pharmacyclics
An AbbVie Company

Statistical Analysis Plan Addendum for Final CSR

A Randomized, Double-Blind, Placebo-Controlled, Phase 3 Study of Ibrutinib or Placebo in Combination with Rituximab in Subjects with Waldenström's Macroglobulinemia

PCYC-1127-CA

21 October 2019

Statistical Analysis Plan Approval

Protocol Number:	PCYC-1127-CA
Protocol Title:	A Randomized, Double-Blind, Placebo-Controlled, Phase 3 Study of Ibrutinib or Placebo in Combination with Rituximab in Subjects with Waldenström's Macroglobulinemia
SAP Version:	1.0
Date:	21 October 2019

By signing below, all parties accept that the analysis methods and data presentations are acceptable and that this document is final.

LIST OF ABBREVIATIONS

Abbreviation	Definition
AE	adverse event
CSR	clinical study report
ITT	intent-to treat
PD	progressive disease
PFS	progression-free survival
TEAE	treatment-emergent adverse events

Table of Contents

1. INTRODUCTION

This SAP is to describe the final analyses of data in evaluation of efficacy and safety of Study PCYC-1127-CA to support the Final CSR. The final analyses will be based on the final data extract. This SAP describes the additional analyses to the primary CSR SAP, dated 13 October 2017. All analyses that were performed previously for the primary CSR will follow the primary CSR SAP and will not be described again in this document. The additional analyses to support the Final CSR will be described in Section 2 (see summary in appendix Table 1).

2. ADDITIONAL ANALYSES

2.1. Progression-Free Survival after Subsequent Anti-neoplastic Therapy (PFS2)

Progression-free survival after subsequent anti-neoplastic therapy will be defined for all subjects in the ITT population as the duration of time from date of randomization to the date of the earliest of the following 3 types of events:

- PD per investigator response assessment after administration of the first subsequent anti-neoplastic therapy
- Death at any time on study due to any cause
- Initiation of a second subsequent anti-neoplastic therapy

For subjects randomized to placebo + rituximab arm, ibrutinib as next-line therapy will be considered as a subsequent anti-neoplastic therapy in the analysis of PFS2. Subjects who did not experience a PFS2 event will be censored at the later date of last available survival status and last adequate assessment. Subjects with no available survival status information and no adequate assessment after initiation of first subsequent anti-neoplastic therapy, will be censored at date of first subsequent anti-neoplastic therapy + 1.

Subjects without a baseline assessment, or with no adequate post baseline assessment and did not start first subsequent anti-neoplastic therapy, will be censored at the date of randomization.

The same methods used to analyze PFS will be used to analyze PFS2.

2.2. Prevalence of Treatment-Emergent Adverse Events (TEAEs)

The prevalence of TEAE will be analyzed for all combined arms with planned treatment of ibrutinib (Arms A+C). The summaries of disposition, demographic and baseline characteristics for Arms A+C will be presented.

The summaries of exposure, TEAEs, and laboratory data will be presented by ibrutinib exposure intervals of 0-1 year, 1-2 years, 2-3 years, 3-4 years, 4-5 years, and >5 years (may be adjusted based on the actual data to ensure there are relatively enough subjects in the highest yearly interval) and overall.

- The exposure summary table will be presented by the actual time on ibrutinib treatment during each yearly interval.

- The TEAEs summaries (excluding TEAEs leading to deaths and other malignancies) will be presented by ibrutinib treatment duration yearly intervals and included an additional period of up to 30 days after the last dose following discontinuation of ibrutinib to cover the treatment-emergent period.

- All TEAEs leading to deaths will be presented by yearly intervals of study duration.

- Other malignancies will be presented by yearly intervals of study duration, which includes time on study to capture events that occurred after discontinuation of ibrutinib (i.e. beyond the 30-day window).

Note that the lower limit of the yearly interval will be included in the previous interval and the upper limit will be included in the current interval (e.g., a TEAE, laboratory finding, or exposure duration ending at exactly 4 years will be included in the yearly interval 3- 4 years but not the interval 4-5 years). For the determination of period prevalence, TEAEs still ongoing from the previous exposure period year will be included in addition to newly occurring TEAEs in the respective time period. Multiple events of the same preferred term (PT) within the same yearly interval for a subject will be counted only once. Prevalence of a specific TEAE during an interval will be computed from the ratio of all subjects with reported events (continuing from previous yearly interval or new onset in the interval) and the number of subjects whose TEAE period is covering the yearly interval (the end of the TEAE period is later than the beginning of the interval).

Prevalence rates of TEAEs in the overall column will be computed from all subjects who reported at least one occurrence of the specific event during the study.

The plots of diastolic blood pressure and systolic blood pressure over the time for subjects with serious hypertension will be provided.

2.3. Overall Survival (OS) after crossover to Next-line Ibrutinib for Crossover Subjects

OS after crossover to ibrutinib is defined from date of crossover to ibrutinib to the date of death due to any cause. Subjects who are alive or exited the study with unknown death status will be censored at the last date known to be alive. OS distribution is estimated by Kaplan-Meier method. Median OS and landmark estimates will be provided.

APPENDIX

Table 1 Summaries of Data and Statistical Methods

Parameters	Definitions and Statistical Methods
Demographics and baseline Characteristics	Information obtained at study start. Characteristics included age, gender, race, ethnicity, geographic region, weight, height, BSA, number of prior treatment lines, estimated creatinine clearance rate and other baseline characteristics.
Subject disposition	The number of subjects continuing ibrutinib treatment (rollover) and primary reasons for discontinuation of treatment for subjects who discontinued ibrutinib will be provided. Time on study will be summarized based on follow-up time of overall survival, and the median time on study will be estimated using Kaplan-Meier methodology. Information on ibrutinib treatment duration will be presented; the number and percentage of subjects with ibrutinib treatment over time will be summarized in 1-year intervals. Additionally, subject disposition will be summarized for subjects who will be included in each yearly interval.
Exposure to ibrutinib	Exposure will be summarized by treatment duration, cumulative dose received, and average daily dose during each yearly interval. The number of subjects continuing ibrutinib treatment in each yearly interval will be also summarized.
PFS2 assessed by investigator	Duration from date of randomization to the date of the earliest of the following 3 types of events: • PD per investigator response assessment after administration of the first subsequent anti-neoplastic therapy • Death at any time on study from any cause • Initiation of a second subsequent anti-neoplastic therapy The same methods used to analyze PFS will be used to analyze PFS2.
OS for crossover	OS after crossover to ibrutinib is defined from date of crossover to ibrutinib to the date of death due to any cause. Subjects who are alive or exited the study with unknown death status will be censored at the last date known to be alive. OS distribution is estimated by Kaplan-Meier method. Median OS and landmark estimates will be provided.
Treatment-emergent adverse events	All reported AEs satisfying either of the following criteria: • Those events that occurred or worsened after the first dose of ibrutinib, through the treatment phase, and within 30 days following the last dose of ibrutinib or initiation of subsequent antineoplastic therapy, whichever occurred first. • Events with a missing onset date with a resolution date during the ibrutinib treatment phase; or • Events that were considered study drug-related regardless of the start date of the event. All available TEAEs will be summarized in this report. Related TEAEs reported after the TEAE period will not be included in the ibrutinib exposure yearly intervals because subjects were no longer on ibrutinib treatment, but they will be included in the overall column in the TEAE tables. The prevalence of safety data will be summarized by exposure period and included TEAEs, Grade 3 or higher TEAEs, related TEAEs, SAEs, TEAEs leading to deaths, TEAEs leading to treatment discontinuation for subjects who discontinued treatment due to TEAEs. Summary of TEAEs leading to discontinuation of ibrutinib will be counted at the yearly interval when subjects discontinued ibrutinib. Treatment-emergent AEs leading to death will be counted at the yearly interval when the subject died and will be presented by study-duration yearly intervals to capture events that occurred after discontinuation of ibrutinib.
Hemorrhage	Hemorrhagic events will be identified and summarized by hemorrhage SMQ excluding laboratory terms. Major hemorrhage is a subset of hemorrhagic events including the following: Grade 3 or higher nonserious TEAEs, SAEs, and TEAEs (without regard to grade or seriousness criteria) representing CNS hemorrhage. The prevalence of hemorrhage and major hemorrhage TEAEs will be provided.

Hypertension	The prevalence of TEAEs, Grade 3 or higher, and serious hypertension reported after the first dose of ibrutinib by yearly intervals will be provided. Hypertension events will be summarized by hypertension SMQ narrow. A summary of incidence of new onset for hypertension events within each yearly interval will also be provided. Multiple occurrences of the same events within the same yearly interval for a subject will be counted only once at the first onset of the event. The number of subjects at risk for each interval will include all subjects whose TEAE period was covering the yearly interval (the end of the TEAE period was later than the beginning of the interval). The number of subjects with and without TEAE hypertension will be summarized by history of hypertensive disorders/conditions or other relevant risk factors and by age (≥65 and <65 years old). Summaries of time to onset of hypertension Grade 3 and higher and serious events will be provided. Spaghetti plots of over time will be provided for subjects who had serious events of hypertension.
Second Primary Malignancies (Referred to as "Other Malignancies" for the remainder of this analysis)	The prevalence of other malignancies by PT will be provided for both the TEAE period by yearly intervals of ibrutinib exposure and for other malignancies reported beyond 30 days of last ibrutinib dose. For other malignancies, new malignancies will be obtained from the PTs under the categories non-melanoma skin cancer, melanoma skin cancer, and non-skin cancer. Non-skin cancer excludes underlying disease or progressive disease malignancies. Due to their long duration and infrequent resolution, other malignancies will also be reported by incidence of new onset of events within each yearly interval. Multiple occurrences of the same events within the same yearly interval for a subject will be counted only once at the first onset of the event. The number of subjects at risk for each interval include all subjects whose TEAE period is covering the yearly interval (the end of the TEAE period is later than the beginning of the interval).
Cardiac arrhythmias	For cardiac arrhythmias, the following grouped terms for prevalence will be used: PTs for atrial fibrillation and atrial flutter combined as well as the narrow SMQ for ventricular tachyarrhythmias. A summary of the prevalence of TEAEs and Grade 3 or higher TEAE by yearly interval will be provided.
Interstitial lung disease	Interstitial lung disease events will be identified by ILD SMQ-narrow. A summary of the prevalence of TEAEs and Grade 3 or higher TEAEs by yearly interval will be provided.
Clinical laboratory test	The summaries of laboratory data will be presented by ibrutinib exposure intervals of 0-1 year, 1-2 years, 2-3 years, 3-4 years, 4-5 years, and >5 years (may be adjusted based on the actual data to ensure there are relatively enough subjects in the highest yearly interval) and overall.

pharmacyclics®
An AbbVie Company

Statistical Analysis Plan

A Randomized, Double-Blind, Placebo-Controlled, Phase 3 Study of Ibrutinib or Placebo in Combination with Rituximab in Subjects with Waldenström's Macroglobulinemia

PCYC-1127-CA

13 October 2017

Version 2.0

Confidentiality Statement

Revision History

Version Number	Date	Description
1.0	20Apr2016	Original version
2.0	13Oct2017	Amendment summary: This version applies the latest SAP template; structure and contents are updated to align with other studies. Below are the major changes made: 1. Remove the section for objectives of study and add a section for endpoints; remove the section for study dosage and administration and include the dosing information under the section for study design; remove the detailed information in the section for blinding and randomization method and refer to the protocol. 2. Add one secondary endpoint for patient reported outcomes (PRO). Update the ordering of secondary endpoints. Add texts indicating that overall survival will not be tested at interim analysis and will be tested at study closure. 3. Update the subgroup definition table for subgroup analyses. 4. Add one sensitivity analysis for progression-free survival and subgroup analysis for overall response rate. 5. Add hemoglobin, tumor involvement of bone marrow, and lymph nodes and spleen assessment as exploratory endpoints. 6. Add handling of missing data for PRO analysis and exploratory analyses for PRO data. 7. Add the following adverse events: hypertension, interstitial lung disease, severe cutaneous adverse reactions, rash, cardiac arrhythmia excluding atrial fibrillation, infusion related reaction. 8. Remove exposure-adjusted incidence rates.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Definition
AE	adverse event
CMH	Cochran-Mantel-Haenszel
CR	complete response
CRF	case reported form
CRR	clinical response rate
CTCAE	Common Terminology Criteria for Adverse Events
DMC	data monitoring committee
DOR	duration of response
ECOG	Eastern Cooperative Oncology Group
IPSS	International Prognostic Scoring System
IRC	independent review committee
ITT	intent-to treat
IWWM	International Workshop on Waldenström's macroglobulinemia
IWRS	Interactive Web Response System
MMRM	Mixed-Effects Model Repeated Measure
MR	minor response
MRU	medical resource utilization
NCCN	National Comprehensive Cancer Network
ORR	overall response rate
OS	overall survival
PD	progressive disease
PFS	progression-free survival
PR	partial response
PRO	patient reported outcome.
TEAE	treatment-emergent adverse events
TTnT	time to next treatment
VGPR	very good partial response
WM	Waldenström's macroglobulinemia

1. INTRODUCTION

This statistical analysis plan (SAP) lays out key elements including definitions and statistical methods for analysis of data in evaluation of efficacy and safety for the PCYC-1127-CA study. Analyses of pharmacokinetics data and biomarkers, and additional analyses of patient-reported outcomes (PROs) will be addressed in separate documents.

1.1. Study Design

This is a randomized, placebo-controlled, double-blind, Phase 3 study designed to evaluate whether treatment with ibrutinib in combination with rituximab will result in an improvement in progression-free survival (PFS) compared to placebo in combination with rituximab in subjects with WM (untreated and previously treated). In addition, an open-label ibrutinib substudy is included to further investigate the safety and efficacy of ibrutinib monotherapy in subjects with WM who are considered refractory to the last prior rituximab-containing therapy.

A total of 150 subjects were randomized in a 1:1 ratio to receive ibrutinib and rituximab (Treatment Arm A) or placebo and rituximab (Treatment Arm B). 31 subjects were enrolled in the open-label ibrutinib monotherapy substudy (Treatment Arm C). All subjects in Arm A and Arm B are to receive intravenous rituximab 375 mg/m^2 once weekly for 4 consecutive weeks, followed by a second 4-weekly rituximab course after a 3-month interval. In addition, all subjects in Arm A and Arm B, per randomization, are to receive oral study drug (420 mg ibrutinib or matching placebo), administered once daily and continuously until criteria for permanent discontinuation of study drug are met. All subjects in Arm C are to receive oral ibrutinib 420 mg, administered once daily and continuously until criteria for permanent discontinuation of study drug are met.

The study includes a Screening Phase, Treatment Phase and a Follow-Up Phase. All subjects are to be followed for disease progression until a PFS event occurs or until the data cutoff for final analysis.

Access to ibrutinib (cross-over) as next-line therapy for subjects with Independent Review Committee (IRC) confirmed disease progression and treated with placebo in combination with rituximab (Arm B) are provided as outlined in the protocol. Subjects who have symptomatic disease meeting criteria for requiring treatment may be eligible to receive ibrutinib as next-line therapy.

An IRC has been established to conduct response assessment and evaluate disease progression status centrally per the IRC charter.

1.2. Endpoints

Primary Endpoints

The primary endpoint is progression-free survival (PFS), as assessed by the IRC based on the modified Consensus Response Criteria from the VIth International Workshop for WM (NCCN 2014).

Secondary Endpoints

Secondary endpoints which will be tested include the following:

- Overall response rate (ORR) assessed by IRC (≥ partial response [PR]; according to the modified VIth IWWM [NCCN 2014] criteria)
- Time to next treatment (TTnT)
- Rate of sustained hemoglobin improvement
- Proportion of subjects with ≥ 3 points increase from baseline by Week 25 in fatigue experience score from the FACT-An 5 items assessing fatigue symptoms
- Overall survival (OS)

Safety Endpoints

Safety endpoints include adverse events, clinical laboratory tests, vital signs and eye-related symptoms.

Exploratory Endpoints

Exploratory endpoints include PFS by investigator, ORR by investigator, time to sustained hemoglobin improvement, duration of response (DOR), clinical response rate (CRR), medical resource utilization (MRU), IgM, hemoglobin (Hgb), lymph nodes and spleen measures, tumor involvement of bone marrow, EQ-5D-5L visual analog scale and utility score, FACT-An total score and subscale scores.

1.3. Statistical Hypotheses

The primary hypothesis of this study is that the experimental treatment ibrutinib + rituximab compared with placebo + rituximab will significantly improve PFS in subjects with WM.

The statistical hypotheses are as follows:

H_0: The PFS distributions of experimental treatment group, $S_I(t)$, and the control group, $S_C(t)$, are equal at all time points t:

$S_I(t) = S_C(t)$, for all $t > 0$

versus

H_1: The PFS distributions of experimental treatment group, $S_I(t)$, are greater than the control group, $S_C(t)$, at at least one time point t:

$S_I(t) > S_C(t)$, for some $t > 0$

These hypotheses will be tested using log-rank test.

1.4. Sample Size Determination

Treatment Arms A and B

This study was designed to evaluate the effect of treatment on PFS and was powered for this endpoint. The desired operating characteristics for the PFS endpoint were used to determine the study's total sample size and overall duration. The sample size and power calculations were based on a 2-sided log-rank test for PFS and the following considerations:

- 1:1 randomization ratio between 2 treatment arms
- Target hazard ratio of 0.5 with exponential distribution for PFS. Assuming the median PFS for the control arm (rituximab+placebo) was 15 months from randomization (this was based on available published data), a target hazard ratio of 0.5 corresponds to 2-fold increase in median PFS for the treatment arm (rituximab+ibrutinib) relative to the control (i.e., 30 months vs. 15 months, respectively)
- Minimum 80% power
- 2-sided overall significance level of 0.05
- One interim analysis at 70% information

The interim analysis was based on a group sequential design with Lan-DeMets spending function with O'Brien-Fleming boundary. Using the above assumptions and based on an estimated accrual rate of approximately 12 subjects per month, the study was to enroll approximately 150 subjects (about 75 subjects to each arm) to observe 71 events in approximately 27 months from the first subject randomized. Sample size was calculated using the software package, East 6 (Cytel Software Corp, Cambridge, MA).

Open-Label Substudy Treatment Arm C

The sample size for the substudy was not determined according to statistical calculation. 31 subjects were enrolled in this substudy.

1.5. Planned Analyses

A pre-specified interim analysis for the randomized treatment arms (Arms A and B) will be conducted after observing approximately 70% (~50) PFS events (progressive disease or death) based on IRC assessment. The final analysis of PFS will be conducted after approximately 71 PFS events are confirmed by the IRC for the randomized treatment arms (Arms A and B). The ibrutinib open-label monotherapy substudy (Arm C) will be summarized separately at the final analysis.

1.6. Testing Procedure and Level of Significance

Interim Analysis

In order to preserve the study wise 2-sided type I error rate of 0.05, the 2-sided significance level for PFS will be 0.015 at 50 PFS events for the interim analysis, based on O'Brien Fleming boundary. However, the alpha spending for PFS will be determined based on the actual information fraction using O'Brien-Fleming boundary at the time of the interim analysis.

If the primary endpoint achieves statistical significance, tests of secondary endpoints will be performed at the 2-sided significance level of 0.05 in a sequential hierarchical manner based on a closed testing procedure. The secondary endpoints will be ranked in sequence according to the hierarchical order specified below. OS will not be tested at interim analysis and will be tested at study closure.

1.) ORR

2.) TTnT

3.) Rate of sustained hemoglobin improvement

4.) Proportion of subjects with ≥ 3 points increase from baseline by Week 25 in fatigue experience score

5.) OS

Final Analysis

The 2-sided significance level for the final analysis of primary endpoint (PFS) will be adjusted to account for interim alpha spending so the overall 2-sided significance level for the study will be preserved at 0.05. Tests of secondary endpoints will be performed using the same approach for the interim analysis.

1.7. Blinding and Randomization Methods

1.7.1. Blinding Method

<u>Treatment Arms A and B</u>

This is a double-blind study. The blinding method is described in protocol section 5.1 Randomization and Blinding.

Open-Label Substudy Treatment Arm C

The substudy is open-label.

1.7.2. Randomization Method

For Treatment Arms A and B, central randomization was implemented in this study. The randomization scheme was implemented within the Interactive Web Response System (IWRS).

Randomization was stratified using the following stratification factors and subjects were randomized in a 1:1 ratio to receive either ibrutinib + rituximab or placebo + rituximab within each randomization stratum:

a) WM International Prognostic Scoring System (IPSS) assessed at screening (low vs. intermediate vs. high)

b) Number of prior systemic treatment regimens (0 vs. 1-2 vs. ≥3)

c) Eastern Cooperative Oncology Group [ECOG] (0-1 vs. 2)

1.8. Independent Data Monitoring Committee

The safety and the interim analysis of this study are monitored by an independent data monitoring committee (DMC). The independent DMC may recommend stopping the study for efficacy if the pre-specified stopping boundary is crossed at the interim analysis. In addition to the ongoing safety monitoring and planned interim analyses for efficacy, periodic safety review meetings are planned. The plan for monitoring subject safety and evaluating efficacy, and the roles and responsibilities of the independent DMC, are detailed in the independent DMC Charter.

2. GENERAL ANALYSIS CONSIDERATION

Time to event or duration of event endpoints will be based on the actual event date (or censoring date) rather than visit number or visit label. Missing efficacy or safety data will not be imputed unless otherwise specified.

In general, the baseline value is defined as the last valid measurement on or prior to the first dose of study treatment (ibrutinib/placebo, rituximab). For by-visit analysis, visit windows will be used to associate assessment with a scheduled visit and will be created in reference to the date of first dose of study treatment to assign visit number based on assessment date.

2.1. Analysis Sets

Intent-to-Treat Population

The intent-to-treat (ITT) population includes all randomized subjects (Arms A and B), regardless of the actual treatment received. This population will be the primary population for the summary/analyses of efficacy endpoints, disposition, demographics and baseline characteristics and PRO data.

Safety Population

Safety population includes all subjects who received at least 1 dose of study treatment (rituximab, ibrutinib/placebo). Subjects in this population will be analyzed as randomized for the randomized subjects. The safety population will be used to summarize the safety (including dosing) data.

For the open-label substudy (Arm C), disposition, demographics and baseline characteristics, efficacy, PRO and safety outcome will be summarized descriptively based on the safety population.

2.2. Definition of Subgroups

Subgroup analyses 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	D, E, S
Sex	Male, Female	D, E, S
Prior treatment history as recorded on CRFs	Previously untreated, Previously treated	D, E, S
Race	White, Non-White	E
Geographic region	US, Non-US	E
Baseline ECOG as recorded on CRFs	0-1, 2	E
Baseline IgM	<40, ≥ 40 g/L	E
Baseline hemoglobin (Hgb)	≤ 11, >11 g/dL	E
Baseline β2-microglobulin	≤3, >3 mg/L	E
WM IPSS as recorded on CRFs	Low, Intermediate, High	E
MYD88 L265P mutational status	Mutated, No mutation	E

Subgroup	Definition of Subgroup	Analysis Type
Concomitant use of moderate and/or strong CYP3A inhibitor	Yes, No	S
Baseline creatinine clearance	<60, ≥60 mL/min	S
Baseline liver function based on NCI criteria	Normal, Abnormal (Mild, Moderate, Severe)	S

analysis type D= demographic and baseline disease characteristics

analysis type E= efficacy (Treatment Arms A and B); Considering small sample size in one or more strata, results for race, geographic region, ECOG, MYD88 L265P mutational status may be presented separately.

analysis type S= safety (adverse events)

IWRS= Interactive Web Response System; ECOG= Eastern Cooperative Oncology Group

2.3. Handling of Missing Data for PRO Analysis

Missing Items within Domains

FACT-An scores are derived per FACT-An Scoring Guidelines (Version 4). Fatigue subscale score is derived per FACIT-Fatigue Subscale Scoring Guidelines (Version 4) from the 13 fatigue related items of FACT-An. Similarly, fatigue experience score is calculated from the 5 items (HI7, HI12, An1, An2, An5) assessing fatigue symptoms and fatigue impact score is calculated from the 8 items (An3-An4, An7-An8, An12, An14-An16) assessing the impact of fatigue.

If there are missing items, subscale scores are prorated. This is done by multiplying the sum of the item scores by the number of items in the subscale, then dividing by the number of items answered. A subscale score is calculated only if >50% of the items were answered. The total score is calculated as the sum of the prorated subscale scores. Total scores are calculated only if the overall item response rate is >80% (i.e. at least 22 of 27 items answered for FACT-G total score, and at least 38 of 47 items answered for FACT-An total score).

3. SUBJECT INFORMATION

3.1. Subject Disposition

Subject disposition will be summarized by treatment arm. Subject enrollment will be summarized by region, country, and investigator.

Time on study is defined in the same way as overall survival with reversed censoring, i.e., subjects who died will be censored at death date. The Kaplan-Meier method will be used to estimate the median time on study.

3.2. Demographics and Baseline Characteristics

Baseline characteristics and demographic information at baseline will be summarized with descriptive statistics by treatment arm.

3.3. Prior and Concomitant Medications

Medications will be coded to a generic name and an Anatomical Therapeutic Chemical (ATC) class per the World Health Organization (WHO) drug dictionary. Concomitant medications will be summarized by therapeutic class and preferred term and by treatment arm. Concomitant medications are defined as medications that were taken at any time on treatment (i.e. from the date of the first dose of study treatment through the date of the last dose of study treatment). The following concomitant medications will be summarized separately: growth factors, blood supportive products and immunoglobulin, CYP3A inhibitors/inducers, anticoagulants and/or antiplatelets.

3.4. Extent of Exposure to Study Treatment

Exposure to study treatment will be summarized by treatment arm. Descriptive statistics will be provided for treatment duration and dosing information (e.g. total cumulative dose administered, relative dose intensity, dose reduction due to adverse events) for all study treatments.

3.5. Previous Treatment History and Subsequent Antineoplastic Therapies

Previous treatment history and subsequent WM antineoplastic agents will be summarized separately.

4. EFFICACY AND SAFETY ANALYSES

4.1. Efficacy Analyses

Treatment Arms A and B

Efficacy endpoints and analysis methods for the randomized treatment arms (Arms A and B) are summarized in Table 2.

The following two randomization stratification factors will be used for the stratified analysis/test: WM IPSS assessed at screening (low, intermediate, high) and number of prior systemic treatment regimens (0, ≥1). Due to limited number of subjects with ≥3 prior systemic treatment regimens, subjects with 1-2 or ≥3 prior systemic treatment regimens will be categorized into one subgroup (i.e. ≥1). All the stratified tests will be based on the randomization stratification factors as recorded in IWRS.

Open-Label Substudy Treatment Arm C

For the open-label Arm C, the efficacy outcome will be analyzed separately from the randomized treatment arms. Time to event variables (e.g. PFS, TTnT, OS, DOR) will be analyzed using Kaplan-Meier estimate. For categorical variables (e.g. ORR, rate of sustained hemoglobin improvement, CRR), 95% confidence intervals (CIs) will be calculated based on the exact binomial distribution. For continuous variables (e.g. IgM, lymph node and spleen measures, tumor involvement of bone marrow), descriptive statistics will be calculated. No comparator analysis will be done with Arms A or B.

Note that date of randomization is not applicable for Arm C subjects, therefore date of the first dose of ibrutinib will be used in calculating efficacy variables.

Table 2: Summary of Efficacy Analyses to be Performed for Arms A and B

Endpoint	Definition	Analysis Methods
Primary Endpoint		
PFS assessed by IRC	Time from the date of randomization to the date of first IRC-confirmed disease progression (PD) 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. For subjects who do not have IRC-confirmed PD and not known death as of clinical data cutoff, PFS is censored at the date of the last evidence of no progression by IRC.	The treatment effect of ibrutinib + rituximab (Arm A) compared to placebo + rituximab (Arm B) will be tested with a stratified log rank test adjusted for 2 randomization stratification factors. The hazard ratio and its 95% confidence interval (CI) based on a Cox regression model stratified by the 2 randomization stratification factors will be calculated. Kaplan-Meier analysis for PFS distribution; median PFS estimated with 2-sided 95% CI. Sensitivity Analyses: • Subjects who received subsequent antineoplastic therapy are censored at the last disease assessment showing no evidence of PD before the use of subsequent therapy: same analyses as above. • Subjects who missed ≥2 consecutively planned disease assessments immediately before IRC-confirmed PD or death are censored at the last disease assessment prior to documented PD or death: same analyses as above. • Investigator assessed PFS: same analyses as above. • IRC-assessed PFS by unstratified log-rank test and unstratified Cox regression model. Subgroup Analyses: Hazard ratio and its 95% CI based on unstratified Cox regression model for each subgroup.
Secondary Endpoints		
ORR assessed by IRC	Proportion of subjects achieving a best overall response of confirmed complete response (CR), very good partial response (VGPR), or PR per the IRC assessment at or prior to initiation of subsequent antineoplastic therapy.	ORR will be compared between the two randomized treatment arms (ibrutinib + rituximab vs. placebo + rituximab) using Cochran-Mantel-Haenszel (CMH) chi-square test, adjusted for the 2 randomization stratification factors. 95% CI for the rate ratio will be calculated. Sensitivity Analyses: Investigator assessed ORR: same analyses as above. Subgroup Analyses: Rate ratio and its 95% CI for each subgroup of selected variables (e.g. prior treatment history, MYD88 L265P mutational status).
Time to next treatment	Time from the date of randomization to the start date of any subsequent WM treatment.	Same as PFS: stratified log-rank test, stratified Cox regression model, and Kaplan-Meier analysis.

Endpoint	Definition	Analysis Methods
(TTnT)	Subjects without subsequent treatment are censored at the date of the last study visit.	
Rate of sustained hemoglobin improvement	Proportion of subjects achieving a sustained improvement in hemoglobin (Hgb) at or prior to initiation of subsequent antineoplastic therapy. Hgb improvement is defined as an increase of ≥ 2 g/dL over baseline regardless of baseline value, or an increase to >11 g/dL with a ≥0.5 g/dL improvement if baseline is ≤ 11 g/dL. Sustained Hgb improvement is defined as improvement that is sustained continuously for ≥ 56 days (8 weeks) without blood transfusion or growth factors.	The two randomized treatment arms will be compared (ibrutinib + rituximab vs. placebo + rituximab) using chi-square test. 95% CI for the rate ratio will be calculated. Subgroup Analysis: Chi-square test for subjects with Hgb ≤ 11 g/dL at baseline.
Proportion of subjects with ≥ 3 points increase from baseline by Week 25 in fatigue experience score	Proportion of subjects with a ≥ 3 points increase from baseline in fatigue experience score at any time point at or prior to initiation of subsequent antineoplastic therapy by Week 25. Fatigue experience score is calculated from the FACT-An 5 items (HI7, HI12, An1, An2, An5) assessing fatigue symptoms.	The two randomized treatment arms will be compared (ibrutinib + rituximab vs. placebo + rituximab) using chi-square test. 95% CI for the rate ratio will be calculated.
Overall survival (OS)	Time from the date of randomization to the date of death from any cause. For subjects who were not known to have died at or prior to the clinical cutoff date, OS is censored on the date last known alive.	Unstratified log-rank test, unstratified Cox regression model, and Kaplan-Meier analysis. Exploratory analysis adjusting the crossover effect may be performed, if needed.
Exploratory Endpoints		
Time to sustained hemoglobin improvement	Time from the date of randomization to the initial date of sustained hemoglobin improvement.	Descriptive summary statistics for subjects who achieved sustained hemoglobin improvement.
DOR by IRC, DOR by investigator	Time from the date of initial documentation of a response (PR or better) to the date of first documented evidence of PD or death for responders (PR or better). The same censoring rules for PFS are used for DOR.	Kaplan-Meier analysis
CRR by IRC, CRR by investigator	Proportion of subjects achieving a best overall response of confirmed CR, VGPR, PR or MR at or prior to initiation of subsequent antineoplastic therapy.	Same analyses as for ORR

Endpoint	Definition	Analysis Methods
IgM, Hgb	Value and change from baseline	Descriptive summary statistics
Lymph nodes and spleen	Sum of product of diameters for target nodal lesions, spleen depth and spleen volume	Descriptive summary statistics
Tumor involvement of bone marrow	Cellularity and intertrabecular space	Descriptive summary statistics
Medical resource utilization	Number of plasmapheresis, number of emergency room visits, number and days of hospitalizations	Descriptive summary statistics
EQ-5D-5L visual analog scale and utility score, FACT-An total score and subscale scores	Value and change from baseline, proportion of subjects achieving clinically meaningful improvement, time to improvement. Clinically meaningful improvement is defined as an increase of ≥7 points for visual analog scale (or ≥ 0.08 points for utility score, ≥7 points for FACT-An total score, ≥6 points for anemia subscale score, ≥3 points for fatigue subscale score) prior to initiation of subsequent antineoplastic therapy.	Descriptive summary statistics. The proportion of subjects achieving clinically meaningful improvement will be compared between the two randomized treatment arms using chi-square test. 95% CI for the rate ratio will be calculated. For each parameter, change from baseline up to Week 49 will be analyzed using the Mixed-Effects Model Repeated Measure (MMRM). The model includes the baseline scores, baseline ECOG (0, 1-2), prior treatment history on CRF (previously untreated, previously treated), WM IPSS on CRF (low, intermediate, high) as covariates; treatment, time point and treatment-by-time point interaction as fixed effects; and subjects as random effect. An unstructured (co)variance structure will be used to model the within-subject error. Kenward-Roger's approximation will be used to estimate denominator degrees of freedom. The 2-sided 95% CIs will be calculated for the least square means.

4.2. Safety Analyses

Safety data will be summarized by treatment arm. Table 3 summarizes the safety analyses to be performed for all treatment arms.

Adverse events (AEs) will be coded in accordance with the Medical Dictionary for Regulatory Activities (MedDRA). Severity of AEs will be graded by the investigator according the NCI-CTCAE v4.03.

All laboratory values will be converted to and reported as international standard (SI) units. In general, only data from the central laboratory will be summarized and analyzed. Laboratory parameters will be graded using the NCI CTCAE v4.03.

Unless otherwise specified, only baseline and post-baseline values collected during the treatment-emergent period will be included in the safety analysis. In general, the treatment-emergent period is defined as the period from the date of the first dose of study treatment (rituximab, ibrutinib/placebo) up to 30 days after the date of the last dose of study treatment or the day before initiation of subsequent antineoplastic therapy (including crossover ibrutinib), whichever comes first. The treatment-emergent period for crossover ibrutinib is defined as the period from the date of the first dose of ibrutinib after crossover up to 30 days after the date of the last dose of ibrutinib or the day before initiation of subsequent antineoplastic therapy, whichever comes first.

The treatment-emergent adverse events (TEAEs) are those events that occur or worsen during the treatment-emergent period or that are related to the study treatment.

Table 3: Summary of Safety Analyses to be Performed for all Treatment Arms

Endpoint	Definition	Analysis Methods
Adverse events	Treatment-emergent adverse events (TEAEs), serious adverse events (SAEs), grade 3 or worse TEAEs, related TEAEs, TEAEs leading to treatment discontinuation, TEAEs leading to dose reduction, TEAEs leading to death, other AEs (e.g. bleeding events, major hemorrhage, hypertension, interstitial lung disease (ILD), severe cutaneous adverse reactions (SCAR), rash, cardiac arrhythmia excluding atrial fibrillation, infusion-related reactions), other malignancies, TEAEs in the crossover period, TEAEs in a fixed period (e.g. during the first 9 months) if needed	Descriptive summary statistics and/or listings
Clinical laboratory tests	Worst post-baseline toxicity grade, creatinine clearance, abnormal uric acid, liver function abnormalities	Descriptive summary statistics and/or listings; shift from baseline in creatinine clearance, shifts from baseline in aspartate aminotransferase (AST), alanine aminotransferase (ALT) and bilirubin toxicity grade
Vital signs, eye-related symptoms	Blood pressure, heart rate, temperature, respiratory rate, weight, proportion of subjects with new or worsening eye-related symptoms	Descriptive summary statistics by treatment arm and visit. A data listing will be provided for abnormal ophthalmologic exam findings.

5. MODIFICATION OF ANALYSIS TO THE PROTOCOL

Below is the major change made to the analyses in the protocol:

- Proportion of subjects with ≥ 3 points increase from baseline by Week 25 in fatigue experience score is added as one of the secondary endpoints to evaluate the treatment effect on fatigue symptoms.

In addition, Hgb, lymph nodes and spleen measures, and tumor involvement of bone marrow are included for exploratory analyses. For the exploratory endpoint time to sustained hemoglobin improvement, descriptive statistics will be calculated for subjects who achieved sustained hemoglobin improvement. Kaplan-Meier analysis and log-rank test will not be performed. For PRO data, exploratory analyses will be conducted for rate of clinically meaningful improvement and time to improvement.

6. REFERENCES

FACT-An Scoring Guidelines (Version 4). www.facit.org

FACIT-Fatigue Subscale Scoring Guidelines (Version 4). www.facit.org