AcertaPharma

ACE-CL-309: Statistical Analysis Plan

A Randomized, Multicenter, Open-Label, Phase 3 Study of Acalabrutinib (ACP-196) versus Investigator's Choice of Either Idelalisib Plus Rituximab or Bendamustine Plus Rituximab in Subjects with Relapsed or Refractory Chronic Lymphocytic Leukemia

Version 1.0 – 23 March 2018
Version 2.0 – 23 January 2019
Version 3.0 – 6 March 2019

Prepared by: PPD 6 MAR 2019
PPD Date
PPD Biostatistics

Approved by: PPD 6 MAR 2019
PPD Date
PPD Clinical Development

Approved by: PPD 6 MAR 2019
PPD Date
PPD Biometrics

CONFIDENTIAL

This document (and all copies thereof) is the property of Acerta Pharma and contains valuable trade secrets and confidential information, including proprietary technical and business information. By accepting this document, you agree that the Confidential Information contained herein will not be disclosed to others without prior written authorization from Acerta Pharma and shall only be used for the purposes intended by the document.

SAP Revision History:

Version/Date	Summary of Major Changes and Rationale
Version 1.0, 23 March 2018	
Version 2.0, 23 January 2019	1. Added a table for planned PFS analyses, the efficacy stopping boundary, and the estimated timing of these analyses – Section 2.4 2. Clarified the hierarchical testing procedure for the key secondary endpoints – Section 2.6 3. Removed crossover population – Section 3.3 4. Clarified the definition of treatment-emergent period and treatment emergent adverse events – Section 4.1.6, Section 8.1.1 5. Added the rule for collapsing stratification factors – Section 7 6. Clarified the censoring rule for primary analysis for IRC-assessed PFS – Section 7.1.1 7. Updated the planned sensitivity and subgroup analyses for primary efficacy endpoint – Section 7.1.3, Section 7.1.4 8. Updated the censoring rule for the primary and sensitivity analysis for overall survival – Section 7.2
Version 3.0, 6 March 2019	1. Updated the censoring rule for primary analysis and sensitivity analysis for IRC-assessed PFS – Section 7.1.1, Section 7.1.3 2. Added the operational definition of the last adequate IRC assessment – Section 7.1.1

CONFIDENTIAL

This document (and all copies thereof) is the property of Acerta Pharma and contains valuable trade secrets and confidential information, including proprietary technical and business information. By accepting this document, you agree that the Confidential Information contained herein will not be disclosed to others without prior written authorization from Acerta Pharma and shall only be used for the purposes intended by the document.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS AND DEFINITIONS

11q del	chromosome deletion 11q22.3
17p del	chromosome deletion 17p13.1
AE(s)	adverse event(s)
ALC	absolute lymphocyte count
BID	twice per day
CLL	chronic lymphocytic leukemia
CR	complete remission (response)
CRi	CR with incomplete bone marrow recovery
CT	computed tomography
CTCAE	Common Terminology Criteria for Adverse Events
DMC	Data Monitoring Committee
DOL	duration of lymphocytosis
DOR	duration of response
ECOG	Eastern Cooperative Oncology Group
eCRF	electronic case report form
EDC	electronic data capturing
CCI	[redacted]
FACIT-Fatigue	Functional Assessment of Chronic Illness Therapy-Fatigue
HR	hazard ratio
IGHV	immunoglobulin heavy-chain variable
INV	Investigator
IRC	Independent Review Committee
IPD	important protocol deviation
ITT	intent-to-treat
IV	intravenous

IWCLL	International Workshop on Chronic Lymphocytic Leukemia
IXRS	interactive voice/web response system
MedDRA	Medical Dictionary for Regulatory Activities
MRD	minimal residual disease
CCI	[redacted]
NCI	National Cancer Institute
nPR	nodular partial remission (response)
ORR	overall response rate
OS	overall survival
PD	progressive disease
PFS	progression-free survival
PR	partial remission (response)
PRL	partial response except for lymphocytosis
PRO	patient-reported outcome
RBC	red blood cell
R/R	relapsed/refractory
SAE(s)	serious adverse event(s)
SAP	statistical analysis plan
SD	stable disease
TEAE(s)	treatment-emergent adverse events
WHO	World Health Organization

1 INTRODUCTION

This statistical analysis plan (SAP) provides details of the statistical analyses that have been outlined for Study ACE-CL-309 Protocol Amendment 5, dated 17 November 2017, A Randomized, Multicenter, Open-Label, Phase 3 Study of Acalabrutinib (ACP-196) versus Investigator's Choice of Either Idelalisib Plus Rituximab or Bendamustine Plus Rituximab in Subjects with Relapsed or Refractory Chronic Lymphocytic Leukemia.

This study is being conducted under the sponsorship of Acerta Pharma. The sponsor is conducting statistical programming and analyses using dummy treatment code before database lock and study "unblinding" to the Sponsor (see Section 2.7.2). All safety and interim efficacy analyses for independent Data Monitoring Committee (DMC) review will be conducted in an unblinded fashion under contract by an independent contract organization. The scope of this plan includes the analyses that are planned and will be executed by Acerta Pharma's Biometrics department or designee after the study is "unblinded" to the Sponsor.

2 STUDY DESCRIPTION

2.1 Study Design

This randomized, global (approximately 150 sites), multicenter, open-label, Phase 3 study will evaluate the efficacy and safety of acalabrutinib monotherapy versus idelalisib/rituximab or bendamustine/rituximab in subjects with relapsed or refractory (R/R) chronic lymphocytic leukemia (CLL).

Approximately 306 eligible subjects will be randomized in a 1:1 ratio into 2 arms (n=153 each) to receive either:

- Arm A: Acalabrutinib 100 mg PO twice per day (BID) until an unacceptable drug-related toxicity occurs or until disease progression.
- Arm B: Before randomization, Investigator's choice of:
  - Idelalisib 150 mg PO BID administered until disease progression or unacceptable toxicity in combination with ≤8 IV infusions of rituximab.
  - Bendamustine 70 mg/m^2 IV (Day 1 and 2 of each cycle) in combination with rituximab IV (375 mg/m^2/500 mg/m^2) on Day 1 of each cycle for up to 6 cycles.

Eligible subjects will be randomized based on the following stratification factors:

- Presence of 17p del (yes versus no)
- Eastern Cooperative Oncology Group (ECOG) performance status (0 or 1 versus 2)
- Number of prior therapies (1, 2, or 3 versus ≥4)

Subjects randomized to Arm B who have sponsor-confirmed disease progression will be eligible to receive crossover treatment with single-agent acalabrutinib at 100 mg BID until disease progression or unacceptable toxicity. For analysis purpose, for subjects randomized to Arm B and crossed over to receive single agent of acalabrutinib, the period from randomization to the day prior to the onset of crossover (i.e., the day prior to the first dose of acalabrutinib), is defined as the main study period; the period from the onset of crossover (i.e., the first dose date of acalabrutinib) to end of study is defined as the crossover period. For all other subjects (i.e., subjects randomized to Arm A and subjects randomized to Arm B who did not cross over), there is only the main study period.

One interim efficacy analysis is planned for the study.

An Independent Review Committee (IRC) will conduct the response evaluations per International Workshop on Chronic Lymphocytic Leukemia (IWCLL) 2008 criteria (Hallek 2008) with incorporation of the clarification for treatment-related lymphocytosis (Cheson 2012)—hereafter referred to as IWCLL 2008 criteria, in accordance with its charter.

An independent DMC will be formed and constituted according to regulatory agency guidelines. Detailed information regarding the composition of the DMC and detailed DMC procedures will be provided in the DMC charter. The DMC will review the safety data periodically and the planned interim efficacy analysis results, both in an unblinded fashion. The DMC will follow the charter to provide recommendations to the sponsor after each DMC review.

2.2 Study Objectives

2.2.1 Primary Objectives

To evaluate the efficacy of acalabrutinib monotherapy (Arm A) compared with idelalisib/rituximab or bendamustine/rituximab (Arm B) based on IRC assessment of progression-free survival (PFS) per IWCLL 2008 criteria in subjects with R/R CLL

2.2.2 Secondary Objectives

To evaluate Arm A compared with Arm B in terms of:

- Investigator (INV)-assessed PFS per IWCLL 2008 criteria
- INV- and IRC-assessed overall response rate (ORR) per IWCLL 2008 criteria (defined as the proportion of subjects who achieve a best response of CR, CRi, nPR, or PR)
- Overall survival (OS)
- PROs by Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT Fatigue)
- INV- and IRC-assessed duration of response (DOR; defined as the time from the first documentation of objective response to the earlier time of disease progression or death from any cause)
- TTNT (defined as the time from randomization to institution of non-protocol specified treatment for CLL)

2.2.3 Safety Objectives

Incidence of AEs and serious adverse events (SAEs)

2.2.4 CCI [redacted]

CCI [redacted]

2.3 Power and Sample Size Justification

The study is expected to enroll approximately 306 subjects with a 1:1 randomization ratio between Arm A and B.

CCI [redacted]

2.4 Interim Analysis

The sponsor will monitor the number of PFS events closely. CCI [redacted] A snapshot of the database will occur after all data through the visit cutoff date have been entered and cleaned. The interim analysis will be conducted based on a snapshot of the database, which includes all available PFS events up to the visit cutoff date.

The interim analysis will be conducted using Lan and DeMets alpha-spending function with O'Brien-Fleming boundaries (O'Brien and Fleming 1979; Lan and DeMets 1983). The table

below presents a summary of the planned PFS analyses, the efficacy stopping boundary, and the estimated timing of these analyses.

CCI

In the event that the actual number of IRC-assessed PFS events is different from the estimated CCI at interim analysis, the efficacy stopping boundary for the interim and final analyses will be determined based on the aforementioned alpha-spending function and the actual number of IRC-assessed PFS events observed at the time of the analyses.

An independent Statistical Data Analysis Center will conduct the unblinded interim analysis for review by the DMC. The DMC will operate independent of the sponsor and the clinical investigators. The sponsor will remain blinded with respect to the interim analysis results unless the DMC recommends trial stopping and the Sponsor accepts the recommendation.

If the criterion for early efficacy is met at the time of the interim analysis, the DMC may recommend stopping the study in accordance with the terms of the DMC charter. If the DMC recommends stopping early for efficacy, the interim analysis of efficacy will be the final analysis.

If the primary efficacy endpoint achieves statistical significance, then selected secondary endpoints will be tested in a manner that will preserve the overall Type I error rate at the 2-sided significance level of 0.05 (see Section 2.6 for details).

The DMC members will use their expertise, experience, and judgment to evaluate the safety data from the trial on a regular basis and to recommend to the sponsor whether the trial should continue or be stopped early for safety. No formal statistical rules recommending early stopping for safety are planned.

2.5 Final Analysis

If the study does not cross the boundary at interim analysis, the study will proceed to final analysis. CCI [redacted] Database lock will occur after all data through the visit cutoff date have been entered and cleaned. The final analysis will be conducted based on the final locked database.

2.6 Multiplicity Adjustments

To control the overall Type I error at 0.05 level, the Lan-DeMets alpha spending function based on the O'Brien-Fleming boundary is used to split α into α_1 and α_2 for interim and final analyses, respectively. The nominal α_1 and α_2 levels will be determined based on the actual information fraction at the time of the interim analysis. If the primary endpoint achieves statistical significance, tests of key secondary endpoints of IRC-assessed ORR and OS will be performed in a sequential hierarchical manner based on a closed testing procedure specified below:

1. IRC-assessed ORR
2. OS

If the primary endpoint of IRC-assessed PFS achieves statistical significance at interim analysis, the IRC-assessed ORR will be tested at an α level of 0.05, given that almost all responses will have been observed at that time (thus the interim and final analyses of IRC-assessed ORR should be the same).

If the IRC-assessed ORR achieves statistical significance, the OS will be tested at the same α level spent for the primary endpoint of IRC-assessed PFS at interim and final analyses, respectively.

2.7 Randomization and Blinding

2.7.1 Randomization

Approximately 306 eligible subjects will be randomized in a 1:1 ratio into the two treatment arms. Central randomization will be implemented using Interactive voice/Web Response System (IXRS). Subjects will be randomized based on the following stratification factors:

- Presence of 17p deletion (yes versus no)
- ECOG Performance Status (0 or 1 versus 2)
- Number of prior therapies (1, 2, or 3 versus ≥4)

The randomization code will be controlled through a centralized procedure and will not be known to study and site personnel or the subject before the subject is randomized into the study.

2.7.2 Blinding

In this open-label study, neither the subjects nor the investigators will be blinded to treatment. However, access to treatment randomization by IXRS will be controlled so that the sponsor's staff overseeing the conduct of the study or analyzing/summarizing data will not have aggregated summary by randomized treatment arm before “unblinding.” Dummy treatment code will be used to set up statistical programming for final analyses.

Response assessment will be performed centrally by the IRC. An Independent Review Charter for IRC will be created to describe details of data review, data flow, and work flow.

3 ANALYSIS POPULATIONS

3.1 Intent-to-Treat Population

The intent-to-treat (ITT) population is defined as all randomized subjects, to be analyzed according to the arm to which they were randomly assigned, following “intent-to-treat” principle. All efficacy analyses, except OS, will be performed for the ITT population during the main study period. OS will be analyzed based on the ITT population during the whole study period—that is, main study period + crossover period. In addition, the ITT population will be used to summarize disposition, demographics, and baseline disease characteristics.

3.2 Safety Population

The safety population consists of all subjects who received any amount of study drug. Safety analyses will be performed in safety population and will be analyzed as treated in main study period and crossover period separately.

4 GENERAL CONVENTIONS

Continuous data will be summarized using descriptive statistics (number of observations, mean, standard deviation, median, 25th and 75th percentiles, minimum, and maximum). Frequencies and percentages will be used for summarizing categorical (discrete) data.

CI, when presented, will generally be constructed at the 95% level. For binomial variables, the normal approximation methods will be employed unless otherwise specified.

Calculation of time to event or duration of event endpoints will be based on the study day of the event or censoring date rather than visit number or visit label. Missing efficacy or safety data will not be imputed unless otherwise specified.

The following rules will be used for the days to months/years conversion:

- 1 cycle = 28 days = 4 weeks;
- 1 month = 30.4375 days;
- 1 year = 365.25 days.

All summaries will be presented by treatment arm unless otherwise specified. Selected subject level data will be presented in data listings by subject number.

4.1 Definition

4.1.1 Baseline and Post-Baseline Value

For safety parameters (including all laboratory parameters), the baseline value is defined as the last measurement prior to the first administration of study drug. A post-baseline value is defined as a measurement taken after the first administration of study drug.

For efficacy parameters, the baseline value, except for laboratory parameters, is defined as the last measurement taken prior to or on the date of randomization; the post-baseline value is defined as a measurement taken after the date of randomization. For laboratory parameters used for response assessment and PROs, baseline values will be defined in reference to first dose date of study drug.

For assessments on the day of first dose where time is not captured, a nominal pre-dose indicator, if available, will serve as sufficient evidence that the assessment occurred prior to first dose. Assessments on the day of first dose where neither time nor a nominal pre-dose indicator

are captured will be considered prior to first dose if such procedures are required by the protocol to be conducted before first dose.

For subjects who were not treated, the baseline is defined as the last valid assessment within the study.

4.1.2 Definition of Study Day

For efficacy data, the study day will be calculated in reference to the date of randomization. Study Day 1 for efficacy data analysis is defined as the date of randomization. For assessments that occur on or after randomization, study day is defined as (date of assessment – date of randomization + 1). For assessments that occur prior to randomization, study day is defined as (date of assessment – date of randomization).

For safety data, study day will be calculated in reference to the first dose date of study drug. Study Day 1 for safety data analysis is defined as the first dose date of study drug. For assessments that occur on or after first dose date, study day is defined as (date of assessment – first dose date + 1). For assessments that occur prior to first dose date, study day is defined as (date of assessment – first dose date).

There is no Study Day 0 for efficacy or safety data analyses.

4.1.3 Definition of Prior and Concomitant Therapy

For the purpose of inclusion in prior and/or concomitant medication or therapy tables, incomplete medication or radiation start and stop dates will be imputed as detailed in Section 4.2. Based on actual or imputed start and stop dates:

- Prior medications/radiation therapies are defined as therapies with a start date occurring before the date of first dose of study treatment.
- Concomitant medications/procedures/radiation therapies are defined as therapies that were taken between the date of the first dose of study drug and 30 days after the last dose of study drug or the first date starting new anticancer therapy, whichever is earlier.
- If medications/procedures/radiation therapies started before first dose of study drug and continued after the subject started receiving study treatment, they will be classified as both prior and concomitant.

4.1.4 Duration of Treatment

Duration of treatment will be calculated from the date of the first dose of study drug to the date of the last dose of study drug, as follows:

> Duration of treatment = date of last dose of study drug - date of first dose of study drug + 1 day.

4.1.5 Time on Study

For subjects in Arm B who crossed over, time on study will be separated into two parts, time on main study period and time on crossover period, as follows:

> Time on study = time on main study period + time on crossover period.

Time on main study period will be calculated from the date of randomization to the first dose of acalabrutinib, as follows:

> Time on main study period = the first dose date of acalabrutinib – date of randomization.

Time on crossover period will be calculated from the first dose of acalabrutinib to the study exit date, as follows:

> Time on crossover period = study exit date – the first dose date of acalabrutinib + 1 day

For other subjects, there is only main study period, time on study will be calculated from the date of randomization to the study exit date, as follows:

> Time on study = time on main study period = study exit date - date of randomization + 1 day.

4.1.6 Treatment-Emergent Period

The treatment-emergent period is defined as the period of time from the date of the first dose of study drug through 30 days after the date of the last dose of study drug or the first date starting new anticancer therapy, whichever is earlier.

4.2 Imputation Rules for Missing and Partial Data

In general, other than partial dates, missing data will not be imputed and will be treated as missing. The algorithms for imputation of partial dates vary depending upon the parameter.

Imputation of partial dates will be made for adverse event (AE) onset and stop dates, start and end dates of concomitant medication, start date of subsequent anticancer therapy, date of initial

diagnosis, and death date. If dates are completely missing, no imputation will be made. For any partial date with missing year, no imputation will be made.

The general rule for imputation is:

- If only day is missing, then the 15th of the month will be used.
- If only year is present, then June 30th will be used.

If such imputation date for initial diagnosis is on or after date of first dose, then date of first dose – 1 will be used. If such imputed date for subsequent anticancer therapies is before date of last dose, then date of last dose + 1 will be used.

If the imputed date is for an AE start date and is in the same year and month as before the first dose date, then the first dose date will be used; if the imputed AE start date is after the AE end date, then the AE end date will be used. If the imputed date is for an AE start date and is in the same year and month as but after the last dose date + 30 days, then the last dose date + 30 days will be used.

If the imputed date is for an AE end date and is after the death date, then the death date will be used; if the imputed AE end date is before the AE start date, then the AE start date will be used.

4.2.1 Death Dates

If death year and month are available but day is missing:

- If mmyyyy for last contact date = mmyyyy for death date, set death date to the day after the last contact date.
- If mmyyyy for last contact date < mmyyyy for death date, set death date to the first day of the death month.
- If mmyyyy for last contact date > mmyyyy for death date, data error and do not impute.

If both month and day are missing for death date or a death date is totally missing, do not impute and censor the subject survival time.

4.2.2 Date Last Known Alive

- If year and month are available but day is missing, set date to the 1st of the month.
- If both month and day are missing, set date to January 1st of the year

4.2.3 Laboratory Values

- Laboratory values below the lower level of quantification (Q) that are reported as "<Q" or "≤Q" in the database will be imputed by Q x 0.99 for analysis purposes. However, the original value will be reported in the Listings.
- Laboratory values above the upper level of quantification (Q) that are reported as ">Q" or "≥Q" in the database will be imputed by Q x 1.01 for analysis purposes. However, the original value will be reported in the Listings.

4.3 Software

Sample size calculation was performed using East 6 (Version 6.3.1). Statistical analyses and data summary will be conducted using SAS Version 9.4 or higher.

5 STUDY POPULATION SUMMARIES

5.1 Disposition

Subject disposition will be summarized categorically and will include the number and percentage of subjects in the ITT and Safety populations.

Study treatment completion status, reasons for study treatment discontinuation, study follow-up completion status, reason for study follow-up discontinuation, time (month) on treatment, and time (month) on study will also be summarized categorically.

5.2 Protocol Deviations

Important protocol deviations (IPDs) will be identified and listed separately by study center and subject. IPD categories, subcategory codes, and descriptions will be defined by IPD guidance and used during the course of the study. The sponsor will review IPDs throughout the study prior to database lock. The final IPD list will be used to produce the IPD summary and listing.

5.3 Demographic and Baseline Characteristics

5.3.1 Demographics and Baseline Characteristics

The following demographic and baseline characteristics will be summarized:

- Sex (male, female)
- Age (continuous)
- Age group (<65, ≥65)
- Age group (<75, ≥75)
- Ethnicity
- Race
- Geographic region: North America, Australia and New Zealand, Western Europe, Central and Eastern Europe, and Asia
- Height (cm), weight (kg), and body surface area (BSA) (m^2)
- ECOG performance status at baseline

The age on the signed informed consent will be used for analysis.

5.3.2 Baseline Disease Characteristics

The following baseline disease characteristics will be summarized:

- Time (month) from initial diagnosis to randomization
- Bulky Disease (<5 cm, ≥5 cm)
- Rai stage (0, I, II, III, and IV)
- Binet Stage (A, B, and C)
- Chromosomal abnormality
- 17p del (yes, no)
- 11q del (yes, no)
- TP53 mutation (yes, no)
- 17p del and TP53 mutation (yes, no)
- 17p del or TP53 mutation (yes, no)
- IGHV (mutated, unmutated)
- Complex karyotype (yes, no)
- 17p del, TP53 mutation, or 11q del (yes vs. no)
- 17p del, TP53 mutation, 11q del, or unmutated IGHV (yes, no)
- Cytopenia at baseline
  - Neutropenia - ANC ≤ l.5 x 10^9/L (yes, no)
  - Anemia - Hgb ≤ 11g/dL (yes, no)
  - Thrombocytopenia: platelet counts ≤100 x 10^9/L (yes, no)
  - Any of the above (yes, no)
- β2 Microglobulin (mg/L) group at Baseline (≤3.5 mg/L, >3.5 mg/L)
- Absolute lymphocyte counts (ALC)
- ANC
- Platelet counts (PLAT)
- Hemoglobin concentration (HGB)
- Prior RBC transfusion within 28 days before randomization
- Prior platelet transfusion within 28 days before randomization

- Creatinine clearance <60 ml/min (yes, no)
- Constitutional Symptoms – weight loss, fever, night sweats, fatigue
- Number of lines of prior therapy

5.4 Medical History

General medical history data will be coded per Medical Dictionary for Regulatory Activities (MedDRA), summarized by system organ class and preferred term, and presented as a data listing.

6 TREATMENTS AND MEDICATIONS

6.1 Prior and Concomitant Medications

Medications recorded on the electronic case report forms (eCRFs) will be coded using the WHO Drug Dictionary. Prior and concomitant medications will be summarized by Anatomical Therapeutic Chemical Classification (ATC level 2 and level 5). Each subject will be counted only once for each therapeutic class, each generic name, and overall.

6.2 Prior and Subsequent Anticancer Therapies

Prior anticancer therapies will be coded and summarized based on the line of the therapy and type of therapy. The subsequent anticancer therapy taken between the end of study drug and the end of study will be summarized separately.

6.3 Study Treatment Exposure

Study treatment exposure during main study period will be summarized as follows:

- Duration of exposure:
  - Acalabrutinib and idelalisib (months): (last dose date – first dose date + 1) / 30.4375
  - Bendamustine and rituximab (days): (last dose date – first dose date + 28)
- Average daily dose (mg) (for acalabrutinib and idelalisib): calculated as (total dose received [mg] / duration of exposure [day])
- Relative dose intensity for acalabrutinib: calculated as (total cumulative dose received [mg] / (duration of exposure [days] * 100 [mg] * 2) *100)
- Relative dose intensity for idelalisib: calculated as (total cumulative dose received [mg] / {duration of exposure [days] * 150 [mg] * 2} *100).
- Relative dose intensity for bendamustine and rituximab: calculated as (total cumulative dose received [mg/kg] / total dose prescribed [mg/kg] from Cycles 1 to 6 * 100).

Study treatment exposure during crossover period will be summarized as follows:

- Duration of exposure: (last dose date of acalabrutinib – first dose date of acalabrutinib + 1) / 30.4375
- Average daily dose (mg): calculated as (total dose received [mg] / duration of exposure [day])

- Relative dose intensity: calculated as (total cumulative dose received [mg] / {duration of exposure [days] * 100 [mg] * 2} *100)

6.4 Study Treatment Modifications

For subjects randomized to Arm A, dose withholding is defined as missing dose for ≥7 consecutive days, dose reduction is defined as taking lower dose level (100 mg QD) for ≥3 consecutive days. Subjects with dose withholding and reduction will be summarized by percentage of subjects and by descriptive statistics.

For subjects randomized to Arm B, any dose modifications with rituximab, bendamustine, or idelalisib will be summarized by percentage of subjects and by descriptive statistics.

7 EFFICACY ANALYSES

All efficacy analyses, except OS, will be performed for the ITT population during the main study period. OS will be analyzed based on the ITT population during the whole study period—that is, main study period + crossover period. All efficacy analyses will be performed at the 2-sided significance level.

The following three randomization stratification factors (collected via IXRS) will be used for the stratified analyses: presence of 17p del (yes, no), ECOG status (2 vs. ≤1) and number of prior therapies (1–3 vs. ≥4). For the primary efficacy analysis of IRC-assessed PFS, if there is at least one stratum that has fewer than 2 events (where a stratum is defined as stratification factor 1 * stratification factor 2 * stratification factor 3), stratification factors should be collapsed until all strata have at minimum 2 events for the primary endpoint. The stratification factors will be collapsed in the following order:

1. ECOG status (2 vs. ≤1)
2. Presence of 17p del (yes vs. no)

If there is still at least one stratum that has fewer than 2 events after collapsing the 2 stratification factors above, an unstratified analysis (equivalently as to collapse all 3 stratification factors) will be performed as the primary analysis.

The final pooling strategy as described above for the primary efficacy analysis of IRC-assessed PFS will be applied for all stratified analyses.

7.1 Primary Efficacy Endpoint and Analysis

7.1.1 Primary Efficacy Endpoint

The primary efficacy endpoint is progression-free survival (PFS), defined as the time from date of randomization to the date of first IRC-assessed disease progression or death due to any cause, whichever comes first.

The primary efficacy endpoint of IRC-assessed PFS events will be analyzed as follows:

Table 2 IRC-assessed PFS

Situation	Date of Progression or Censoring	Outcome
PFS events include death or first IRC-assessed disease progression that occurred at or prior to the data analysis cutoff date.		
Death before first disease assessment	Date of death	Event
IRC-assessed PD or death between scheduled visit	Earliest date of IRC-assessed PD or death	Event
All other cases will be censored as follows:		
No baseline tumor assessments	Randomization	Censored
No adequate post-baseline assessment	Randomization	Censored
No IRC-assessed PD or death at the time of data cutoff (including subjects who had PD or died after data cutoff)	Date of last adequate IRC assessment before data cutoff	Censored
No IRC-assessed PD or death before withdrew consent or lost to follow-up	Date of last adequate IRC assessment before data cutoff	Censored
No IRC-assessed PD or death before start of subsequent anticancer therapy	Date of the last adequate IRC assessment before start of subsequent anticancer therapy	Censored
IRC-assessed PD or death after start of subsequent anticancer therapy	Date of the last adequate IRC assessment before start of subsequent anticancer therapy	Censored
IRC-assessed PD or death after 2 or more consecutively missed visits	Date of last adequate IRC assessment before the consecutively missed visits	Censored

PD=progressive disease; IRC=independent review committee.

PFS will be calculated as date of disease progression or death (censoring date for censored subjects) – randomization date + 1.

The last adequate IRC assessment is defined as the last IRC-assessed overall response that is not ‘UNK’ (unknown), as defined in the IRC charter.

7.1.2 Primary Efficacy Analysis

The primary efficacy analysis is to compare PFS as assessed by IRC between Arms A and B in ITT population.

The primary efficacy analysis will be performed using a stratified log rank test adjusting for randomization stratification factors. The estimate of the hazard ratio (Arm B/Arm A) and its corresponding 95% CI will be computed using a Cox Proportional Hazards model stratified by randomization stratification factors. Randomization stratification factors will be based on the data recorded in IXRS.

Kaplan-Meier (KM) curve will be used to estimate the distribution of PFS. PFS rate based on KM point estimate and its corresponding 95% CI will be calculated at selected timepoints for each treatment arm. A summary of PFS events will be provided by treatment arm.

7.1.3 Sensitivity Analyses

The following sensitivity analyses are planned for IRC-assessed PFS in support of the primary analysis:

- Unstratified analysis
- The progression-free survival (PFS) will be analyzed as the time from date of randomization to the date of first IRC-assessed disease progression or death due to any cause, whichever comes first, regardless of the use of subsequent anticancer therapy; i.e., subjects will not be censored at the last adequate IRC assessment prior to the subsequent anticancer therapy.
- IRC-assessed PD or death after 2 or more consecutively missed visits will be included as a PFS event
- Subjects with IPD will be excluded from the analysis

7.1.4 Subgroup Analysis

Subgroup analyses will be performed using potential prognostic variables at screening or baseline listed below to investigate the consistency and robustness of the primary analysis:

- Presence of 17p del (yes vs. no)
- ECOG status (0, 1 vs. 2)
- Number of prior therapies (1-3 vs. ≥4)
- Age group (<65 vs. ≥65)

- Age group (<75 vs. 7≥75)
- Sex (Male vs. Female)
- Race (White vs. Non-White)
- Geographic region (North America vs. Australia, New Zealand vs. Western Europe vs. Central and Eastern Europe vs. Asia)
- Rai Stage at screening (Stage 0-II vs. III-IV)
- Bulky disease (<5 cm vs. ≥5 cm)
- β2-microglobulin at baseline (≤3.5 mg/L vs. >3.5 mg/L)
- Presence of 11q del (yes vs. no)
- TP53 mutation (yes vs. no)
- 17p del and TP53 mutation (yes vs. no)
- 17p del or TP53 mutation (yes vs. no)
- IGHV (mutated vs. unmutated)
- 17p del, TP53 mutation, or 11q del (yes vs. no)
- 17p del, TP53 mutation, 11q del, or unmutated IGHV (yes vs. no)

The subgroup analyses for the stratification factors (presence of 17p del, ECOG status, and number of prior therapies) will be based on the values entered into the IXRS; all other factors will be based on values recorded on the eCRF.

The HR (Arm A versus Arm B) with its 2-sided 95% CI will be calculated based on an unstratified Cox regression model for each subgroup. The HRs and their 2-sided 95% CIs will be displayed for all subgroups graphically in a forest plot.

Additional subgroup analysis may be provided as appropriate.

7.2 Secondary Efficacy Endpoints and Analyses

7.2.1 IRC-Assessed Overall Response Rate (ORR)

Best overall response is defined as the best response as assessed by IRC on or before the initiation of subsequent anticancer therapy. If a subject takes any subsequent anticancer therapy, best overall response will be re-derived based on all IRC timepoint-by-timepoint assessments prior to the use of subsequent anticancer therapy. Categories for response

assessments include CR, CRi, nPR, PR, PRL, stable disease (SD), and PD. ORR is defined as the proportion of subjects achieving a best overall response of complete remission (CR), complete remission with incomplete bone marrow recovery (CRi), nodular partial remission (nPR), or partial remission (PR) at or before initiation of subsequent anticancer therapy.

Best overall response will be summarized by number and percentage of subjects for each response category. ORR will be summarized by number and percentage of subjects, and its corresponding 95% CI will be calculated based on normal approximation (using Wilson's score). ORR will be analyzed using the Cochran-Mantel-Haenzel (CMH) test adjusting for randomization stratification factors.

Descriptive statistics will be provided for best overall response for each treatment arm. The number and proportion of subjects within each category of response will be presented. The proportion will be estimated by dividing the number of subjects within each category of response by the number of subjects. Each subject will be counted within only one response group, with the best response during the study as the classification group.

ORR including PR with lymphocytosis (PRL) assessed by IRC will also be analyzed with the same analysis method used for ORR. Lymphocytosis is defined in Section 8.3.2.

The concordance between INV-assessed and IRC-assessed best overall response will be summarized by treatment arm.

7.2.2 Overall Survival (OS)

OS is defined as the time from date of randomization to death due to any cause. Subjects who were not known to have died prior to the analysis data cutoff date will be right-censored as follows:

Table 3 Overall Survival

Situation	Censoring Date
Lost to follow-up immediately after randomization	Randomization date
Not known to have died at or prior to analysis data cutoff date	Date last known alive before analysis data cutoff date

OS will be calculated as death date (or censoring date) - randomization date + 1.

OS will be analyzed in the same fashion as that for primary efficacy endpoint as described in Section 7.1.

A sensitivity analysis for OS will be conducted in which Arm B subjects who crossed over to receive acalabrutinib will be censored at the day prior to first dose of acalabrutinib.

7.2.3 INV-Assessed PFS

INV-assessed PFS is defined as time from randomization until disease progression (assessed by the Investigator per IWCLL 2008 criteria) or death from any cause, whichever occurs first. Analysis methods for INV-assessed PFS will be similar to IRC-assessed PFS as described in Section 7.1. The concordance between INV-assessed and IRC-assessed PD will be summarized by treatment arm.

7.2.4 INV-Assessed Overall Response Rate (ORR)

INV-assessed ORR will be summarized and analyzed similar to IRC-assessed ORR as described above.

7.2.5 INV- and IRC-Assessed Duration of Response (DOR)

DOR is defined as the interval from the first documentation of CR, CRi, PR, or nPR to the first documentation of disease progression or death from any cause, whichever is earlier. Following first documentation of CR, Cri, PR, or nPR, subjects who do not have disease progression and are alive by the analysis data cutoff date will be censored by the same censoring rule as for PFS.

DOR determined by IRC and by investigators will be analyzed in the same fashion as that for primary efficacy endpoint as described in Section 7.1.

7.2.6 Time to Next Treatment (TTNT)

TTNT is defined as the time from date of randomization to date of institution of non-protocol specified treatment for CLL (or first dose date of acalabrutinib for Arm B subjects who crossed over to receive acalabrutinib) or death due to any cause, whichever comes first. Subjects who do not have the above specified events prior to the data cutoff date will be censored at the date of last visit. TTNT will be calculated as: (earlier date of institution of non-protocol specified treatment for CLL or death due to any cause) – date of randomization + 1. For censored subjects, date of last visit will replace earlier date of institution of non-protocol specified treatment for CLL or death due to any cause in the calculation.

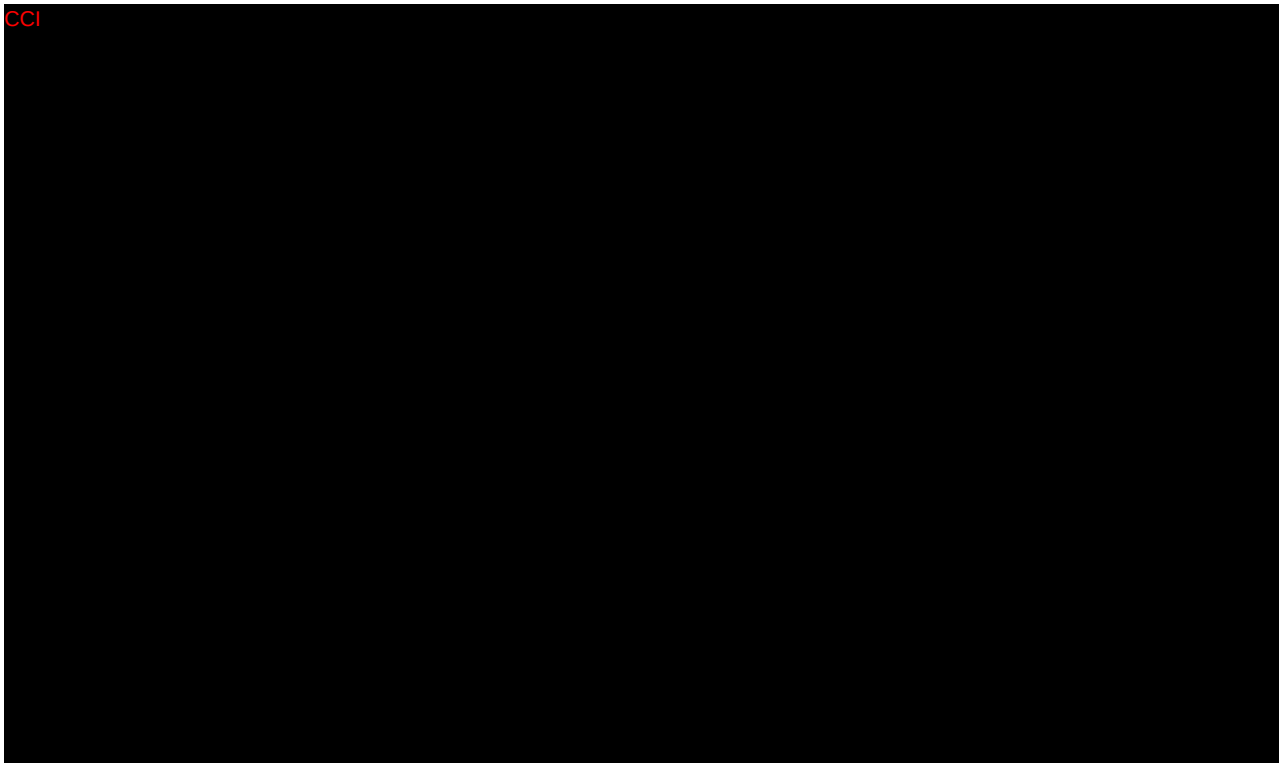
TTNT will be analyzed in the same fashion as that for primary efficacy endpoint as described in Section 7.1.

7.2.7 PRO by FACIT-Fatigue

Analysis of patient-reported outcome (PRO) endpoints will be specified in a separate SAP.

7.3 CCI

CCI


CCI

8 SAFETY SUMMARIES

Safety analyses will be performed using the safety population unless specified otherwise. Data collected during the main study period and crossover period will be summarized separately.

Safety and tolerability will be assessed by the incidence of treatment-emergent adverse events (TEAEs), changes in laboratory parameters and vital signs from baseline, and ECOG performance status.

8.1 Adverse Events

The investigator will grade AEs according to the National Cancer Institute (NCI) Common Terminology Criteria for Adverse Events (CTCAE) v4.03. Each AE verbatim term will be coded to a system organ class (SOC) and a preferred term (PT) using the Medical Dictionary for Regulatory Activities (MedDRA).

8.1.1 Treatment Emergent Adverse Events

A TEAE is defined as any event with an onset date on or after the first dose date of study drug or any ongoing event that worsens in severity after the first dose date of study drug, and prior to 30 days after the date of the last dose of study drug or the first date starting new anticancer therapy, whichever is earlier. Any AEs with onset date after the start of subsequent anticancer therapy will not be considered as a TEAE. Only TEAEs will be tabulated in summary tables. For the purpose of determining a TEAE, incomplete onset dates will be imputed as detailed in Section 4.2.

All TEAEs will be summarized by treatment arm as treated. For each treatment arm, AE incidence will be summarized with frequency and percentage by system organ class and Preferred Term, and the denominator for the AE incidence will be based on the number of subjects treated in that treatment arm, unless otherwise specified. In addition, AE incidence will also be summarized by severity and relationship to study drug. Relation to study drug is recorded on the eCRF per the investigator's judgment.

Subjects with multiple occurrences of events for a given Preferred Term, system organ class, or overall will be counted only once at the maximum severity and strongest relationship to study drug, respectively, for each preferred term and system organ class.

In addition, Grade 3 or Grade 4 TEAEs, TEAEs that led to permanent study drug treatment discontinuation, TEAEs that led to dose reduction, serious TEAEs, TEAEs of clinical interest, and TEAEs that resulted in death will be summarized by treatment arm as treated.

8.1.2 Adverse Events of Clinical Interest

AEs of clinical interest include cardiac events, cytopenia, hemorrhage, hepatic events, hypertension, infection, interstitial lung disease/pneumonitis, tumor lysis syndrome, and secondary malignancy. Major hemorrhage is defined as any hemorrhagic event that is reported as serious or Grade ≥3 in severity or that is a CNS hemorrhage (any severity grade). Secondary malignancies are defined as new malignant tumors, including solid tumors, skin malignancies, and hematologic malignancies. The secondary malignancies are to be reported for the duration of study treatment and during any protocol-specified follow-up periods, including post-progression follow-up for OS. These events will be summarized similar to TEAEs by treatment arms.

8.2 Deaths

All reported subject deaths will be summarized by treatment arm. A by-subject listing that includes date of death and cause of death will be provided.

8.3 Laboratory Assessments

8.3.1 Data Summary Methods

All laboratory values will be converted to and reported as SI units and classified as normal, low, or high based on the normal ranges provided by the central laboratory. In general, only data from the central laboratory will be summarized and analyzed. Hematologic parameters, including platelet counts, hemoglobin, and neutrophils, will be assessed by both the grading scale for hematologic toxicity in CLL studies in the IWCLL 2008 guidelines and the grading scale in NCI CTCAE v4.03. All other gradable laboratory parameters will be graded using the NCI CTCAE v4.03.

Per the grading scale in the IWCLL guidelines, 1) ANC: Both baseline grade and post-baseline grade are defined based on absolute values and 2) hemoglobin and platelet: baseline grade is not applicable (no criterion is provided to define baseline grade) and post-baseline grade is based on percentage decrease from baseline value.

Gradable parameters that have criteria available for both low and high values (e.g., hypercalcemia for a high value of calcium and hypocalcemia for a low value of calcium) based on the NCI CTCAE v4.03 will be summarized for both criteria (low and high). Subjects will be counted only once for each criterion/direction. The same subject can be counted for both criteria if the subject has laboratory values meeting each criterion. Subjects meeting the

criteria for Grade 1 or higher for the high direction will be summarized under Grade 0 when summarizing the low direction and vice versus.

For each laboratory parameter, the baseline laboratory value/grade is defined as the last laboratory value/grade collected on or prior to the date of the first dose of study drug. The change from baseline to post-baseline value/grade will be summarized for each parameter.

For liver function tests, laboratory value abnormalities will be summarized for the following categories: ALT or AST >3 x ULN and total bilirubin >2 x ULN same visit (biochemical criteria for Hy's law).

8.3.2 Analysis of Lymphocytosis

For all subjects with baseline and any post-baseline ALC measurements, ALC at peak summary will be provided by treatment arm. Median percentage change in ALC from baseline along with its 95% CI will also be displayed graphically over time.

Lymphocytosis is defined as an elevation in ALC of ≥50% compared with baseline and a post-baseline assessment >5,000/μL in the peripheral blood. The number and percentage of subjects with at least one occurrence of lymphocytosis will be summarized. ALC at peak and time to peak ALC for subjects who have lymphocytosis will be summarized by descriptive statistics.

Duration of lymphocytosis (DOL) is defined as the duration of time from the earliest date on which the ALC value met the lymphocytosis criteria at a post-baseline assessment to the earliest date on which a subsequent ALC value met the resolution criteria. Resolution of lymphocytosis is defined as 1) a decrease of ALC value to the baseline level or lower or 2) an achievement of ALC value that is below 5,000/μL in the peripheral blood, whichever occurs first.

> DOL = Earliest date of meeting resolution criteria - Earliest date of meeting lymphocytosis criteria + 1.

Subjects who developed lymphocytosis but whose lymphocytosis was not resolved prior to the analysis cutoff date will be censored at the last sample date with non-missing ALC value at or prior to the analysis cutoff date. KM curves will be used to estimate the distribution of DOL. The 50th percentile of KM estimates along with its two-sided 95% CI will be used to estimate the median DOL.

8.3.3 Analysis of Serum Immunoglobulins

Serum immunoglobulins (IgA, IgG, and IgM) will be summarized using descriptive statistics at each scheduled post-baseline timepoint. An additional IgG summary to exclude subjects who received IV immunoglobulin on the study will be provided.

8.4 ECOG Performance Status

The ECOG performance status grade ranges from 0 to 5. The ECOG performance score and change from baseline will be summarized by descriptive statistics over time.

8.5 Vital Signs and Weight

Body temperature, heart rate (beats/min), systolic blood pressure (mmHg), diastolic blood pressure (mmHg), respiratory rate (breaths/min) and weight for each parameter will be summarized over time by treatment arms.

8.6 Electrocardiogram

Subject incidence of electrocardiogram (ECG) abnormality at baseline and post baseline visits will be summarized by treatment arms. Number and percent of subject with QTcB or QTcF values >480 msec at screening visit will be summarized in a table.

9 REFERENCES

Common Terminology Criteria for Adverse Events (CTCAE) Version 4.03, USDHHS, NIH, NCI; publish date May 28, 2009 (v4.03: June 14, 2010).

Hallek M, et al. Guidelines for the diagnosis and treatment of chronic lymphocytic leukemia: a report from the International Workshop on Chronic Lymphocytic Leukemia updating the National Cancer Institute–Working Group 1996 guidelines. Blood 2008 111(12):5446-56.

Cheson BD, Byrd JC, Rai KR, et al. Novel targeted agents and the need to refine clinical end points in chronic lymphocytic leukemia. J Clin Oncol 2012;30:2820-2.

O'Brien PC and Fleming TR. A Multiple Testing Procedure for Clinical Trials, Biometrics 1979 35(3):549-56

Lan KG and DeMets DL. Discrete Sequential Boundaries for Clinical Trials, Biometrika 1983 70(3):659-63