib after HSCT.
- To compare non-relapse mortality (NRM) and event-free survival (EFS) (where events include relapse, death, stopping therapy and administration of donor lymphocyte infusion (DLI) or new therapy for suspicion of disease) in participants treated with gilteritinib as maintenance therapy after HSCT compared to those treated with placebo.
- To compare 6-month cumulative incidence of grades II-IV and III-IV acute graft-versus-host disease (GVHD) and 12-month and 24-month cumulative incidence of mild, moderate, and severe chronic GVHD in participants treated with gilteritinib as maintenance therapy after HSCT compared to those treated with placebo.
- To examine the effect of pre- and post-transplant MRD on RFS and OS. Relationship of minimal residual disease (MRD), as determined using a next-generation sequencing (NGS) platform specific to FLT3/ITD mutations, with RFS and OS.
- To compare incidence and severity of infection in participants treated with gilteritinib as maintenance therapy after HSCT compared to those treated with placebo.

3.1.3 Exploratory Objectives:

The exploratory objectives are to:

- Examine health-related quality of life (HQL)
- Compare leukemia-specific health-related quality of life (as measured by Functional Assessment of Cancer Therapy-Leukemia [FACT-Leu]) in participants treated with gilteritinib as maintenance therapy after HSCT compared to those treated with placebo.
- Assess the effect of gilteritinib as maintenance therapy after HSCT compared to those treated with placebo on disease- and treatment-related symptoms as well the proximal and distal impacts of changes in symptoms (e.g., functioning, and health-related quality of life [HQL]) as measured by Functional Assessment of Cancer Therapy – Bone Marrow Transplant (FACT-BMT), FACT-Leu and EuroQol Group-5 Dimension –5 Level (EQ-5D-5L)
- Examine healthcare resource utilization.

- Estimate the proportion of participants who are willing and able to be randomized to maintenance and describe the frequencies of events that preclude randomization.
- Assess gilteritinib pharmacokinetics using population pharmacokinetic modeling.
- Evaluate FLT3 mutation status and other potential genomic and/or other biomarkers that may correlate to treatment outcome
- To compare GVHD-free relapse-free survival (GRFS), acute GVHD-free relapse-free survival (aGRFS) and chronic GVHD-free relapse-free survival (cGRFS) in participants treated with gilteritinib as maintenance therapy after HSCT compared to those treated with placebo

3.2 Study Design

The study is a double-blind, placebo-controlled, randomized Phase III, multi-center trial comparing gilteritinib as maintenance therapy vs placebo, in FLT3/ITD AML participants in first morphologic CR who are undergoing an HSCT. Randomization will start between days 30 and 90 after HSCT, after the participant has engrafted. The target number of participants randomized is 346, 173 for each arm. To account for an anticipated 35% dropout between registration and randomization, 532 participants are expected to register at the time of transplant in order to get the targeted number randomized. The trial will be conducted at approximately 200 centers in North America, Europe, South America, Central America, Asia/Pacific and rest of world.

This trial will have a three-step enrollment process:

- Screening segment will occur after informed consent is signed
- Registration will occur prior to HSCT after all registration eligibility criteria are met
- Randomization will occur after HSCT once the participant is ready to begin the study drug and after all randomization eligibility criteria are met

After engraftment, participants will undergo a BM aspirate and/or biopsy to ensure continued CR status (Engraftment is defined as absolute neutrophil count (ANC) ≥ 500 cells/µL and platelets ≥ 20000/µL on 3 consecutive measurements (each occurring at least 1 day apart)). The participant must not have had a platelet transfusion within 7 days prior to the first measurement. BM aspirate and/or biopsy must be performed ≤ 30 days prior to randomization and can be repeated if necessary to meet the 30 day requirement. Aspirate is required and BM biopsy in addition is preferred. If aspirate is unobtainable (e.g., dry tap), BM biopsy is required. Once aspirate and/or biopsy confirm continued morphological CR status, participants can then proceed to randomization.

Randomization should occur anytime between days 30 to 90 after day zero (the day of allograft infusion) of HSCT. BM aspirate and/or biopsy after engraftment after HSCT must be done to confirm ongoing remission and must be performed ≤ 30 days prior to randomization. Please note that the recommended window to randomize participants is between 30 to 45 days after HSCT, but the protocol allows randomization up to day +90 after HSCT. Study drug should be started within 24 hours of randomization. Day 1 of the treatment period is defined as the date of randomization. Study drug will be administered on

an outpatient basis but may be continued in the inpatient setting if participants require inpatient evaluation. No investigational or commercial agents or therapies other than study drug may be administered with the intent to treat the participant's malignancy.

Randomized participants will return to the clinic for study visits while on study drug for up to 24 months (730 days) after initiation of study drug.

The primary analysis will occur 2.5 years after the last participant is randomized (i.e., the data cutoff date is 07Jan2023) or when approximately 122 events have been observed for the primary endpoint, whichever occurs first. If RFS is not statistically significant or both RFS and OS are statistically significant at the primary analysis, the study including follow-up, will be stopped. If RFS is statistically significant but OS is not statistically significant at the primary analysis, participants will be followed for relapse and survival for a total of 5 years after the last participant is randomized.

For an individual participant, study duration will last for 2.5 to 5.5 years if the RFS is not statistically significant or if both the RFS and OS are statistically significant at the primary analysis; otherwise, study duration for an individual participant will last up to 5 to 8 years.

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.

Participants will be randomized at a ratio of 1:1 between the treatment arm and the placebo arm using permuted blocks of random sizes. Randomization will be stratified by conditioning regimen intensity (myeloablative vs reduced intensity/non-myeloablative), time from transplant to randomization (30-60 days vs 61-90 days), and the presence of MRD (present vs absent/indeterminate) based on the pre-transplant BM aspirate. Date of randomization is the date the participant is randomized in IRT.

4 SAMPLE SIZE

The primary analysis will be done using stratified log-rank tests, with the randomization factors used as stratification variables. RFS in the control group is assumed to be 67% at 1 year, 59% at 2 years, and 55% at 3 years, based on CIBMTR data on participants with FLT3/ITD mutation transplanted in CR1 who were alive and progression free at 60 days. A total of 122 events provides 85% power to detect a HR of 0.57 (corresponding to a 15% difference in 2-year RFS) with two-sided significance level of 0.05. This assumes that the survival function in the control group has piecewise constant hazard functions in each interval, that the treatment arm has proportional hazards with a HR of 0.57, and maximum follow up of 3 years (Note that CIBMTR data indicates that very few events occur after 3 years so there is little benefit in longer follow up). Assuming approximately 2 years of accrual and 5% drop out rate per year, 346 participants need to be randomized to ensure a high likelihood of obtaining 122 events.

As of February 15, 2023, 103 RFS events by BERC were observed after 2.5 years from the last participant randomized (i.e., the data cutoff date is 07Jan2023). This will provide approximately 78.6% power with two-sided significance level of 0.05.

Table 1 Summary of Number of Deaths, Stopping Boundary and Power for OS Testing

			Primary analysis (2.5 years after last randomization)			Final analysis (5 years after last randomization)				
Observed # deaths at primary analysis[a]	Predicted # deaths at final analysis[a]	IF[b]	2-sided p-value boundary	approx. observed HR at boundary	power	Observed # deaths at final analysis[c]	IF[b] actually used at primary analysis	Adjusted 2-sided p-value boundary[d]	approx. observed HR at boundary	power
86	100	0.860	0.0313	0.629	57.0%	95	0.905	0.0440	0.662	68.2%
86	100	0.860	0.0313	0.629	57.0%	96	0.896	0.0434	0.662	68.6%
86	100	0.860	0.0313	0.629	57.0%	97	0.887	0.0428	0.663	68.9%
86	100	0.860	0.0313	0.629	57.0%	98	0.878	0.0422	0.663	69.3%
86	100	0.860	0.0313	0.629	57.0%	99	0.869	0.0417	0.664	69.6%
86	100	0.860	0.0313	0.629	57.0%	100	0.860	0.0412	0.665	70.0%
86	100	0.860	0.0313	0.629	57.0%	101	0.851	0.0407	0.665	70.3%
86	100	0.860	0.0313	0.629	57.0%	102	0.843	0.0402	0.666	70.6%
86	100	0.860	0.0313	0.629	57.0%	103	0.835	0.0397	0.667	71.0%
86	100	0.860	0.0313	0.629	57.0%	104	0.827	0.0393	0.667	71.3%
86	100	0.860	0.0313	0.629	57.0%	105	0.819	0.0389	0.668	71.6%

a: The observed number of deaths at primary analysis and predicted number of deaths will be updated at the finalization of the SAP for the primary analysis.

b: IF=information fraction.

c: The observed number of deaths at final analysis will be updated at the finalization of the SAP for the final analysis.

d: adjusted p-value boundary is based on the actual information fraction used at the primary analysis per Wassmer and Brannath, 2016.

As February 15, 2023, 86 OS events were observed after 2.5 years from the last participant randomized (i.e., the data cutoff date is 07Jan2023). It is predicted the study would have 100 OS events by 5 years after the last participant randomized and the actual observed number of OS events may be slightly different from 100. In such case, the study will use 100 OS events to calculate the information fraction to calculate the p-value boundary at the primary analysis. O'Brien-Fleming type alpha-spending function [Lan & DeMets, 1983] will be utilized to control the overall 2-sided 0.05 significance level for the OS primary and final analyses as shown in Table 1. For example, with 86 observed OS events at the primary analysis and 104 observed OS events at the final analysis, the study has 57.0% power and 71.3% power to detect a difference in OS between gilteritinib and placebo with the assumption of HR=0.6 at the 2-sided p-value boundary of 0.0313 for the primary analysis and adjusted 2-sided p-value of 0.0393 per Wassmer and Brannath (2016) for the final analysis, respectively.

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.

Intent-to-Treat Analysis Set (ITT) will be used for efficacy and biomarker analysis. Safety Analysis Set (SAF) will be used for the analyses of safety variables. Pharmacokinetic Analysis Set (PKAS) will be used for pharmacokinetic analyses. The data from all randomized participants will be included in the data tables, listings, and figures.

5.1 Intent-to-Treat Analysis Set (ITT)

The intent-to-treat analysis set (ITT) will consist of all participants who are randomized and will be used for efficacy analyses. Participants will be included in the treatment group to which they are randomized. Participants who are randomized but do not start therapy will not be replaced.

The ITT 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 consists of all participants who took at least 1 dose of study drug (gilteritinib or placebo). Participants 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 population administered at least 1 dose of study drug (gilteritinib), have at least 1 measurable concentration datum and for whom both the date and time of dosing on the day of sampling and the date and time of PK sampling is known. Additional participants may be excluded from the PKAS at the discretion of the pharmacokineticist.

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

6 ANALYSIS VARIABLES

6.1 Efficacy Endpoints

6.1.1 Primary Efficacy Endpoint

The study has intercurrent events like early discontinuation of the study treatment or initiation of subsequent AML therapy or missing visit for disease assessment, while the study continues to collect relapse and death information regardless of having the intercurrent events or not. The study has missing data from discontinuation from the study before relapse due to participant withdrawal/lost of follow-up.

The primary efficacy endpoint is Relapse-free survival (RFS) as assessed by the blinded endpoint review committee (BERC). Treatment policy ignoring the intercurrent events is used to define the primary efficacy endpoint. For each participant, RFS is defined as the time from the date of randomization until the date of documented morphological relapse, or death from any cause, whichever occurs first.

Morphological relapse is defined as documentation of any of the following events:

- Bone marrow blasts ≥ 5% (not attributable to regenerating BM)
- Any circulating blasts (not attributable to regenerating BM or growth factors)
- Presence of extramedullary blast foci per Revised International Working Group (R-IWG) criteria
- The earliest date of any of the relapse event will be used for RFS.

For a participant who is not known to have relapse or death, RFS is censored at the date of last follow-up for relapse assessment. Last follow-up for relapse assessment date refers to last date among the following assessments: bone marrow aspiration or biopsy, peripheral blood assessment of blast, or a follow-up visit where no relapse was indicated. A participant who has no post-baseline disease assessment will be censored at the randomization date.

Only relapses or deaths occurring on or prior to the cutoff date are counted as RFS events. Participants with relapses or deaths after the cutoff date will be censored at the last follow-up for relapse assessment date on or before cutoff date.

Alternative definition of RFS will also be used for sensitivity analyses:

- RFS will be defined similarly as above except that a participant is censored at the last follow-up for disease assessment when relapse or death is documented after more than 1 missed disease assessment, considering hypothetical policy for missing visits and treatment policy ignoring other intercurrent events.
- RFS will be defined similarly as above except that a participant is censored at the end of planned treatment considering on-treatment policy.
- RFS will be defined similarly as above by using investigator assessed relapse.

Table 2 RFS Definition

	Primary (By BERC)	**Sensitivity 1 (Missing >1 assessment)**	**Sensitivity 2 (censored at planned EoT)**	**Sensitivity 3 (By investigator)**
Situation	**Event/Censor; RFS date**	**Event/Censor; RFS date**	**Event/Censor; RFS date**	**Event/Censor; RFS date**
Death or relapse on or before analysis cutoff date	Event; Earlier date of relapse by BERC or death	Event; Earlier date of relapse by BERC or death If relapse after ≥2 missed assessments and patient is alive, Censor; Last follow-up for relapse assessment date	Event; Earlier date of relapse by BERC or death If relapse or died after planned EoT, Censor; Date of planned EoT	Event; Earlier date of relapse by investigator or death
Death and relapse after analysis cutoff date	Censor; Last follow-up for relapse assessment date before cutoff date	Censor; Last follow-up for relapse assessment date before cutoff date	Censor; Last follow-up for relapse assessment date before cutoff date	Censor; Last follow-up for relapse assessment date before cutoff date
No RFS event	Censor; Last follow-up for relapse assessment date before cutoff date	Censor; Last follow-up for relapse assessment date before cutoff date	Censor; minimum(Last follow-up for relapse assessment date before cutoff date, EoT date)	Censor; Last follow-up for relapse assessment date before cutoff date
No RFS event and no disease assessment	Censor; Randomization date	Censor; Randomization date	Censor; Randomization date	Censor; Randomization date

RFS = Date of Event or Censor – Date of Randomization +1

Date of EoT = last dose date

The additional alternative definition of RFS below will be used for sensitivity analysis to assess the impact of Covid-19 if at least 2.5% of the participants are impacted.

- RFS will be defined similarly as the primary definition above except that a participant is censored at the last follow-up for relapse assessment if the assessment was delayed or missing due to Covid-19.
- RFS will be defined similarly as the primary definition above except that a participant is censored at the end of treatment if the reason for treatment discontinuation is due to Covid-19.
- RFS will be defined similarly as the primary definition above except that a participant is censored at the date of Covid-19 diagnosis or death date due to Covid-19, whichever occurs first.

- RFS will be defined similarly as the primary definition above except that a participant is censored at the date of Covid-19 diagnosis or death date due to Covid-19 or at the end of treatment if the reason for treatment discontinuation is due to Covid-19 or at the last follow-up for relapse assessment if the assessment was delayed or missing due to Covid-19, whichever occurs first.

6.1.2 Secondary Efficacy Endpoints

6.1.2.1 Key Secondary Efficacy Endpoint

- Overall survival (OS)

OS is defined as the time from the date of randomization until the documented date of death from any cause per treatment policy by ignoring the intercurrent events. All events of death will be included, regardless of whether the event occurred while the participant is still taking study drug or placebo or after the participant discontinue study drug. Participants who are still alive at the time of analysis will be censored at the last day known to be alive or at the analysis cutoff date, whichever is earlier. Participants who died after the analysis cut-off date will be censored at the analysis cut-off date. The date of last known alive is the latest date that the participant is confirmed to be alive across the entire database.

OS (in days) is calculated as:

(Date of death or known to be alive) – (Date of randomization) +1.

Table 3 OS Definition

Situation	Date of Event or Censor	Outcome
Death before or on analysis cutoff date	Date of death	Event
Death after analysis cutoff date	Analysis cutoff date	Censor
Last known alive is before or on cutoff date	Last known alive date	Censor
Last known alive is after cutoff date	Analysis cutoff date	Censor

OS = Date of Event or Censor – Date of Randomization +1

The alternative definition of OS below will be used for sensitivity analysis to assess the impact of Covid-19 if at least 2.5% of the participants are impacted.

- OS will be defined similarly as the primary definition above except that a participant is censored at the end of treatment if the reason for treatment discontinuation is due to Covid-19.
- OS will be defined similarly as the primary definition above except that a participant is censored at the date of Covid-19 diagnosis or death date due to Covid-19, whichever occurs first.
- OS will be defined similarly as the primary definition above except that a participant is censored at the date of Covid-19 diagnosis or death date due to Covid-19 or at the end of treatment if the reason for treatment discontinuation is due to Covid-19, 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, or premature discontinuation of the treatment, or initiation of other anti-leukemic treatment, or death from any cause, whichever occurs first.

See Section 6.1.1 for the definition of relapse. Anti-leukemic treatment is defined as hypomethylating agents, chemotherapy, oral anticancer agents, DLI or cellular therapies given because of detectable disease (via cytogenetics, flow cytometry or PCR), not meeting R-IWG criteria for relapse.

If a participant experiences relapse or death, the participant 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 participant discontinues the treatment or initiates other anti-leukemic treatment, the participant 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 date of start of anti-leukemic treatment, respectively.

If multiple events occur on the same date, take the order of "relapse", "death", "treatment discontinuation", and "initiation of other anti-leukemic treatment".

For a participant who is not known to have relapse, death, treatment discontinuation or initiation of other anti-leukemic treatment, EFS is censored at the last follow-up for relapse assessment or randomization date, whichever occurs last

Table 4 EFS Definition

Situation	Date of Event or Censor	Outcome
Death or relapse or premature discontinuation of the treatment, or initiation of other anti-leukemic treatment on or before analysis cutoff date	Date of relapse, death, premature discontinuation of treatment or initiation of other anti-leukemic treatment	Event
Death and relapse and premature discontinuation of the treatment and initiation of other anti-leukemic treatment after analysis cutoff date	Last follow-up for relapse assessment date before cutoff date	Censor
No EFS event	Last follow-up for relapse assessment date before cutoff date	Censor
No EFS event and no disease assessment	Randomization date	Censor

EFS = Date of Event or Censor – Date of Randomization +1

- Time to Non-Relapse Mortality (NRM)

Time to NRM is defined similar to RFS (see Section 6.1.1) with event of interest as the participants who died without documentation of morphological relapse. Competing event will be the documentation of relapse. For a participant who is not known to have relapse or death, time to NRM is censored at the date of last follow-up for relapse assessment before cutoff date.

- Time to Relapse

Time to relapse is defined similar to RFS (see Section 6.1.1) with event of interest as the participants who had documentation of morphological relapse. Competing event will be the death without documentation of relapse. For a participant who is not known to have relapse or death, time to relapse is censored at the date of last follow-up for relapse assessment before cutoff date.

- Time to Treatment-Emergent Acute GVHD II-IV

Time to treatment-emergent acute GVHD II-IV is defined as the time from the date of randomization until the date of treatment-emergent acute GVHD II-IV, or treatment-emergent death from any cause, whichever occurs first. Participants who are still alive and have no acute GVHD event at the time of analysis will be censored at the last day of the aGVHD assessment or at the analysis cutoff date, whichever is earlier.

The event of interest will be the participants who had treatment-emergent acute GVHD II-IV. Events include the following:

- Development of new grade II-IV acute GVHD after randomization
- Worsening in the grade of acute GVHD at the time of randomization by one point (e.g., grade I acute GVHD at the time of randomization, progressing to grade II or more after randomization)

Competing event will be the treatment emergent death without documentation of acute GVHD II-IV event. An additional analysis will be done treating treatment emergent death and treatment emergent relapse without documentation of acute GVHD II-IV event as competing events. Definition of relapse follows Section 6.1.1.

Grading of acute GVHD will be derived by consensus grading (Przepiorka 1995) per BMT-CTN manual of procedures (MOP).

- The acute GVHD algorithm calculates the grade based on the organ (skin, GI and liver) stage and etiology/biopsy reported on the weekly GVHD form.
- If none of the etiologies for skin, upper GI, lower GI, or liver are reported as GVHD, then the overall grade is 0
- If multiple etiologies are specified for lower GI or liver, the organ system will be downstaged by 1.
- If an upper GI biopsy is negative, upper GI symptoms are downstaged
- If GVHD is not listed as an etiology for upper GI then upper GI symptoms are downstaged
- Each organ contributes to the overall grade; while to get an overall grade, it does not necessarily need all organ symptoms. Different organ/stage determine different grade.

Table 5 Organ Stage for GVHD grading

Stage	Skin	Lower GI	Upper GI	Liver
1	Maculopapular Rash <25% BSA	Diarrhea 500-999ml/day	Persistent Nausea, Vomiting, or Anorexia	Bilirubin 2-3 mg/dl
2	Maculopapular Rash 25-50% BSA	Diarrhea 1000-1500 ml/day	---	Bilirubin 3.1-6 mg/dl
3	Maculopapular Rash > 50% BSA	Diarrhea > 1500 ml/day	---	Bilirubin 6.1-15 mg/dl
4	Generalized Erythroderma > 50% BSA **plus** Bullous Formation and Desquamation > 5% BSA	Severe Abdominal Pain With or Without Ileus or Grossly Bloody Stool	---	Bilirubin > 15 mg/dl

Table 6 Acute GVHD Grade

Grade	Skin	Lower GI	Upper GI	Liver
I	Stage 1-2	0	0	0
II	Stage 3 or	Stage 1 or	Stage 1 or	Stage 1
III	---	Stage 2-4	---	Stage 2-3
IV	Stage 4	---	---	Stage 4

- Time to Treatment-Emergent Acute GVHD III-IV

Time to treatment-emergent acute GVHD III-IV is defined as the time from the date of randomization until the date of treatment-emergent acute GVHD III-IV, or treatment emergent death from any cause, whichever occurs first. Participants who are still alive and no treatment-emergent acute GVHD event at the time of analysis will be censored at the last day of the aGVHD assessment or at the analysis cutoff date, whichever is earlier.

The event of interest will be the participants who had treatment-emergent acute GVHD III-IV. Event of interest include the following:

- Development of new grade III-IV acute GVHD after randomization
- Worsening in the grade of acute GVHD at the time of randomization by one point (e.g., grade II acute GVHD at the time of randomization, progressing to grade III or more after randomization)

Competing event will be the treatment emergent death without documentation of acute GVHD III-IV. An additional analysis will be done treating treatment emergent death and treatment emergent relapse without documentation of acute GVHD III-IV event as competing events.

- Time to Treatment-Emergent Chronic GVHD

Time to treatment-emergent chronic GVHD is defined as the time from the date of randomization until the date of treatment-emergent chronic GVHD, or treatment emergent death from any cause, whichever occurs first. Participants who are still alive and no treatment-emergent chronic GVHD event at the time of analysis will be censored at the last day of the chronic GVHD assessment or at the analysis cutoff date, whichever is earlier.

The event of interest will be the participants who had treatment-emergent chronic GVHD. Event of interest include the following:

- Development of new mild, moderate or severe chronic GVHD after randomization
- Worsening in the grade of chronic GVHD at the time of randomization by one point (e.g. mild chronic GVHD at the time of randomization, progressing to moderate or severe chronic GVHD after randomization)

Competing event will be the treatment emergent death without documentation of chronic GVHD. An additional analysis will be done treating treatment emergent death and treatment emergent relapse without documentation of chronic GVHD event as competing events.

- Time to Post-Transplant Eradication of MRD

Time to post-transplant eradication of MRD is defined only on participants with positive MRD status at baseline. It is the time from the date of randomization until the eradication of MRD, or death from any cause, whichever occurs first. Participants who are still alive at the time of analysis will be censored at the last MRD assessment or at the analysis cutoff date, whichever is earlier. Participants without post-baseline MRD assessment will be censored at the randomization date.

The event of interest will be the participants who had MRD eradication. Competing event will be the death during MRD assessment period without documentation of MRD eradication.

- Time to Post-Transplant Detection of MRD

Time to post-transplant detection of MRD is defined only on participants with negative MRD status at baseline. It is the time from the date of randomization until the detection of MRD, or death from any cause, whichever occurs first. Participants who are still alive at the time of analysis will be censored at the last MRD assessment or at the analysis cutoff date, whichever is earlier. Participants without post-baseline MRD assessment will be censored at the randomization date.

The event of interest will be the participants who had positive MRD. Competing event will be the death during MRD assessment period without detection of MRD.

- Minimal Residual Disease (MRD)

MRD will be analyzed by a Sponsor-designated central laboratory. A peripheral blood sample is not acceptable for MRD assessment.

FLT3/ITD mutation ratio will be measured in relation to total FLT3. For a participant 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 "Present" if $\log_{10}$-transformed overall FLT3/ITD mutation ratio is greater than -4; otherwise, the presence of MRD will be "Absent". For participants with multiple ITD mutations detected, the ratios will be summed.

Additional absolute cut points (e.g., $\log_{10}$-transformed overall FLT3/ITD to total FLT3 ratio of -5, -6, etc.) may be evaluated, as well as the $\log_{10}$ transformed change in FLT3/ITD to total FLT3 ratio at each visit compared to baseline.

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.

- Non-randomization

These are the participants who were registered but not randomized

- 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 designed to measure leukemia-specific signs, symptoms and the impact of leukemia on participants. The 44-item scale has global and domain scores including physical well-being, social/family well-being, emotional well-being, functional well-being and additional leukemia-specific concerns.

- Functional Assessment of Cancer Therapy-Bone Marrow Transplant (FACT-BMT)

The FACT-BMT comprised of a general core questionnaire, the Functional Assessment of Cancer Therapy-General (FACT-G), which evaluates the HQL of participants receiving treatment for cancer, and a transplant-specific module, bone marrow transplant (BMT) Concerns, that addresses disease and treatment-related questions specific to BMT. The

FACT-G consists of four subscales developed and normed in cancer participants: Physical Well-being, Social/Family Well-being, Emotional Well-being, and Functional Well-being. Each subscale is positively scored, with higher scores indicating better functioning. The FACT-BMT Total, which is the grand total of all items in the FACT-G and BMT modules, will be used as the outcome measure in summarizing the FACT-BMT data.

- Healthcare resource utilization

Healthcare resource utilization includes hospitalization or emergency department visits (duration, underlying reasons, and setting specified: general ward, ICU, etc.), antibiotic use, and oral and IV medication for AEs.

- Acute GVHD-free relapse-free survival (aGRFS)

Time to aGRFS is defined as the time from the date of randomization until the date of treatment-emergent acute GVHD III-IV, or morphological relapse (BERC) or death from any cause, whichever occurs first. Participants who are still alive and no aGRFS event at the time of analysis will be censored at the last follow-up for relapse assessment or at the analysis cutoff date, whichever is earlier.

Treatment-emergent acute GVHD III-IV event is defined as:

- Development of new grade III-IV acute GVHD after randomization
- Worsening in the grade of acute GVHD at the time of randomization by one point (e.g., grade II acute GVHD at the time of randomization, progressing to grade III or more after randomization)
- Chronic GVHD-free relapse-free survival (cGRFS)

Time to cGRFS is defined as the time from the date of randomization until the date of treatment-emergent chronic GVHD that requires new or additional immunosuppressive treatment, or morphological relapse (BERC) or death from any cause, whichever occurs first. Participants who are still alive and no cGRFS event at the time of analysis will be censored at the last follow-up for relapse assessment or at the analysis cutoff date, whichever is earlier.

Treatment-emergent chronic GVHD event is defined as:

- Development of new mild, moderate or severe chronic GVHD that requires new or additional immunosuppressive treatment after randomization
- Worsening in the grade of chronic GVHD that requires new or additional immunosuppressive treatment at the time of randomization by one point (e.g., mild chronic GVHD at the time of randomization, progressing to moderate or severe chronic GVHD after randomization)
- GVHD-free relapse-free survival (GRFS)

Time to GRFS is defined as the time from the date of randomization until the date of treatment-emergent acute GVHD III-IV, or treatment-emergent chronic GVHD that requires new or additional immunosuppressive treatment, or morphological relapse (BERC) or death from any cause, whichever occurs first. Participants who are still alive and no GRFS event at

the time of analysis will be censored at the last follow-up for relapse assessment or at the analysis cutoff date, whichever is earlier.

Treatment-emergent acute GVHD III-IV event is defined in the aGRFS definition.
Treatment-emergent chronic GVHD event is defined in the cGRFS definition.

6.2 Safety Variables

Safety endpoints such as AEs, laboratory tests, vital signs, ECGs, physical examination and Karnofsky performance status are secondary endpoints of the study.

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 is marked as an event occurred after the first dose of the study drug, then the AE will be considered treatment emergent. If the AE occurs on Day 1 and is marked as an event occurred before the first dose of the study drug, then the AE will not be considered treatment emergent. If a participant 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 7.10.1.
- 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 or Important Medical Event (IME).
- 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
  Below is a table of the laboratory tests that will be performed during the conduct of the study. Refer to the Schedule of Assessments for study visit collection dates.

  Additional laboratory tests should be performed according to institutional standard of care. Clinical significance of out-of-range laboratory findings is to be determined and documented by the investigator or delegated sub-investigator who is a qualified physician.

Table 7 Laboratory Assessments

Panel/Assessment	Parameters to be analyzed	NCI-CTCAE Grading (Hyper/Hypo)
Hematology	Hematocrit (Hct)	N/A
	Hemoglobin (Hgb)	Both
	Red Blood Cell Count (RBC)	N/A
	White Blood Cell Count (WBC)	Both
	Platelets	Hypo
	Neutrophils	Hypo
	Eosinophils	N/A
	Basophils	N/A
	Lymphocytes	Both
	Monocytes	N/A
Biochemistry (fasting)	Sodium (Na)	Both
	Magnesium (Mg)	Both
	Creatine Phosphokinase (CK)	Hyper
	Potassium (K)	Both
	Corrected Calcium (CCa)	Both
	Chloride (Cl)	N/A
	Phosphate (P)	Hypo
	Creatinine (Cr)	Hyper
	Glucose (Gl)	Both
	Blood Urea Nitrogen (BUN)	N/A
	Alkaline Phosphatase (ALP)	Hyper
	Aspartate Aminotransferase (AST)	Hyper
	Alanine Aminotransferase (ALT)	Hyper
	Bilirubin Total (TBL)	Hyper
	Total Protein (TP)	N/A
	Albumin (Alb)	Hypo
	Bicarbonate (HCO3)	N/A
	Serum HCG for female participants of childbearing potential	N/A
	Thyroid Stimulating Hormone (TSH)	N/A
	Free Thyroxine (Free T4)	N/A

- Vital signs (systolic and diastolic blood pressures (mmHg))
- 12-lead electrocardiogram (ECG)
- Karnofsky performance status scores

6.3 Pharmacokinetic Variables

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

Pharmacokinetic samples for gilteritinib will be collected within 0.5 hours before drug administration.

6.4 Other Variables

- Body Mass Index (BMI)

  BMI = weight (kg) / [height (m)]^2

- Time from CR to randomization (days)

  Time from CR to randomization = Date of randomization – date of first CR

- Time from transplant to ANC recovery (days)

  Time from transplant to ANC recovery = Date of ANC recovery – date of transplant

- Time from diagnosis to transplantation (days)

  Time from diagnosis to transplantation = Date of transplantation – date of diagnosis + 1

- 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 (mg) 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 dose intensity = 120 mg/day

- Relative dose intensity (%)

  Relative dose intensity = dose intensity (mg/day)/planned daily dose (mg/day)*100%

- Previous and concomitant medications

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

  Concomitant medication is defined as medication administered 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 participants (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 participants with no missing data, i.e., will add up to 100%.

Summaries based on ITT (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 or PKAS will be presented by actual treatment received.

All statistical comparisons will be made using 2-sided tests at α=0.05 significance level unless specifically stated otherwise. All null hypotheses will be of no treatment effect difference between gilteritinib and placebo, all alternative hypotheses will be there is treatment effect difference between gilteritinib and placebo, 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.

Study day for safety assessments (e.g., laboratory assessment, onset of adverse events, vital signs, etc.) will be calculated in reference to the first dose date. For assessments conducted before the first dose, study day will be calculated as (assessment date – first dose date). For assessments conducted on or after the first dose, study day will be calculated as (assessment date – first dose date + 1).

Study day for efficacy assessments (bone marrow assessment, overall survival, relapse-free survival, etc.) will be calculated in reference to the randomization date. For assessments conducted before randomization, study day will be calculated as (assessment date – randomization date). For assessments conducted on or after randomization, study day will be calculated as (assessment date – randomization date + 1).

For efficacy evaluation, baseline is defined as the last available measurement from Day -14 to before randomization. For safety evaluation, baseline is defined as the last available measurement from day -14 to before first dose. For MRD assessment, baseline is defined as the last available measurement from day -30 to before randomization. Unless otherwise specified, all summaries will be presented by treatment arm.

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

7.2 Study Population

7.2.1 Disposition of Participants

The following participant data will be presented:

- Number and percentage of participants with informed consent, discontinued before registered, registered, discontinued before randomization, randomized (overall only);
- Number and percentage of randomization failure participants, by primary reason for randomization failure for randomization failure participants only;

For all randomized participants, the following participant data will be presented by treatment group:

- Number and percentage of participants in each analysis set;
- Number and percentage of participants who completed or discontinued treatment by primary reason for treatment discontinuation;
- Number and percentage of participants who completed the 30-day follow-up evaluation by status;
- Number and percentage of participants who completed the long term follow-up evaluation by status.

7.2.2 Protocol Deviations

Protocol deviations (PDs) as defined in the study protocol (Appendix C of the protocol) will be assessed for all randomized participants. The number and percentage of participants meeting any criteria will be summarized for each criterion and overall, by treatment group and total as well as by study site. Participants deviating from a criterion more than once will be counted once for the corresponding criterion. Any participants who have more than one protocol deviation will be counted once in the overall summary. A data listing will be provided by site and participant.

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 participants allocated to treatment in each country and site will be presented by treatment group.

Descriptive statistics for age, weight, body mass index (BMI), height at study entry, time from CR to randomization, time to ANC recovery and time from diagnosis to transplantation will be presented. Frequency tabulations for conditioning regimen intensity, time from transplant to randomization, pre-transplant MRD status, sex, ethnicity, race, age group, Karnofsky performance score (<90 vs >=90), recipient CMV status, donor/recipient CMV status, HSCT-specific comorbidity index (0 vs 1-2, vs >=3), donor type/HLA matching(matched sibling, unrelated donor (URD), mismatched URD, related haploidentical, or umbilical cord blood), graft type (umbilical cord, BM, PB), prophylaxis regimen, latest status of acute GVHD prior to randomization(0-1 vs 2 vs 3-4, yes vs no) and latest status of chronic GVHD prior to randomization (yes vs no) and cytogenetic risk 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 ongoing medical conditions, Common Terminology Criteria for Adverse Events (CTCAE) grade 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/ingredients (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 PT/ingredients. Participants 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 participants who took each medication.

Frequency tabulations for treatment used for GVHD will be presented by treatment group.

7.2.5 Prior AML Therapy

Frequency tabulations of participants with prior AML therapy, regimen, type of treatment, prior use of FLT3 inhibitor, and first 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.

7.2.6 New Anti-Leukemic Therapy

Frequency tabulations of participants with new anti-leukemic therapy, regimen, type of treatment, best response to therapy and reason for initiating therapy will be presented by treatment group.

New anti-leukemic therapy will also be summarized by PT by treatment group and presented in decreasing order of frequency based on the total number of participants who took each medication.

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 participants 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 20 months
  - at least 20 months, less than 24 months
  - 24 months or more
  - Unknown.

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

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

7.3.2 Relative Dose Intensity

RDI will be examined for participants in the SAF whose total study drug count and first and last days of treatment are known.

RDI will be summarized in two ways for the SAF:

- Descriptive statistics will be presented by treatment arm.
- RDI will be categorized according to the following categories by treatment regimen:
  - less than 50%
  - at least 50%, less or equal to 80%
  - greater than 80%
  - Unknown.

7.4 Analysis of Efficacy

To address multiplicity, a gatekeeping testing strategy will be used for RFS (primary efficacy endpoint) and OS (key secondary endpoint). RFS will be tested once at 2-sided significance level of 0.05. Only if RFS is significant, will hypothesis testing for OS primary and final analyses be performed. An O'Brien-Fleming type alpha-spending function [Lan & DeMets, 1983] will be utilized to control the overall 2-sided 0.05 significance level for the OS primary and final analyses. Other secondary endpoints' testing will not be multiplicity adjusted.

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 ITT to compare the RFS as assessed by BERC between gilteritinib and placebo. The null and alternative hypotheses for this comparison are:

- H_{01}:RFS in the gilteritinib arm is not different to RFS in the placebo arm
- H_{11}:RFS in the gilteritinib arm is different from RFS in the placebo arm.

The primary analysis will be performed 2.5 years after the last patient was randomized or when approximately 122 RFS events have been observed, whichever occurs first. Comparison of gilteritinib and placebo will be tested at 2-sided significance level of 0.05

The distribution of RFS will be estimated for each treatment arm using Kaplan-Meier methodology and compared between gilteritinib and placebo using log-rank test stratified by randomization strata as described in Section 3.3. In addition, stratified Cox proportional hazard model will be used to estimate the hazard ratio and the corresponding 95% confidence interval. The same stratification factors will be applied to both the stratified log-rank test and the stratified Cox proportional hazards model. The Cox proportional hazard model will be included as a sensitivity analysis for testing H_{01}. The stratification factors may be pooled in the stratified analysis per Section 7.10.5.

Kaplan-Meier survival plots will be used to describe the RFS in each treatment group. Median RFS and RFS probabilities at 1, 2 and 3 years with 95% confidence interval will be estimated from the Kaplan-Meier plots.

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

In the event that the MRD status is not available at the time of randomization, participants with unknown MRD status will be grouped with negative MRD status for stratification purposes. The MRD detection threshold is set to 10^{-4} (ratio of FLT3-ITD to total FLT3). The presence of MRD will be "Present" if $\log_{10}$-transformed overall FLT3/ITD mutation ratio is greater than -4; otherwise, the presence of MRD will be "Absent". Additional thresholds may be explored.

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:

- The primary analysis repeated on the ITT using investigator assessed relapse.
- The primary analysis repeated on the ITT with the addition of GVHD prophylaxis (CNI (CsA or Tac) + MTX +/- Others , CNI (CsA or Tac) + MMF +/- Others , Other) as a stratification factor. The stratification factors may be pooled in the stratified analysis per Section 7.10.5.
- Unstratified log-rank test and Cox proportional hazards model on the ITT
- Stratified log-rank test and Cox proportional hazards model on the ITT where RFS is censored at the last evaluable disease assessment without relapse when relapse or death is documented after more than one missed disease assessment.
- Stratified log-rank test and Cox proportional hazards model on the ITT where RFS is censored at the end of treatment when participant ended the treatment early
- If there is a difference in the stratification values between IRT and the CRF, primary analysis may be repeated on the ITT using CRF stratification values.
- Restricted mean survival time (RMST) analysis will be used for RFS on the ITT with 24 months as threshold. Restricted mean survival time will be estimated with its 95% CI. Treatment difference of RMST and its 95% CI and the corresponding p-value will be provided.
- The primary analysis repeated on the ITT for alternative definitions of RFS taking into account the impact of Covid-19, if at least 2.5% of the participants are impacted.

To investigate non-proportionality between the treatment arms, the LOG(–LOG) survivor function versus time curves will be plotted. The scaled Schoenfeld residuals by time plot will be examined for evidence of a non-zero correlation, which indicates non-proportionality. Cox models stratified on treatment may be explored given evidence of non-proportional hazards, in order to generate adjusted RFS curves by treatment group, which can be compared between the groups at 2 and 3 years.

7.4.2 Analysis of Secondary Endpoints

7.4.2.1 Key Secondary Efficacy Analysis

OS will be a key secondary endpoint, with explicit control of the type I error rate through a gatekeeper approach. Formal significance testing of OS will be conducted if the RFS comparison is statistically significant. Otherwise, survival analyses will be considered exploratory.

Primary analysis for OS will be performed when the primary analysis for RFS is statistically significant at 2-sided $\alpha = 0.05$. If the primary analysis for RFS is statistically significant while the primary analysis for OS is not statistically significant, a final analysis for OS will occur 5 years after the last participant was randomized.

OS will be analyzed on the ITT 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 not different to OS in the placebo arm
- H_{12}:OS in the gilteritinib arm is different from OS in the placebo arm.

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 Cox proportional hazard model will be included as a sensitivity analysis for testing H_{02}. The stratification factors may be pooled in the stratified analysis per Section 7.10.5.

Kaplan-Meier curves will be used to describe the OS in each arm. Median OS time and OS probabilities at 1, 2, 3, 4 and 5 years along with 95% confidence interval will be estimated from the Kaplan-Meier curves.

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

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

- The primary analysis repeated on the ITT with the addition of GVHD prophylaxis (CNI (CsA or Tac) + MTX +/- Others , CNI (CsA or Tac) + MMF +/- Others , Other) as a stratification factor. The stratification factors may be pooled in the stratified analysis per Section 7.10.5.
- Unstratified log-rank test and Cox proportional hazards model on the ITT
- If there is a difference in the stratification values between IRT and the CRF, primary analysis may be repeated on the ITT using CRF stratification values
- The primary analysis repeated on the ITT for alternative definitions of OS taking into account the impact of Covid-19, if at least 2.5% of the participants are impacted.

To investigate non-proportionality between the treatment arms, the analyses for RFS as assessed by BERC will be carried out for OS.

Subsequent anti-cancer therapies will be collected and summarized by treatment arm within drug class. Imbalance in the use of subsequent anti-cancer therapy may potentially bias the inference in OS. The following methods may be used to assess the impact of the use of subsequent anti-cancer therapy in OS.

- Censor OS at the last contact date prior to the initiation of subsequent anti-cancer therapy
- Rank-preserving structural failure time method (RPSFT) to correct for confounding introduced by the change of treatment including new anti-leukemic therapy. The RPSFT model is based on an accelerated failure time model and uses a structural assumption of time-proportionality instead of a proportional hazards assumption as used in the Cox model
- Inverse probability of censoring weights (IPCW).

7.4.2.2 Other Secondary Efficacy Analysis

EFS

EFS will be analyzed in the same manner as RFS in the ITT using the stratified log-rank test with the same strata. 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, survival rates at 1, 2 and 3 years if appropriate and 95% confidence interval will be estimated from the Kaplan-Meier curve. In addition, reason for participant starting new therapies and type of new therapies will be summarized.

NRM

Incidence of NRM will be estimated using the cumulative incidence function, treating relapse/progression as a competing risk. Incidence of NRM will be compared between the treatment arms using a Fine and Gray model, adjusting for strata variables of conditioning regimen intensity (myeloablative vs reduced intensity/non-myeloablative), time from transplant to randomization, and pre-transplant MRD status. Cumulative incidence at 1, 2 and 3 years, subdistribution hazard ratio and its 95% confidence interval will be produced.

Relapse

Incidence of relapse will be estimated using cumulative incidence function, treating death without relapse as a competing risk. Incidence of relapse/progression will be compared between the treatment arms using a Fine and Gray model, adjusting for strata variables of conditioning regimen intensity (myeloablative vs reduced intensity/non-myeloablative), time from transplant to randomization, and pre-transplant MRD status. Cumulative incidence at 1, 2 and 3 years, subdistribution hazard ratio and its 95% confidence interval will be produced.

Acute GVHD II-IV

Cumulative incidence of new onset acute GVHD or worsening in the grade of acute GVHD at the time of randomization by one point (e.g., grade II acute GVHD at the time of randomization, progressing to grade III or more after randomization) will be estimated from the time of randomization using the cumulative incidence function, treating treatment emergent death prior to acute GVHD II-IV as the competing risk. Cumulative incidence of acute GVHD II-IV will be compared between treatment arms using a Fine and Gray model, adjusting for strata variables of conditioning regimen intensity (myeloablative vs reduced intensity/non-myeloablative), time from transplant to randomization, and pre- transplant MRD status. An additional analysis will be performed treating treatment emergent death and treatment emergent relapse without documentation of an acute GVHD event as competing risks. Cumulative incidence at 100 and 180 days, subdistribution hazard ratio and its 95% confidence interval will be produced.

Acute GVHD III-IV

Cumulative incidence of new onset acute GVHD III-IV or worsening in the grade of acute GVHD at the time of randomization by one point (e.g., grade III acute GVHD at the time of randomization, progressing to grade IV or more after randomization) will be estimated from

the time of randomization using the cumulative incidence function, treating treatment emergent death prior to acute GVHD III-IV as the competing risk. Cumulative incidence of acute GVHD III-IV will be compared between treatment arms using a Fine and Gray model. Cumulative incidence at 100 and 180 days, subdistribution hazard ratio and its 95% confidence interval will be produced. An additional analysis will be performed treating treatment emergent death and treatment emergent relapse without documentation of an acute GVHD event as competing risks.

Chronic GVHD

Cumulative incidence of chronic GVHD from the time of randomization will be estimated using the cumulative incidence function, treating treatment emergent death prior to chronic GVHD as the competing risk. Cumulative incidence of chronic GVHD will be compared between treatment arms using a Fine and Gray model, adjusting for strata variables of conditioning regimen intensity (myeloablative vs reduced intensity/non-myeloablative), time from transplant to randomization, and pre- transplant MRD status. Cumulative incidence at 1 and 2 years, subdistribution hazard ratio and its 95% confidence interval will be produced. In addition, distribution of maximum grade of chronic GVHD will be tabulated. An additional analysis will be performed treating treatment emergent death and treatment emergent relapse without documentation of a chronic GVHD event as competing risks.

Eradication of MRD

The time until eradication of MRD in participants who have detectable MRD prior to randomization will also be described and compared between treatment arms using Gray's test; death during MRD assessment period without eradication of MRD will be treated as a competing risk. Cumulative incidence at 1 and 2 years, subdistribution hazard ratio from an unadjusted Fine and Gray model, and its 95% confidence interval will be produced.

Detection of MRD

The cumulative incidence of detection of MRD in participants who are MRD undetectable prior to randomization will be described in each group and compared between treatment arms using Gray's test; death during MRD assessment period without detection of MRD will be treated as a competing risk. Cumulative incidence at 1 and 2 years, subdistribution hazard ratio and its 95% confidence interval will be produced.

MRD as prognostic endpoint for RFS and OS

Subgroup analysis for RFS and OS using pre-HSCT MRD status (present vs absent/unknown) and MRD status at randomization (present vs absent) will be performed. Additional analysis may be performed to investigate MRD as prognostic endpoint for RFS/OS.

7.4.3 Analysis of Exploratory Endpoints

FLT3/ITD allelic frequencies will be analyzed using the diagnostic specimens when available.

FLT3 mutation status at relapse will be summarized by the number and percentage of participants by treatment group on ITT

Incidence of resource utilization which includes hospitalization (duration, underlying reasons, and setting specified: general ward, ICU, etc.), antibiotic use, and medication for AEs will be summarized by treatment group on ITT. The difference between treatment groups will be tested using CMH test while controlling of randomization stratification factors.

Resource utilization counts including duration of hospital stays, duration of medications and duration of antibiotic use will be summarized by treatment group using descriptive statistics (mean, standard deviation, minimum, maximum and median) on ITT. For the duration of hospital stays, missing end date will be imputed as the start date if the type of encounter is outpatient/day clinic or emergency department visit without admission. Otherwise, missing end date will be imputed using the last dosing date + 7 days. 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 ITT. 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. 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 ITT. 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 factors.

FACT-BMT score will be summarized by treatment group using descriptive statistics (mean, standard deviation, minimum, maximum and median) on ITT. 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 factors.

FACT-Leu and FACT-BMT instruments share set of questions for PWB, SWB, EWB, FWB, and FACT-G total score and were collected only once. These parameters will be displayed on both FACT-Leu and FACT BMT summary tables.

The completion and compliance rates of the EQ-5D-5L, FACT-BMT, and FACT-Leu will be summarized at each visit and overall. Additional exploratory analyses of resource utilization, patient reported outcomes (FACT-BMT, FACT-Leu) and health outcome (EQ-5D-5L) will be performed and the details will be included in a separate Patient Reported Outcome Statistical Analysis Plan.

aGRFS, cGRFS and GRFS will be analyzed on the ITT in the same manner as RFS using the stratified log-rank test with the same strata as described in Section 7.4.1.1.

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).

An overview table to report the number and percentage of participants and an overview table to report number of events and events adjusted by patient years of drug exposure 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,
- 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 participants with TEAEs and the number of events and events adjusted by patient years of 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,
- TEAE leading to dose reduction,
- Drug related TEAE leading to dose reduction,
- TEAEs leading to dose interruption,
- Drug-related TEAEs leading to dose interruption,
- TEAEs excluding serious adverse events that equal to or exceed a threshold of 10% 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
- Grade 3 or higher TEAEs,
- Drug-related Grade 3 or higher TEAEs.

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

- TEAE
- Serious TEAE
- TEAE leading to withdrawal of treatment,
- Drug related-TEAE leading to withdrawal of treatment,
- TEAE leading to dose reduction,
- Drug-related TEAE leading to dose reduction,
- TEAE leading to dose interruption,
- Drug-related TEAE leading to dose interruption,
- Grade 3 or higher TEAE,
- Drug-related Grade 3 or higher TEAE

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

The number and percentage of participants 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 participant count, if a participant has multiple TEAEs with the same SOC or PT, but with differing severity grade or relationship, then the participant 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 participant 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.

The number and percentage of participants with treatment-emergent adverse events of interest, as classified by SOC and PT will also be summarized by treatment group and overall. The list of adverse events of interest to be summarized may change during the course of the study due to ongoing pharmacovigilance. It will be finalized before the database hard lock.

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

7.5.2 Infection

The number of infections and the number of participants experiencing infections will be tabulated by type of infection, severity, and time period after transplant. The cumulative incidence of CTCAE grade 3 to 5 infections, treating treatment emergent death without infection (grade 5) as a competing event, will be compared between the treatment arms using Gray's test. Participants who are still alive and no infection event at the time of analysis will be censored at the end of treatment plus 30 days or at the analysis cutoff date, whichever is earlier.

7.5.3 Clinical Laboratory Evaluation

Quantitative clinical laboratory variables, i.e., hematology, biochemistry, will be summarized using mean, standard deviation, minimum, maximum and median for each treatment group at each visit. Additionally, a within-participant change will be calculated as the post-baseline measurement minus the baseline measurement and summarized in the same way. Changes in laboratory values will be summarized for baseline values versus minimum, maximum and last post-baseline values. Plots of median lab values at each scheduled assessment time by treatment group may be provided for each laboratory parameter.

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 participant can be counted for both values if the participant has different laboratory values meeting each criterion. NCI-CTCAE grade of laboratory evaluations will be summarized by number and percentage of participants for each visit. Shift tables of NCI-CTCAE grade change from baseline to worst post-baseline grade for selected laboratory parameters will also be presented. The number and percentage of participants 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).

The list of laboratory parameters to be summarized may change during the course of the study due to ongoing pharmacovigilance. It will be finalized before the database hard lock.

Laboratory data will be displayed in listings.

7.5.3.1 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 participant's worst post-baseline record 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 participants with potentially clinically significant values in liver enzymes and total bilirubin will be presented by treatment group.

7.5.4 Vital Signs

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 participant for each treatment group. The results of vital signs may be displayed in figures.

The following potentially clinically significant criteria are defined:

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

7.5.5 Electrocardiograms (ECGs)

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 participants 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 participant.

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 frequency of participants 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.

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

Number and percent of participants with 12 lead ECG abnormalities as well as number and percent of participants 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.

7.5.6 Pregnancies

A detailed listing of all pregnancies will be provided.

7.5.7 Karnofsky Performance Status Scores

Karnofsky performance status scores will be summarized by treatment group and visit. Distribution of participants with scores <90, 90-100 will be tabulated by treatment group and visit.

7.6 Analysis of PK

PK analysis will be conducted on the PKAS.

Trough gilteritinib concentrations for a given timepoint will be excluded from the analysis if it meets at least one of the following exclusion criteria:

- Discontinuation of study drug prior to day of sample collection.
- Change in dose from 120 mg prior to day of sample collection (For participants that return to 120 mg dose, minimum of 14 days of dosing required prior to day of sample collection).
- Pause of ≥ 2 days in 14 days prior to day of sample collection (Note for C1D8 sample, exclude if pause of ≥ 1 day in 7 days prior to day of sample collection).
- When the time of PK sample collection occurs after dose administration on that day.

Trough gilteritinib concentrations will be summarized 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. In addition, mean (SD) plot of trough gilteritinib concentration by visit from 0 to 6 months will be produced.

ASP2215 C_{trough} group is defined based on median ASP2215 C_{trough} C1D15 value. If individual subject's C_{trough} value is greater or equal to median ASP2215 C_{trough} value, then it's defined as high ASP2215 C_{trough}, otherwise it is defined as low ASP2215 C_{trough}.

Demographic and other baseline characteristics will be summarized by ASP2215 C_{trough} C1D15 group using descriptive statistics.

KM plot of RFS and separately for OS by median ASP2215 C_{trough} C1D15 group will be provided.

Listing of individual participant trough gilteritinib concentration by visit will also be provided. 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.

7.7 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	< 50 years ≥ 50 - <60 years ≥ 60 years
Sex	Female Male
MRD status at randomization	Present Absent
Pre-transplant MRD status at enrollment from IRT	Present Absent/Indeterminate
Region	North America Europe Asia/Pacific/ Rest of World
Conditioning regimen intensity from IRT	Myeloablative Reduced intensity/non-myeloablative
GVHD prophylaxis	CNI (CsA or Tac) + MTX +/- Others CNI (CsA or Tac) + MMF +/- Others Other
Time from transplant to randomization from IRT	30 – 60 days 61 – 90 days
Cytogenetic Risk Status	Favorable Intermediate Unfavorable Other/Unknown
CR status prior to randomization	CR CRp+CRi
Prior use of FLT3 inhibitor	Yes No

Race	White Black or African American Asian Other/Missing
HSCT-Specific Comorbidity Index	0 1-2 >=3
Baseline Karnofsky Performance Score	>=90 <90

Subgroup analysis for incidence of FLT3 mutation at relapse (present/absent) will also be provided.

Additional subgroup analysis based on specific regions such as Japan may also be performed.

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

Policies and composition of the DSMB are described in the BMT CTN's Manual of Procedures. No interim efficacy or futility analyses are planned. RFS with immature data may be participant to overestimation; the data are not robust and is rarely reproducible when more mature data are available. For this reason, early evaluation may lead to stopping the trial too early, and therefore the DSMB will not stop the trial due to differences in RFS. Toxicity, AEs, and other safety endpoints will be monitored regularly and reported to the DSMB at each meeting.

Monitoring of a key safety endpoint will be conducted monthly, and if rates significantly exceed pre-set thresholds, the NHLBI will be notified so that the DSMB can be advised. The stopping guideline serves as trigger for consultation with the DSMB for additional review and are not formal "stopping rules" that would mandate automatic closure of study enrollment. Guidelines for safety monitoring is being outlined in Section 5.3.1 of the protocol.

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

7.10.1 Missing Data

It should be noted that date conventions were used so that dates could be consistently entered through the EDC system. Dates that were deemed critical had an associated field added so that the site could indicate if the day or month was unknown and entered utilizing the convention.

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.

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:

		Stop Date						
		Complete: *yyyymmdd*		**Partial: *yyyymm***		**Partial: *yyyy***		**missing**
Start Date		**< 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

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.

7.10.2 Outliers

All values will be included in the analyses.

7.10.3 Visit Windows

Not applicable. Nominal visits will be used in the by visit summary. Values from unscheduled visits will be included in the summary of extreme cases (e.g., summary of worst post-baseline, summary of minimum post-baseline, summary of maximum post-baseline). For efficacy endpoints, all values (scheduled and unscheduled) will be included in the analysis.

7.10.4 Blinding

This is a double-blind study. Participants will be randomized to receive gilteritinib or placebo in a double-blind fashion such that neither the investigator, Sponsor's study management team, clinical staff, nor the participant will know which agent is being administered. The randomization number will be assigned based on information obtained from the Interactive Response Technology (IRT) system and Advantage eClinical[SM].

The site and the BERC members will also be blinded to the MRD status used for stratification.

The randomization list and study medication blind will be maintained by the IRT system. The Data Safety Monitoring Board (DSMB) will be provided access to the dosing assignment for periodic review of the masked data as documented in the DSMB Charter and the study specific DSMB Charter Addendum. The DSMB may elect to break the mask.

The Sponsor may break the treatment code for participants who experience a suspected unexpected serious adverse reaction (SUSAR), in order to determine if the individual case or a group of cases requires expedited regulatory reporting. Individual Emergency Codes will be provided to the limited staff who are responsible for breaking the codes for all SUSAR cases for reporting purposes.

The treatment code for each randomized participant will be provided by the IRT in the event of medical emergency requiring knowledge of the treatment assigned to the participant. The time, date, participant number and reason for obtaining any of these codes, and therefore breaking the blind, must be documented in the study file. They must only be requested by the investigator or other persons designated as sub-investigators. No participants or other study personnel will be made aware of the treatment given to any participant unless a medical emergency necessitates such disclosure.

Unblinding of the study drug should only be considered for participant safety and/or evidence of documented morphological relapse contingent upon knowing the blinded study drug assignment.

- Unblinding for patient safety by the investigator or designated sub-investigator must be reported immediately to the Sponsor (Astellas Medical Monitor) and must include an explanation of why the study medication was unblinded. If possible, the Sponsor should be contacted prior to unblinding of the study drug.

- Unblinding for documented relapse by the investigator or designated sub-investigator must be reported to the Sponsor (Astellas Medical Monitor), including an explanation and evidence of relapse prior to unblinding of the study drug. The Astellas Medical Monitor must provide approval for unblinding for relapse prior to unblinding.

The study will be unblinded after database hard lock for the primary analysis.

7.10.5 Pooling of Strata

In the RFS, OS and EFS 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 MRD status pre-HSCT, conditioning regimen, and time from HSCT to randomization until the issue is resolved or the normal (un-stratified) Cox proportional hazard model is applied. For stratified analysis with GVHD prophylaxis regimen as one of the stratification factors, the order of pooling will be MRD status pre-HSCT, conditioning regimen, time from HSCT to randomization and GVHD prophylaxis regimen.

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.

8 DOCUMENT REVISION HISTORY

Version	Date	Changes	Comment/rationale for change
1.0	21-Mar-2017	NA	Document finalized
2.0	21-Feb-2023	Changed the term subject to participant	To match the terminology in the protocol
		Section 1 Updated to clarify the timing of SAP finalization.	To match the language from the SAP template
		Section 2 Flow chart was updated Table 1 Schedule of Assessments was deleted	To match the flow chart in protocol version 4 Schedule of assessment from the protocol was referred to ensure consistency.
		Section 3.1.2 – added an objective for infection endpoint	To reflect protocol version 4 objectives
		Section 3.1.3 – added an objective for FLT3 mutation status, aGRFS, cGRFS and GRFS	To reflect protocol version 4 objectives
		Section 3.2 – the text for study design was updated The timing of the primary and final analyses was clarified	To reflect the changes in the study design and the timing of the analyses made in protocol version 4
		Section 4 – added the estimated statistical power for RFS and OS after 2.5 years from the last patient randomized. Additionally, the estimated power for OS after 5 years from the last patient randomized was also included.	To provide the estimated power for RFS and OS for primary analysis and power for OS for final analysis
		Section 5 – per protocol set was deleted. In addition, reference to analysis set classification meeting was deleted	Per protocol set will not be used for any summaries or analyses Analysis sets will be derived programmatically, classification meeting is not needed
		Section 5.3 Pharmacokinetics Analysis Set (PKAS) definition was updated to clarify both dosing date/time and sampling date/time must be available.	To clarify the detailed PKAS criteria.
		Section 6.1.1 to clarify the censoring rules for RFS is up to follow-up visit where no relapse was indicated	To clarify RFS censoring rule

		Table 1 was updated to include different sensitivity analysis for RFS Alternative definitions for RFS accounting for the impact of Covid-19 were added	To provide the derivation of RFS for different sensitivity analyses Alternative definitions were added to assess the impact of Covid-19.
		Section 6.1.2.1: Added the definition of last known alive date Alternative definitions for OS accounting for the impact of Covid-19 were added	To clarify the derivation of last known alive date Alternative definitions were added to assess the impact of Covid-19.
		Section 6.1.2.2 Updated the censoring rules for EFS Added the censoring rule for NRM and relapse aGVHD II-IV, aGVHD III-IV and cGVHD – modified the censoring rule time to MRD eradication and detection – modified the censoring rule	To match the censoring rule for RFS To provide details on the censoring rules for NRM and relapse To clarify that the censoring rule is up to GVHD assessment only instead of last known alive date To clarify that the censoring rule is up to MRD assessment only instead of last known alive date
		Section6.1.3 – aGRFS, cGRFS and GRFS were added as exploratory endpoints	To reflect the addition of these endpoints in protocol version 4
		Section 6.2 and added Section 10.2 to clarify Adverse events of special safety interest (AESI) searching strategy	To provide details on the search strategy for AESI
		Section 7.1 Added the statement regarding null and alternative hypotheses Computing environment was updated	 To clarify the default null and alternative hypotheses being tested To reflect the latest SAS version we have in the computing environment

		Baseline definition of MRD was added	To clarify that MRD assessment has a different window for baseline derivation
		Sections 7.2.5 and 7.2.6 were added to have summaries for prior AML therapy and new ant-leukemic therapy	Prior AML therapy and new anti-leukemic therapy will be summarized
		Section 7.2.3 was updated to add subgroup of cytogenetic risk status for demographics and subgroup analysis.	Cytogenetic risk status will be summarized and analyzed
		Section 7.3.1 Exposure time categorization was updated in months	To change the unit in months instead of days
		Section 7.4 Detailed description on how to address multiplicity was added	Clarified the testing procedure to address multiplicity
		Section 7.4.1 Added the timing of the primary analysis for RFS Clarified the stratification factors for analysis Clarified the grouping for GVHD prophylaxis Deleted PPS Added RMST analysis for RFS Added sensitivity analysis for RFS taking into account Covid-19 impact.	To clarify the different analyses for RFS
		Section 7.4.2 Added the timing of the primary analysis for OS. Added the alpha spending for OS primary and final analyses Clarified the stratification factors for analysis Clarified the grouping for GVHD prophylaxis	To clarify the different analyses for OS

		Deleted PPS Added sensitivity analysis for OS taking into account Covid-19 impact.	
		Section 7.4.2.2 aGVHD II-IV, aGVHD III-IV and cGVHD – clarified the definition of the event of interest. In addition, both relapse and death were added as competing risks Updated the analysis for MRD as prognostic factor	To clarify the analysis for aGVHD II-IV, aGVHD III-IV and cGVHD To clarify the analysis for MRD as prognostic factor
		Section 7.4.3 Added the analyses for aGRFS, cGRFS and GRFS Added the imputation rule for duration of hospitalization	To clarify the analyses to be made for the new exploratory endpoints aGRFS, cGRFS and GRFS To provide guidance on how to compute the duration if end date is missing
		Updated Section 7.5.1 to include additional summaries of AEs and AE summaries by patient-year	To clarify the different summaries for AE
		Section 7.5.2 – added the censoring rule for participant who did not experience infection	To provide details on the censoring rule for infection
		Updated Section 7.5.3 to add the grade 3 and 4 summary for selected lab test parameters	To add new summary for critical lab results
		Updated Section 7.5.5 to add ECG abnormalities	To add new summary for ECG abnormalities
		Section 7.6 Added the criteria for excluding records from the analysis Added more details on the different analysis for PK	To clarify exclusion criteria for analysis To clarify the different analysis for PK
		Section 7.6.8 Added the following subgroups Cytogenetic risk, CR status prior to randomization, prior use of FLT3 inhibitor, race, HSCT-specific	Additional subgroups were included in the analyses

		comorbidity index and baseline Karnofsky performance score	
		Updated Section 7.10.4 for handling unblinding	Updated per protocol amendment 2
		Added section 7.10.5	To clarify the strata pooling strategy for efficacy analysis
		Updated section 10.2 search strategy for adverse event of special interest	To implement updated search strategy for adverse event of special interest

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, Willman CL, et al. Revised Recommendations of the International Working Group for Diagnosis, Standardization of Response Criteria, Treatment Outcomes, and Reporting Standards for Therapeutic Trials in Acute Myeloid Leukemia. J Clin Oncol. 2003;21(24):4642-4649.

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)

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

Wassmer G and Brannath W, Group Sequential and Confirmatory Adaptive Designs in Clinical Trials, 2016

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	Data Science Astellas Pharma Global Development, Inc.
PPD	Division of Biostatistics Medical College of Wisconsin

Contributors and Reviewers

PPD	Data Science Astellas Pharma Global Development, Inc.
PPD	Medical Science Astellas Pharma Global Development, Inc.
PPD	Medical Science Astellas Pharma Global Development, Inc.
PPD	Clinical Sciences Astellas Pharma Global Development, Inc.
PPD	Clinical Pharmacology & Exploratory Development Astellas Pharma Global Development, Inc.

PPD	Clinical Pharmacology & Exploratory Development Astellas Pharma Global Development, Inc.
PPD	Health Economics & Outcomes Research– Oncology Medical Affairs Astellas Pharma Global Development, Inc.
PPD	Data Science Astellas Pharma Global Development, Inc
PPD	Division of Hematology and Oncology Medical College of Wisconsin
PPD	Division of Hematologic Malignancies Johns Hopkins University
PPD	Department of Hematology/Oncology Biostatistics Massachusetts General Hospital

Author and Approver Signatories

(E-signatures are attached at end of document)

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

PPD, Division of Biostatistics, Medical College of Wisconsin, was the study statistician for this study and the primary author of this Statistical Analysis Plan

PPD

PPD, Data Science was the global statistical lead (GSTATL) for this project and biostatistics peer reviewer of this Statistical Analysis Plan

This Statistical Analysis Plan was approved by:

PPD

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