Statistical Analysis Plan

A Randomized, Double-blind, Placebo-controlled Phase 3 Study of the Bruton's Tyrosine Kinase (BTK) Inhibitor, PCI-32765 (Ibrutinib), in Combination with Bendamustine and Rituximab (BR) in Subjects With Newly Diagnosed Mantle Cell Lymphoma

Protocol Number: PCI-32765MCL3002; Phase 3

JNJ-54179060 (Ibrutinib)

Status:	Approved
Date:	22 December 2021
Prepared by:	Janssen Research & Development, LLC
Document No.:	EDMS-ERI-130634570

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

TABLE OF CONTENTS

AMMENDMENT HISTORY

December 20, 2021

Added post hoc power analysis for OS including information fraction for interim updates.

November 14, 2021

Updated further analysis plan for OS.

July 13, 2021

Clarified Histology Group and TP53 Mutation categories. Added "s" to the abbreviation of MIPI, for simplified mantel cell lymphoma international prognostic index.

Administrative/Consistency:

- Accepted two tracked changes remaining in prior approved version ("mutation" versus "mutated").
- Minor updates for font size and punctuation.
- Corrected spacing (e.g., "5cm" changed to "5 cm").
- Abbreviation DOR changed to DoR, aligning with format of "DoR" and "DoCR" and text.
- Added abbreviation for EuroQol five-dimensional questionnaire (EQ-5D-5L).

July 8, 2021

Updated references and corrected grammar, punctuation, and formatting. Clarified "sponsor".

Capitalization of Ibrutinib corrected to be lowercase, except when the first word in a sentence and in the cover page.

Changed treatment "arm" to "group" for consistency and CSR/manuscript(s) standards.

SAP reconciled against protocol and DPS, requiring additions/deletions.

Renamed and updated Section 8 (formerly COVID-19 analyses) to address significant changes from protocol.

Analysis Sets: Added

- Definition for PK population.
- Statement for region- and country-specific reproduction of analysis/outputs for regulatory submissions, per request.

Removed the reporting disposition for region, country, site.

Updates associated with primary or key secondary endpoints:

- Added the definition of per-protocol as an Analysis Set, as well as updating Table 3 and Section 5.2.3.2.6 as a Supplementary analysis.
- Updated hierarchy of testing for key secondary endpoints, per Janssen and Pharmacyclics joint agreement.
- Replaced "final" analysis with "primary" analysis.
- Updated Table 3 – added "Supplementary" to "Sensitivity" efficacy analyses, adding COVID related analyses to the list.
- Incorporated estimand framework for primary endpoint.
- Added a statement for the hierarchical rules associated with statistical tests for key secondary endpoints.
- Re-ordered hierarchical order of statistical tests.

- Added COVID-19 to section covering OS.
- Clarified use of investigator assessed (versus IRC assessed) for study endpoints.
- Clarified analysis of MRD negativity rate. Specified exclusion criteria for sample inclusion relative to CR subjects and sample processing requirements.
- Updated time to next treatment to include hyphens, for consistency within and across other documents.

Subgroup Analysis (Section 2.9):

- Added forest plot for OS and ORR subgroup analyses.

- Added TP53 and histology subgroup analysis. A separate forest plot will contain these two variables for endpoints for PFS, OS, and ORR.

COVID-19 analyses are included in each appropriate section (versus Section 8.0).
Removed the word "improvement" in Section 5.3.6 (TTW).

Moved to other sections: details to analyzing AE data from the Follow-Up Phase to the AE section, time to response and duration of response.

Other updates for AEs:

- Removed analysis for duration of lympohocycits.
- Updated test for ATC and generic term in AE section.
- Added incidence and prevalence.
- MedDRA updated to 24.0.
- Updated AESI to be "AEI" (AEs of interest) and noted that the list will be updated at time of final DB lock.

Expanded Appendix table for timing of PROs to include post-PD collection during follow-up. Clarified restart of timing based on date of PD and all other rows indicate pre-PD data collection. Corrected calculations presented in the table for rows associated with Week 12.

Miscellaneous Additions: "under accelerated approval" in the Introduction, mention of region-specific subgroup analyses for regulatory submission purposes, summary of annual incidence rates, demographics and analysis of TP53 mutation status (PFS) and Histology Group, NE to the order of response categories used for analysis, duration of CR (DoCR), summary and listing of subsequent anti-cancer treatment, and analysis of FACT-G in the PRO section.

Miscellaneous Clarifications/Updates: major protocol deviations are reviewed and classified by clinical team, groupings for concomitant medications, use of non- versus un-stratified, use of electronic case report forms (eCRFs), imputation rule for last dose date, replaced "body system" with "system organ class", changed "Ibr" to be consistent with CSR outputs (i.e., "Ibrutinib" or "ibrutinib"), and updated text associated with the FACT-Lym subscale score to the standard reference of "FACT-LymS".

October 20, 2020

Added descriptive summary of COVID-19 pandemic impact and sensitivity and/or supplementary analysis of efficacy endpoint.

Updated interim analysis boundaries based on observed PFS events (information).

Other editorial changes for clarity.

ABBREVIATIONS

AE(s)	adverse event (s)
AEI	AEs of interest
ALT	alanine aminotransferase
ALC	absolute lymphocyte counts
ANC	absolute neutrophil count
aPTT	activated partial thromboplastin time
AST	aspartate aminotransferase
BR	bendamustine and rituximab
BSA	body surface area
BTK	Bruton's tyrosine kinase
CI	confidence interval
CNS	central nervous system
CR	complete response
CrCL	creatinine clearance
CRF	case report form
CT	computed tomography (scan)
DMC	data monitoring committee
DoCR	duration of complete response
DoR	duration of response
ECG	electrocardiogram
ECOG	Eastern Cooperative Oncology Group
eCRF	electronic case report form
EQ-5D	EuroQol
EQ-5D-5L	EuroQol five-dimensional questionnaire
FACT-G	functional assessment of chronic illness therapy general
FACT-Lym	functional assessment of cancer therapy-lymphoma
Hgb	hemoglobin
Ig	immunoglobulin
INR	international normalized ratio
ITT	intent-to-treat (population)
IPCW	inverse probability of censoring weighting
LDH	lactic acid dehydrogenase
LymS	lymphoma specific additional concerns subscale
MCL	mantle cell lymphoma
MedDRA	Medical Dictionary for Regulatory Activities
sMIPI	simplified MCL international prognostic index
MRD	minimal residual disease
MRI	magnetic resonance imaging
NCI CTCAE	National Cancer Institute Common Terminology Criteria for Adverse Events
ORR	overall response rate
OS	overall survival
PD	progressive disease
PET	positron emission tomography
PFS	progression-free survival
PFS2	progression-free survival after initiation of subsequent antineoplastic therapy
PR	partial response
PRO	patient-reported outcome(s)
PT	preferred term
RPSFTM	rank-preserving structural failure time models
SAP	statistical analysis plan
SD	stable disease
SMQ	standardized MedDRA query
SOC	system organ class
TEAE	treatment-emergent adverse event
TOI	trial outcome index score

TTNT	time-to-next treatment
TTW	time to worsening
WBC	white blood cell (count)

1. INTRODUCTION

Ibrutinib (JNJ-54179060: PCI-32765) is a Bruton's tyrosine kinase (BTK) inhibitor currently being co-developed by Janssen Research & Development, LLC (JRD) and Pharmacyclics, LLC. Janssen Research & Development, LLC (JRD) is the study sponsor. It has been approved by the US Food and Drug Administration as monotherapy for patients with previously treated mantle cell lymphoma (MCL), under accelerated approval, and in the EU for patients with relapsed or refractory MCL. The MCL3002 study is designed to evaluate whether the addition of ibrutinib to standard treatment (bendamustine and rituximab [BR] followed by rituximab [R] maintenance) will improve the outcome of subjects with newly diagnosed MCL. This statistical analysis plan (SAP) contains definitions of analysis sets, derived variables, data handling rules and statistical methods for the analysis of efficacy, safety, and patient reported outcomes. Details of the study conduct and data collection can be found in the study protocol amendment INT-7 and electronic case report forms (eCRFs). Any changes to the protocol analysis plan, including additional analyses are documented here.

1.1. Trial Objectives

- The primary objective of the study is to evaluate whether the addition of ibrutinib to BR will result in prolongation of investigator-assessed progression-free survival (PFS) in subjects with newly diagnosed MCL who are 65 years of age or older.
- Secondary objectives include evaluation of:
    - Overall survival (OS)
    - Complete response (CR) rate (investigator assessed)
    - Overall response rate (ORR; CR+ partial response [PR]) (investigator assessed)
    - Patient-reported lymphoma symptoms and concerns as measured by the lymphoma (Lym) subscale of the Functional Assessment of Cancer Therapy-Lymphoma (FACT-LymS)
    - Rate of minimal residual disease (MRD)-negative response
    - Time-to-next treatment (TTNT)
    - Duration of response (DoR)
    - Safety of Ibrutinib when combined with BR
    - The pharmacokinetics of ibrutinib and explore the potential relationships between ibrutinib metrics of exposure with relevant clinical, pharmacodynamic, or biomarker information.

1.2. Trial Design

This is a randomized, double-blind, placebo-controlled, multicenter Phase 3 study to compare the efficacy and safety of ibrutinib in combination with BR versus placebo + BR in subjects with newly diagnosed MCL who are 65 years of age or older. Approximately 520 subjects will be randomized in a 1:1 ratio and stratified by simplified MCL international prognostic index (sMIPI) score (low risk [0-3] vs. intermediate risk [4-5] vs. high risk [6-11]).

Subject eligibility will be determined up to 30 days prior to randomization. The Treatment Phase will extend from randomization until discontinuation of all study treatment or study end. A cycle is defined as 28 days. All subjects will receive open-label BR chemoimmunotherapy for a maximum of 6 cycles; subjects with a CR or PR will continue to receive open-label rituximab (R) maintenance every second cycle for a maximum of 12 additional doses (i.e., Cycles 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, and 30). In addition to the BR therapy, all subjects will receive blinded study drug (ibrutinib or placebo). Subjects randomized to treatment group A will receive placebo capsules, and subjects randomized to treatment group B will receive ibrutinib capsules. Study drug will be administered once daily continuously until disease progression, unacceptable toxicity, or study end.

Subjects with stable disease (SD) after initial chemoimmunotherapy (BR+ibrutinib/placebo) should continue treatment with ibrutinib/placebo until disease progression, unacceptable toxicity, or study end. Subjects with progressive disease (PD) must discontinue all study treatment. For subjects who discontinue background BR therapy and do not have PD, treatment with study drug (ibrutinib/placebo) will continue until disease progression or unacceptable toxicity. Placebo will be stopped when the study is unblinded for the clinical cutoff for the primary analysis of PFS.

Radiological assessments (computed tomography [CT]/magnetic resonance imaging [MRI]) will be performed at Screening, and every 12 weeks in the first 12 months after the start of study treatment. Positron emission tomography (PET) scan is optional at Screening but mandatory at the time of maximal tumor reduction, defined as the time of CR or when 2 consecutive CT scans show no further tumor reduction, and at suspected disease progression, if a new lesion was detected on CT at suspected disease progression. After the first 12 months, scans will be performed every 16 weeks until disease progression or the clinical cutoff for the primary analysis of PFS, whichever comes first. Subjects who discontinue treatment prior to disease progression (for other reasons such as an adverse event [AE]) must continue to have regularly scheduled CT scans/efficacy assessments.

Four analyses are planned. The first 3 analyses will be based on approximately 50%, 68%, and 100% of the total planned 265 PFS events, respectively. The interim analyses and the primary analysis of PFS will take place at these 3 clinical cutoffs, respectively; subject treatment assignment will be unblinded for the primary analysis of PFS. The last clinical cutoff will occur at the end of the study, when 60% of the randomized subjects have died, or the sponsor terminates the study, whichever comes first. Assessment of tumor response and progression will be conducted in accordance with the Revised Response Criteria for Malignant Lymphoma [1].

After the clinical cutoff for the primary analysis of PFS, all subjects without PD will continue disease assessments according to standard of care until disease progression.

1.3. Statistical Hypotheses for Trial Objectives

The primary hypothesis of the study is that ibrutinib in combination with BR compared with placebo + BR will prolong PFS in subjects with newly diagnosed MCL who are 65 years of age or older.

The statistical hypotheses are as follows:

H0: The PFS distributions of the 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

H1: The PFS distributions of the 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

This study is designed to evaluate the effect of treatment on PFS and is powered for this endpoint. The sample size for the study is calculated based on the following assumptions:

1. 1:1 randomization ratio between 2 treatment groups
2. Target hazard ratio of 0.7. Assuming the median PFS for the control group (BR+placebo) is 42 months from randomization, a target hazard ratio of 0.7 corresponds to an 18-month increase in median PFS for the treatment group (ibrutinib+BR) relative to the control (i.e., 60 months vs. 42 months, respectively)
3. Approximately 77% power
4. One-sided overall significance level of 0.025
5. Two interim analyses for both efficacy and futility at approximately 50% and 68% of the planned PFS events, respectively, with the stopping boundary based on the O'Brien-Fleming [2] spending function

Using the above assumptions, the study will enroll approximately 520 subjects (about 260 subjects in each group) to observe 265 events. As of Amendment INT-4, based on actual enrollment, the current projections are to reach 265 events at 72 months after the first patient was enrolled.

1.5. Randomization and Blinding

Central randomization will be implemented in this study. Subjects will be randomly assigned to 1 of 2 treatment groups 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 the simplified sMIPI score (low risk [0-3] vs. intermediate risk [4-5] vs. high risk [6-11]).

This randomization scheme is implemented within the interactive web response system (IWRS) that dictates the treatment assignment and matching study drug kits for the subject throughout the Treatment Phase (Section 2.5). Subjects randomized to treatment are not to be replaced with another subject, should they discontinue treatment for any reason.

2. GENERAL ANALYSIS DEFINITIONS

2.1. Study Treatment and Study Medication

For the purpose of the analysis, the term "study treatment" refers to all 3 drugs, ibrutinib/placebo, bendamustine, and rituximab. The term "study medication or study drug" refers to ibrutinib/placebo.

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 classified according to the assigned treatment group, regardless of the actual treatment received. This population will be used for all efficacy analyses, analyses of disposition, demographic, baseline disease characteristics and patient-reported outcome (PRO) endpoints.

- **Per-protocol (PP) population:** all randomized subjects except for those as described below:

  - did not meet all inclusion and exclusion criteria
  - did not receive the treatment to which they were randomized
  - had less than 75% of study treatment compliance
  - had no post-baseline efficacy assessment and stay on study for more than 175 days

  Subjects in this analysis set will be analyzed according to the treatment to which they are randomized.

- **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.

- **Pharmacokinetic-evaluable Population**: defined as subjects who have received at least 1 dose of ibrutinib/placebo and have at least 1 pharmacokinetic sample obtained posttreatment.

Additional analysis sets for subjects enrolled within specific regions or counties may be used for regional regulatory filing purposes, as requested.

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 treatment. For subjects who have been randomized but not treated with any dose, the randomization date will be used as the reference date for the 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 the protocol and recorded on the eCRFs will be used in the statistical analyses.

2.5. Visit Windows

Visit window will be based on phases and cycles:

- **Screening Phase:** prior to the first dose of study treatment.

- **Treatment Phase:** Between the date of the first dose of study treatment and the date of the last dose of study treatment. As noted above, the term "study treatment" refers to all of the drugs used in the combination treatment.

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

- **Follow-up Phase:** After the End-of-Treatment Phase until the study cutoff. A cycle-based analysis may be performed for safety parameters during the treatment up to the date of the last dose of study treatment + 30 days or the end of treatment visit, whichever comes later.

2.6. Study Day and Cycle Day

Assessments will be presented chronologically by cycle day or study day as follows:

For efficacy data, the randomization date is considered as the reference date (Day 1). For safety data, the date of the first dose of study treatment (the 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 treatment (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. Pooling Algorithm for Analysis Centers

The data from all investigative sites will be pooled for all analyses.

2.8. 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 the date of the initial diagnosis according to the following rules. The start date will be imputed before the 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 the 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 the 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 the date of the last dose, then the date of the last dose +1 will be used.
- If the last dose end date is missing, the end of treatment date will be used as the last dose date.
- 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 the last dose date +1 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 the first dose date, but the imputed day is before the day of the first dose date, then the first dose date will be used, or

    if it is in the same year and month as the first dose date, but the day after the last dose date + 30 days, then the last dose date + 30 days will be used.

  - If the 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 is after the date of death, then the date of death will be used.

  - If the imputed AE start date is in the same month and year but after the first subsequent therapy start date, then the first 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 date of death 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 medications.

2.9. Definition of Subgroups

Table 1: Subgroup Definition

Subgroup	Definition of Group	Analysis Type
Sex	Male, Female	E, S
Race	White, Non-White	E
Race	White, Asian, Others	S
Age	<70, ≥70	E, S
Baseline ECOG	0 vs. 1 or 2	E, S
Simplified MCL international prognostic index (sMIPI)	Low risk [0-3], Intermediate risk [4-5], High risk [6-11]	E
Tumor bulk (largest diameter)	≥ 5 cm, < 5 cm	E

Baseline TP53[a]	Unmutated vs. Mutated	E
Histology Group[b]	Blastoid/Pleomorphic vs. Non-Blastoid/Non-Pleomorphic[d] vs. Unknown	E
Hepatic Impairment[c]	Normal vs. Mild or Moderate	S
Renal function impairment	CRCL≥ 60 mL/min vs. CRCL<60mL/min	S
Concomitant use of CYP3A inhibitor	Any vs. None; Strong vs. Other	S

E= efficacy (primary, other endpoints such as OS and ORR); S= Safety (AEs);

a. An exploratory analysis of the effect of treatment on PFS by TP53 mutation status will be performed in subjects having known baseline TP53 mutation status (i.e., subset of the ITT population).
b. An exploratory analysis of the effect of treatment on PFS by histology will be performed.
c. Hepatic impairment is based on NCI ODWG Liver function Classification.
d. Includes diffuse, nodular, and other histology types.

2.10. Exposure Related Definitions

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

For ibrutinib/placebo, dosing information includes the total dose received (the sum of the total dose), the total prescribed dose (treatment duration in days multiplied by 560), the average dose level per administration (the ratio of the total dose received and the treatment duration) and the relative dose intensity (the total dose received divided by the total expected dose level up to the last dose date).

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

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

Dose reduction for ibrutinib/placebo due to AE is defined as prescribing a lower dose level due to AE (420 mg/day [level 1 reduction] or 280 mg/day [level 2 reduction] or 140 mg/day [level 3 reduction]). For each subject, use the highest level of dose reduction to summarize categories - 'One dose reduction', 'Two dose reductions' and 'Three dose reductions'.

2.11. Other General Definitions

2.11.1. Year and Month

1 year equals 365.25 days, and 1 month equals 30.4375 days.

2.11.2. Age

Age in years will be calculated at the date of informed consent signed.

2.11.3. Time from Initial Diagnosis to Randomization

Time from initial diagnosis to randomization in months will be calculated as (date of randomization - date of initial diagnosis)/30.4375, and the result will be rounded to the first decimal place. A partially missing initial diagnosis date will be imputed based on the rules provided in Section 2.78 of the SAP.

2.11.4. Tumor Bulk and Tumor Burden at Baseline

Tumor bulk is defined as the largest diameter of a measurable tumor lesion assessed by investigator.

Tumor burden is defined as the sum of the product of diameters of all measurable lesions at baseline by investigator.

2.11.5. Date of Response/Best Response

Date of first observed response/best response (CR or PR).

2.11.6. Date of Progression

Date of first observed progression.

3. INTERIM ANALYSIS AND DATA MONITORING COMMITTEE REVIEW

3.1. Data Monitoring Committee

An independent data monitoring committee (DMC) is established to monitor data on an ongoing basis to ensure the safety of the subjects enrolled in this study. At the interim analyses, the DMC may make recommendations regarding study continuation if the pre-specified stopping boundary is crossed for efficacy or futility. In addition to the planned interim analyses, the committee will meet periodically to review ongoing safety data. After the review, the DMC will make recommendations regarding the conduct of the study. The details are provided in the DMC charter.

3.2. Interim Analysis

Two interim analyses were planned using group sequential testing design, scheduled to be conducted when approximately 50% and 68% of the PFS events (PD or death) have occurred.

The O'Brien-Fleming boundaries [2] as implemented by Lan-DeMets spending function is used to control the one-sided Type I error of 0.025 for the comparison of the 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.

Table 2: Stopping Boundaries and Corresponding Observed p-Values and Hazard Ratios

Analysis	PFS Events (info)	Anticipated Time to Analysis	Anticipated Enrollment (n)	Cumulative Error		Stopping Boundary			
				Alpha	Beta	p-value (one-sided)		Observed Hazard Ratio	
				Superiority	Futility	Superiority	Futility	Superiority	Futility
Interim 1	134 (50%)	40 months	520	0.002	0.088	<0.0016	≥0.2429	<0.599	≥ 0.886

Actual	140 (52.8%)			0.002	0.098	<0.002	≥0.212	<0.613	≥ 0.873
Interim 2	180 (68%)	50 months	520	0.007	0.141	<0.0060	≥0.1180	<0.686	≥0.837
Actual	186 (70.2%)			0.007	0.151	<0.007	≥0.104	<0.695	≥0.830
Primary	265 (100%)	72 months	520	0.025	0.225	<0.0229	≥0.0229	<0.781	≥ 0.781
Actual				0.025	0.229	<0.023	≥0.023	<0.780	≥ 0.780
Alpha is Type I error of 0.025 (one-sided). Superiority and futility boundaries are based on O'Brien-Fleming [2] boundaries. Assuming enrollment rate 35/month.									

4. SUBJECT AND TREATMENT INFORMATION

Analyses of disposition, demographic and baseline disease characteristics will be conducted on the ITT population. Analyses of the extent of exposure and concomitant therapy 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. Disposition Information

Disposition information will be summarized for the ITT population. The number of subjects ongoing and discontinued study treatment as well as their reasons for treatment discontinuation will be summarized. Disposition due to COVID-19 will also be summarized and listed.

Descriptive statistics will be provided for the time on study. Time on study is defined the same way as OS with reversed censoring, i.e., a subject who died will be censored at the date of death. 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

All demographic and baseline characteristics will be summarized for the ITT population using descriptive statistics. Subject enrollment will be summarized by region, site, and country. The following demographics and baseline disease characteristics information will be summarized for ITT subjects:

- Demographics: Age (continuous and grouped as <70 or ≥70 years), sex (male or female), race, region, ethnicity
- Baseline disease characteristics: Time from the initial diagnosis to randomization, Serum β_2 microglobulin, stage of MCL at study entry (II or III or IV), type of histology, simplified sMIPI score (low risk vs. intermediate risk vs. high risk), Eastern Cooperative Oncology Group (ECOG) performance status score, endoscopy, hepatic impairment, renal function

impairment, histology group (grouped as Blastoid/Pleomorphic vs. Non-Blastoid/Non-Pleomorphic vs Unknown), and TP53 (Unmutated vs. Mutated)

- Extent of disease: Number of lesions, extranodal disease, tumor bulk (grouped as <5 cm, ≥5 cm), tumor burden, baseline lymphoma symptoms, bone marrow involvement, and baseline lactic acid dehydrogenase (LDH)
- Hematology: hemoglobin (Hgb), white blood cell count (WBC), absolute neutrophil count (ANC), and platelets
- Chemistry: sodium, potassium, creatinine clearance (CrCL), creatinine, alanine aminotransferase (ALT), aspartate aminotransferase (AST), alkaline phosphatase, lactic acid dehydrogenase (LDH), total bilirubin, albumin, calcium, magnesium, phosphate, and uric acid
- Vital signs: height (cm), weight (kg), temperature (C°), systolic blood pressure, diastolic blood pressure, heart rate, body surface area (BSA)
- Coagulation: Prothrombin international normalized ratio (INR), and Activated partial thromboplastin time (aPTT)

4.3. Concomitant Medications

Use of concomitant therapies other than antineoplastic agents or other systemic therapies for MCL between treatment start and the end of the study will be provided by ATC class and drug generic term. Multiple medication usage by a subject will be counted only once for that therapeutic class. 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 30 days after the last dose of study treatment). However, subgroup analysis by CYP3A inhibitors will use a different time window (concomitant medication starting date within the range of the first dose date and last dose date of study treatment).

Following concomitant medications of special interest will also be provided:

- Anti-coagulation and anti-platelets
- CYP3A inhibitors/inducers
- Growth factors/cytokines: Referred to medications coded to ATC level 3 Text = "other anti-anemic preparations" or ATC level 4 Text = "colony stimulating factors"
- Transfusions: Referred to those medications that are coded to B05AX

4.4. Protocol Eligibility and Major Protocol Deviations

Subjects with major protocol deviations will be listed by treatment group. Protocol deviations will be based on clinical review primarily on the following aspects (but not limited to): (1) eligibility criteria, (2) treatment compliance, (3) subject safety (4) efficacy assessment deviation. 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. The major protocol deviations will be reviewed and/or classified based on clinical review of the protocol deviations.

Major protocol deviations due to COVID-19 will be summarized and listed by treatment group. Listing of minor protocol deviations due to COVID-19 will also be provided. Summary table and listing of study assessment compliance will be presented.

4.5. Extent of Exposure

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

The number and percentage of subjects with dose reductions due to AEs will be summarized. Subjects with dose interruptions (missing dose ≥7 consecutive days) will also be presented.

Details of the exposure-related definition is specified in Section 2.10.

4.6. Medical History

Abnormal medical history findings reported by investigator will be summarized by system organ class (SOC) and preferred term (PT).

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.

Table 3: Summary of Efficacy Analyses to be Performed

Endpoint	Analysis	Analysis Method	Population
Primary			
PFS assessed by investigator	Primary	Stratified log-rank test, stratified Cox regression model, all PD and death considered as events regardless of subsequent antineoplastic therapy; evaluation of censoring reasons	ITT
	Sensitivity/Supplementary	• Unstratified log-rank test, Unstratified Cox regression model; • Subjects who received subsequent therapy are censored at the last disease assessment showing no evidence of PD before the use of subsequent anticancer therapy • Subjects censored at the last disease assessment before PFS event if they progress or die after missing >=2 planned consecutive disease assessments • IRC assessed PFS by stratified log-rank test and stratified Cox regression model; concordance rate between the IRC-determined PD and investigator-determined PD • Subject censored at the last disease assessment before pre-PD death due to COVID-19. (This supplementary analysis would only be conducted if subject (pre-PD) death due to COVID-19 is >5% of total PFS events.) • Subjects restricted to the defined PP analysis set. (This supplementary analysis would only be conducted if the PP population is less than 90% subjects in the ITT population.)	ITT, unless otherwise noted.
	Subgroup (See Section 2.9)	Unstratified Cox regression model within each subgroup	ITT
	Covariate-adjusted analysis of PFS	A selected set of potential prognostic variables obtained at or before baseline as covariates in unstratified Cox regression models: sex, race, age group, baseline ECOG, sMIPI score, tumor bulk	ITT
Key Secondary			
CR rate		CMH Chi-square test	ITT
Rate of Minimal Residual Disease-Negative Response		CMH Chi-square test	ITT

Endpoint	Analysis	Analysis Method	Population
Overall Survival		Stratified log-rank test, stratified Cox regression model	ITT
Overall Response Rate		CMH Chi-square test	ITT
Time-to-Next Treatment		Stratified log-rank test, stratified Cox regression model	ITT
Patient-Reported Outcome		Descriptive summary, mixed model with repeated measurements, stratified log-rank test, stratified Cox regression model	ITT

5.1. Analysis Specifications

5.1.1. Level of Significance

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

Statistical inference on the primary endpoint, PFS, at the interim and primary analyses will be conducted at a one-sided significance level of 0.025, under group sequential testing design per O'Brien-Fleming [2] boundaries, as specified in Section 3.

If the primary endpoint achieves statistical significance, tests of key secondary endpoints will be performed at the two-sided significance level of 0.05 in a sequential hierarchical manner based on a closed testing procedure. The hypothesis for a secondary endpoint will be tested only if the null hypotheses for the primary endpoint and for the preceding secondary endpoints are rejected. The key secondary endpoints will be ranked in sequence according to the hierarchical order specified below:

1. Complete response (CR) rate (investigator assessed)
2. Time-to-next treatment (TTNT)
3. Overall survival
4. Overall response rate (CR+PR) (investigator assessed)
5. Rate of MRD-negative response
6. Time to worsening (TTW) in the Lym subscale of FACT-Lym (FACT-LymS)

5.1.2. Data Handling Rules

Unless specified otherwise, missing values will not be imputed.

5.1.3. General Analysis Considerations

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.

5.2. Primary Efficacy Endpoint

5.2.1. Progression Free Survival

The primary endpoint of PFS is assessed by the investigator. PFS is defined as the duration from the date of randomization to the date of disease progression or relapse from CR (as assessed by the investigator) or the date of death, whichever occurs first. Subjects who are progression free and alive will be censored at the time of their last disease assessment. Subjects with no post-baseline disease assessment will be censored on the randomization date.

5.2.2. Primary Estimand

Primary Trial Objective: To demonstrate the superiority of ibrutinib+BR to that of placebo+BR, including R maintenance, in terms of PFS in subjects with newly diagnosed MCL.

Estimand Scientific Question of Interest: What is the effect on PFS of assigning subjects to ibrutinib+BR vs. BR?

This primary estimand is the main clinical quantity of interest to be estimated in this study, which is defined by the following five attributes [3]:

Population: subjects with newly diagnosed MCL

Treatment: ibrutinib+BR vs. placebo+BR, both including R maintenance therapy for patients achieving response

Variable: PFS (PD is assessed by the investigator)

Population-level summary: Kaplan-Meier estimates of PFS, hazard ratio of ibrutinib+BR vs. placebo+BR

Intercurrent events and handling strategies: treatment discontinuation, use of subsequent anticancer therapy, death due to COVID-19

Intercurrent Events	**Name of Strategy for Addressing Intercurrent Events and Its Description**
Treatment discontinuation (due to AE or other reasons other than AE or worsening of disease)	**Treatment policy strategy:** use time to PD or death, regardless of whether or not treatment discontinuation had occurred.
Use of subsequent anticancer therapy	**Treatment policy strategy:** use time to PD or death, regardless of whether or not used subsequent anti-cancer therapy. **Hypothetical strategy:** subjects are censored at the last disease assessment showing no evidence of PD before the use of subsequent anti-cancer therapy.
Death due to COVID-19	**Composite variable strategy:** consider (pre-PD) death as a PFS event. **Hypothetical strategy:** subjects are censored at the last disease assessment before (pre-PD) death due to COVID-19.

5.2.3. Analysis Methods

5.2.3.1. Primary Analysis

Assumptions:

- Non-informative censoring assumed for all types of censoring
- Distinct baseline hazard for each stratum, common proportional hazard ratio across strata

Primary Estimator:

- A stratified Cox regression model with study intervention as the sole explanatory variable will be performed, with sMIPI as a stratification factor.
- Hazard ratio and its 95% CIs will be estimated.
- The treatment policy strategy is adopted for handling the intercurrent events of treatment discontinuation and use of subsequent anti-cancer therapy. The composite variable strategy is adopted for handling the intercurrent events of pre-PD death (PFS event) due to COVID-19.

Decision making will be based on the stratified log-rank test for statistical significance. Kaplan-Meier method will be used to estimate the distribution of PFS for each treatment group. The non-stratified Cox regression model may be used to analyze the treatment effect on PFS after adjusting for covariates (selected demographics and baseline characteristics) as appropriate. The potential prognostic factors can be found in Section 2.9.

5.2.3.2. Sensitivity and Supplementary Analyses

5.2.3.2.1. Sensitivity Analysis of Homogeneous Baseline Hazard

Assumptions:

- Identical baseline hazard for each stratum and common proportional hazard ratio across strata.

Sensitivity Estimator:

- A unstratified Cox regression model with study intervention as the sole explanatory variable will be performed.

5.2.3.2.2. Sensitivity Analysis of Disease Assessment Follow up

Assumptions:

- Non-informative censoring assumed for all types of censoring.

Sensitivity Estimator:

- A stratified Cox regression model with study intervention as the sole explanatory variable will be performed, subjects will be censored at the last disease assessment if they progress or die after missing ≥ 2 consecutive planned disease assessment visits.

5.2.3.2.3. Sensitivity Analysis of PD Assessment

Assumptions:

- No potential reader-evaluation bias in PD assessment.

Sensitivity Estimator:

- A stratified Cox regression model with study intervention as the sole explanatory variable will be performed, with sMIPI as a stratification factor.
- The concordance rate between the IRC-determined PD and investigator-determined PD will be evaluated. The number and percentage of PD events and non-PD cases determined by investigator and by IRC will be cross-tabulated.

5.2.3.2.4. Supplementary Analysis of Estimand 2:

Estimand 2 is defined to support the primary estimand. The only attribute that changes from the definition of the primary estimand is how the handling strategy is adopted for the use of subsequent anti-cancer therapies.

Estimand 2:

- Hypothetical strategy: if all subjects had continued treatment as planned and had not used any subsequent anti-cancer therapies.
- Under the estimand 2, time to PD or death will be censored at the last disease assessment showing no evidence of PD before the use of subsequent anti-cancer therapy. Same analyses described in primary estimator will be applied.

5.2.3.2.5. Supplementary Analysis of Estimand 3

Estimand 3 is defined to support the primary estimand. The only attribute that changes from the definition of the primary estimand is how the handling strategy is adopted for the death due to COVID-19:

Estimand 3:

- Hypothetical strategy: if all subjects had continued treatment as planned and had not died from COVID-19.
- Under the estimand 4, time to PD or death will be censored at the last disease assessment before pre-PD death due to COVID-19. Same analyses described in primary estimator will be applied, however, this supplementary analysis would only be conducted if subject (pre-PD) death due to COVID-19 is >5% of total PFS events.

5.2.3.2.6. Supplementary Analysis of Estimand 4

Estimand 4 is defined to support the primary estimand. The only attribute that changes from the definition of the primary estimand is the analysis set from ITT to PP population.

Estimand 4:

- Under the estimand 4, same analyses described in primary estimator will be applied, however, this supplementary analysis would only be conducted if the PP population is < 90% of the ITT population.

5.2.3.3. 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 a forest plot.

5.3. Key Secondary Efficacy Endpoints

5.3.1. Complete Response

Complete response rate is defined as the proportion of subjects who achieve CR (based on investigator assessment) on or prior to the initiation of subsequent anticancer therapy. The rates will be compared between groups using Cochran-Mantel-Haenszel (CMH) Chi-square methodology, adjusted for the stratification factor sMIPI score used at the randomization. Time to CR will be analyzed for subjects with CR and is defined as the interval between the date of randomization and the date of initial documentation of a CR. Duration of complete response (DoCR) is defined as the interval between the date of initial CR until disease progression or death from any cause, whichever occurs first, for subjects who achieved CR. DoCR will be summarized descriptively using the Kaplan-Meier method, and no inferential comparison will be made between treatment groups. Both the CR rate and DoCR by IRC assessment will also be summarized.

5.3.2. Time-to-Next Treatment (TTNT)

Time-to-next treatment (TTNT) is measured from the date of randomization to the start date of any anti-lymphoma treatment subsequent to study treatment. Stem cell transplant is included as at type of anti-lymphoma subsequent treatment. Subjects without subsequent treatment will be censored at the date of the last contact.

TTNT will be analyzed using the same analysis methods as used for PFS. The number of subjects with subsequent anti-cancer therapy will be summarized by therapy type. A listing of subjects with subsequent anti-cancer therapy will also be provided.

5.3.3. Overall Survival

Overall survival (OS) is defined as the time from the date of randomization to the date of death from any cause. If the subject is alive, or the vital status is unknown, the subject will be censored at the date the subject was last known to be alive.

5.3.3.1. Post-hoc Power Analysis of OS

In addition to the OS analysis with the hierarchical testing at the time of primary analysis of PFS, patients will continue to be followed, and the OS analysis will be updated periodically (when 50%, 55% and 60% of randomized subjects have died) until the study end, as specified in the protocol. No formal hypothesis testing will be performed. Post-hoc power analysis of OS under various scenarios of hazard ratios (HRs) and median OS (with the assumption of exponential distribution) were conducted at the time of primary analysis of PFS. As shown in Table 4, the power for testing the OS effects with HRs of 0.80 and 0.85 are approximately 50% and 30%, respectively, at the protocol specified study end, when 60% of subjects have died (314 OS events). Although the additional analyses of OS were not intended to be part of group sequential testing to control for overall Type I error, information fraction was calculated based on the number of subjects have died at a given analysis. Assuming the number of OS events at the study end is 314, and noting at the time of the primary PFS analysis 40.3% of subjects had died (approximately an information fraction of 67.2% for OS), for subsequent planned OS analyses, when 50% and 55% of subjects

have died, these would correspond to information fractions of approximately 83.3% and 91.7%, respectively.

Table 4. Post-hoc power analysis of OS

Hazard Ratio	Median OS for Placebo+BR	Median OS for Ibrutinib+BR	No. of OS Events at Study End	Power for OS (%)	Fixed Sample Size	Study Duration (months)	OS Event Projection at Primary Analysis of PFS
0.8	72	90	314	50.6	523	116.1	249
0.85	76.5	90	314	30.1	523	119.4	244
0.8	84	105	314	50.6	523	133.8	223
0.85	89.25	105	314	30.1	523	137.7	218
0.8	96	120	314	50.6	523	151.6	201
0.85	102	120	314	30.1	523	156.0	197

1. The post-hoc power analysis is calculated for a fixed sample size design of 523 subjects with no formal interim analysis. 2. The observed average accrual rate of 27.5 per month and variance of log hazard ratio under null hypothesis are used.

The conditional probabilities of observing HR < 1 at future additional OS analyses under various assumptions of HRs for the remainder of study were estimated [4] based on the observed HR for OS at the time of primary analysis of PFS, see Table 5.

Table 5. Conditional probability of HR < 1 at further OS analyses

Observed HR at Primary Analysis of PFS (211 OS Events)	Assumed HR for the Remainder of Study	Probability of HR < 1 with Additional OS Analysis (%) 262 OS Events[1] (83.3% information)	288 OS Events[1] (91.7% information)	314 OS Events[1] (100% information)
1.07 (95% CI 0.81, 1.40)	0.8	44.0	56.3	64.2
1.07 (95% CI 0.81, 1.40)	0.85	36.5	47.3	54.5
1.07 (95% CI 0.81, 1.40)	0.9	29.8	38.9	45.1
1.07 (95% CI 0.81, 1.40)	0.95	24.1	31.5	36.4
1.07 (95% CI 0.81, 1.40)	1	19.3	25.0	28.8

[1] The specified number of 262, 288, 314 OS events are approximately 50%, 55%, 60% of the number of subjects randomized in the study.

5.3.3.2. Primary Estimand of OS

The primary estimand of OS is defined by the following five attributes:

Population: subjects with newly diagnosed MCL

Treatment: ibrutinib+BR vs. placebo+BR both including R maintenance therapy for patients achieving response

Variable: OS

Population-level summary: Kaplan-Meier estimates of OS, hazard ratio of ibrutinib+BR vs. placebo+BR

Intercurrent events and handling strategies: treatment discontinuation, use of subsequent anticancer therapy (including treatment cross-over to ibrutinib or other BTK inhibitors as monotherapy), death due to COVID-19, death due to competing risks

Intercurrent Events	**Name of Strategy for Addressing Intercurrent Events and Its Description**
Treatment discontinuation (due to AE or other reasons other than AE or worsening of disease)	**Treatment policy strategy:** use time to death, regardless of whether or not treatment discontinuation had occurred
Use of subsequent anti-cancer therapy	**Treatment policy strategy:** use time to death, regardless of whether or not subsequent anti-cancer therapy had been used **Hypothetical strategy:** if all subjects had continued treatment as planned and had not used any subsequent anti-cancer therapies (or had not received ibrutinib or other BTK inhibitors monotherapy)
Death due to COVID-19	**Composite variable strategy:** consider death due to COVID-19 as an OS event **Hypothetical strategy:** subjects are censored at the date of death due to COVID-19
Death due to competing risks	**Composite variable strategy:** consider death due to reasons other than progressive disease or treatment-emergent adverse events as an OS event **Hypothetical strategy:** consider death due to reasons other than progressive disease or treatment-emergent adverse events as a competing risk event

Primary Estimator:

- A stratified Cox regression model with study intervention as the sole explanatory variable will be performed, with sMIPI as a stratification factor. Un-stratified log-rank test and un-stratified Cox regression model will not be performed as number of events is more than 10% of the ITT analysis set.
- Hazard ratio and its 95% CIs will be estimated.
- The treatment policy strategy is adopted for handling the intercurrent events of treatment discontinuation and the use of subsequent anti-cancer therapy. The composite variable strategy is adopted for handling the intercurrent events of death due to competing risk including COVID-19.

The un-stratified Cox regression model may be used to analyze the treatment effect on OS after adjusting for covariates (selected demographics and baseline characteristics) as appropriate. The forest plot will be used to assess OS within factors listed in Section 2.9.

5.3.3.3. Supplementary Analysis of Estimand 2:

Estimand 2 is defined to support the primary estimand. The only attribute that changes from the definition of the primary estimand is how the handling strategy is adopted for the use of subsequent anti-cancer therapies.

Estimand 2:

- Hypothetical strategy: if all subjects had continued treatment as planned and had not used any subsequent anti-cancer therapies (or had not used ibrutinib or other BTK inhibitors as monotherapy).
- Under the estimand 2, the inverse probability of censoring weighting (IPCW) analysis [5] will be used to adjust for the use of subsequent anti-cancer therapies. The rank-preserving structural failure time models (RPSFTM) [6] will be used to adjust for treatment crossover to ibrutinib or other BTK inhibitors as monotherapy from the placebo+BR arm.

5.3.3.4. Supplementary Analysis of Estimand 3:

Estimand 3 is defined to support the primary estimand. The only attribute that changes from the definition of the primary estimand is how the handling strategy is adopted for the death due to COVID-19:

Estimand 3:

- Hypothetical strategy: if all subjects had continued treatment as planned and had not died from COVID-19.
- Under the estimand 3, the time to death will be censored at the date of death due to COVID-19. This supplementary analysis would only be conducted if subject deaths due to COVID-19 are greater than 5% of the total OS events.

5.3.3.5. Supplementary Analysis of Estimand 4:

Estimand 4 is defined to support the primary estimand. The only attribute that changes from the definition of the primary estimand is how the handling strategy is adopted for the death due to reasons other than progressive disease or treatment-emergent adverse events:

- Estimand 4: Hypothetical strategy: if all subjects had continued treatment as planned and had not died from competing risks.
- Under the estimand 4, death will be distinguished by causes of death: due to progressive disease or treatment-emergent adverse events vs. competing risks (including COVID-19). Cumulative incidence function with the associated Gray's test [7] and competing risks regression analysis [8] will be used for the competing risk analysis.

5.3.4. Overall Response Rate

Overall response rate (ORR) is defined as the proportion of subjects who achieve either CR or PR as the best overall response as assessed by investigator. The order of the overall response category is: CR>PR>SD>PD>NE. The maximum category of overall response is the best overall response. The overall response category will be derived based on a response assessment performed on or before the start of subsequent anti-cancer therapy. Subjects without documented subsequent anti-

cancer therapy (and/or the start date of anti-cancer therapy is missing) will be considered as not having received subsequent anti-cancer therapy.

ORR will be estimated according to the crude proportion of confirmed responders (PR or better) based on the best overall response and summarized by treatment group. ORR will be compared between treatment groups using CMH test adjusted for the randomization stratification factor. The relative risk and its 95% CI are to be presented as well. A subgroup analysis will be performed using a forest plot for efficacy factors listed in Section 2.9, including the binomial estimate of ORR and 95% CIs for each pre-defined subgroup.

Duration of response (DoR) is defined as the interval between the date of initial documentation of a response including PR and the date of first documented evidence of PD or death. The censoring rule for DoR is the same as PFS. Only subjects who achieved a PR or better will be included in the analysis of duration of response. DoR will be summarized descriptively using the Kaplan-Meier method, and no inferential comparison will be made between treatment groups.

Time to response will be analyzed for subjects with CR/PR and is defined as the interval between the date of randomization and the date of initial documentation of a response. Descriptive summaries will be provided for the time to response.

For ORR based on IRC assessment, the same analysis methods as used for the investigator assessment of overall response will be used.

5.3.5. 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, (i.e., <1 MCL cell per 10,000 leukocytes for detection using the MRD assay), as assessed by flow cytometry of a bone marrow and/or peripheral blood sample, and whose response is CR. Within the same visit, if both the bone marrow and the peripheral test results are available, the positive bone marrow test overrules the negative peripheral blood test. Subjects with missing MRD data are considered non-responders. The CMH chi-square test adjusted for the stratification factor will be used to compare the MRD negativity rate between treatment groups, chi-square test will be used for un-stratified tests, and the Fisher's exact test may be used if the rate in any treatment group is too small.

Samples excluded from MRD analysis include those meeting the following criteria:

- Assay processed > 5 days from collection date
- Collected either without investigator assessed CR occurring or > 30 days prior to first occurrence of investigator assessed CR
- Collected after investigator assessed PD
- MRD samples taken after global INT-6 implementation date (16Aug2019)

5.3.6. Time to Worsening in the Lym Subscale of FACT-Lym

Time to worsening in the Lym subscale of the FACT-Lym (FACT-LymS) is defined as the interval from the date of randomization to the start date of the worsening through study discontinuation. Worsening is defined by a 5 or greater point reduction from baseline in the FACT-LymS. Death (due to any cause) and missing data due to very ill subjects as noted on the eCRF will also be considered as worsening. Subjects who have not met the definition of worsening will be censored at the last PRO assessment. Subjects with no baseline assessment or on-study assessment will be censored at the date of randomization.

Further details are provided in Section 7.

5.4. Exploratory Efficacy Endpoint

PFS2 is defined as the duration from the date of randomization to the date of the earliest of the following 3 types of events: PD that occurs after the first subsequent anti-cancer therapy, death at any time on study from any cause, the start of second subsequent anti-cancer therapy. Those who do not receive subsequent antineoplastic therapy or do not experience the event as specified above are censored at the last investigator assessed disease assessment. PFS2 will be analyzed using the ITT population and using the same analysis method as used for PFS.

6. SAFETY

Safety will be analyzed using the incidence and severity of AEs and laboratory tests. Analyses will be based on the safety analysis set (Section 2.2). Descriptive statistics and frequency tables will be reported and used for all safety data. Inferential statistics are not planned to be performed on safety data. Summary table and listing for treatment-emergent COVID-19 AEs will also be presented.

The baseline value for safety analyses is defined as the value collected at the time closest to and prior to the start of study treatment. For the analysis of lab measurements for safety parameters by cycle, mean values within each cycle will be used. For the analysis of lab grade for safety parameters by cycle, the worst grade within each cycle will be used. Annual incidence and prevalence rates of safety events will be reported.

Unless otherwise stated, safety data will be summarized by treatment group.

6.1. Adverse Events

Category	Analysis	Sorted By	Drug-Related TEAE
General	Overall summary		✔
	Treatment-emergent AEs (TEAEs)	SOC+ PT; PT toxicity grade + PT toxicity grade	✔ ✔ ✔
	Serious TEAEs	SOC+ PT + toxicity grade; PT+ SOC Toxicity grade +PT	✔

Category	Analysis	Sorted By	Drug-Related TEAE
	Grade 3 or worse TEAE	PT+ SOC PT	✔ ✔
	TEAEs leading to treatment discontinuation (Ibrutinib/placebo, BR)	PT + SOC PT	
	TESAEs leading to treatment discontinuation (Ibrutinib/placebo, BR)	PT+SOC	
	TEAEs leading to death	PT + SOC PT	
	Grade 3 or worse serious TEAE	PT+SOC	
	TEAEs leading to dose reduction/ (Ibrutinib/placebo, BR; reduction is only for Ibrutinib/placebo/B)	PT + SOC PT	✔ ✔
	TESAEs leading to dose reduction (Ibrutinib/placebo/B)	PT+SOC	
	AEs of Interest (Hemorrhagic events)	PT	
	Other safety observations (e.g., other malignancies)[a]	PT	
	Other safety observations (i.e., not noted above)[a]	PT + toxicity grade; SOC+PT + toxicity grade	
	Deaths during treatment emergent period and within 30 days of last dose	Reason for death	
Subgroup (See Section 2.9)	Overall summary		
	TEAEs	SOC+ PT	
Exposure adjusted incidence rate	Overall summary		
	TEAEs	SOC+ PT+ toxicity grade; PT+ toxicity grade	
	Serious TEAEs	SOC+ PT + toxicity grade; PT+ toxicity grade	

a) To be finalized at the date of final database lock. See Section 6.2.

6.1.1. All Adverse Events

The verbatim terms used in the eCRFs by investigators to identify AEs will be coded using the Medical Dictionary for Regulatory Activities (MedDRA 24.0). The severity of AEs is assessed using the National Cancer Institute Common Toxicity Criteria for Adverse Events (NCI CTCAE) Version 4.03.

Drug-related AEs are those assessed by investigator as being possible, probable, or very likely related to study drug.

Treatment-emergent AEs (TEAEs) will be summarized and are defined as those events that 1) occur after the first dose of study treatment, through the treatment phase, and for 30 days following the last dose of study treatment or until subsequent anti-cancer therapy if earlier; 2) any event that is considered study treatment-related regardless of the start date of the event; or 3) any event that is present at baseline but worsens in severity or is subsequently considered drug-related by the investigator.

TEAEs will be summarized by SOC and PTs, by NCI CTCAE toxicity grade, by relationship to study drug, and by action taken. For each TE AE, 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). The same summary will be provided for serious TE AEs, and drug-related serious TEAEs, as well as TEAEs leading to treatment discontinuation, death, and dose reduction.

6.2. Adverse Events of Interest

AEs of Interest (AEIs) and other safety observations (e.g., other malignancies) will be summarized by treatment group.

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 ≥ 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 by investigators for the duration of study treatment and during any protocol-specified follow-up periods including post-progression follow-up for OS. Summary tables for other malignancies occurring during the study and during the treatment-emergent reporting period by PT will be provided, categorized by 3 groups: non-melanoma skin cancer, melanoma skin cancer, and non-skin cancer.

In addition, AEIs classified as Other Safety Observations at the date of the final database lock will be summarized.

6.3. Deaths

Incidences of deaths during the treatment emergent period and within 30 days after last dose are to be reported, along with the primary cause of death.

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 (EAIR) may also be summarized if the median treatment duration of one group is 30% longer than that of the other group. An overall summary of EAIRs will be presented. EAIRs for AEs and serious AEs will also be summarized by SOC and PT.

The analysis restricts 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 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 will be used for 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 patient 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 patients as described before, i.e.,

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

whereby the TEAE enters the sum in the numerator unweighted ($TEAE_i$ =1, otherwise $TEAE_i$ =0), and 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 patient per SOC only, and only one (the corresponding) exposure time in the denominator (where the denominator in the $EAIR_{SOC}$ depends on the number of PTs per SOC):

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

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 patient only, and only one (the corresponding) exposure time in the denominator (where the denominator in the $EAIR_{global}$ depends on the overall number of PTs):

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

Whereby, $TEAE_i$ represents the first TEAE among all TEAEs of patient i and t_i as before the time when the TEAE occurs (non-censored data) or the 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 the last dose or the start of subsequent anticancer therapy, whichever is earlier, will be reported in SI units. Applicable laboratory results will be graded according to NCI CTCAE version 4.03. Note, toxicity grading for creatinine increase will be based on the NCI CTCAE v4.03 criteria but limited only to the part based on the upper limit of normal (ULN); the other part that is based on the change from baseline will not be used for toxicity grading. Toxicity grading for all other laboratory parameters will be based on the NCI CTCAE v4.03 criteria as is. Generic normal ranges will be applied whenever reference ranges are not available. The following laboratory tests will be analyzed:

- Hematology: hemoglobin, WBC, ANC, absolute lymphocyte counts (ALC), and platelets
- Chemistry: sodium, magnesium, creatinine, CrCL, AST, ALT, alkaline phosphatase, LDH, total bilirubin, albumin, potassium, calcium, phosphate, uric acid
- Other: IgG, IgM, IgA

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 parameters. Worst toxicity for hematology and clinical chemistry during treatment will also be summarized for both groups.

6.5.1. Creatinine Clearance

CrCL is calculated using the Cockcroft-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

A shift summary from baseline to worst on-treatment toxicity grade for CrCL will be provided.

6.5.2. Analysis of Lymphocytosis

A descriptive summary of ALC, with and without lymphocytosis, time to lymphocytosis, and duration of lymphocytosis will be provided by treatment. Lymphocytosis is defined as ALC increasing ≥ 50% from baseline and achieving a level ≥ $5x10^9$/L. For subjects with lymphocytosis, resolution of lymphocytosis is defined as 1) a decrease of the ALC value to baseline level or lower, or 2) or an achievement of an ALC value that is < $5x10^9$/L, whichever occurs first. Time to peak ALC is defined as the number of weeks from the first dose date of study drug to the highest ALC value within the first 9 months of the first dose of study drug. Time to lymphocytosis is defined as the interval between the date of the first dose and the date the subject had lymphocytosis for the first time and will be presented in weeks.

6.6. Electrocardiogram (ECG)

Abnormal electrocardiogram (ECG) findings will be summarized. An abnormal finding is considered to be treatment-emergent if it occurred during the treatment and up to 30 days after the last dose of study treatment or the start of the subsequent therapy, whichever is earlier.

6.7. Other Safety Parameters

Frequencies of ECOG scores will be reported over time. Descriptive statistics of the change in ECOG scores from baseline will also be provided.

7. PATIENT-REPORTED OUTCOME

Patient-reported outcome assessments using the Functional Assessment of Cancer Therapy – Lymphoma (FACT-Lym) [9] and the five level EuroQol five-dimensional questionnaire (EQ-5D-5L) will be analyzed to examine if the subjects' perspective of response to treatment are accompanied by measurable changes in disease symptoms and health status. Time to clinically meaningful worsening in the FACT-Lym lymphoma subscale is also included as one of the key secondary objectives of this study.

The FACT-Lym and EQ-5D-5L questionnaires will be administered on Day 1 of every cycle during the first 6 cycles, then every 12 weeks in the first 12 months after the first dose of study treatment. Thereafter, PROs will be performed every 16 weeks. During the first 6 cycles, if a PRO assessment was conducted, but the cycle was subsequently delayed, the PRO assessment should be repeated on Day 1 of the cycle when the treatment is resumed. The FACT-Lym will be performed until disease progression, death, or clinical cut-off, whichever comes first. The EQ-5D-5L will be performed until death or study end.

For both instruments, 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. Compliance rates for both instruments will also be summarized. Compliance rates defined as the number of questionnaires received as a percentage of the number expected per protocol-specified collection schedule are assessed for FACT-Lym and EQ-5D (EuroQol).

For the FACT-Lym, the primary interest is the lymphoma symptoms subscale (FACT-LymS), however, other FACT subscales will also be examined, including:

- The four individual domain scores of the FACT-G (Physical Well-being, Family/Social Well-being, Emotional Well-being, Functional Well-being)
- The Trial Outcome Index score (TOI): Physical and Functional domains of the FACT-G items plus FACT-LymS)
- FACT-G total score (total of Physical, Family/Social, Emotional, and Functional Well-being subscales)
- The FACT-Lym total score (FACT-G plus FACT-LymS)

In the event of missing item(s) in a FACT-Lym assessment, the subscale scores will be computed according to the FACT scoring and administration guidelines by multiplying the sum of the subscale by the number of items in the subscale, then dividing by the number of items actually answered.

Time to clinically meaningful worsening and improvement in the FACT-LymS subscale of the FACT-Lym is defined as the interval from the date of randomization to the start date of the worsening and improvement separately. Instrument validation has found the minimum important difference (MID) for the FACT-LymS to be in the range of 3 to 5 points, with 5 being the more conservative [9, 10]). Worsening and improvement is defined by a 5 or greater point reduction and increase from baseline in the Lym lymphoma subscale, respectively [9, 11]. Death and missing data due to being very ill as noted on the eCRF will also be considered as worsening. Subjects who have not met the definition of worsening will be censored at the last FACT-Lym assessment. Subjects with no worsening event and with no baseline assessment or on-study assessment will be censored at the date of randomization. Time to worsening in FACT-LymS will be estimated using Kaplan-Meier methods. 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. The best and worst change of the FACT-LymS TOI score and FACT-Lym total score from baseline to post-baseline may be explored using waterfall plots.

Exploratory analyses may be conducted to examine time to worsening according to clinical response status (i.e., responder vs. non-responder) or by patient baseline characteristics of interest (i.e., ECOG performance status).

For the EQ-5D-5L and FACT-LymS, a mixed effects model with repeated measures analysis will be conducted estimating the change from baseline at each time point between two treatments. ITT subjects who have a baseline value and at least 1 post-randomization value are included in the analysis. Change from baseline will be fitted to a mixed -effects model including subjects as a random effect, and baseline value, treatment indicator, time, treatment-by-time interaction, and stratification factors as fixed effects.

8. SIGNIFICANT CHANGES FROM PROTOCOL

Primary analysis is updated with the framework of estimand and summary of COVID-19 impact has been added. Post-hoc power analysis of OS has also been added.

REFERENCES

1. Cheson BD, Pfistner B, Juweid ME, et al. Revised response criteria for malignant lymphoma. Journal of Clinical Oncology 2007; 25: 579-586.
2. O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. Biometrics 1979; 35(3):549-556.
3. ICH E9 (R1) Addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials, November 2019.
4. Wiener, L.E., Ivanova A., Koch G.C. Methods for clarifying criteria for study continuation at interim analysis. Pharmaceutical Statistics 2020; 19:720–732.
5. Robins, J.M., Finkelstein, D.M. Correcting for Noncompliance and Dependent Censoring in an AIDS Clinical Trial with Inverse Probability of Censoring Weighted (IPCW) Log Rank Tests. Biometrics 2000; 56(3):779-788.
6. Robins, J.M., Tsiatis, A.A. Correcting for non-compliance in randomized trials using rank preserving structural failure time models. Communications in Statistics-Theory and Methods 1991; 20(8):2609-2631.
7. Gray, R.J. A class of *k*-sample tests for comparing the cumulative incidence of a competing risk. Annals of Statistics 1988; 16, 1141-1154.
8. Fine, J.P. and Gray, R.J. A proportional hazards model for the subdistribution of a competing risk. Journal of the American Statistical Association 1999; 94, 496-509.
9. Carter GC, Liepa A, Zimmerman AH, Morschhauser F. Validation of the Functional Assessment of Therapy – Lymphoma (FACT-LYM) in patients with relapsed/refractory mantle cell lymphoma. Poster presented at: Annual Meeting of American Society of Hematology, December 7, 2008; San Francisco, California.
10. Hlubocky F.J., Webster K., Cashy J., Beaumont J., and Cella D., "The Development and Validation of a Measure of Health-Related Quality of Life for Non-Hodgkin's Lymphoma: The Functional Assessment of Cancer Therapy—Lymphoma (FACT-Lym)," Lymphoma, vol. 2013, Article ID 147176, 9 pages, 2013. doi:10.1155/2013/147176.
11. Cella D, Webster K, Cashy J, et al. Development of a Measure of Health-Related Quality of Life for Non-Hodgkin's Lymphoma Clinical Research: The Functional Assessment of Cancer Therapy – Lymphoma (FACT-Lym) [abstract]. ASH. 2005; 106. Abstract 750.

ATTACHMENT 1: VISIT WINDOW FOR PRO ANALYSIS

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

Week Number Pre-PD	**Start Day**	**Target Day**	**End Day**
Baseline (weeks 1-4)	-30	1	1
Cycle 2 (weeks 5-8)[a]	2	29	42
Cycle 3 (weeks 9-12)[a]	43	57	70
Cycle 4 (weeks 13-16)[a]	71	85	98
Cycle 5 (weeks 17-20)[a]	(X-1)*28-13=99	(X-1)*28+1=113	(X-1)*28+14=126
Cycle 6 (weeks 21-24)[a]	127	141	183
Every 12 weeks up to 12 months			
Week 33[b]	183	225	266
Week 45[b]	(Y-1)*7-41=265	(Y-1)*7+1=309	(Y-1)*7+42=350
Week 57[b]	349	393	449
Every 16 weeks thereafter			
Week 73[c]	449	505	560
Week 89[c]	(Y-1)*7-55=561	(Y-1)*7+1=617	(Y-1)*7+56=672
Week 105[c]	673	729	784
Week 121[c]	785	841	896
Week 137[c]	897	953	1008
Week 153 and up [c]	1009	1065	1120
Week Number Post-PD[d]	**Start Day**	**Target Day**	**End Day**
Week 16[e]	50	106	161
Week 32[e]	(Y-1)*7-55=162	218	(Y-1)*7+56=273
Week 48[e]	274	330	385

a. Target day in cycle = (X-1)*28+1 where X is cycle number, (X-1)*28-13 for start day and (X-1)*28+14 for end day.
b. Target day in week= (Y-1)*7+1 where Y is week number, (Y-1)*7-40 for start day and (Y-1)*7+41 for end day.
c. Target day in week = (Y-1)*7+1 where Y is week number, (Y-1)*7-55 for start day and (Y-1)*7+56 for end day.
d. At PD, sites should attempt to administer the EQ-5D-5L in survival follow-up every 16 weeks (up to 3 times), unless death or study end occurs first.
e. Using date of PD as the relative date for counting number of weeks: Target date post-PD in week = (Y-1)*7+1 where Y is week number, (Y-1)*7-55 post-PD for start day and (Y-1)*7+56 post-PD for end day.