Janssen Research & Development

Statistical Analysis Plan

Randomized, Double-blind, Placebo-controlled Phase 3 Study of Ibrutinib, a Bruton's Tyrosine Kinase (BTK) Inhibitor, in Combination with Bendamustine and Rituximab (BR) in Subjects With Relapsed or Refractory Chronic Lymphocytic Leukemia/Small Lymphocytic Lymphoma

Protocol PCI-32765CLL3001; Phase 3

JNJ 54179060 (Ibrutinib)

Date: 9 February 2015
Prepared by: Janssen Research & Development, LLC
Document No.: EDMS-ERI-99931663

Compliance: The study described in this report was performed according to the principles of Good Clinical Practice (GCP).

TABLE OF CONTENTS

AMENDMENT HISTORY

ABBREVIATIONS

AE(s)	adverse event (s)
AESI	adverse events of special interest
ALT	alanine aminotransferase
ALC	absolute lymphocyte counts
ANC	absolute neutrophil count
ANCOVA	analysis of covariance
aPTT	activated partial thromboplastin time
AST	aspartate aminotransferase
BR	bendamustine and rituximab
BSA	body surface area
BTK	Bruton's tyrosine kinase
CBC	complete blood count
CI	confidence interval
CIT	chemoimmunotherapy
CLL	chronic lymphocytic leukemia
CR	complete response
CRF	case report form
CRi	complete response with an incomplete marrow recovery
CT	computed tomography (scan)
CYP	cytochrome P450
DBP	diastolic blood pressure
DOL	duration of lymphocytosis
DOR	duration of response
DMC	Data Monitoring Committee
EAIR	exposure-adjusted incidence rate
ECG	electrocardiogram
ECOG	Eastern Cooperative Oncology Group
eCRF	electronic case report form
EORTC QLQ-C30	European Organization for Research and Treatment of Cancer Quality of Life Questionnaires Core 30
EQ-5D	EuroQoL Five-Dimension.
FACIT-Fatigue	Functional Assessment of Chronic Illness Therapy-Fatigue
FISH	fluorescence in situ hybridization
HDL	high-density lipoproteins
Hgb	hemoglobin
Ig	immunoglobulin
IgA	immunoglobulin A
IgG	immunoglobulin G
IgM	immunoglobulin M
IgVH	immunoglobulin variable heavy gene
INR	international normalized ratio
IPE	Iterative Parameter Estimate
IRC	independent review committee
ITT	intent-to-treat (population)
IV	intravenous
IVIG	intravenous immunoglobulin
IWCLL	International Workshop on Chronic Lymphocytic Leukemia Criteria
IWRS	Interactive Web Response System

JRD	Janssen Research & Development, LLC
LDH	lactic acid dehydrogenase
LLN	lower limit of normal
MedDRA	Medical Dictionary for Regulatory Activities
MRI	magnetic resonance imaging
MMRM	Mixed Model for Repeated Measures
MRD	Minimal residual disease
NCI CTCAE	National Cancer Institute Common Terminology Criteria for Adverse Events
NE	not evaluable
NED	no evidence of disease
NHL	non-Hodgkin's lymphoma
nPR	nodular partial response
ORR	overall response rate
OS	overall survival
PCYC	Pharmacyclics Inc.
PD	progressive disease
PET	positron emission tomography
PFS	progression-free survival
PLT	platelets
PR	partial response
PRL	partial response with lymphocytosis
PRO	patient-reported outcome(s)
QLQ	quality of life questionnaire
SAE	serious adverse event
SAP	statistical analysis plan
SBP	systolic blood pressure
SD	stable disease
SLL	small lymphocytic lymphoma
SOC	system organ class
TEAE	treatment-emergent adverse events
TTR	time to response
VAS	Visual Analogue Scale
ULN	upper limit of normal
UNK	Unknown
WBC	white blood cell (count)
WHO	World Health Organization

Definitions of Terms

Study medication or study drug	Ibrutinib/placebo
Study treatment	Bendamustine hydrochloride, rituximab, and ibrutinib/placebo

1. INTRODUCTION

This clinical study is a part of a comprehensive ibrutinib clinical development plan to evaluate the safety and efficacy of ibrutinib for subjects with B-cell malignancies. This randomized, double-blind, placebo-controlled Phase 3 study is designed to evaluate the efficacy and safety of ibrutinib, a Bruton's tyrosine kinase (BTK) inhibitor, in combination with bendamustine and rituximab (BR) in subjects with relapsed or refractory B-cell chronic lymphocytic leukemia (CLL)/ small lymphocytic lymphoma (SLL).

The purpose of the statistical analysis plan (SAP) is to lay out key elements including definition and statistical methods for the planned analyses for the primary, secondary and safety endpoints.

1.1. Trial Objectives

1.1.1 Primary Objective

The primary objective is to determine whether the addition of ibrutinib to BR significantly improves progression-free survival (PFS) compared with BR in subjects with relapsed or refractory CLL/SLL.

1.1.2 Secondary Objectives

The secondary objectives are to evaluate:

- the safety of ibrutinib in combination with BR
- the overall response rate (ORR; complete response [CR] + complete response with incomplete marrow recovery [CRi] + partial response [PR] + nodular partial response [nPR])
- the overall survival (OS)
- the rate of minimal residual disease (MRD)-negative response
- improvement in hematologic parameters (hemoglobin, neutrophil count, platelet count)
- improvement of disease-related symptoms (fatigue, night sweats, weight loss, fever, and abdominal discomfort due to splenomegaly)
- patient-reported symptoms, functional status, and well-being as measured by European Organization for Research and Treatment of Cancer Quality of Life Questionnaire (EORTC-QLQ) C30, EORTC QLQ CLL 16, EQ-5D-5L, and Functional Assessment of Chronic Illness Therapy (FACIT)-Fatigue Scale

- To characterize the pharmacokinetics (PK) of ibrutinib and explore its potential effect on bendamustine and rituximab PK, the potential relationships between ibrutinib metrics of exposure with relevant clinical or biomarker information

- To examine biomarkers related to B-cell receptor (BCR) and compensatory signaling pathways and explore their association with resistance to ibrutinib treatment

1.2. Trial Design

This is a randomized, double-blind, placebo-controlled, multicenter Phase 3 study to determine the benefits and risks of combining ibrutinib and BR in subjects with relapsed or refractory B-cell CLL/SLL following at least 1 line of prior systemic therapy.

All subjects will receive background therapy with BR for a maximum of 6 cycles. A cycle will be defined as 28 days, with the exception of Cycle 1, which will be 29 days to allow for rituximab dosing prior to bendamustine and study medication. Approximately 580 subjects will be randomized in a 1:1 ratio to either Treatment Arm A (placebo) or Treatment Arm B (ibrutinib 420 mg). Study medication will be administered orally once daily on a continuous schedule.

Stratification factors will include whether refractory to purine analog therapy (yes or no) and number of prior lines of therapy (1 or >1). For stratification, refractory is defined specifically as a failure to respond (ie, progressive disease [PD] or stable disease [SD]) to a purine analog or a recurrence of disease within 12 months of the last dose of purine analog therapy. Separate lines of therapy are defined as single or combination therapy regimens that are either separated by disease progression or refractory disease or by a treatment-free interval of 6 months or greater. In addition, a change of regimen due to insufficient response may be considered a separate line of therapy.

Subject participation will include a Screening Phase, a Treatment Phase, and a Follow-up Phase. The Screening Phase will be up to 30 days prior to randomization. The Treatment Phase will extend from randomization until study treatment (bendamustine, rituximab, and ibrutinib/placebo) discontinuation. Subjects will receive both BR and ibrutinib or placebo for the first 6 cycles, and thereafter will receive study medication until disease progression or unacceptable toxicity. For subjects who discontinue BR prior to completing 6 cycles, treatment with study medication will continue. The Follow-up Phase will begin once a subject discontinues study treatment. Subjects who discontinue for reasons other than disease progression will continue to have disease evaluations according to the Time and Events Schedule. The Follow-up Phase will continue until death, loss to follow up, consent withdrawal, or study end, whichever occurs first. Study end is defined as when either 80% of the subjects have died or 4 years after the last subject is randomized into the study, whichever occurs first.

Assessment of tumor response and progression will be conducted in accordance with the International Workshop on Chronic Lymphocytic Leukemia (IWCLL) 2008 Guidelines (Hallek 2008). The investigator will evaluate sites of disease by radiological imaging, physical examination, or other procedures as necessary, and review of hematology and clinical chemistry results. A central laboratory will perform complete blood count (CBC) testing. The primary efficacy analysis of PFS will be based on assessment by an Independent Review Committee (IRC). Patient-reported symptoms, functional status, and well-being will also be measured.

During the study, safety evaluations will include adverse event (AE) monitoring, physical examinations, concomitant medication usage, and clinical laboratory parameters (hematology, chemistry, coagulation). At each site visit, subjects will be evaluated for toxicity. Blood samples will be drawn for assessment of PK and biomarker parameters. All study evaluations will be conducted according to the Time and Events Schedule.

An independent Data Monitoring Committee (DMC) will be formed and constituted according to regulatory agency guidelines. One interim analysis is planned for the study. Detailed information regarding the composition and procedures of the DMC and interim analysis are provided in Section 3 and a separate DMC charter.

At investigator's request, subjects who have IRC-confirmed disease progression and meet the IWCLL criteria for requiring subsequent anti-CLL therapy and other eligibility criteria may be considered for next-line ibrutinib treatment. If upon unblinding by a medical monitor separate from the study team it is determined that the subject was randomized to placebo (Treatment Arm A), the subject will be permitted to cross over to receive ibrutinib 420 mg orally, daily. Subjects will continue to be followed up for assessment of response and other study endpoints as specified in the Time and Events Schedule for next-line ibrutinib. Open-label next-line treatment with ibrutinib will continue until disease progression, unacceptable toxicity, withdrawal from study, or until the study end, whichever occurs earlier.

1.3. Statistical Hypotheses for Trial Objectives

The primary hypothesis of this study is that the experimental treatment ibrutinib + BR compared with placebo + BR will significantly improve PFS, in subjects with relapsed or refractory CLL/SLL.

The statistical hypotheses are as follows:

H_0: The PFS distributions of experimental treatment group, $S_T(t)$, and the placebo group, $S_P(t)$, are equal at all time points t:

$S_T(t) = S_P(t)$, for all $t > 0$

versus

H_1: The PFS distributions of experimental treatment group, $S_T(t)$, are greater than the placebo group, $S_P(t)$, at least one time point t:

$S_T(t) > S_P(t)$, for some $t > 0$

These hypotheses will be tested using stratified log-rank test and assessed within the context of a group sequential testing design as described in Sections 3.

1.4. Sample Size Justification

A recent Phase 2 study reported median PFS in subjects with relapsed or refractory CLL treated with combination of BR is 15 months [1]. It is assumed that the PFS follows an exponential distribution with a constant hazard ratio.

Approximately 580 subjects (290 per treatment group) will be randomized to observe 342 PFS events. The study is designed to detect a hazard ratio of 0.7 for the ibrutinib + BR group relative to placebo + BR group (corresponding to an improvement of 43% in median PFS, e.g. from 15 months to 21.5 months) with 90% power at a 1-sided significance level of 0.025, using a group sequential testing design as described in Section 3. Assuming the enrollment rate is on average 32 subjects per month, the total study duration is expected to be approximately 33 months with 18 months of enrollment and 15 months of follow up in order to observe 342 PFS events.

1.5. Randomization

Central randomization will be implemented in this study. Subjects will be randomly assigned to 1 of 2 treatment groups in a 1:1 ratio based on a computer-generated randomization schedule prepared before the study by or under the supervision of the sponsor. The randomization will be balanced by using randomly permuted blocks and will be stratified by purine analog refractory status (yes vs. no), and prior lines of therapy (1 vs. >1). These stratification factors represent prognostic factors that may potentially affect treatment outcome and are included to eliminate bias and to increase the precision of overall treatment effect estimates. The interactive web response system (IWRS) will assign a unique treatment code, which will dictate the treatment assignment and matching study medication kit for the subject. The requestor must use his or her own user identification and personal identification number when contacting the IWRS, and will then give the relevant subject details to uniquely identify the subject.

This is a double-blind study. The block size will be chosen to minimize the chance of accidental unblinding while sufficiently controlling for potential imbalance between treatment groups, and will be kept confidential as part of the randomization schedule. The randomization schedule will be prepared by an independent statistician not otherwise involved with this study, and will be implemented within IWRS.

1.6. Blinding Procedures

It is the intent for all subjects, investigators, and study team members from the sponsor to remain blinded to treatment group assignment until the completion of the study and database is finalized. The IRC who performs tumor assessment will be blinded to study treatment as well. The exceptions are:

- The independent DMC and the independent biostatistician and statistical programmer from Statistical Support Group (SSG) who are responsible for preparing interim tables, listings, and graphs for DMC review. Unblinding procedures and the control of the unblinded data are described in the DMC charter.
- Personnel performing blood serum concentration assays and analysis for PK.
- Unblinded safety representative and Ethics Committee for serious adverse event (SAE) reporting.

It is planned in the protocol to analyze the data when the prespecified required number of PFS events is observed. While the study is ongoing, the results may require unblinding for review by independent DMC, and potentially by regulatory agencies. An unblinding plan including the names of the members and the process will be prepared prior to performing the analysis.

Protocol amendment INT-3 described the circumstances and process for crossover and access to next-line ibrutinib for those subjects who have IRC confirmed disease progression and were previously on placebo with bendamustine and rituximab. The implementation of crossover aims to protect the primary endpoint by mitigating potential early dropout to access ibrutinib through other mechanisms. The steps taken to protect the integrity of the key efficacy endpoints during conduct of the crossover phase of the study and to minimize access to unblinded information are described in a separate document.

2. GENERAL ANALYSIS DEFINITIONS

2.1. Study Medication

For the purpose of analysis, the term “study medication” refers to all 3 drugs, ibrutinib/placebo, bendamustine and rituximab.

2.2. Analysis Sets

The analysis sets are:

- Intent-to-Treat (ITT) population: defined as all subjects randomized into the study and who will be analyzed according to assigned treatment group, regardless of the actual treatment received. This population will be used for all analyses of efficacy and patient-

reported outcome (PRO) endpoints, analyses of disposition, demographic, and baseline disease characteristics.

- Safety population: defined as all randomized subjects who receive at least one dose of ibrutinib or placebo. This population will be used for all safety analyses and analyses of exposure. All subjects will be analyzed according to the treatment which they actually received.

- Per-protocol (PP) population: defined as all randomized subjects who undergo at least one post-baseline disease assessment (post-baseline tumor assessment by independent radiology reviewers/IRC) or died within 2 months and do not have major protocol violations as described below:

    a. did not meet all entry criteria

    b. did not receive the treatment to which they were randomized

    c. had less than 67% of study medication compliance

    d. Received prohibited concomitant medications; specifically, other antineoplastic therapy during the treatment phase.

  Subjects in this population will be analyzed according to the treatment to which they are randomized. This population will be used for sensitivity analysis for PFS.

2.3. Baseline Definitions

Unless specified otherwise, the baseline value is defined as the last non-missing value collected on or before the administration of the first dose of study medication. For subjects who have been randomized but not treated with any dose, randomization date will be used as the reference date for baseline value calculation.

2.4. Cycles

In defining treatment cycles in a combination therapy setting, the entire combination must be taken into account. Therefore the nominal cycles as defined in protocol and recorded on the Case Report Form (CRF) will be used in the statistical analyses.

2.5. Visit Windows

Visit windowing will be based on phases and cycles:

Screening: Before the first dose of study medication.

Treatment: Between the date of first dose of study medication and the date of the last dose of study medication. As noted above, the term “study medication” refers to all of the drugs used in the combination treatment.

The Treatment Phase will be subdivided by cycles, based on the nominal treatment cycles for the combination therapy as recorded on the CRF.

End of treatment: Between date of last dose of study medication + 1 and date of last dose of study medication+ 30 days. The assessments performed during the 'End-of-Treatment Visit' will be included in this phase.

Follow-up: After the end of treatment until the study cut off.

Cycle-based analysis may be performed for safety parameters during the treatment up to date of last dose +30 days or End-of-Treatment visit whichever comes later. For the analysis of lab grade by cycle, worst grade within each cycle will be used.

2.6. Study Day and Cycle Day

Assessments will be presented chronologically by study day or cycle day as described below.

For efficacy data, the randomization date is considered as the reference date (Day 1). For safety data, date of first dose of study medication (randomization date for subjects who have been randomized but not treated) will be used as the reference date (Day 1).

Reference date (Day 1) = randomization date (for efficacy data), or first dose date of study medication (for safety data)

Study Day = assessment date – reference date + 1 for assessment performed on or after the reference date; assessment date – reference date for assessment performed before the reference date.

Cycle Day = assessment date - date of the first dose for the cycle + 1.

2.7. Missing and Partial Dates

In general, imputation of missing dates will be made for AE onset date, AE resolution date, date of death, start and end dates of prior, concomitant and subsequent therapies, and date of initial diagnosis according to the following rules. Start date will be imputed before end date.

- If date is completely missing, no imputation will be made.
- If year is missing, no imputation will be made.
- If only year is present but month and day are missing, then June 30th will be used.
- If only day is missing but year and month are available, then the 15th of the month will be used.

However, the above imputations will be modified by the following rules:

- For initial diagnosis if such imputed date is on or after the randomization date, then randomization date - 1 will be used.
- If such imputed date for prior therapies or initial diagnosis is on or after the randomization date, then randomization date - 1 will be used. If such imputed date for subsequent therapies is before date of last dose, then date of last dose +1 will be used.
- The imputed start date for subsequent therapies will be adjusted sequentially using the following steps:
    - If the imputed start date is before the treatment discontinuation date or last dose date if no treatment discontinuation date but in the same year and month, then the treatment discontinuation date or last dose date if no treatment discontinuation date will be used.
    - If subsequent therapy end date is not missing and is before the imputed subsequent therapy start date, then the subsequent therapy end date will be used as the start date.
- If the imputed date is for a date of death and is before the last date that the subject is known to be alive, the latter date will be used.
- The imputed AE start date will be adjusted sequentially using the following steps:
    - If the imputed date is in the same year and month as but day before the first dose date, then the first dose date will be used, or if it is in the same year and month as but day after the last dose date + 30 days, then the last dose date + 30 days will be used.
    - If AE end date is not missing and the imputed AE start date is after the AE end date, then the AE end date will be used.
    - If the imputed AE start date and is after date of death, then date of death will be used
    - If the imputed AE start date is in the same month and year but after the 1st subsequent therapy start date, then 1st subsequent therapy start date 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, or if the imputed AE end date is before the AE start date, then the AE start date will be used.

- The AE imputation rule will be used for concomitant medication.

2.8. Treatment-Emergent Period

In general, the treatment-emergent period is defined as the time from first dose date through 30 days after last dose date, or day before subsequent antineoplastic therapy, whichever occurs first.

For subjects in placebo group who receive cross-over ibrutinib therapy, treatment-emergent period before cross-over ibrutinib therapy is defined as the time from first dose date through 30 days after last dose date, or day before subsequent antineoplastic therapy, or day before first dose date of crossover ibrutinib, whichever occurs first. Treatment-emergent period for cross-over ibrutinib therapy is defined as the time from the first dose date of crossover ibrutinib therapy through 30 days after last dose date.

2.9. Definition of Subgroups

Subgroup analysis will be performed for the selected variables to assess the internal consistency of the treatment benefit and/or safety. The subgroup variables and the cutoff values are subject to change if warranted to better represent the data.

Table 1. Subgroup Definition

Subgroup	Definition of Subgroup	Analysis Type
Age	<65, ≥65 yrs.	D, E, S
Sex	Male, Female	D, E, S
Race	White, Non-White	D, E, S
Geographic region	North America, Europe, Latin America, and Asia	D, E, S
Diagnosis	CLL, SLL	D, E, S
Rai Stage at screening	Stage 0-II, III-IV	E
Refractory to prior purine analogue therapy	Refractory to prior purine analogue therapy, Yes (defined as no response or failure within12 months), No Refractory to prior purine analogue therapy (Response lasting ≥12 months)	E

Subgroup	Definition of Subgroup	Analysis Type
Baseline ECOG	0, 1	E
Bulky Disease	Yes (LDi ≥5 cm), no (LDi <5 cm),	E
Chromosome 11q Deletion	Yes, no	E
Elevated LDH at Baseline	No (<350 U/L), Yes (≥350 U/L)	E
Cytopenias at Baseline	Yes, no	E
Serum β2 –microglobulin	≤3.5 mg/L, > 3.5 mg/L	E
IgVH	Mutated, Unmutated	E
ZAP-70	Elevated, Not Elevated	E
CD38	Positive (≥ 30%), negative (<30%)	E
complex karyotype	Yes, no	E
Refractory to purine analog therapy as recorded in IWRS	Yes, no	E
Prior lines of therapy as recorded in IWRS	1, >1	E
Concomitant use of any CYP3A inhibitor	Yes, No	S
Concomitant use of strong CYP3A inhibitor	Yes, No	S
Cytopenia is defined as yes if platelet count ≤ 100,000/uL, Hgb ≤ 11g/dL, or ANC ≤ 1500/uL is observed. analysis type D= demographic and baseline disease characteristics analysis type E= efficacy (PFS, OS, ORR) analysis type S= safety (adverse events) ECOG= Eastern Cooperative Oncology Group; IgVH= immunoglobulin variable heavy gene; IWRS= Interactive Web Response System; LDH= lactic acid dehydrogenase; LDi=longest diameter.		

2.10. Exposure Related Definitions

Treatment duration is the interval between date of first dose and end-of-treatment date. If date of end of treatment is not available, use last available treatment date.

Number of treatment cycles = the last cycle number - the first cycle number + 1.

For ibrutinib/placebo, dosing information includes total dose received (the sum of total dose), average dose level per administration (the ratio of total dose and treatment duration) and relative dose intensity (the ratio of average dose level per administration and 420).

For background therapy (BR), dosing information includes total dose received (the sum of actual dose administered), average dose level per administration (the ratio of total dose and treatment duration in cycles), and relative dose intensity:

$$\frac{Sum\ of\ actual\ dose\ administered}{planned\ total\ cumulative\ dose\ for\ the\ regimen}$$

Dose reduction for ibrutinib/placebo is defined as prescribing lower dose level (280 mg/day [level 1 reduction] or 140 mg/day [level 2 reduction]). For each subject, use the highest level of dose reduction to summarize categories - 'One dose reduction' and 'Two dose reductions'.

3. INTERIM ANALYSIS AND DATA REVIEW COMMITTEE

3.1. Data Monitoring Committee

An independent DMC was formed to monitor data on a regular basis to ensure the safety of the subjects in this study, assess the evidence of benefit or adverse effects of ibrutinib, and to monitor the overall conduct of the study. At interim analysis, the DMC may make recommendations regarding study continuation if the prespecified stopping boundary is crossed for efficacy or futility. In addition to the planned interim analysis, the DMC will meet periodically to review the cumulative safety data throughout the trial. The DMC safety review will focus on deaths, treatment discontinuations, SAEs and Grade 3/4 AEs and AEs of special interest, and to identify any potential added toxicity when ibrutinib is combined with BR. Details are provided in the DMC Charter.

3.2. Interim Analysis

One interim analysis is planned using group sequential testing design. It is scheduled to be conducted when 50% of the PFS events (or approximately 171 PFS events) have occurred.

The O'Brien-Fleming boundaries [2] as implemented by Lan-DeMets spending function were used to control the 1-sided Type I error of 0.025 and Type II error of 0.10 for the comparison on PFS endpoint. The efficacy and futility monitoring plan is summarized in Table 2. The actual type I error rates to spend at each analysis may differ depending on the actual information fraction. The sample size and interim analysis boundaries are calculated using the software package East 5.4 (Cytel Software Corp., Cambridge, MA).

Table 2. Stopping Boundaries and Corresponding Observed Difference

Analysis	PFS Events (info)	Anticipated Time to Analysis§	Anticipated Enrollment (n)	Cumulative Error		Stopping Boundary			
				Alpha	Beta	p-value (1-sided)		Observed Hazard Ratio	
				Efficacy	Futility	Efficacy	Futility†‡	Efficacy	Futility
Interim	171 (50%)	19 month	580	0.002	0.020	<0.0015	≥0.3947	<0.634	≥ 0.960
Final	342 (100%)	33 month	580	0.025	0.100	<0.0245	≥0.0245	<0.807	≥ 0.807

§Assuming enrollment rate 32/month

† CP1 (conditional power assuming observed hazard ratio at interim is true) is 1.2%

‡ CP2 (conditional power assuming the original hazard ratio assumption is true) is 42.2%.

Type I error of 0.025 (1-sided), Type II error of 0.10, efficacy and futility boundaries are based on O'Brien-Fleming boundaries.

Observed hazard ratio are approximates

4. SUBJECT AND TREATMENT INFORMATION

All statistical analyses will be performed using statistical analysis system (SAS®). Analyses of disposition, demographic, baseline disease characteristics and prior and concomitant therapy will

be conducted on the ITT population. Analyses of treatment compliance and extent of exposure will be conducted on the safety population. No statistical testing is planned.

Unless otherwise specified, all continuous endpoints will be summarized using descriptive statistics, which will include the number of subjects with a valid measurement (n), mean, standard deviation (SD), median, minimum, and maximum. All categorical endpoints will be summarized using frequencies and percentages. Percentages will be calculated by dividing the number of subjects with the characteristic of interest by the number of subjects in the analysis population.

4.1. Subject Disposition

Disposition information will be summarized for the ITT population and safety population. Subject enrollment will be summarized by region, country, and investigator. The number of subjects undergoing, discontinuing and completing the study treatment as well as their reasons for treatment discontinuation will be summarized.

Descriptive statistics will be provided for time on study. Time on study is defined the same way as OS with reversed censoring, i.e., subject who died will be censored. Based on this definition, time on study is the same as length of follow up. The Kaplan-Meier method will be used to estimate the median time on study.

4.2. Demographics and Baseline Characteristics

Subject demographics and baseline disease characteristics will be summarized using descriptive statistics.

- Demographics and baseline characteristics: age, sex, race, ethnicity, geographic region, height (cm), weight (kg), systolic blood pressure/diastolic pressure (SBP/DBP (mmHg), body surface area (m^2)
- Baseline disease characteristics: time from initial diagnosis to randomization, time from progression/relapses since last line of treatment to randomization, diagnosis (CLL, SLL), baseline Eastern Cooperative Oncology Group (ECOG, baseline Rai stage, baseline Binet stage, bulky disease, cytopenia at baseline, 11q deletion, immunoglobulin variable heavy gene (IgVH) status, elevated lactate dehydrogenase (LDH) at baseline, serum β2 – microglobulin
- Hematology: leukocytes, lymphocytes, hemoglobin, platelets, neutrophils
- Chemistry: creatinine, LDH, uric acid, liver function tests (alkaline phosphatase, aspartate aminotransferase [AST], alanine aminotransferase [ALT], total bilirubin), and electrolytes (sodium, magnesium, potassium, and phosphate).

4.3. Prior and Concomitant Medications

For summarization purposes, medications will be coded to a generic term based on the World Health Organization (WHO) dictionary. Medications administered prior to the first dose of study medication will be considered prior medications. Concomitant therapies include those taken on or after first dose date through the date of last dose of study drug. Using this definition, a medication can be classified as both prior and concomitant.

Incidence of prior and concomitant medications will be summarized by Anatomical Therapeutic Chemical (ATC) class and drug generic term. Prior anticancer therapy will be summarized by type (e.g. radiotherapy, surgery, chemotherapy). Best response to last line of prior therapy will also be summarized. Concomitant medications of special interest will be provided: CYP3A/4 inhibitors and inducers, growth factors, transfusions, anticoagulation and antiplatelets.

4.4. Protocol Deviations

Subjects with eligibility and major protocol deviations will be listed by treatment group.

Protocol deviations will be based on clinical review, but not limited to, the following aspects: (1) eligibility criteria, (2) patient safety, (3) efficacy assessment deviation, (4)treatment compliance . Protocol deviations will be closely monitored during the execution of the study and the final set of protocol deviation criteria will be finalized before database lock.

4.5. Extent of Exposure

Descriptive statistics (n, mean, standard deviation, median, and range) will be provided for total number of cycles, treatment duration and dosing information for all study medications (ibrutinib/placebo, BR).

The number and percentage of subjects with dose reduction and dose interruption will be summarized. In addition, subjects with dose modifications and reasons for dose modifications will be summarized.

Details of exposure related definition is specified in Section 2.10.

5. EFFICACY

Analysis of efficacy endpoints will be conducted on the ITT population. Table 3 summarizes the efficacy endpoints and analysis methods to be performed. The analyses on PROs are detailed in Section 7. Analyses to support HEMAR are described in Appendix 1.

Table 3. Summary of Efficacy Analyses to be Performed

Endpoint	Analysis	Analysis Method	Population
Primary			

Endpoint	Analysis	Analysis Method	Population
PFS assessed by IRC	Primary	Stratified log-rank test, stratified Cox regression model, all PD and death considered as events regardless of subsequent antineoplastic therapy.	ITT; PP
	Sensitivity	1) Alternative censoring rule 1: subjects who received subsequent antineoplastic therapy are considered to have had a PFS event at the initiation of subsequent therapy 2) Alternative censoring rule 2: subjects who received subsequent antineoplastic therapy are censored at the last disease assessment showing no evidence of PD before the use of subsequent therapy 3) Alternative censoring rule 3: subjects will be censored at the last disease assessment if they progress or die after missing ≥2 planned disease assessment visits. 4) Non-stratified log-rank test, non-stratified Cox regression model; evaluation of censoring reasons 5) Investigator assessed PFS by stratified log-rank test and stratified Cox regression model; concordance rate between the IRC-determined PFS and investigator-determined PFS	ITT
	Subgroup	Non-stratified Cox regression model within each subgroup	ITT
Secondary			
Overall Survival	Primary	Stratified log-rank test, stratified Cox regression model, all death considered as events regardless of any subsequent antineoplastic therapy including crossover	ITT
	Subgroup	Non-stratified Cox regression model within each subgroup	ITT
	Sensitivity	1. Which censor subject at the time of crossover for placebo subjects; 2. Which censor ibrutinib subjects at subsequent therapy.	ITT
Overall Response Rate	Primary	IRC assessed ORR CMH chi-square test, Logistic regression model	ITT subjects with measurable disease at baseline
	Sensitivity	Investigator assessed ORR by CMH chi-square test, logistic regression model	ITT subjects with measurable disease at baseline
	Subgroup	Logistic regression model	ITT subjects with measurable disease at baseline

Endpoint	Analysis	Analysis Method	Population
Rate of MRD Negative Response		chi-square test	ITT
Sustained Hematologic Improvement		chi-square test	ITT; ITT subjects with cytopenia at baseline
Disease-Related Symptom Improvement		Descriptive summary	ITT
Patient-Reported Outcome		Descriptive summary, mixed model with repeated measurements	ITT
Other			
Subsequent Therapy		Descriptive summary	ITT
Beta2-microglobulin, IgA, IgG, and IgM		Descriptive summary	ITT

5.1. Analysis Specifications

5.1.1. Level of Significance

In general, all tests will be performed at a 2-sided significance level of 0.05, unless otherwise specified. All interval estimation will be reported using 2-sided 95% confidence intervals (CIs).

Statistical inference on the primary endpoint, PFS, at interim and final analysis will be conducted at 1-sided significance level of 0.025, under group sequential testing design per O'Brien-Fleming boundaries, as specified in Section 3.

Testing of Secondary Endpoints

The secondary endpoints are to be tested sequentially at the nominal 0.05 significance level (2-sided). The order of these endpoints is as follows:

1. ORR
2. OS
3. MRD negative rate
4. Time to improvement in FACIT fatigue score
5. Rate of sustained hemoglobin improvement
6. Rate of sustained platelet improvement

A secondary hypothesis will be tested if and only if the primary hypothesis is rejected along with all the secondary hypotheses that precede it.

5.1.2. Data Handling Rules

Unless specified otherwise, missing values will not be imputed.

5.1.3. General Analysis Considerations

Descriptive statistics and subject listings will be used to summarize the data. For continuous variables, number of observations, means, standard deviations, medians, and ranges will be used. For discrete variables, frequency will be summarized. For time-to-event variables, Kaplan-Meier estimates will be provided.

Unless otherwise specified, disease progression and disease response will be based on assessments from an IRC, according to principles adapted from 2008 International Working Group for CLL (IWCLL) Guidelines (Hallek 2008) [3]. Disease evaluations will include: physical examination, CBC, computed tomography (CT) scan of the neck, chest, abdomen, and pelvis, and bone marrow aspirate. Criteria for response categories, as well as the process and convention of the IRC are detailed in a separate IRC charter.

5.2. Primary Efficacy Endpoint

5.2.1. Progression Free Survival (PFS) based on IRC

The primary efficacy endpoint, PFS, is defined as the time between the date of randomization and the date of disease progression (as assessed by the IRC) or date of death due to any cause, whichever occurs first, regardless of the use of subsequent antineoplastic therapy prior to documented PD or death. The primary efficacy analysis will be based on the PFS determined by IRC and will be performed on the ITT population.

5.2.2. Primary Analysis Methods for PFS

Kaplan-Meier method will be used to estimate the distribution of PFS for each treatment group. The median PFS and its 95% CI will be provided. PFS rate with 95% CI at selected landmark points will be provided. The stratification factors to be used in the analysis are: refractory to purine analog therapy (yes vs. no), and number of prior lines of therapy (1 vs. >1) used in the randomization. The treatment effect of ibrutinib compared to placebo based on PFS will be tested with a stratified log-rank test. The hazard ratio for ibrutinib relative to placebo and its associated 95% CI will be calculated based on the stratified Cox proportional hazards model by the stratification factor at randomization.

For primary analysis of PFS, all disease progression (assessed by IRC) or death documented in the study will be considered as events, including those that occur after subsequent antineoplastic therapy is started or study medication is discontinued. Subjects who are progression free and alive will be censored at the date of last disease assessment. Subjects with no baseline or any postbaseline disease assessment will be censored at the date of randomization.

5.2.3. Sensitivity Analysis of PFS

Sensitivity analysis will be performed for PFS as follows:

1. Use of subsequent antineoplastic therapy prior to documented PD or death

Two alternative censoring rules will be performed for subjects who start subsequent antineoplastic therapy prior to first documentation of disease progression or death due to any cause:

a. subjects will be considered to have had a PFS event at the initiation of subsequent therapy.

b. subjects will be censored at the last disease assessment showing no evident of PD before the use of subsequent antineoplastic therapy.

2. Disease assessment follow up: Subjects will be censored at the last disease assessment if they progress or die after missing ≥2 planned disease assessment visits.

3. Non-stratified log-rank test and non-stratified Cox regression model will be conducted for PFS based on IRC assessment in ITT population. The reason for censoring will be summarized for PFS.

4. PFS based on PD assessed by investigators: For this sensitivity analysis, the same censoring rules and analysis methods used for primary analysis of IRC reported PFS will be used. The concordance rate between the IRC-determined PFS and investigator-determined PFS will be evaluated. The number and percentage of PD events and censored cases determined by investigator and by IRC will be cross-tabulated.

5.2.4. Subgroup Analysis of PFS

Subgroup analysis will be performed for the selected potential prognostic variables (as listed in Section 2.9) to assess the consistency and robustness of the treatment benefit for PFS. The non-stratified log-rank test analysis method for PFS will be used for each of the subgroup analysis. Median PFS with 95% CI, and hazard ratio between the two treatment groups within each subgroup and their 95% CI will be calculated using non-stratified Cox regression model. Subgroup analysis will be presented graphically in forest plot.

5.3. Secondary Endpoints

5.3.1. Overall Survival

Overall survival is defined as the time from date of randomization to date of death from any cause. For primary analysis of OS, all deaths documented in the study will be considered as events, including those that occur after the date of first dose of ibrutinib for crossover subjects.

Survival time of living subjects will be censored on the last date a subject is known to be alive or lost to follow-up.

Overall survival will be analyzed using the same analysis methods as used for primary analysis of PFS (Section 5.2.2).

The same subgroup analysis used for PFS will be performed for OS if the number of events within each subgroup is sufficient.

5.3.2. Sensitivity Analysis of Overall Survival

Sensitivity analysis will be performed for OS as follows:

1. Which censor subject at the time of crossover for placebo patients;
2. Which censor ibrutinib patients at subsequent therapy.

5.3.3. Overall Response Rate

Overall response rate is defined as the proportion of subjects achieving a best overall response of either CR, CRi, nPR, or PR as evaluated by IRC.

All randomized subjects with a measurable disease at baseline will be included in this analysis. Subjects with missing post-randomization data are considered non-responders.

Overall response rate will be estimated according to the crude proportion of confirmed responders (PR or better) based on the best overall response and summarize by treatment arm. Overall response rate will be compared between treatment arms using Cochran-Mantel-Haenszel (CMH) chi-square test, adjusted for the two randomization stratification factors. Logistic regression analysis will also be performed to estimate an odds ratio and its associated 95% CI between the 2 treatment groups.

Subjects who achieved PR by all parameters with the exception of reduction in ALC are considered PR with lymphocytosis (PRL). The ORR including PRL (per IRC assessment) will be summarized.

For overall response rate based on investigators' assessment, same analysis methods used for IRC reported overall response will be used.

Subgroup analysis for ORR may be performed if the number of responders within each subgroup is sufficient.

5.3.4. Rate of Minimal Residual Disease-Negative Response

Rate of MRD-negative response is defined as the proportion of subjects who reach MRD-negative disease status (<1 CLL cell per 10,000 leukocytes), as assessed by flow cytometry of a bone marrow aspirate/biopsy or peripheral blood sample. For subjects with both bone marrow aspirate/biopsy and peripheral blood sample available at the same visit, the bone marrow

aspirate/biopsy shall take precedence. All randomized subjects will be included in this analysis. Subjects with missing MRD data are considered non-responders. Chi-square test will be used for rate of MRD-negative response. Fisher's exact test may be used if the rate in any treatment group is small.

5.3.5. Sustained Hematologic Improvement

Sustained hematologic improvement is defined as hematological improvement that is sustained continuously for ≥56 days without blood transfusion or growth factors:

- Hemoglobin >11 g/dL if baseline ≤11 g/dL or increase ≥2 g/dL over baseline;
- Platelet counts >100 × 10^9/L if baseline ≤ 100 × 10^9/L or increase ≥50% over baseline;

Proportion of subjects achieving sustained hematological improvement will be summarized by treatment group. This proportion will be compared for the two treatment arms using chi-square test.

The analysis will be conducted on 1) ITT subjects 2) ITT subjects with cytopenia(s) at baseline. Subjects with cyctopenia(s) at baseline without any post-baseline hematologic assessment will be considered as having no improvement.

5.3.6.

5.3.7. Disease-Related Symptom Improvement

The most common disease-related symptoms associated with CLL/SLL include fatigue, weight loss, fevers, night sweats, and abdominal discomfort/splenomegaly. Disease-related symptoms will be assessed by investigator according to common terminology criteria for adverse events (CTCAE) criteria. Descriptive summary and shift tables will be provided at each scheduled time point.

5.4. Other Efficacy Endpoints

- Number of subjects with subsequent antineoplastic therapy will be summarized by therapy type. Listing of subjects with subsequent antineoplastic therapy will also be provided.
- Time-to-next therapy and time-on-next therapy will be analyzed for subjects who have started subsequent antineoplastic therapy. Time-to-next therapy is defined as the time from the date of randomization to the start date of any subsequent antineoplastic therapy. Time-on-next therapy is calculated as the time interval between the start of subsequent antineoplastic therapy and the end of subsequent antineoplastic therapy (progressive disease, death, or the start of next line of subsequent therapy, whichever is earlier). Data will be censored at the last visit date, for subjects still on the same subsequent therapy at the clinical cutoff.

- PFS2 will be analyzed for ITT subjects. It is defined as the time interval between the date of randomization and date of event, which is defined as progressive disease as assessed by investigator that starts after the next line of subsequent antineoplastic therapy (including cross-over to ibrutinib), death from any cause, or the start of the second subsequent antineoplastic therapy if no progressive disease is recorded. Those who do not receive subsequent therapy or do not experience the event as specified above are censored at the last disease assessment without second disease progression.
- Beta2-microglobulin, Serum immunoglobulin (IgA, IgG, and IgM) will be descriptively summarized.

6. SAFETY

Safety will be analyzed using the incidence and severity of AEs, laboratory tests, and electrocardiogram (ECG) measurements. Unless specified otherwise, all safety analyses will be based on the safety analysis set. Descriptive statistics will be reported for all safety data. Inferential statistics are not planned to be performed on safety data.

The baseline value for safety analysis is defined as the value collected at the time closest to and prior to the start of study medication.

Unless otherwise stated, safety data will be summarized by treatment arm as treated.

6.1. Adverse Events

Table 4. Summary of Adverse Event Analyses to be Performed

Category	Analysis	Sorted By	Cut off	Drug-Related TEAE
General	Overall summary			✔
	TEAEs	SOC+ PT+ toxicity grade; PT+ toxicity grade	10%, 5%	✔
	Serious TEAEs	SOC+ PT + toxicity grade; PT+ toxicity grade	2%	✔
	Grade 3 or worse TEAE	PT+ toxicity grade	2%	✔
	TEAEs leading to treatment discontinuation (ibrutinib/placebo, BR)	PT + toxicity grade		✔
	TEAEs leading to death	PT + toxicity grade		
	TEAEs leading to dose modification or modification (ibrutinib/placebo, BR)	PT + toxicity grade		

Category	Analysis	Sorted By	Cut off	Drug-Related TEAE
	AEs of clinical interest (Hemorrhagic events)	PT + toxicity grade		
	Other safety observations (e.g. other malignancies, eye disorder)	PT + toxicity grade		
	Deaths within 30 days of last dose	Reason for death		
	Atrial Fibrillation and Atrial Flutter: overall summary			
Subgroup	Overall summary			
	TEAEs	SOC+ PT+ toxicity grade		
Crossover Period	Overall summary			✔
	TEAEs	SOC+ PT+ toxicity grade		✔
	Serious TEAEs	SOC+ PT + toxicity grade		
Exposure adjusted incidence rate	Overall summary			
	TEAEs	SOC+ PT+ toxicity grade; PT+ toxicity grade	1 per 100 patient month	
	Serious TEAEs	SOC+ PT + toxicity grade; PT+ toxicity grade	0.1 per 100 patient month	

6.1.1. All Adverse Events

The verbatim terms used in the CRF by investigators to identify AEs will be coded using the Medical Dictionary for Regulatory Activities (MedDRA). These coded AE terms are referred to as preferred terms (PT); classification into System Organ Class (SOC) is a result of the coding process.

Treatment-emergent AEs (TEAEs) are defined as 1) those that occur in TEAE period as defined in Section 2.8; 2) present before first dose, but worsened in toxicity grade during treatment; 3) had missing start date and its end date is during the treatment; 4) was a drug-related event. Drug-related AEs are those assessed by investigator as being possible, probable or very likely related to study drug. To determine TEAE, partially missing AE start dates will be imputed according to the rules stated in Section 2.7.

For each TEAE, the percentage of subjects who experience at least 1 occurrence of the given event will be summarized. Tables will be sorted by frequency in incidence (the highest to lowest incidence).

For subjects using CYP3A inhibitor during treatment-emergent period, all TEAEs will be summarized by SOC, PT, maximum severity, any CYP3A inhibitor (Yes vs No), and strong CYP3A inhibitor (Yes vs No).

For subjects who received cross-over ibrutinib therapy, all TEAEs collected in the treatment-emergent period of cross-over ibrutinib therapy will be summarized.

6.2. Adverse Events of Clinical Interest and other safety observations

Hemorrhagic events will be identified by hemorrhage Standardized MedDRA Query [SMQ] excluding laboratory terms and be tabulated. Major hemorrhage is a subset of hemorrhagic events which are grade $\geq$ 3 or serious or belong to central nervous system (CNS) hemorrhage/hematoma.

Other malignancies: are defined as new malignant tumors including solid tumors, skin malignancies and hematologic malignancies and are to be reported for the duration of study treatment and during any protocol-specified follow-up periods including post-progression follow-up for overall survival.

Adverse events of interest and other safety observations will be summarized by treatment arms.

6.3. Deaths

A summary of the number of deaths during the treatment phase and up to 30 days after last dose will be provided, along with the primary cause of death. In particular, frequencies of deaths due to study treatment-related adverse events will also be reported. A death is study medication-related death if the primary cause is a drug related AE.

6.4. Exposure-Adjusted Incidence Rates

6.4.1. Restriction on the First Event

To adjust for unequal lengths of study treatment duration among subjects, and potentially between treatment groups, exposure adjusted incidence rate may also be summarized if the median treatment duration of one group is 20% longer than that of the other group.

The analysis restricts on the occurrence of the first event per subject and ignores the existence of later (multiple) events as these cannot be assumed to occur independent of previous events (e.g. subjects suffering from infections may have in general a higher risk of having other complications and may even have a higher risk of getting other infections). For this reason the exposure-adjusted incidence rate (EAIR) should be interpreted as 'rate until the first event

occurs'. Rates estimated from several subjects can be averaged on the level of a PT, of an SOC, or on a global level.

6.4.2. Duration of Exposure: Censored & Non-censored

The EAIR for a subject is derived from the duration of treatment exposure for that subject. When averaging incidence rates, a subject's duration of exposure is given either by

a. time to the earliest onset of the TEAE (non-censored data), or
b. total duration of treatment, for subjects who did not experience the AE in question (censored data).

Depending on whether a subject has an AE or not, the duration of exposure enters the denominator in its non-censored or censored form, respectively.

6.4.3. Incidence Rate per Subject

The incidence rate for a specific event of a subject is the reciprocal of time t when the first event occurs:

$$EAIR_i = \frac{1}{t_i}.$$

6.4.4. Average EAIR per Preferred Term

The *EAIR* for a specific PT is an average over all subjects as described before, i.e.

$$EAIR_{PT} = \frac{\sum_{i=1}^{n} TEAE_{PT,i}}{\sum_{i=1}^{n} t_{PT,i}},$$

whereby

a. the TEAE enters the sum in the nominator unweighted ($TEAE_i$ =1, otherwise $TEAE_i$ =0), and
b. the duration of exposure enters the denominator as:

$$t_i = \begin{cases} time\ of\ TEAE \text{ if occurring (non-censored data)} \\ total\ duration\ of\ treatment \text{ if no event occurs (censored data)} \end{cases}$$

6.4.5. Average EAIR per SOC

The average *EAIR* per SOC considers the first event per subject per SOC only, and only one (the corresponding) exposure time in the denominator:

$$EAIR_{SOC} = \frac{\sum_{i=1}^{n} TEAE_{SOC,i}}{\sum_{i=1}^{n} t_{SOC,i}},$$

Note: This EAIR is an incidence rate *per SOC*.

6.4.6. Average EAIR on a Global Level

The average *EAIR* on a global level considers the overall first event per subject only, and only one (the corresponding) exposure time in the denominator:

$$EAIR_{global} = \frac{\sum_{i=1}^{n} TEAE_i}{\sum_{i=1}^{n} t_i},$$

whereby $TEAE_i$ represents the first TEAE among all TEAEs of subject i and t_i as before (time when TEAE occurs (non-censored data) or total duration of treatment if no event occurs (censored data).

6.5. Clinical Laboratory Tests

Laboratory data of hematology and serum chemistry up to 30 days after last dose or the end of treatment visit date, whichever is later, will be reported in SI units.

For hemoglobin, absolute neutrophil count (ANC), and platelets, toxicity will be assessed by the grading scale for hematologic toxicity in CLL studies in the IWCLL 2008 guidelines. Other laboratory results will be graded according to NCI-CTCAE version 4.03. Note that toxicity grading for creatinine increase will be based on the NCI CTC v4.03 criteria, but limited only to the part based on the upper limit of normal (ULN), the other part, that is based on change from baseline, will not be used for toxicity grading. Generic normal ranges will be applied whenever reference ranges are not available.

The following laboratory tests will be analyzed:

- Hematology: hemoglobin, white blood cell (WBC), ANC, absolute lymphocyte count (ALC), and platelets
- Coagulation (screening only): Prothrombin international normalized ratio (INR), activated partial thromboplastin time (aPTT)
- Chemistry: alkaline phosphatase, ALT, AST, bilirubin (total), creatinine, creatinine clearance (CrCl), LDH, magnesium, potassium, sodium, and uric acid

Descriptive statistics (mean, standard deviation, median and range) will be calculated for the raw data and for their changes from baseline at each time point of assessment as well as for the changes from baseline to the last value. Parameters will be summarized by toxicity grade. Change from baseline to the worst grade during the treatment will be provided as shift tables for selected parameters. In addition, treatment-emergent worsening in toxicity grade will be summarized for selected hematology and chemistry parameters.

Liver function abnormality by Hy's Law: For subjects with any elevated AT (AST or ALT) of >3xULN, ALP <2xULN, and associated with an increase in bilirubin (total) ≥2xULN, a listing

for all subjects with all such records will be provided and a summary table of the number of such subjects will be provided by treatment arm.

The frequencies of abnormal treatment emergent uric acid will be summarized by treatment arm.

6.5.1. Creatinine Clearance

Creatinine clearance (CrCl) is calculated using the Cockroft-Gault formula:

$$CrCl_{(est)} = \frac{(140 - age[yr])(lean\ body\ wt[kg])}{(72)(serum\ creatinine[mg/dL])} \times 0.85 (if\ female)$$

for males, the factor is 1 instead of 0.85

6.5.2. Analysis of Lymphocytosis

For all subjects with baseline and any post-baseline ALC measurements, a summary of peak ALC values will be provided by treatment arm.

Lymphocytosis is defined as an elevation in ALC of ≥50% compared to baseline and to $\geq 5x10^9$/L (5,000/μL) at a post-baseline assessment. The number of subjects with at least one occurrence of lymphocytosis will be summarized. For subjects with lymphocytosis, 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 $5x10^9$/L (5,000/μL), whichever occurs first.

The following analyses will be conducted for subjects with lymphocytosis by treatment arm:

- ALC at peak and time to peak ALC within the time period from the first dose of study drug to 9 months of first study treatment (Study day 270), 30 days following the last dose of study drug or initiation of subsequent antineoplastic therapy, whichever the earliest,
- Time to lymphocytosis (TTL) is defined as the time from first dose date of study treatment to the first post-baseline ALC which meets the lymphocytosis criteria and will be summarized descriptively for subjects who developed lymphocytosis.
- 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. Subjects who have developed lymphocytosis but not recovered will be censored at the last available ALC measurements at or prior to initiation of subsequent antineoplastic therapy.

6.6. Electrocardiogram

QT prolongation and other clinically significant ECG abnormalities will be summarized and listed. Descriptive statistics will be calculated for the ECG parameters at baseline.

The ECG parameters that will be summarized are heart rate, RR interval, QT interval, and QTc. The QTc will be computed using Bazett's correction, i.e., QTcB = $QT/\sqrt{RR}$, and Fridericia's correction, i.e., QTcF = $QT/\sqrt[3]{RR}$.

All treatment-emergent abnormal findings will be tabulated, displaying the number of subjects with abnormal findings after dosing up to the end (Day 28) of the last cycle. An abnormal finding is considered to be treatment-emergent if it occurred during treatment and up to 30 days after the last dose.

7. PATIENT-REPORTED OUTCOMES

Patient-reported outcomes will be measured by 4 questionnaires as described below. Refer to the Study PCI-32765CLL-3001 protocol for samples of PRO scales and data collection schedule.

7.1.1. EORTC QLQ-C30

EORTC-QLQ-C30 includes 30 separate items resulting in 5 functional scales (physical functioning, role functioning, emotional functioning, cognitive functioning, and social functioning), 1 Global Health Status scale, 3 symptom scales (fatigue, nausea and vomiting, and pain), and 6 single items (dyspnea, insomnia, appetite loss, constipation, diarrhea, and financial difficulties). Scores will be derived using validated scoring algorithms according to EORTC QLQ-C30 Scoring Manual (3rd Edition, 2001). EORTC QLQ-C30 improvement/worsening is defined as ≥10 points for each assessment post baseline.

7.1.2. EORTC QLQ-CLL 16

The EORTC QLQ-CLL 16 is a 16-item disease-specific module that comprises 5 domains of patient-reported health status important in CLL. There are three multi-item scales that include fatigue (2 items), treatment side effects and disease symptoms (8 items), and infection (4 items), and 2 single-item scales on social activities and future health worries.

7.1.3. EQ-5D-5L

The EQ-5D is a 5-item questionnaire and a "thermometer" visual analogue scale ranging from 0 (worst imaginable health state) to 100 (best imaginable health state). The scores for the 5 separate questions are categorical and should not be analyzed as ordinal numbers. EQ-5D improvement/worsening is defined as ≥7 points for each assessment post baseline.

The scores for the 5 categorical dimensions will be used to compute a single utility score ranging from zero (0.0) to 1 (1.0) representing the general health status of the individual. The United Kingdom weights will be used to generate patient utilities from the 5 dimensions of the EQ-5D in this study.

7.1.4. FACIT-Fatigue

The FACIT-Fatigue Scale measures fatigue severity and impact on daily activities. It includes 13 items that assess tiredness, weakness and difficulty conducting usual activities due to fatigue. Scores range from 0 to 52 with high scores indicating less fatigue. FACIT improvement/worsening is defined as ≥3 points change for each assessment post-baseline.

7.1.5. Exploratory analyses with Patient-reported outcome Questionnaires

In addition to the planned analyses on the domains and scales from each of the patient-reported outcome questionnaires, exploratory analyses at the individual item level will be conducted to better understand domains that show significant effects or trends toward significance. Since the EORTC QLQ-CLL 16 is disease-specific, it is anticipated that most of the exploratory individual item analyses will be conducted on that questionnaire, including, but not limited to the following items: "Did you have night sweats?", "Did you feel lethargic?", "Have you felt 'slowed down'?", "Were you limited in planning activities, for example meeting friends, in advance?". However, individual item analyses may also be conducted on the EQ-5D-5L, EORTC QLQ-C, FACIT-Fatigue, depending upon the pattern of findings at the domain or scale level. In each case where an individual item analysis is conducted, the analytic approach will replicate that described for the scale and domain scores.

7.1.6. Analysis Methods

Descriptive statistics (number of observations, mean, standard deviation, median, minimum, maximum) at baseline and each post-baseline time point is reported by treatment group for FACIT-Fatigue total score, each scale of EORTC QLQ-C30 and EORTC QLQ-CLL 16, and EQ 5D visual analogue scale and weighted utility score. These descriptive statistics will also be summarized by the response categories contributing to the ORR. A responder analysis and presentation by minimal clinical difference will also be provided.

Compliance

Compliance rates defined as the number of questionnaires received as a percentage of the number expected per protocol-specified collection schedule are assessed for FACIT-Fatigue, EORTC QLQ-C30, EORTC QLQ-CLL 16, and EQ-5D.

Mixed Model for Repeated Measures

For each of the PRO measures of interest, a mixed-effects model with repeated measures (MMRM) analysis is conducted estimating change from baseline at each time point. ITT subjects who have a baseline value and at least 1 post-randomization value are included in the analysis.

For each scale, change from baseline is fitted to a mixed-effects model including subjects as a random effect, and baseline value, treatment, time, treatment-by-time interaction, and randomization stratification factors as fixed effects.

Time to Improvement and Deterioration Analysis

Using a threshold value for improvement and deterioration specific to the PRO measures of interest, time to improvement and time to deterioration will be compared between treatment groups in ITT subjects. Subjects who have not experienced event at the time of analysis will be censored on the last known date without event. Subjects with no on-study assessment or no baseline assessment will be censored at date of randomization.

Distributions of time-to-event variables will be estimated using the Kaplan-Meier product-limit method. Median times to event with 2-sided 95% confidence intervals will be estimated. The stratified log-rank test accounting for stratification factors will be used for treatment comparison. A stratified Cox proportional-hazards model will provide estimates of hazard ratios with 95% confidence intervals.

Analysis Visits

To determine the scheduled time points, all PRO scores are to be assigned to a particular time window for a scheduled time point based on the rules presented in Table 5. In the case that more than 1 score is found with a time window, the score closest to the window center will be used in the analysis. In the case that there are 2 values that are equidistant from the center, the value prior to the center will be used.

Details for target date for each scheduled visit and time interval between scheduled visits are provided in the table below:

Table 5. Visit Windows for PRO Assessments

EORTC QLQ-C30 and EQ-5D-5L

Analysis visit (cycle)	Start	Target	End
Baseline	-30	1	1
Cycle 3	2	57	85
Cycle 5	86	113	141
Cycle 7	142	169	211
Cycle 10	212	253	295

EORTC QLQ-CLL 16 and FACIT-Fatigue

Analysis visit (cycle)	Start	Target	End
Baseline	-30	1	1
Cycle 2	2	29	57
Cycle 4	58	85	113
Cycle 6	114	141	169
Cycle 8	170	197	225
Cycle 10	226	253	295

All Scales

Analysis visit (cycle)	Start	Target	End
Cycle 13	296	337	379
Cycle 16	380	421	463
Cycle 19	464	505	547
Cycle 22	548	589	631
Cycle 25	632	673	715
Cycle 28	716	757	799
Cycle 31	800	841	883
Cycle 34	884	925	967
Cycle 37	968	1009	1051

Each cycle is 28 days

8. PHARMACOKINETICS AND PHARMACODYNAMICS ANALYSIS

Details of analysis plan in PK/PD and results will be presented in a separate report.

9. BIOMARKER ANALYSIS

Details of analysis plan in biomarker and results will be presented in a separate report.

10. Changes to Protocol Specified Analyses

The following analyses are different from those described in the protocol:

- Hemoglobin >11 g/dL if baseline ≤11 g/dL or increase ≥2 g/dL over baseline;
- Platelet counts >100 × 10^9/L if baseline ≤ 100 × 10^9/L or increase ≥50% over baseline;
- The term used for the secondary objective MRD-negative remissions is changed to MRD-negative response based on clinical interpretation.
- The definition for major hemorrhage is updated in alignment with the related post marketing commitment on major bleeding and allows consistent analysis for the ibrutinib clinical development program. The updated definition is based on searching and subtyping grade ≥ 3 or serious or central nervous system (CNS) hemorrhage/hematoma by hemorrhage SMQ excluding laboratory terms. The following types of events in protocol definition were removed because they are grade ≥ 3 or serious hemorrhage events according to NCI CTCAE criteria: intraocular bleeding causing loss of vision, the need for a transfusion of two or more units of red cells or an equivalent amount of whole blood, hospitalization or prolongation of hospitalization. Intracranial hemorrhage in protocol is included in the CNS hemorrhage/hematoma as a component of major hemorrhage.

11. REFERENCES

1. Fischer K, Cramer P, Busch R, et al. Bendamustine combined with rituximab in patients with relapsed and/or refractory Chronic Lymphocytic Leukemia: A multicenter phase II trial of the German Chronic Lymphocytic Leukemia Study Group. J Clin Oncol. 2011;29:3559-3566.
2. O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. Biometrics. 1979;35(3):549-556.
3. Hallek M, Cheson BD, Catovsky D, 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; 111:5446-5456.
4. Branson M and Whitehead J. Estimating a treatment effect in survival studies in which patients switch treatment. Stat Med 2002; 21:2449-2463.

12. Appendix 1: Additional Exploratory Analysis to Support HEMAR

HEMAR analyses will be presented in a separate report.

12.1. Subgroup Definition

The dataset will also be analyzed by a revised definition of refractory and relapse whereby refractory is defined as disease progression ≤ 6 months and relapse is defined as disease progression >6 months.

Based on this definition the below subgroup analysis will be performed for selected variables as described in sections 5.2 for the primary endpoint; 5.3 for secondary endpoints; section 6 for safety endpoints and section 7 for patient-reported outcomes.

Subgroup	Definition of Subgroup	Analysis Type
Age	<65, ≥65	D, E, S
Sex	Male, Female	D, E, S
Race	White, Non-White	D, E, S
Geographic region	North America, Europe, Latin America, and Asia	D, E, S
Diagnosis	CLL, SLL	D, E, S
Rai Stage at screening	Stage 0-II, III-IV	E
Failure to respond to a purine analogue or a recurrence of disease after purine analog therapy	No prior PA, within12 months, ≥12 months	E
Baseline ECOG	0, 1	E
Bulky Disease	Yes (LDi ≥5 cm), no (LDi <5 cm),	E
Chromosome 11q Deletion	Yes, no	E
Elevated LDH at Baseline	No (<350 U/L), Yes (≥350 U/L)	E
Cytopenias at Baseline	Yes, no	E
Serum β2 –microglobulin	≤3.5 mg/L, > 3.5 mg/L	E
IgVH	Mutated, Unmutated	E
ZAP-70	Elevated, Not Elevated	E
CD38	Positive (≥ 30), negative (<30)	E
complex karyotype	Yes, no	E
Refractory to purine analog therapy as recorded in IWRS	Yes, no	E
Prior lines of therapy as recorded in IWRS	1, >1	E
Concomitant use of any CYP3A inhibitor	Yes, No	S
Concomitant use of strong CYP3A inhibitor	Yes, No	S

Subgroup	Definition of Subgroup	Analysis Type
Cytopenia is defined as yes if platelet count ≤ 100,000/uL, Hgb ≤ 11g/dL, or ANC ≤ 1500/uL is observed. analysis type D= demographic and baseline disease characteristics analysis type E= efficacy (PFS, OS, ORR) analysis type S= safety (adverse events) ECOG= Eastern Cooperative Oncology Group; IgVH= immunoglobulin variable heavy gene; IWRS= Interactive Web Response System; LDH= lactic acid dehydrogenase; LDi=longest diameter.		

For all p-interaction values ≤0.20 detailed information on each endpoint will be provided.

12.2. Further endpoints for analyses (OR, RR, RD)

- Disease-related symptoms that started or worsened from the first dose
    - Fatigue
    - Weight loss
    - Anorexia
    - Night Sweats
    - Abdominal Pain
    - Fever

12.2.1. Subgroups for further endpoints from market access analysis

Endpoints of interest for market access will be analyzed also by pre-specified subgroups as described below:

Subgroup	Definition of Subgroup	Analysis Type
Age	<65, ≥65	D, E, S
Sex	Male, Female	D, E, S
Race	White, Non-White	D, E, S
Geographic region	North America, Europe, Latin America, and Asia	D, E, S
Diagnosis	CLL, SLL	D, E, S
Rai Stage at screening	Stage 0-II, III-IV	E

Subgroup	Definition of Subgroup	Analysis Type
Refractory to prior purine analogue therapy	Refractory to prior purine analogue therapy, Yes (defined as no response or failure within12 months), No Refractory to prior purine analogue therapy (Response lasting ≥12 months)	E
Baseline ECOG	0, 1	E
Bulky Disease	Yes (LDi ≥5 cm), no (LDi <5 cm),	E
Chromosome 11q Deletion	Yes, no	E
Elevated LDH at Baseline	No (<350 U/L), Yes (≥350 U/L)	E
Cytopenias at Baseline	Yes, no	E
Serum β2 –microglobulin	≤3.5 mg/L, > 3.5 mg/L	E
IgVH	Mutated, Unmutated	E
ZAP-70	Elevated, Not Elevated	E
CD38	Positive (≥ 30), negative (<30)	E
complex karyotype	Yes, no	E
Refractory to purine analog therapy as recorded in IWRS	Yes, no	E
Prior lines of therapy as recorded in IWRS	1, >1	E
Concomitant use of CYP3A inhibitor	Yes, No	S
Concomitant use of strong CYP3A inhibitor	Yes, No	S
Cytopenia is defined as yes if platelet count ≤ 100,000/uL, Hgb ≤ 11g/dL, or ANC ≤ 1500/uL is observed. analysis type D= demographic and baseline disease characteristics analysis type E= efficacy (PFS, OS, ORR) analysis type S= safety (adverse events) ECOG= Eastern Cooperative Oncology Group; IgVH= immunoglobulin variable heavy gene; IWRS= Interactive Web Response System; LDH= lactic acid dehydrogenase; LDi=longest diameter.		

12.3. Cross-over correction for the Overall Survival Endpoint

The following constitutes an example for a IPCW-correction on the primary endpoints overall survival (OS) and progression-free survival (PFS).

Database and variables

Database with 485 subjects providing information on overall and progression-free survival with variables

1. TIME_TO_DEATH, censor variable STATUS_DEATH
2. TIME_TO_PFS, censor variable STATUS_PFS

Covariates considered by the IPCW approach:

1. AMLtype (Acute Myeloid Leukemia)
2. ECOG (Cooperative Oncology Group performance status)

3. AGE (Age)
4. BMBLAST (Bone Marrow Blasts, % of abnormal leukemia cells)
5. RACE
6. GENDER
7. CGRISK (Cytogenetic risk, based on chromosomal abnormalities)
8. TRTMNT (2 Treatment arms)

Tables delivered

- Table 2.1: Overall survival, ITT analysis set
- Table 2.2: Progression-free survival, ITT analysis set

IPCW approach and modification

Step 1: Create a panel data with monthly records per patient + new variable: cross over y/n per record. The panel data contain follow-up times from randomization until failure (death) or informative censoring (crossover) in monthly intervals.
Step 2: Multiple logistic regression: model the probability to cross over dependent on variables assumed to cause informative censoring (see Section 0). The inverse of the prediction yields the weights for Step 3.
Step 3: Weighted Cox regression (alternatively a parametric approach: logistic, exponential, …)

Applying the IPCW method as described by Robins and Finkelstein (2000) on the present data is complicated by the following two reasons:

- The assumption of proportional hazards required by the Cox regression (Step 3) is partially negated in the present data because the survival curves of the treatment arms cross (see CSR Figures 3 to 7, p. 90 and following pages).
- Computing CIs from methods involving sample size inflating procedures like inverse probability weighting requires sophisticated approaches like, for instance, bootstrap procedures.

Considering these two points the original IPCW approach has been modified in Step 3 by replacing the weighted Cox regression by a weighted logistic regression modeling the time-dependent probability of OS and PFS. This corresponds to a parametric approach based on the logistic model as the regression is applied on the panel data. Thus, the modified IPCW approach presented here involved the following 3 steps:

Step 1: Create a panel data with monthly records per patient + new variable: cross over y/n per record.
Step 2: Multiple logistic regression to estimate the time-dependent probability to cross over dependent on covariates AMLtype, ECOG, AGE, BMBLAST, RACE, GENDER, CGRISK, and TRTMNT.
Step 3: Logistic regression on panel data yielding a time-dependent odds ratio, interpreted as a hazard ratio.
Step 4: The results of logistic regression, Step 2, are

Effect	Estimate	Lower 95%	Upper 95%

		CL	CL
AMLTYPE [de novo vs secondary]	0.779	0.719	0.845
ECOG [0-1 vs 2]	1.491	1.359	1.637
AGE	0.922	0.915	0.929
BMBLAST [<= 30 vs >30]	0.915	0.839	0.998
RACE [non-white vs white]	1.189	1.064	1.328
SEX [female vs male]	1.320	1.221	1.426
CGRISK [intermediate vs poor]	1.456	1.342	1.580
month	1.059	1.055	1.063
Treatment [Dacogen vs TC]	0.788	0.730	0.850

Conclusions & recommendations

The hazard ratios of 0.702 and 0.659 for OS and PFS, respectively, and their significant CIs resulting from the modified IPCW analysis performed here can support the hypothesis of a treatment benefit in favor of Dacogen with respect to OS and PFS. This, however, should be considered in light of the following two restrictions, and leads to respective recommendations:

Possibility of overcompensation by weights

The weights resulting from Step 2 (logistic regression) adopt values of up to 60, as already described in the provided master thesis (see above, Section). This might revert the original intention to compensate for informative censoring into the opposite, i. e. into overcompensation of informative censoring, as shown for Weibull distributed survival times.
The present approach has reduced the risk of potential overcompensation by replacing the Cox regression by a logistic regression applied on panel data. The risk of overcompensation is reduced here because the weights modify the dispersion parameter of the binomial model underlying the logistic regression (in contrast to the Cox regression where the weights modify the location of the survival curves).
Recommendation: The present analysis could be complemented by non-parametric methods (e. g. Rank preserving structural failure time model, RPSFT) which do not depend on distribution assumptions and are less affected by weighting bias.

Modification of IPCW approach by parametric analysis

The Cox regression within the IPCW approach performed here (Step 3) has been replaced by a logistic regression. This is acceptable for the reasons outlined above. However, in contrast to the non-parametric Cox regression, the logistic regression is a parametric method based on the assumption that OS and PFS times follow a logistic distribution. Consequently, in the event that the logistic distribution is not the appropriate model for the distribution for times of OS and/or PFS, the present analysis has limited validity. On the other hand, as non-parametric methods based on a proportional hazards assumption also have limitations with respect to the present data (e.g. crossing of survival curves), the methodological spectrum for the current data set becomes narrow.

Recommendation: If a parametric approach should be avoided the aforementioned RPSFT method could become relevant for the analysis of these data in a twofold manner.

Final Statement

The IPCW method employed in this analysis is a powerful procedure for adjusting for non-informative censoring (or switching of treatments), it is, however, not without its limitations. Nonetheless, the analysis appears to be supportive in the hypothesis that the true treatment effect is better than that observed in the ITT analysis.