STATISTICAL ANALYSIS PLAN

NCT Number: NCT03589326

Study Title: A Phase 3, Randomized, Open-label, Multicenter Study Comparing Ponatinib Versus Imatinib, Administered in Combination With Reduced-Intensity Chemotherapy, in Patients With Newly Diagnosed Philadelphia Chromosome Positive Acute Lymphoblastic Leukemia (Ph+ ALL)

Study Number: Ponatinib-3001

Statistical Analysis Plan Version and Date:

Version 3.0: 20-September-2022

STATISTICAL ANALYSIS PLAN

STUDY NUMBER: Ponatinib-3001

A Phase 3, Randomized, Open-label, Multicenter Study Comparing Ponatinib Versus Imatinib, Administered in Combination with Reduced-Intensity Chemotherapy, in Patients with Newly Diagnosed Philadelphia Chromosome–Positive Acute Lymphoblastic Leukemia (Ph+ ALL)

PHASE 3

Version**: Final 3.0**

Date**:** 20 September 2022

Prepared by:

[redacted] PhD
[redacted] Oncology Statistics

Based on:

Protocol Version: Amendment 10

Protocol Date: 20 October 2021

1.1 Approval Signatures

Study Title: A Phase 3, Randomized, Open-label, Multicenter Study Comparing Ponatinib Versus Imatinib, Administered in Combination with Reduced-Intensity Chemotherapy, in Patients with Newly Diagnosed Philadelphia Chromosome–Positive Acute Lymphoblastic Leukemia (Ph+ ALL)

Approvals:

[redacted] PhD [redacted] Oncology Statistics	Date

2.0 TABLE OF CONTENTS

LIST OF IN-TEXT TABLES

LIST OF IN-TEXT FIGURES

3.0 LIST OF ABBREVIATIONS

Abbreviation	Term
ABI	ankle-brachial index
AE(s)	adverse event(s)
AESI(s)	adverse event(s) of special interest
ALL	chromosome-positive acute lymphoblastic leukemia
ALP	albumin, alkaline phosphatase
ALT	alanine aminotransferase
ANC	absolute neutrophil count
AOE(s)	arterial occlusive event(s)
AST	aspartate aminotransferase
BCR	breakpoint cluster region
BCR-ABL	breakpoint cluster region-Abelson
BP	blood pressure
BSA	body surface area
CI(s)	confidence intervals
CMH	Cochran-Mantel-Haenszel
CML	chronic myeloid leukemia
CNS	central nervous system
CR	complete remission (complete response)
Cri	incomplete complete remission
EAIR	Exposure-adjusted incidence rates
ECG(s)	Electrocardiogram(s)
ECHO	echocardiogram
ECOG	Eastern Cooperative Oncology Group
EFS	event-free survival
EQ-5D-5L	EuroQOL-5 Dimension-5 Level
EOT	end of treatment
FA	final analysis
FACT-Leu	Functional Assessment of Cancer Therapy – Leukemia
HR	hazard ratio
HLT	High Level Term
HRQOL	health-related quality of life
HSCT	hematopoietic stem cell transplantation
IA	interim analysis
IDMC	independent data monitoring committee
IPCW	Inverse Probability Censoring Weighting
IS	International Scale
IXRS	interactive voice/web response system
ITT	intent-to-treat
K-M	Kaplan-Meier

Abbreviation	Term
LDH	lactate dehydrogenase
LVEF	left ventricular ejection fraction
MedDRA	Medical Dictionary for Regulatory Activities
MR3	molecular response 3-log reduction (BCR-ABL1/ABL1 ≤0.1%)
MR4	molecular response 4-log reduction (BCR-ABL1/ABL1 ≤0.01%)
MR4.5	molecular response 4.5-log reduction (BCR-ABL1/ABL1 ≤0.0032%)
MRD	minimal residual disease
MRU	medical resource utilization
MSM	Marginal Structural Model
MUGA	Multiple-Gated Acquisition
NCI CTCAE	National Cancer Institute Common Terminology Criteria for Adverse Events
ORR	overall response rate
OS	overall survival
PD	progressive disease (disease progression)
Ph+ ALL	Philadelphia chromosome–positive acute lymphoblastic leukemia
PIF	primary induction failure
PK	Pharmacokinetic
PP	per-protocol
PT	Preferred Term
QD	once daily
QTcF	QT interval corrected per Fridericia method
SAE(s)	serious adverse event(s)
SAP	statistical analysis plan
SD(s)	standard deviation(s)
SOC	System Organ Class
TEAE(s)	treatment-emergent adverse event(s)
TKI	tyrosine kinase inhibitor
VTE(s)	venous thrombotic/embolic event(s)
WHO	World Health Organization

4.0 OBJECTIVES

4.1 Primary Objectives

The primary objective of the study is to compare the efficacy of ponatinib versus imatinib, administered as first-line therapy in combination with reduced-intensity chemotherapy, in patients with newly diagnosed Ph+ ALL, as measured by the MRD-negative CR rate at the end of induction (see Table 5.a for the definitions of MRD negativity and CR).

4.2 Secondary Objectives

4.2.1 Key Secondary Objectives

The key secondary objective is to compare event-free survival (EFS) between the 2 cohorts.

4.2.2 Other Secondary Objectives

Other secondary objectives are:

- To compare the rates of CR and incomplete CR (CRi) between the 2 cohorts, at the end of Cycle 1, the end of Cycle 2, the end of induction, and during or at the end of consolidation (end of Cycle 9 or end of study treatment, whichever occurs earlier).
- To compare the rates of MR3, MR4, and MR4.5 between the 2 cohorts, at the end of Cycle 1, the end of Cycle 2, the end of induction, and during or at the end of consolidation (end of Cycle 9 or end of study treatment, whichever occurs earlier).
- To compare the rates of primary induction failure (PIF) and overall response rate (ORR) between the 2 cohorts, at the end of induction.
- To compare rates of MRD-negative CR at multiple intervals after the end of induction.
- To determine the duration of MRD-negative CR in each of the 2 cohorts.
- To determine the duration of CR in each of the 2 cohorts.
- To compare the time to treatment failure between the 2 cohorts.
- To compare the duration of MR4.5 between the 2 cohorts, in patients who achieved MR4.5.
- To compare outcomes in patients with and without HSCT, between the 2 cohorts.
- To compare OS between the 2 cohorts.
- To collect plasma concentration-time data to contribute to population pharmacokinetic (PK) and exposure-response analyses of ponatinib.

4.3 Safety Objectives

The safety objectives are:

- To characterize the rates of AEs/SAEs, AOEs, venous thrombotic/embolic events (VTEs), and other safety outcomes of interest in the 2 cohorts, using multiple methods.
- To compare the tolerability between the 2 cohorts, including the rates of discontinuation, dose reductions, and dose interruptions due to AEs.

4.4 Exploratory Objectives

The exploratory objectives are:

- To compare patient-reported quality of life (Functional Assessment of Cancer Therapy – Leukemia [FACT-Leu] and EuroQOL-5 Dimension-5 Level [EQ-5D-5L]) results between the 2 cohorts.
- To compare medical resource utilization (MRU) results between the 2 cohorts.
- To compare the time to start of alternative therapy between the 2 cohorts.
- To compare the time to HSCT between the 2 cohorts.
- To explore biomarkers of disease sensitivity and resistance to ponatinib and imatinib and/or biomarkers affecting ponatinib efficacy or safety.

4.5 Study Design

4.5.1 Study Design

This phase 3 study is designed as an open-label, multicenter, randomized comparison of the TKIs ponatinib versus imatinib, when administered as first-line therapy in patients aged ≥18 years with newly diagnosed Ph+ ALL. The TKIs will be administered in combination with 20 cycles of a reduced-intensity chemotherapy regimen (including 3 cycles of induction therapy, 6 cycles of consolidation therapy, and 11 cycles of maintenance therapy), followed by single-agent therapy with ponatinib or imatinib, to be administered continuously. Patients will remain on study treatment until they are deceased, have failed to achieve the primary endpoint at the end of induction (patients who do not achieve the primary endpoint may remain on study drug, at the investigator's discretion, if they have achieved CR at the end of induction), have experienced relapse from CR or have progressive disease, have an unacceptable toxicity, have withdrawn consent, have proceeded to HSCT or alternative therapy, or until the sponsor terminates the study, whichever occurs first.

The primary endpoint of this study is MRD-negative CR at the end of induction (defined in Table 5.a). Patients who achieve the primary endpoint will continue in the study in the consolidation and maintenance phases followed by a single-agent therapy phase. Patients who achieve CR but do not achieve the primary endpoint at the end of induction may, at the investigator's discretion, continue on study treatment. All patients who do not achieve CR at the

end of induction will be discontinued from study drug. For all discontinued patients, the patient's treating physician should consider alternative therapy options.

Upon enrollment, patients will be randomized in a 2:1 ratio of ponatinib:imatinib to be taken throughout the study, beginning on Cycle 1 Day 1. Patients randomized to Cohort A (ponatinib) will receive 30 mg of oral ponatinib QD, which will be reduced to 15 mg if MRD-negative CR is achieved at the end of induction. If a patient loses MRD negativity after dose reduction to 15 mg, re-escalation to 30 mg may be considered after discussion with the sponsor's medical monitor/designee. Dose reductions to 10 mg of ponatinib QD may be considered for safety reasons after discussion with the sponsor's medical monitor/designee (see Protocol Section 8.4.1). For patients in the ponatinib cohort who achieve CR but do not achieve the primary endpoint at the end of induction and who continue in the study at the investigator's discretion, the dose of ponatinib will be reduced, as described above, at any time when the patient achieves MRD-negative CR and re-escalated, as described above, upon loss of response. Patients randomized to Cohort B (imatinib) will receive 600 mg of oral imatinib QD. Intrathecal therapy will be performed twice per month for the first 6 cycles for central nervous system (CNS) disease prophylaxis. At the end of the 20 cycles, all patients remaining on study will remain on ponatinib or imatinib (administered as a single agent).

MRD status will be measured using qPCR-based tests validated for the ability to detect breakpoint cluster region-Abelson (BCR-ABL1)/ABL1 levels with a minimal sensitivity of 0.01%, with MRD negativity defined as ≤0.01% BCR-ABL1/ABL1. Separate tests will be used to assess the *p210* and *p190* variants of BCR-ABL1 (see Protocol Section 4.2.3), which comprise >95% of the variants present in adult patients with Ph+ ALL. For the p210 test, BCR-ABL1/ABL1 levels will be reported on the International Scale (IS) with traceability to the World Health Organization (WHO) first International Genetic Reference Panel. For the p190 test, for which there is no internationally available reference material, the raw ratio of BCR-ABL1/ABL1 levels will be reported.

The key secondary endpoint for this study is EFS. Other secondary endpoints will include rates of CR and CRi at the end of Cycle 1, Cycle 2, the end of induction, and during or at the end of consolidation (end of Cycle 9 or end of study treatment, whichever occurs earlier); rates of MR3, MR4, and MR4.5 at the end of Cycle 1, the end of Cycle 2, the end of induction and during or at the end of consolidation (end of Cycle 9 or end of study treatment, whichever occurs earlier); rates of MRD-negative CR at multiple intervals after the end of induction; rates of PIF and ORR at the end of induction; duration of MRD-negative CR; duration of CR; time to treatment failure; duration of MR4.5 in patients who achieved MR4.5; OS and rate of relapse from CR for on-study patients with and without HSCT, and OS.

Safety and tolerability parameters will be assessed in both cohorts, including incidence of all AEs, SAEs, AOEs, and VTEs; rates of discontinuation, dose reductions, and dose interruptions due to AEs; incidence of death while on treatment, and changes from baseline in vital signs and laboratory test results. Plasma concentration-time data will also be collected for patients receiving ponatinib.

Exploratory endpoints will include change from baseline in patient-reported quality-of-life; MRU assessments; time to start of alternative therapy; time to HSCT; and biomarkers of disease sensitivity and resistance to ponatinib and imatinib and/or biomarkers affecting ponatinib efficacy or safety.

4.5.2 Randomization and Stratification

The randomization scheme will be generated by an independent statistician who is not on the study team. Before dosing, a randomization number will be assigned to each patient. The randomization assignment will be implemented by an interactive voice/web response system (IXRS).

Eligible patients will be randomized in a 2:1 ratio to receive ponatinib or imatinib treatment arms, stratified by ages: 18 through <45 years; ≥45 through <60 years; and ≥60 years.

5.0 ANALYSIS ENDPOINTS

5.1 Primary Endpoints:

The primary endpoint is MRD-negative CR at the end of induction (see Table 5.a for the definitions of MRD negativity and CR).

5.2 Secondary Endpoints:

5.2.1 Key Secondary Endpoints

The key secondary endpoint is:

- EFS, defined as the dates of randomization until:
  - Death due to any cause.
  - Failure to achieve CR by the end of induction.
  - Relapse from CR.

5.2.2 Other Secondary Endpoints

Other secondary endpoints (defined in Table 5.a) are:

- CR and CRi rates at the end of Cycle 1, the end of Cycle 2, the end of induction, and during or at the end of consolidation (end of Cycle 9 or end of study treatment, whichever occurs earlier).
- Molecular response rates (MR3, MRD negativity [MR4], and MR4.5) at the end of Cycle 1, the end of Cycle 2, the end of induction, and during or at the end of consolidation (end of Cycle 9 or end of study treatment, whichever occurs earlier).
- Rates of PIF and ORR at the end of induction.

- Rates of MRD-negative CR at multiple intervals after the end of induction.Duration of MRD-negative CR.
- Duration of CR.
- Time to treatment failure.
- Duration of MR4.5 in patients who achieved MR4.5.
- OS and rate of relapse from CR for on-study patients with and without HSCT.
- OS.

Table 5.a Definitions of Efficacy Response Criteria

Term	Definition
CNS-1	CNS-1: No lymphoblasts in the CSF regardless of WBC count.
CNS-2	WBC count <5 leukocytes/µl in the CSF with the presence of blasts.
CNS-3 [a]	WBC count of ≥5 leukocytes/µl with the presence of blasts.
CNS disease remission	No lymphoblasts in CSF regardless of WBC count in a patient with CNS-2 or CNS-3 at diagnosis.
CNS relapse	Development of CNS-3 status or development of clinical signs of CNS leukemia (eg, facial nerve palsy, brain/eye involvement, hypothalamic syndrome).
CR	Complete remission; meeting all the following for at least 4 weeks (ie, no recurrence): • No circulating blasts and <5% blasts in the BM. • Normal maturation of all cellular components in the BM. • No extramedullary disease (CNS involvement, lymphadenopathy, splenomegaly, skin/gum infiltration, testicular mass). • ANC >1000/µl (or >1.0 × 10^9/L). • Platelets >100,000/µl (or >100 × 10^9/L).
CRi	Hematologic complete remission with incomplete hematologic recovery. Meets all criteria for CR except platelet count and/or ANC.
Duration of CR	The interval between the first assessment at which the criteria for CR are met until the time at which relapse from CR occurs.
Duration of MR4.5	The interval between the first assessment at which the criteria for MR4.5 are met until the earliest date at which loss of MR4.5 occurs.
Duration of MRD negativity	The interval between the first assessment at which the criteria for MRD negativity are met until the earliest date at which loss of MRD negativity occurs or relapse from CR occurs.
Duration of MRD-negative CR	The interval between the first assessment at which the criteria for MRD-negative CR are met until the earliest date at which loss of MRD negativity or relapse from CR occurs.

Table 5.a Definitions of Efficacy Response Criteria

Term	Definition
EFS	Event-free survival (EFS), defined as the dates of randomization until: • Death due to any cause. • Failure to achieve CR by the end of induction. • Relapse from CR.
Loss of MR3	An increase to >0.1% BCR-ABL1/ABL1.
Loss of MR4.5	An increase to >0.0032% BCR-ABL1/ABL1. This result must be confirmed at the subsequent visit, unless it is associated with loss of MR3 or relapse from CR.
Loss of MRD negativity	An increase to ≥0.01% BCR-ABL1/ABL1. This result must be confirmed within 4 weeks with either a BM aspirate (optional) or peripheral blood, unless it is associated with loss of MR3 or relapse from CR.
MR3	Molecular response 3-log reduction (≤0.1% BCR-ABL1/ABL1), or undetectable BCR-ABL1 transcripts in cDNA with ≥1000 ABL1 transcripts.
MR4.5	Molecular response 4.5-log reduction (≤0.0032% BCR-ABL1/ABL1), or undetectable BCR-ABL1 transcripts in cDNA with ≥32,000 ABL1 transcripts.
MRD-negative CR	Meeting the criteria for both MRD negativity and CR.
MRD negativity (MR4)	≤0.01% BCR-ABL1/ABL1, or undetectable BCR-ABL1 transcripts in cDNA with ≥10,000 ABL1 transcripts. Also referred to as MR4.
ORR	Overall response rate: CR + CRi.
OS	Overall survival. The interval between randomization and death due to any cause.
PD	Progressive disease. Increase of at least 25% in the absolute number of circulating or BM blasts or development of extramedullary disease.
PIF	Primary induction failure: Patients who received treatment for ALL but never achieved CR or CRi by the end of induction. PIF is not limited by the number of unsuccessful treatments; this disease status only applies to recipients who have never been in CR or CRi.
Relapse from CR	Reappearance of blasts in the blood or BM (≥5%) or in any extramedullary site after a CR.
Time to treatment failure	Time to end of study-randomized treatment (except for HSCT without loss of MRD-negative CR) due to safety and/or efficacy reasons.

Abbreviations: ANC, absolute neutrophil count; BCR-ABL, breakpoint cluster region-Abelson; BM, bone marrow; CNS, central nervous system; CSF, cerebrospinal fluid; CR, complete remission; CRi, incomplete blood count recovery; HSCT, hematopoietic stem cell transplant; MR3, molecular response 3-log reduction (BCR-ABL1/ABL1 ≤0.1%); MR4, molecular response 4-log reduction (BCR-ABL1/ABL1 ≤0.01%); MR4.5, molecular response 4.5-log reduction (BCR-ABL1/ABL1 ≤0.0032%); MRD, minimal residual disease; ORR, overall response rate; OS, overall survival; PD, progressive disease; Ph+ ALL, Philadelphia chromosome–positive acute lymphoblastic leukemia; RBC, red blood cell; WBC, white blood cell.

[a] If the patient has leukemic cells in the peripheral blood and the lumbar puncture is traumatic and WBC ≥5/μL in CSF with blasts, then compare the CSF WBC/RBC ratio to the blood WBC/RBC ratio. If the CSF ratio is at least 2-fold greater than the blood ratio, then the classification is CNS-3; if not, then it is CNS-2.

5.3 Safety Endpoints:

The safety endpoints are:

- Incidence and exposure-adjusted incidence rates of AOEs, VTEs, AEs, and SAEs, in each of the 2 cohorts.
- Incidence of dose reductions, interruptions, and discontinuations due to AEs, in each of the 2 cohorts.
- Incidence of death on treatment, in each of the 2 cohorts.
- Changes from baseline in vital signs (including systolic and diastolic BP, and heart rate) and clinical laboratory test results, in each of the 2 cohorts.

5.4 Pharmacokinetic Endpoint

The PK endpoint is plasma concentration-time data to contribute to population PK and exposure-response analyses of ponatinib.

5.5 Exploratory Endpoints

The exploratory endpoints are (see Table 5.a for the definitions):

- Change from baseline in patient-reported quality of life (FACT-Leu and EQ-5D-5L).
- MRU assessments.
- Time to start of alternative therapy.
- Time to start of HSCT.
- Biomarkers of disease sensitivity and resistance to ponatinib and imatinib and/or biomarkers affecting ponatinib efficacy or safety.

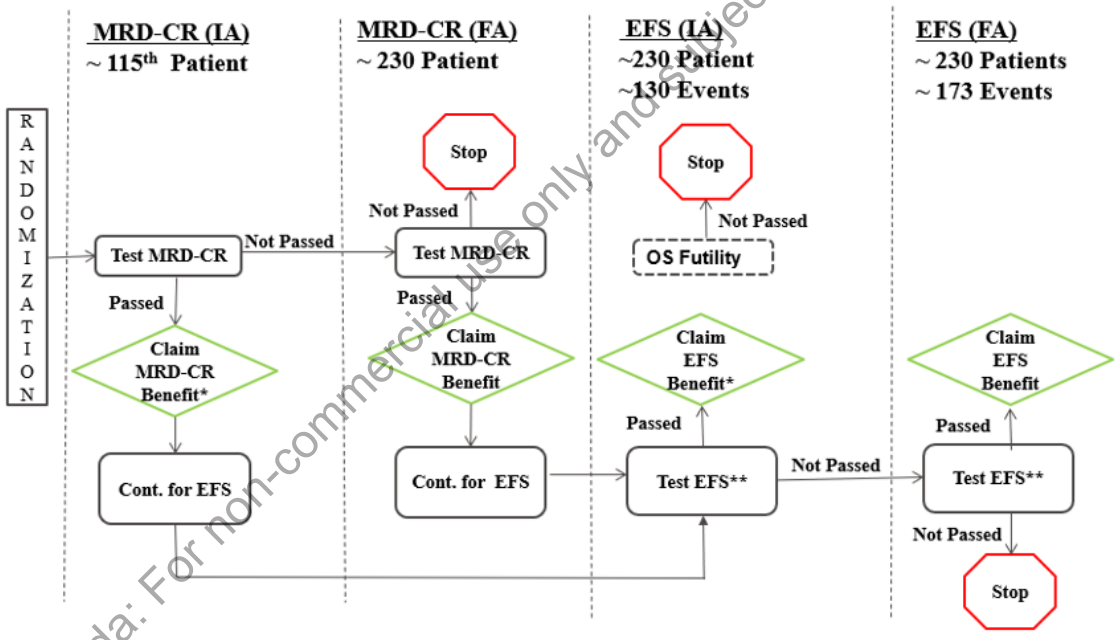
6.0 DETERMINATION OF SAMPLE SIZE

The study employed group sequential design for the primary endpoint MRD-negative CR, the key secondary endpoint EFS, and other endpoints (Duration of CR, Duration of MRD-negative CR, ORR and OS).

Assuming an effect size ranging from 20% to 28% (40 - 48% vs. 20% MRD-negative CR rates for the active and control arms, respectively), an upfront committed sample size of approximately 230 patients (approximately 153vs 77 for the active and control arms, respectively, based on a 2:1 allocation ratio) will provide 84% to 98% power for MRD-negative CR final analysis using the efficacy boundary of 0.036 according to the the group sequential testing procedure with IA performed from 116 patients [7]. The O'Brien-Fleming alpha spending function (the Lan-DeMets method [1]) will be implemented to determine the significance level at IA and FA for the primary endpoint, with an overall type I error rate at a 2-sided 0.05 level.

Inference for the key secondary endpoint of EFS will be conducted at α = 5% level only if the primary endpoint is met either at IA or FA for the MRD-negative CR (Figure 6.a). Based on 3-year EFS data observed from various phase 2 studies [2,3], effect size is assumed as 67% vs 46% for EFS at year 3 for the active and control arms, respectively, or HR = 0.516 for non-HSCT patients. The effect size is assumed as 53% and 40% for EFS at year 3 for active and control arms, respectively, or HR = 0.693 for patients who are undertaking HSCT. Also, it is assumed that 50% and 45% of patients from active and control arms will undertake HSCT, respectively. Based on simulation studies, approximately 230 patients will be enrolled to collect long-term EFS data. Among these 230 patients, approximately 173 events need to be accumulated at FA so that the power will be approximately 80% for the EFS endpoint. It is expected that the time of EFS will be approximately 8.5 years after first patient has been enrolled.

Figure 6.a Statistical Analysis Schema



*If efficacy boundary is crossed at IA, it is the final analysis and no formal hypothesis testing will be performed.
**If the efficacy boundary is crossed at either IA or FA for EFS, the following endpoints will be tested in the order using the same boundaries: a) Duration of CR; b) ORR; c) Duration of MRD-negative CR; d) OS.

Abbreviations: Cont., continue; CP, conditional probability; CR, complete response; EFS, event-free survival; FA, final analysis; IA, interim analysis; LPI, last patient in; MRD, minimal residual disease.

7.0 METHODS OF ANALYSIS AND PRESENTATION

7.1 General Principles

In general, summary tabulations will be presented by treatment arm, and will be displayed by the number of observations, mean, standard deviation (SD), median, minimum, and maximum for continuous variables, and the number and percent per category for categorical data. The Kaplan-Meier (K-M) survival curves and 25th, 50th (median), and 75th percentiles will be provided along with their 2-sided 95% confidence intervals (CIs) for time-to-event data.

Where appropriate, variables will be summarized descriptively by study visit. The denominator for the proportion will be based on the number of subjects who provided non-missing responses to the categorical variable.

A windowing convention will be used to determine the analysis value for a given study visit for observed data analyses.

All available efficacy and safety data will be included in data listings and tabulations as needed. Data that are potentially spurious or erroneous will be examined under the auspices of standard data management operating procedures.

Baseline values are defined as the last observed value before the first dose of study medication.

Screen failure subjects will be presented.

All statistical analyses will be conducted using SAS® Version 9.4, or higher.

7.1.1 Definition of Study Days

- Study Day 1 is defined as the date on which a subject is administered with their first dose of the medication. After Study Day 1, Study Day = Date of event - Date of the first dose + 1 day.
- Prior to Study Day 1, Study Day = Date of the first dose - Date of event.

7.1.2 Conventions for Missing/Partial Dates in Screening Visit

The following rules apply to dates recorded during the screening visits.

- If only the day-component is missing, the first day of the month will be used if the year and the month are the same as those for the first dose of study drug. Otherwise, the fifteenth will be used.
- If only the year is present, and it is the same as the year of the first dose of study drug, the fifteenth of January will be used unless it is later than the first dose, in which case the date of the first of January will be used, unless other data indicate that the date is earlier.
- If only the year is present, and it is not the same as the year of the first dose of study drug, the fifteenth of June will be used, unless other data indicates that the date is earlier.

7.1.3 Conventions for Missing Adverse Event Dates

AEs with start dates that are completely or partially missing will be analyzed as follows:

- If month and year are known but day is missing
  - If month and year are the same as month and year of first dose date, then impute to first dose date.
  - If month and year are different than month and year of first dose date, then impute to first date of the month.
- If year is known but day and month are missing
  - If year is same as year of 1st dose date, then 1st dose date will be used instead.
  - If year is different than year of 1st dose date, then 1st of January of the year will be imputed.
- If all is missing, then it is imputed with 1st dose date.

Imputing missing AE start date is mandatory. After the imputation, all imputed dates are checked against the start dates to ensure the stop date does not occur before start date. If the imputed stop date occurs prior to start date, then keep the imputed date same as the start date.

AEs with stop dates that are completely or partially missing will be analyzed as follows:

- If "ongoing" is checked, no imputation is necessary.
- If month and year are known but day is missing, the last day of the month will be imputed
- If year is known, but day and month are missing,
  - If YYYY < year of last dose, then 31st of December will be imputed.
  - If YYYY = year of last dose, then 31st of December will be imputed.
  - If YYYY > year of last dose, then 1st of January will be imputed.
- If all are missing, then impute date to 31st of December, in the year of last dose.

Imputing missing AE stop date is not mandatory if AE is regarded as ongoing. However, if it is to be done, the rules are outlined above. If subject dies, then use death date for AE stop date.

After the imputation, all imputed dates are checked against the start dates to ensure the stop date does not occur before start date. If the imputed stop date occurs prior to start date, then keep the imputed date the same as the start date.

7.1.4 Conventions for Missing Concomitant Medication/Therapy Dates

Concomitant medications/therapies with start dates that are completely or partially missing will be analyzed as follows:

- If month and year are known, but day is missing, then impute day to first of the month
  - If year is known, but day and month are missing, then 1st of January of the year will be imputed.
- If all is missing, then impute date to Date of Birth (DOB)
  - If DOB is not available but age is available, then estimate DOB by using screening date and age (age = screening date – DOB).

Concomitant therapies with stop dates that are completely or partially missing will be analyzed as follows:

- If "ongoing" is checked, no imputation is necessary.
- If month and year are known but day is missing, the last day of the month will be imputed
- If year is known, but day and month are missing,
  - If YYYY < year of last dose, then 31st of December will be imputed
  - If YYYY = year of last dose, then 31st of December will be imputed
  - If YYYY > year of last dose, then 1st of January will be imputed
- If all is missing, then impute date to 31st of December in the year of last dose

Imputing missing concomitant therapies is optional. However, if it is to be done, the rules are outlined above. If subject dies, then use death date for concomitant therapies stop date. After the imputation, all imputed dates are checked against the start dates to ensure stop date does not occur before start date. If the imputed stop date occurs prior to start date, then keep the imputed date same as the start date.

7.1.5 Conventions for Missing Subsequent Medication/Therapy Dates

Subsequent therapies with start dates that are completely or partially missing will be analyzed as follows:

- When month and year are present and the day of the month is missing,
  - If the onset month and year are the same as the month and year of last dose with study drug, the day of last dose + 1 will be imputed.
  - If the onset month and year are not the same as the month and year of last dose with study drug, the first day of the month is imputed.

- When only a year is present,
  - If the onset year is the same as the year of last dose with study drug, the date of last dose + 1 will be imputed.
  - If the onset year is not the same as the year of last dose with study drug, the first day of the year is imputed.
- If no components of the onset date are present the date of last dose + 1 will be imputed.

7.2 Analysis Populations

The Analysis Populations will include the following.

7.2.1 Intent-to-Treat Analysis Population

The intent-to-treat (ITT) analysis set is defined as all patients who are randomized. Patients will be analyzed according to the treatment they were randomized to receive, regardless of any errors of dosing.

7.2.2 Per-Protocol Analysis Population

The per-protocol (PP) population is a subset of the ITT population. The PP population consists of all patients who do not violate the terms of the protocol in a way that would affect the study outcome significantly, as determined by the sponsor's medical monitor/designee. All decisions to exclude patients from the PP population will be made before the database lock for the analyses.

The PP population will be used as a sensitivity analysis of the ITT population for the efficacy endpoints as needed if more than 5% of patients from the ITT population are excluded from this analysis.

7.2.3 Safety Analysis Population

The safety population is defined as all patients who are randomized to the ponatinib or imatinib arm and receive at least 1 dose of any study drug. Patients will be analyzed according to the treatment actually received. That is, patients who receive any dose of ponatinib will be included in the ponatinib arm, and patients who receive any dose of imatinib will be included in the imatinib arm, regardless of their randomized treatment.

Patients at sites in Japan are assigned to the ponatinib arm only. These nonrandomized patients will be analyzed separately.

7.3 Disposition of Subjects

Dispositions of patients will be summarized using number and percentage of patients based on the ITT analysis set by each treatment cohort and all patients combined for the following patient disposition data. All percentages will be based on the number of patients in the ITT analysis set.

- Patients who are in the PP analysis population.

- Patients who are in the safety analysis population.
- Patients who are on study treatment.
- Patients who discontinued from study treatment.
- Primary reason for discontinuing study treatment.
- Patients who are on study.
- Follow-up status.

7.4 Demographic and Other Baseline Characteristics

Demographic and baseline characteristics will be summarized using frequency distributions (ie, number and percentage of patients) for categorical data and summary descriptive statistics (ie, n, mean, SD, median, minimum, and maximum) for continuous data, based on the ITT analysis set for each treatment cohort and for all patients combined.

Demographic data will also be presented in a by-patient listing. Baseline demographic data to be evaluated will include age at informed consent, age category (ie, 18≤age<45 years, 45≤age<60 years, and age≥60 years), sex, race, ethnicity, height, weight, body surface area (BSA), and other parameters as appropriate. No inferential statistics will be generated.

Throughout this study, baseline assessments are defined as those performed at the closest time before the start of study drug administration.

Baseline characteristics include Eastern Cooperative Oncology Group (ECOG) score, time from initial diagnosis of PH+ ALL to first dose date prior anti-cancer regimen, Framingham Score, BCR-ABL transcript type, baseline white blood count, baseline homoglobin, baseline platelet, baseline blast count, extramedullary disease and other parameters as appropriate.

7.5 Medical and Surgical History

Medical history will be coded using the Medical Dictionary for Regulatory Activities (MedDRA). Medical history will be summarized by MedDRA System Organ Class (SOC), and Preferred Term (PT) using number and percentage of patients based on the ITT analysis set for each treatment cohort. Patients with the same medical history more than once will have that medical history counted only once within each SOC, and once within each PT.

Surgical history will not be coded. Surgical history will be summarized using number and percentage of patients based on the ITT analysis set for each treatment cohort, as approximate.

Medical and surgical history will be presented in a by-patient listing.

Family medical history will also be presented in a by-patient listing.

7.6 Prior and Concomitant Medications and Concomitant Procedures

Prior and concomitant medications will be coded using the World Health Organization (WHO) Drug Dictionary. Prior and concomitant medications will be summarized by WHO therapeutic

class and standard drug names using number and percentage of patients based on the ITT analysis set for prior medications and the safety analysis set for concomitant medications for each treatment cohort.

Prior and concomitant medications will also be presented in a by-patient listing.

Prior radiation, and prior anticancer therapy will be summarized using number and percentage of patients based on the ITT analysis set for each treatment cohort, as appropriate.

Concomitant procedures will not be coded, but will be presented in a by-patient listing.

7.7 Study Drug Exposure and Compliance

Extent of Exposure:

Parameters pertaining to study drug exposure (ie, duration of exposure, number of days dosed, number of cycles dosed, dose intensity, relative dose intensity, total cumulative dose) will be summarized separately by treatment cohort and overall treatment period. Duration of treatment exposure is defined as the time interval from the first dose to the last dose of study treatment (last dose date – first dose date +1).

Dose intensity in mg/day is calculated as total cumulative dose in mg divided by duration of treatment exposure in day. Relative dose intensity is calculated as total cumulative dose in mg divided by expected total dose × 100%.

The drug exposure will be listed and summarized for the following drugs.

- Ponatinib
- Imatinib
- Vincristine
- Dexamethasone
- Methotrexate
- Cytarabine
- Prednisone
- Intrathecal therapy for CNS disease prophylaxis

Dose adjustment:

Dose adjustment and dose adjustment due to adverse event (AE) will be summarized by treatment cohort for each treatment phase and overall treatment period.

A by-patient listing including extent of exposure and dose adjustment will be presented.

7.8 Efficacy Analysis

The standard *closed* sequential testing procedure will be used for testing the selected efficacy endpoitns with the following testing order.

1. MRD-negative CR rate: it will be tested at the IA or FA at the significance level determined by the O'Brien-Fleming alpha spending function (the Lan-DeMets method [1]) using the group sequential testing approach. At the IA (alpha = 0.022 with an efficacy boundary of 0.022) given 116 patients have been observed, and FA (alpha = 0.028 with an efficacy boundary of 0.036) for MRD-negative CR if number of patients is 230 [7].
2. EFS: EFS will be tested at the IA or FA at the significance level determined by the Gamma Family (-1) alpha spending function using group sequential testing approach [4]. At the IA (alpha = 0.033 with an efficacy boundary of 0.033) and FA (alpha = 0.017 with an efficacy boundary of 0.034) for EFS if observed number of events at IA and FA are 130 and 173, respectively. If the efficacy boundary is crossed at either IA or FA for EFS, the following endpoints will be tested in the order listed below using the same boundaries (0.033 for IA and 0.034 for FA) [8]:
   a) Duration of CR.
   b) ORR.
   c) Duration of MRD-negative CR.
   d) OS.

Therefore, the overall type I error rate for these selected efficacy endpoints is strongly controlled at a 2-sided 0.05 alpha level.

The boundaries for hypotheses testing will be updated according to the observed data in the IA and FA, using the prespecified alpha spending function.

For the secondary endpoint of OS, a futility analysis will be conducted at the time of the IA for EFS. The hazard ratio and corresponding 95% CI for the OS analysis will be calculated and reviewed by the IDMC. If the HR is >1.2, the IDMC will review the totality of the data and provide a recommendation to the sponsor's executive committee regarding study continuation.

All other efficacy endpoints, if tested, will be at a 2-sided alpha level of 0.05.

For MRD-negative CR, the analysis will be based on ITT population who have been identified with BCR-ABL1 dominant variants of p190 or p210. Other efficacy analyses will be conducted in the ITT population, unless otherwise specified.

MRD negativity will be based on the central laboratory results, and CR status will be based on the investigator's assessment verified by the Sponsor that the data reported for bone marrow blast, platelets and neutrophil counts are consistent with the protocol definition for CR in Table 5.a.

7.8.1 Primary Efficacy Endpoint(s)

The primary endpoint is defined as achievement of MRD-negative CR (BCR-ABL/ABL1 ≤0.01% and meeting criteria for CR) at the end of induction (see Table 5.a for endpoint definitions). The analysis of the primary efficacy endpoint will test for differences comparing the proportion of patients who achieve the primary endpoint at the end of induction in the ponatinib arm versus the imatinib arm. The patients who early terminate the study treatment prior to the end of induction will be considered as non-responders.

If a C4D1 visit or assessment is not done or not available (e.g. patient discontinues after C3D28 or has a C4D1 dry tap bone marrow at C4D1), then the next available assessment that is completed within 45 days of C3D1 or within 15 days of C4D1 (i.e. "EOT" or "Unscheduled" visit) will be used.

If a MRD assessment is not available for C4D1 (including above), and if the patient had at least one earlier sample assessed as MRD negative (e.g. C2D1 or C3D1), at least one later sample assessed as MRD negative up to and including C6D1 visit (+7 day window), and there are no intervening MRD positive results, then that patient will be considered to be MRD negative at C4D1. For all MRD analyses conducted in the study, only MRD assessments performed at central laboratories will be used and only from patients where a dominant p190 or p210 transcript was identified by central laboratories using baseline (Screening/C1D1) samples.

There will be one IA and possibly an FA in the study for the primary endpoint of MRD-negative CR in ITT population who have been identified with BCR-ABL1 dominant variants of p190 or p210.

If the MRD-negative CR does not achieve the significance boundary at the IA, the study will continue and the FA will be triggered after the end of induction phase data have been collected for approximately 230 patients.

The primary analysis for MRD-negative CR will be conducted using a Cochran-Mantel-Haenszel (CMH) chi-square test. The CMH chi-square p-value on risk difference between treatment arms will be calculated. The risk difference and relative risk will be presented along with 95% 2-sided confidence intervals.

Sensitivity analyses for the primary endpoint will include:

1. MRD-negative CR will be analyzed in the PP analysis set if more than 5% of patients are excluded from this analysis.
2. After the FA for MRD-negative CR is conducted, an additional sensitivity analysis for the primary endpoint will be retrospectively performed for the first 150 patients who have been enrolled and treated at the end of induction phase.

Other sensitivity analyses will be considered as appropriate.

Subgroup analyses will be performed for the primary endpoint relative to the baseline randomization stratification factor (age); additional age category (18 through <60 years; ≥60 years), demographic data, such as gender (male, female), race (white; non-white), region (north

America; south America; Europe; APAC); and baseline disease characteristics including BCR-ABL1 Transcript Type (P190; P210) and ECOG status (0; 1 or 2), as appropriate.

7.8.2 Secondary Efficacy Endpoint(s)

7.8.2.1 Key Secondary Efficacy Endpoint

The key secondary endpoint is EFS, defined as the dates of randomization until:

- Death due to any cause.
- Failure to achieve CR by the end of induction.
- Relapse from CR.

EFS will be tested only if the primary endpoint comparison achieves statistical significance at the IA or FA for MRD-negative CR. EFS endpoint will be tested at the 5% level at IA or FA for EFS, per the closed sequential testing procedure, to maintain the family-wise type I error rate at 5% level.

One IA and one FA will be planned for EFS. When approximately 130 EFS events are observed (75% of the total 173 expected EFS events), an IA will be performed. The FA will be performed when approximately 173 EFS events have been observed. A 2-sided, stratified log-rank test will be used to compare the treatment groups with respect to PFS at a 2-sided alpha level of 0.05 for ITT population. In addition, an unadjusted stratified Cox model will be used to estimate the hazard ratio and its 95% CIs for the treatment effect using the stratification factor. The K-M survival curves and K-M median PFS (if estimable), along with their 2-sided 95% CIs, will also be provided for each treatment group. The test significance for the IA and FA of EFS will be determined using Gamma Family (-1) boundaries. Based on the projected number of EFS events, the formal hypothesis testing will be stopped for overwhelming efficacy if the 2-sided p-value crosses the efficacy boundary of 0.033 at IA. The final analysis will be tested at 2-sided alpha level of efficacy boundary 0.034 (corresponding to nominal alpha of 0.017).

The primary analysis for EFS will be based on time-to-event analysis. Since it is expected that a subset of patients who achieve MRD-negative CR after the induction phase will proceed to HSCT, the number of events needed for EFS analysis may change depending on how HSCT cases are handled in the EFS. The primary analysis of EFS will not consider censoring at the time of HSCT or initiation of alternative therapy. Other details regarding the handling of missing assessments and censoring for EFS analysis are presented in Table 7.a.

Table 7.a Censoring Rules for EFS Primary Analysis Based on FDA Recommendations

Situation	Date of Progression or Censoring	Outcome
Death due to any cause	Date of death	Event
Failure to achieve CR by the end of induction	Day 1	Event
Relapse from CR	Date of documented relapse from CR	Event
No post-randomization CR assessments	Day 1	Event
No documented death or relapse	Date of last adequate assessment*	Censored
Not reached the end of induction	Date of last adequate assessment*	Censored
Lost to follow-up, withdraw consent before any documented death or relapse	Date of last adequate assessment*	Censored

* Adequate disease assessment is defined as there is sufficient data to evaluate a patient's disease status.

Sensitivity analyses for EFS will include:

1. EFS will be analyzed in the PP analysis set if more than 5% of patients will be excluded in this analysis.
2. Considering that EFS may be influenced by subsequent therapies administered after failure to achieve or maintain remission, an additional sensitivity analysis that treats an alternative therapy as an event where the start date for the alternative therapy is the event date, will be conducted to obtain a more precise assessment of efficacy for EFS.
3. If there exists time depended confounding factors caused by informative censoring from imbalance in the proportion of HSCT events between the 2 cohorts, Marginal Structural Model (MSM) [5] and Inverse Probability of Censoring Weighted (IPCW) [6] analysis of the EFS endpoint will be considered.
4. If the propotional hazard assumption is violated, non-propotional hazard Cox models will be applied to evaluate HR using piecewise expotential model.

In the MSM and IPCW analyses, in order to derive weights adjusting for the time-fixed and time-varying confounding effects due to taking HSCT, the covariates affecting EFS endpoint will be used. Potential time-fixed covariates and time-varying covariates include demographic data, such as age (18 through <45 years; 45 -<60 years; ≥60 years), gender (male, female), race (white; non-white), region (north America; south America; Europe; APAC); and baseline disease characteristics including BCR-ABL1 dominant transcript (P190; P210) and ECOG status (0; 1 or 2); baseline laboratory parameters such as white blood count, homoglobin, platelet, LDH, peripheral blood blast, bone marrow blast; and time-dependent covariates including SCT status, duration of exposure, relapse status at each study visit, initiation of alternative therapy, and other perimeters as appropriate. The final criteria for selected covariates would need to be statistically have a p-value of less than or equal to 0.15 in the multivariate logistic regression models for weight calculations. If there are more than 5% missing in the baseline covariate, then this covariate will be dropped from the weighting calculation and final model. For both MSM and

IPCW analyses, logistic regression models on repeated measurements will be used to approximate the Cox models in the weight derivations from which stabilized weights will be derived per subject per observation. Adjusted K-M curves will also be presented along with hazard ratios (HRs), 95% confidence intervals for HRs, and adjusted p-values based on MSM and IPCW approaches. SAS proc PHREG procedure with counting process type of data input, which takes multiple observations per subject, will be used as the final Cox model for both MSM and IPCW approaches, where robust variance will be used to accommodate covariance introduced by correlated longitudinal observations within each subjects and other extra variabilities due to departure from model assumptions.

Subgroup analyses will be performed for EFS similar to the primary endpoint.

7.8.2.2 Other Secondary Efficacy Endpoints

The following other secondary endpoints will be analyzed (see Table 5.a for endpoint definitions):

- CR and CRi rates at the end of Cycle 1, the end of Cycle 2, the end of induction, and during or at the end of consolidation (end of Cycle 9 or end of study treatment, whichever occurs earlier).
- Molecular response rates (MR3, MRD negativity [MR4], and MR4.5) at the end of Cycle 1, the end of Cycle 2, the end of induction, and during or at the end of consolidation (end of Cycle 9 or end of study treatment, whichever occurs earlier).
- Rates of PIF and ORR at the end of induction.
- Rates of MRD-negative CR at multiple intervals after the end of induction.
- Duration of MRD-negative CR.Duration of CR.
- Time to treatment failure.
- Duration of MR4.5 in patients who achieved MR4.5.
- OS and rate of relapse from CR for on-study patients with and without HSCT.
- OS.

If the efficacy boundary is crossed at either IA or FA for EFS, the following secondary endpoints will be tested in the order listed below using the same boundaries (0.033 for IA and 0.034 for FA):

a) Duration of CR.
b) ORR.
c) Duration of MRD-negative CR.
d) OS.

The remaining secondary efficacy endpoints will be tested at $\alpha = 0.05$ level in a nonhierarchical fashion without adjustments for multiplicity.

For analysis of time-to-event endpoints (eg, time to treatment failure, OS), 2-sided, stratified log-rank tests will be used to compare the treatment groups with respect to the endpoints. In addition, an unadjusted stratified Cox model will be used to estimate the hazard ratio (HR) and its 95% CIs for the treatment effect using the stratification factors. K-M survival curves and K-M medians (if appropriate and estimable), along with their 2-sided 95% CIs, will also be provided for each treatment group.

For the secondary endpoint of OS, a futility analysis with be conducted at the time of the IA for EFS. The hazard ratio and corresponding 95% CI for the OS analysis will be calculated and reviewed by the IDMC. If the HR is >1.2, the IDMC will review the totality of the data and provide a recommendation to the sponsor's executive committee regarding study continuation.

OS results are expected to be confounded by alternative therapies after patients discontinue from the study assigned drug. Thus, sensitivity analyses, such as Marginal Structural Models (MSM) and Inverse Probability Censoring Weighting (IPCW), will be conducted for OS analysis adjusting for time depending on confounding factors occurring due to taking alternative therapies. With IPCW and MSM analyses, to reduce bias, the following settings will be similar with the EFS analysis including: 1) the list of potential confounders, both baseline and time-dependent, which may impact both OS and censoring outcome, and thus will be included in the initial weighting models; 2) the p-value cut off for confounders remain in the final weighting models will be at 0.15 level; and 3) SAS procedure.

Duration of MRD-negative CR is defined as the time from the date of first documentation of MRD-negativity or CR (whichever comes latest), to the date of first documentation of loss of MRD-negativity (BCR-ABL/ABL1 >0.01%) or relapse from CR, for patients who achieve MRD-negative CR. The primary analysis for duration of MRD-negative CR will be based on time-to-event analysis.

The primary analysis for duration MRD-negative CR will not consider censoring at the time of HSCT or initiation of alternative therapies.

Duration of CR is defined as the time from the date of first documentation of a CR to the date of first documentation of PD for patients who achieved CR. These patients without documentation of PD will be censored at the date of their last response assessment. The primary analysis for duration of CR will be based on time-to-event analysis.

Duration of MR4.5 is defined as the time from the date of first documentation of a MR4.5 to the date of first documentation of loss of MR4.5 for patients who achieved MR4.5. The analysis for duration of MR4.5 will be based on time-to-event analysis. Duration of MRD-negativity will be defined and explored in a similar manner, as appropriate.

The proportion-based other secondary endpoints (eg, CR and CRi rates, proportion of patients received HSCT) will be analyzed in the same fashion as the primary endpoint. The analyses will be conducted using a Cochran-Mantel-Haenszel (CMH) chi-square test. The CMH chi-square p-

value on risk difference between treatment arms may be calculated. The risk difference and relative risk will be presented along with 95% 2-sided confidence intervals.

7.8.3 Additional Efficacy Endpoint(s)

The exploratory endpoints are:

- Time to HSCT.
- Time to start of alternative chemotherapy.
- Change from baseline in patient-reported HRQOL (FACT-Leu and EQ-5D-5L).
- MRU assessments.
- Biomarkers of disease sensitivity and resistance to ponatinib and imatinib.

Further details on the exploratory endpoint analyses will be discussed in the following.

7.8.3.1 Time-to-Next-Treatment and Time-to-HSCT Analyses

Time to subsequent antineoplastic therapy will be defined as the time from randomization to the date of first documentation of subsequent antineoplastic therapy or the last contact date for subjects who never received subsequent antineoplastic therapy.

Likewise, time to HSCT will be defined as the time from randomization to the date of first documentation of HSCT or the last contact date for subjects who did not receive an HSCT.

A Cox regression model with treatment as explanatory variable will be used for the time-to-event analyses. Median will be calculated by K-M method.

7.8.3.2 Patient-Reported Outcomes Analysis

Quality of life and health outcomes measures are being collected using the EQ-5D-5L and FACT-Leu instruments. Means and medians of scores of these questionnaires will be summarized for each cohort by time point, overall, and for each domain. Assessments based on the FACT-Leu will be analyzed to determine if treatments affect all domains.

Analyses of HRQOL scores, including global health status, will be performed using longitudinal models for scores and change from baseline scores. All subscales and individual item scores will be tabulated. Descriptive summaries of observed data will be provided at each scheduled assessment time point.

The manuals published for FACT-Leu will be used for scoring and handling missing data.

EQ-5D-5L scores will be summarized in descriptive statistics for treatment groups. Both utility scores and change from baseline scores will be assessed across time using longitudinal models.

Compliance for EQ-5D-5L and FACT-Leu will also be summarized by number of expected and number and percentage of received by treatment group over time.

PROs by proportion of patients that achieved or did not achieve MRD-negative CR at end of induction will be summarized by treatment group as appropriate.

Patient-reported outcome analysis will use safety population.

7.8.3.3 Health Economics Analysis Using Medical Resource Utilization

Medical resource utilization data will be summarized in descriptive statistics for safety population hospitalization (length of stay, inpatient, outpatient, and reason), number of missing days from work or other activities, by patient and caregiver, and by treatment group.

7.8.3.4 Biomarkers of Disease Sensitivity and Resistance to Ponatinib and Imatinib

The mutation status of BCR-ABL1 and other genes implicated in tumor biology and/or drug metabolism will be determined, as clinically needed, through analyses of tumor cells collected at study entry, on study, and/or at EOT. Analysis methodologies include, but are not limited to, DNA sequencing, digital PCR, and mass spectrometry.

7.9 Pharmacokinetic/Pharmacodynamic Analysis

7.9.1 Pharmacokinetic Analysis

The PK data collected in this study are intended to contribute to future population PK analyses of ponatinib. These population PK analyses may additionally include data collected in other ponatinib clinical studies. The analysis plan for the population PK analysis will be defined separately and the results of these analyses will be reported separately.

Ponatinib plasma concentration-time data will be listed and summarized by time point.

7.9.2 Pharmacodynamic Analysis

Not Applicable.

7.10 Other Analysis

In general, missing or partial dates due to unexpected situations, such as COVID-19 or Ukraine Crisis, will follow the convention without special handling. COVID-19 or Ukraine Crisis impact on the study visit and dosing/laboratory schedule will be tabulated and listed. In addition, if missing data distribution is not balanced between the two arms, sensitivity analysis by excluding patients with missing data due to Covid/Crisis will be carried out.

In addition, the patients who received ponatinib alone post 20 cycles may be analyzed for efficacy and safety endpoints.

7.11 Safety Analysis

The safety analysis will be carried out at interim and final analyses. In addition, an extended safety analysis for the study will be carried out at study completion.

Safety analyses will be based on the safety analysis set. Descriptive statistics (ie, n, mean, SD, median, minimum and maximum for continuous variables, and frequency and percentage of patients for categorical variables) will be used to summarize the safety parameters.

Safety evaluations will be based on incidence, severity, and type of AEs; clinically significant changes or abnormalities in the patient's physical or neurological examinations; vital signs; ECG; ECOG performance status; clinical laboratory test results and other safety parameters.

7.11.1 Adverse Event

7.11.1.1 Adverse Events

AEs will be coded using MedDRA. All AEs will be presented in a by-patient listing. Treatment-emergent adverse events (TEAEs) are defined as any AEs that occur on or after administration of the first dose of any study drug and through 30 days after the last dose of any study drug.

AEs will be tabulated according to MedDRA by SOC and PT and will include the following categories:

- TEAEs.
- Drug-related TEAEs.
- Grade 3 or higher TEAEs.
- Grade 3 or higher drug-related TEAEs.
- The most commonly reported TEAEs (ie, those events reported by ≥10% of all patients).
- SAEs (related and regardless of relationship)
- TEAEs leading to study drug modification and discontinuation.
- TEAEs leading to hospitalization or prolonging hospitalization.
- TEAEs leading to death.
- Adverse events of special interest (AESIs) including Arterial Occlusive Events (AOEs) and Venous Thromboembolic Events (VTEs); and Other AEs of Significance, as appropriate.

Patients with the same AE more than once will have that event counted only once within each SOC and once within each PT.

TEAEs will also be summarized by the National Cancer Institute Common Terminology Criteria for Adverse Events (NCI CTCAE) version 5.0. Patients with the same AE more than once will have the maximum intensity of that event counted within each SOC and once within each PT.

The most commonly reported TEAEs (ie, those events reported by ≥10% of any treatment cohort) will be tabulated by PT. Patients with the same AE more than once will have that event counted only once within each PT.

An overall summary treatment-emergent AE table will include numbers and percentages of patients who had any treatment-emergent AE, drug-related treatment-emergent AE, grade 3 or higher treatment-emergent AE, grade 3 or higher drug-related treatment-emergent AE, serious AE (SAE), drug-related SAE, treatment-emergent AE resulting in discontinuation, and on-study deaths. On-study death is defined as the death that occurs between the first dose of any study drug and within 30 days of the last dose of any study drug.

In addition, TEAEs will be summarized by each treatment cohort. Secondary malignancy will be tabulated and listed as appropriate.

By-patient listing of grade 3 or higher treatment-emergent AE will also be provided, where the cycle day information for the AE onset and end dates will be included in the listing.

Analysis of AOEs and VTEs

Arterial occlusive and venous thromboembolic events with an initial onset date on or after the first dose date will be considered treatment-emergent and summarized. Number and percentages of patients who developed AOEs and VTEs will be summarized for each cohort. These events will be categorized as follows:

- Arterial occlusive events
  - Cardiovascular occlusive events.
  - Cerebrovascular occlusive events.
  - Peripheral vascular occlusive events.
- Venous thrombotic events.

Exposure-adjusted incidence rates (EAIR) of adjudicated AOEs and VTEs will be calculated for each cohort and for all patients. The 95% CI of the EAIR will be computed.

The following additional descriptive analyses will be performed to characterize AOEs and VTEs described above:

- Time to onset: Calculated as date of first event AE- first dose date + 1.
- Dose at onset: Dose of ponatinib/imatinib taken immediately prior to onset of first event.

Analysis of Other AEs of Significance

Categories of AEs will be prospectively defined using Standardized Medical Dictionary for Regulatory Activities (MedDRA) Queries (SMQs) or Modified MedDRA Queries based on SMQs and MedDRA System Organ Classes (SOCs). The AE crude rates, as well as the frequency of occurrence by overall toxicity—categorized by toxicity grades (severity)—will be described for each cohort. Events will also be characterized by time to onset, dose at onset, and duration, as described above. Categories of AEs will include but will not be limited to:

- Cardiac failure
- Arrhythmias including QT prolongation

- Pancreatitis and Amylase or Lipase elevations
- Hepatotoxicity
- Myelosuppression
- Hemorrhage
- Fluid retention
- Hypertension

7.11.1.2 Serious Adverse Events

The number and percentage of patients experiencing at least 1 treatment-emergent SAE will be summarized by MedDRA primary system organ class, and preferred term. Drug-related SAEs will be summarized similarly.

In addition, a by-patient listing of the SAEs will be presented (the patient listing will contain all SAEs regardless of treatment-emergent AE status).

7.11.1.3 Deaths

All deaths will be summarized by treatment arm, including deaths occurring on-study and death during follow-up separately.

A by-subject listing of the deaths will be presented. All deaths occurring on-study and during follow-up will be displayed (regardless of treatment emergent AE status).

7.11.1.4 Adverse Events Resulting in Discontinuation of Study Drug

A by-patient listing of treatment-emergent AEs resulting in discontinuation of study drug regimen will be presented.

7.11.2 Clinical Laboratory Evaluations

For the purposes of summarization in both the tables and listings, all laboratory values will be converted to standardized units. If a lab value is reported using a non-numeric qualifier (eg, less than (<) a certain value, or greater than (>) a certain value), the given numeric value will be used in the summary statistics, ignoring the non-numeric qualifier. If a patient has repeated laboratory values for a given time point, the value from the last evaluation will be used.

Laboratory test results will be summarized according to the scheduled sample collection time point. Change from baseline will also be presented. Unscheduled laboratory test results will be listed and included in laboratory shift tables. The parameters to be analyzed are as follows:

- Hematology: hemoglobin, absolute neutrophil count (ANC), platelets counts, WBC, lymphocytes and leukocytes
- Serum chemistry: blood urea nitrogen (BUN), creatinine, total bilirubin, urate, lactate dehydrogenase (LDH), albumin, alkaline phosphatase (ALP), aspartate aminotransferase

(AST), alanine aminotransferase (ALT), glucose, corrected calcium, sodium, potassium, chloride, lipase and amylase

Shift tables will be constructed for laboratory parameters to tabulate changes from study entry to post study entry either by CTCAE toxicity grade or abnormality. Parameters to be tabulated will include, but not limited to:

- Hematology: ANC, hemoglobin, and platelets.
- Serum chemistry: ALT, AST, ALP, creatinine, total bilirubin, amylase and lipase.

Mean laboratory values and box plots over time for key lab parameters will be produced, including but not limited to ANC, platelets, and liver function tests (ALT/SGPT, AST/SGPT, alkaline phosphatase, and total bilirubin), lipase and amylase.

By-patient listings to be presented include hematology, serum chemistry, urinalysis, urine total protein, and urine creatinine.

7.11.3 Vital Signs

The actual values of vital sign parameters (ie, systolic and diastolic BP, heart rate, respiratory rate, temperature, height, and weight) will be summarized at all planned timepoints by each treatment cohort using descriptive statistics (ie, n, mean, SD, median, minimum and maximum). Change of vital signs from baseline values will also be summarized at all planned timepoints. Vital sign values will also be presented in a by-patient listing.

7.11.4 12-Lead ECG

The absolute values and absolute change from baseline of electrocardiogram (ECG) parameters including ECG ventricular rate, PR interval, RR interval, QRS duration, QT interval, and QT interval corrected per Fridericia method (QTcF) interval will be summarized at each timepoint using descriptive statistics (ie, n, mean, SD, median, minimum and maximum).

QTc interval will be calculated using Fridericia correction, if necessary. The formulas are:

$$QTcF\ (Fridericia) = QT / (RR^{0.33})$$

where RR = 60 / heart rate (bpm)

In addition, a categorical analysis of QTcF intervals will be performed for each time point. The number and percentage of patients in each QTcF interval (<450 msec, 450-480 msec, >480-<500 msec, and ≥500msec) will be summarized at each time point. Categories of changes from baseline (≥30 msec and ≥60 msec) will be summarized as well. Maximum QTcF intervals and maximum changes from baseline will also be summarized similarly in a separate display.

ECG abnormalities will be presented in a data listing.

7.11.5 ECOG Performance Status

ECOG performance status and shifts from baseline to post study entry assessment over time, and ECOG score frequency table over time will be summarized. Shifts from baseline to the worst post study entry score will be tabulated by each treatment cohort.

7.11.6 Other Observations Related to Safety

The ankle-brachial index (ABI), echocardiogram (ECHO) for assessment of left ventricular ejection fraction (LVEF) and Multiple-Gated Acquisition (MUGA) scan will be presented in a data listing.

7.12 Interim Analysis

MRD-negative CR

There would be one IA and possibly an FA in the study for the MRD-negative CR primary endpoint using a group sequential testing approach.

The IA has been performed after the end of induction phase data had been collected for 116 patients. The primary endpoint of MRD-negative CR was first tested at IA with a 2-sided efficacy boundary of 0.022and will be tested at FA with a 2-sided efficacy boundary of 0.036 after the end of induction phase data have been collected for 230 patients.

If the significance boundary is crossed at the FA for MRD-negative CR, then there will be a testing for EFS and other secondary endpoints at a 2-sided alpha level of 0.05 using group sequential testing approach.

The boundaries for hypotheses testing in MRD-negative CR will be updated according to the observed data in the FA, using the prespecified alpha spending function.

EFS

There will be one IA and possibly an FA in the study for the key secondary endpoint EFS using a group sequential testing approach.

When approximately 130 EFS events are observed (75% of the total 173 expected EFS events), an IA will be performed. The IA is expected to occur approximately 5.5 years after the first patient is enrolled. The FA is expected to be performed approximately 8.5 years after the first patient enrolled, when all approximately 173 EFS events have been observed.

The test significance for the IA and FA of EFS will be determined using Gamma Family (-1) boundaries. Based on the projected number of EFS events, the formal hypothesis testing will be stopped for overwhelming efficacy if the 2-sided p-value crosses the efficacy boundary (p=0.033) at IA and this will be the FA for EFS for statistical testing purpose. If EFS does not achieve statistical significance at the IA, the final analysis will be tested at 2-sided alpha level of efficacy boundary 0.034 (corresponding to nominal alpha of 0.017).

If the efficacy boundary is crossed at either IA or FA for EFS, the following secondary endpoints will be tested in the order listed below using the same boundaries (0.033 for IA and 0.034 for FA):

a) Duration of CR.

b) ORR.

c) Duration of MRD-negative CR.

d) OS.

The boundaries for hypotheses testing in EFS will be updated according to the observed data in the IA and FA, using the prespecified alpha spending function.

Overall Survival

For the secondary endpoint of OS, a futility analysis will be conducted at the time of the IA for EFS. The hazard ratio and corresponding 95% CI for the OS analysis will be calculated and reviewed by the IDMC. If the HR is >1.2, the IDMC will review the totality of the data and provide a recommendation to the sponsor's executive committee regarding study continuation.

The analyses for the IA and FA for MRD-negative CR and the IA for EFS will be carried out by an independent statistical team in a manner that maintains the blinding of the study results to the team (see Section 13.4). The IDMC will review both efficacy and safety data at the time of the IA, and will inform the sponsor's executive committee of their recommendation.

7.13 Changes in the Statistical Analysis Plan

Reference materials for this statistical plan include Clinical Study Protocol Ponatinib-3001 amendment 10 (Protocol amendment dated 20 October 2021). Additional major changes include:

- Furthe clarify the following order of secondary endpoints that will be tested at the time of IA for EFS: a) Duration of CR; b) ORR; c) Duration of MRD-negative CR; d) OS.

8.0 REFERENCES

1. Lan K, DeMets DL. Discrete sequential boundaries for clinical trials. Biometrika 1983;70:659-63.
2. Short NJ, Jabbour E, Sasaki K, Patel K, O'Brien SM, Cortes JE, et al. Impact of complete molecular response on survival in patients with Philadelphia chromosome-positive acute lymphoblastic leukemia. Blood 2016;128(4):504-7.
3. Sasaki K, Jabbour EJ, Ravandi F, Short NJ, Thomas DA, Garcia-Manero G, et al. Hyper-CVAD plus ponatinib versus hyper-CVAD plus dasatinib as frontline therapy for patients with Philadelphia chromosome-positive acute lymphoblastic leukemia: A propensity score analysis. Cancer 2016;122(23):3650-6.
4. Hwang I, ShihW, DeCani J Group sequential designs using a family of type I error probability spending functions. Statistics in Medicine 1990; 9(1):1439–1445.

5. Hernán MÁ, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. Epidemiology 2000; 561–570.
6. Robins JM, Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. Biometrics 2000;56(3):779–788.
7. Jennison C, Turnbull B. Group Sequential Methods with Application to Clinical Trials: Chapman & Hall; 2000.
8. Hung HM, Wang SJ, O'Neill R. Statistical considerations for testing multiple endpoints in group sequential or adaptive clinical trials. Journal of Biopharmaceutical Statistics 2007;17(6):1201-10.