pharmacyclics®
An AbbVie Company

Statistical Analysis Plan

A Multicenter, Randomized, Double-blind, Placebo-controlled Phase 3 Study of the Bruton's Tyrosine Kinase (BTK) Inhibitor, Ibrutinib, in Combination with Rituximab versus Placebo in Combination with Rituximab in Treatment Naïve Subjects with Follicular Lymphoma

PCYC-1141-CA

May 13, 2024

Version 7.0

The information in this document is considered privileged and confidential by Pharmacyclics LLC/AbbVie Inc, and may not be disclosed to others except to the extent necessary to obtain Institutional Review Board approval and informed consent, or as required by Federal and State Laws. Persons to whom this information is disclosed must be informed that this information is privileged and confidential and that it should not be further disclosed.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AE	adverse event
ATC	Anatomical Therapeutic Chemical
BOR	best overall response
CI	confidence interval
CMH	Cochran-Mantel-Haenszel
CR	complete response
CRF	case report form
CSR	clinical study report
CTCAE	Common Terminology Criteria for Adverse Events
ECOG	Eastern Cooperative Oncology Group
EAIR	Exposure-adjusted incidence rate
DMC	data monitoring committee
DOCR	duration of complete response
DOR	duration of response
FL	follicular lymphoma
HR	hazard ratio
ITT	intent-to treat
IXRS	interactive voice or web response system
MedDRA	Medical Dictionary for Regulatory Activities
MRD	Minimal residual disease
NCI	National Cancer Institute
ORR	overall response rate
OS	overall survival
PCYC	Pharmacyclics
PD	progressive disease
PFS	progression-free survival
PK	pharmacokinetics
PR	partial response
PT	preferred term
QoL	quality of life
SAP	statistical analysis plan
SOC	system organ class
TEAE	treatment-emergent adverse events
TTP	time to progression

1 INTRODUCTION

The statistical analysis plan (SAP) version 7.0 is based on the protocol amendment 4 dated February 07, 2023. This SAP defines key elements including variable definitions, and statistical methods for analysis of data in evaluation of efficacy and safety of PCYC-1141-CA. Analyses of biomarker and pharmacokinetics (PK) data are not in the scope of this document.

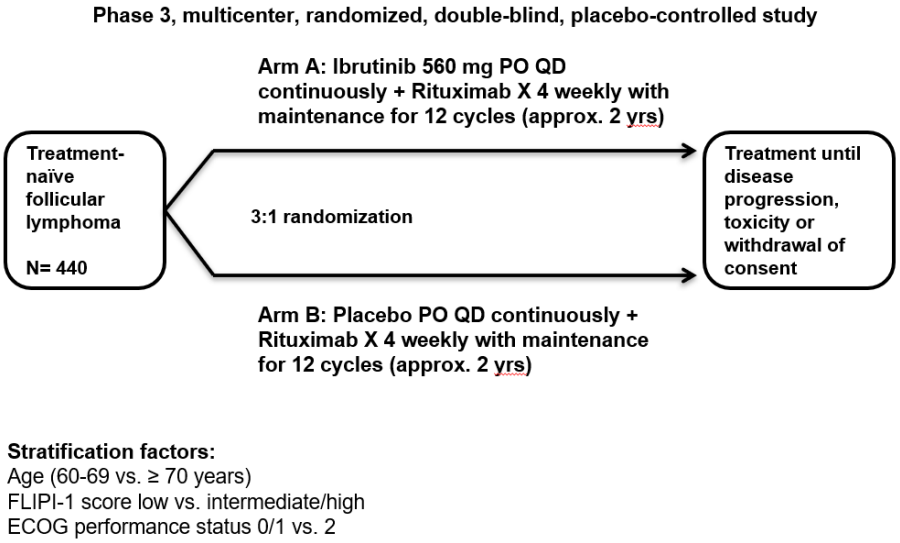
Throughout this SAP, "study treatment" and "study drug" are used interchangeably, and both are referring to ibrutinib, placebo, and/or rituximab.

Analysis methods specified in this document take precedence over those described in the protocol should there be any differences.

1.1 Study Design

This is a randomized, double-blind, placebo-controlled Phase 3 study designed to assess the efficacy and safety of ibrutinib in combination with rituximab compared to placebo in combination with rituximab in treatment naïve subjects with follicular lymphoma (FL).

Figure 1. Study Design



This study evaluates ibrutinib + rituximab (Arm A) vs. placebo + rituximab (Arm B) with 2 interim analyses and the primary analysis planned when the study observes 200 progression-free survival (PFS) events, or if pre-specified boundaries are crossed at an interim analysis. The final analysis for OS is planned when 165 OS events are observed or 5 years after the enrollment of the last patient, whichever is earlier.

1.2 Endpoints

1.2.1 Primary Endpoint

The primary efficacy endpoint is progression-free survival (PFS) as assessed by the investigator. PFS is defined as the time from date of randomization to the date of the first documented progressive disease (PD), as defined on the Revised Response Criteria for Malignant Lymphoma (Cheson 2014, Lugano Classification), or death due to any cause, whichever occurs first.

The estimand corresponding to the primary efficacy endpoint (PFS per investigator assessment) is the treatment effect (hazard ratio: ibrutinib + rituximab versus placebo + rituximab) in all randomized subjects with treatment naïve FL regardless of premature discontinuation of study treatment, as measured by the earlier occurrence of progressive disease per the investigator assessment or death, whichever occurs first (Table 1). The details of handling strategies of the intercurrent events in the PFS primary and supplementary analyses for US and for Other Regions are in Table 2.1 and Table 2.2.

Table 1. Attributes of the Estimand for PFS Analyses

Estimand Label	Attributes of the Estimand				
	Treatment	**Population**	**Endpoint**	**Intercurrent Events**	**Statistical Summary**
PFS (inv)	Ibrutinib + rituximab or placebo + rituximab	all randomized subjects with treatment naïve FL	PFS per investigator assessment	1. Premature discontinuation of study treatment 2. Use of subsequent anti-cancer therapy 3. Death due to COVID-19 without PD	Hazard ratio of PFS between two treatment arms. Kaplan-Meier estimates of the distribution of PFS for each treatment arm, median PFS and their associated CIs.

Table 2.1. Handling Strategies for the Intercurrent Events

Intercurrent Events	Handling Strategy for Addressing Intercurrent Events
Premature discontinuation of study treatment	**Treatment policy strategy:** Ignore the intercurrent events and use time to PD or death, whichever occurs first regardless of discontinuation of study treatment.
Use of subsequent anti-cancer therapy	**Hypothetical strategy:** Subjects are censored at the last adequate disease assessment showing no evidence of PD before the use of subsequent anti-cancer therapy. **Treatment policy strategy:** Ignore the intercurrent events and use time to PD or death, whichever occurs first regardless of use of subsequent anti-cancer therapy.
Death due to COVID-19	**Composite variable strategy:** Consider COVID-19 death without PD as a PFS event. **Hypothetical strategy:** Subjects are censored at the last adequate disease assessment before death due to COVID-19 without PD.

Table 2.2 Intercurrent Events Handling Strategies for US and Other Regions

Intercurrent Events	US & Other Regions
Premature discontinuation of study treatment	Treatment policy strategy
Use of subsequent anti-cancer therapy	Primary analysis: Hypothetical strategy Supplementary analysis: Treatment policy strategy
Death due to COVID-19	Primary analysis: Composite variable strategy Supplementary analysis: Hypothetical strategy

For the PFS primary analysis for US regulatory purposes, in addition to the estimand's strategies for handling intercurrent events described in Table 2.2. Subjects who missed two or more consecutive overall disease assessments will be censored at the last adequate disease assessment before the missing interval, regardless of whether or not a PFS event subsequently occurs (see Appendix 6.1).

Adequate disease assessment is defined as having sufficient evidence (e.g., imaging to support the findings) to correctly indicate that progression has or has not occurred (i.e. overall disease assessment is CR, PR, SD, PD).

Details are specified in Table 6 of Section 4 including sensitivity/supplementary analyses.

1.2.2 Secondary Endpoints

Secondary endpoints include the following:

- Overall response rate (ORR) as assessed by investigator (Cheson 2014)
- Overall survival (OS)
- Infusion-related reaction rate (IRR)
- Duration of response (DOR) as assessed by investigator
- Frequency, severity, seriousness, and relatedness of adverse events (AEs)

The estimands corresponding to the key secondary efficacy endpoints (ORR and OS) are:

- The estimand for the secondary endpoint, ORR, is the treatment effect (rate ratio: ibrutinib + rituximab versus placebo + rituximab) in all randomized subjects with treatment naïve FL, as measured by the best response for each subject per investigator.
- The estimand corresponding to OS is the treatment effect (hazard ratio: ibrutinib + rituximab versus placebo + rituximab) with respect to overall survival in all randomized subjects with treatment naïve FL, as measured by time from randomization to death of any causes.

1.2.3 Exploratory Endpoints

- Health-related quality of life (QOL) as measured by the FACT-Lym
- Pharmacokinetic parameters or metrics of systemic exposure of ibrutinib and rituximab
- Evaluation and/or identification of relevant patient populations defined by biomarker(s)
- Minimal residual disease (MRD)-negative rate defined as the proportion of FL subjects who achieve a CR and who reach MRD-negative disease status
- Complete response rate at 30 months (CR30) as defined by the International Working Group (IWG) Revised Response Criteria for Malignant Lymphoma (Cheson 2014)
- Time to Progression (TTP) as assessed by investigator
- Duration of Complete Response (DOCR) as assessed by investigator

1.3 Sample Size Determination

Sample Size Consideration for PFS

A sample size of approximately 440 subjects (actually enrolled 445 subjects) were randomized to Arms A and B to evaluate whether treatment in Arm A will result in a prolongation of progression-free survival (PFS) when compared to Arm B in treatment naïve subjects with follicular lymphoma. The sample size was determined according to the Group Sequential Design with 2 interim analyses. For this calculation, the references used for the assumptions were Hainsworth 2002, Ghielmini 2004, Zucca 2019, and Fowler 2020. Assuming exponential survival distribution for the PFS events and 66.7% improvement in median PFS of Arm A over Arm B (a hazard ratio [HR] of 0.60), the study has approximately 87% power with 200 observed PFS events using an overall 1-sided significance level of 0.025. For the sample size calculation, a median PFS of 28 months in the control arm, and the actual enrollment rates using a 3-month average until November 2020 and a projected enrollment rate of 8 subjects per month afterwards were used. Two interim analyses were planned with the first interim analysis (IA1) for futility and the second interim analysis (IA2) for efficacy.

IA1 was planned when approximately 55% of the 200 PFS events had occurred. The futility boundary was based on a Rho parameter of 5. Following review of the data at IA1, the DMC suggested continuing the study without modification.

IA2 was planned after approximately 75% of the 200 PFS events (ie, 150 PFS events) were observed and after the last subject had been enrolled, and at least 90% of the subjects had had the opportunity to be followed on study for at least 1 year. The 1-sided α 0.025 for PFS was split: 0.001 allocated to PFS IA2 and 0.024 allocated to primary analysis for PFS, following the Bonferroni rule, Table 3.1. Following review of the data at IA2, the DMC suggested continuing the study without modification.

Power Consideration for ORR

For ORR if we assume a response rate of 0.57 (Zucca 2019) in Arm B and a response rate of 0.73 (PCYC-1125-CA; Fowler 2020) in Arm A, then the power will be approximately 85% for a 2-sided α of 0.05 and a sample size of 440. One interim analysis was planned for ORR. However, since the enrollment has been completed as of July 1, 2021, there is no need to conduct an interim analysis. The primary analysis (i.e., final analysis) of ORR will be carried out at the time of the primary analysis of PFS following the sequential testing procedure, Table 4.1.

Power Consideration for OS

The final analysis for OS is planned when 165 OS events are observed or 5 years after the enrollment of the last patient, whichever is earlier. The study has ~80% power for the OS

analysis at an overall 1-sided alpha of 0.025 when 165 OS events have occurred. An interim analysis for OS will be conducted at the time of the primary analysis for PFS when both PFS and ORR have crossed the superiority boundaries. The OS analysis design is based on the assumption that the 36-month OS landmark estimate is 73% for Arm B (Batlevi 2020) and 82.8% for Arm A (PCYC-1125-CA; Fowler 2020), i.e., a HR of 0.6 and median OS for the control arm of 79.3 months. The 1-sided alpha efficacy boundaries at the interim analysis for OS (at the time of the primary analysis for PFS) and the final analysis for OS were computed based on the Haybittle-Peto boundary ensuring control of overall type-1 error, Table 3.2. The Sponsor may consider extending follow-up to attain more OS events to account for the impact of COVID-19 (FDA guidance, June 2020).

A separate futility analysis for OS was planned to be conducted at the time of the PFS IA2. Another futility analysis will be conducted at the time of the PFS PA. These OS futility assessments will not be included in the sequential testing procedure and will be handled separately based on the futility boundary considerations specified in the section below.

1.4 Planned Analyses

1.4.1 Interim Analyses

Two interim analyses for PFS were planned, and the alpha spending function from the Rho family for futility (IA1) and Bonferroni rule for efficacy (IA2) were planned to be used for the boundaries. The first interim analysis (IA1) for futility including subjects in treatment Arms A and B was planned when approximately 55% of the 200 PFS events had occurred. The futility boundary was based on a Rho parameter of 5. IA1 was carried out by an independent statistical group, and results were reviewed by the DMC at the meeting on December 11, 2020. As a result, the DMC recommended to continue the study without modification.

The second interim analysis (IA2) for efficacy was planned after approximately 150 PFS events (75% of information fraction) were observed. Based on prior FDA advice (response to FDA information request dated December 10, 2020), this analysis was timed to ensure >90% of subjects had had the opportunity to be followed on study for 1 year or longer at the time of database lock. Per FDA guidance to account for the impact on PFS due to COVID-19-related deaths without prior disease progression, the follow-up time was also planned to be extended to attain more PFS events (FDA guidance, June 2020). The futility and efficacy boundaries were generated using EAST version 6.5 and summarized in Table 3.1 below.

Primary endpoint, PFS, and secondary endpoints, ORR, OS and infusion-related reaction rate (IRR) were planned to be analyzed at IA2 following the testing procedure described in Table 4.1 and Table 4.2. The secondary endpoint OS was planned to be analyzed at IA2 for futility assessment only. Following review of the data at IA2, on July 22, 2022, the DMC suggested continuing the study without modification.

Table 3.1. Stopping Boundaries for PFS

Planned PFS Analysis	Information Fraction (IF)	Events	Cum. α Spent	Boundary (1-sided p-value)	
				Efficacy	Futility
Interim Analysis 1 (IA1)	0.55	110	0	NA	0.5596†
Interim Analysis 2 (IA2)*	0.75	150	0.001	0.001#	NA
Primary Analysis (PA)	1	200	0.025	0.024#	0.024

* IA2 planned after 75% IF, last subject enrolled, and 90% of the subjects reaching at least 1 year of study follow up.
\# Efficacy boundary was based on Bonferroni rule splitting the 1 sided α 0.025: 0.001 to IA2 and 0.024 to PA.
† Futility boundary was based on Rho family with parameter 5.

An interim analysis for OS will be conducted at the time of the primary analysis for PFS when both PFS and ORR have crossed the superiority boundaries. Assuming the final OS analysis is based on 165 events, the significance level of the final analysis for OS will be adjusted for the OS interim analysis based on the Haybittle-Peto boundary. The boundaries were generated using EAST version 6.5 and summarized in Table 3.2 below.

For the futility analyses (non-binding) for OS at the time of IA2 and at the time of PFS PA, the hazard ratio futility boundary family was used with HR 1.25 and HR 1.0, respectively, and were calculated using EAST Version 6.5, Table 3.2. The number of OS events was anticipated to be too small to warrant testing for superiority at the IA2 timing. The proof-of-concept (POC) criteria proposed by Wiener et al 2020 and Saville et al 2011 were considered in generating the OS futility boundaries. The following assumptions were used to generate the OS futility boundaries:

- Median OS for control arm 79.29 months per protocol assumption
- HR 0.60 per protocol assumption
- The information fraction for the OS interim futility analyses at the time of IA2 and at the time of the PFS PA are estimated to be ~52% (i.e., ~85 events) and ~85% (i.e., ~140 events), respectively, of the total events of 165 events at the time of OS Final Analysis.
- POC 1-sided kappa 0.25 & overall POC power at least 0.98 are selected to assess a directional trend for futility (no detriment).

Both interim analyses IA1 and IA2 were performed by the independent Statistical Support Group (Cytel), and the results were submitted to the DMC.

Table 3.2. Stopping Boundaries for OS

Planned OS Analysis (timing)	Info. Fraction	Events	Cum. α Spent	Boundary	
				Efficacy (1-sided p-value)	Futility (HR)
Futility Analysis (at PFS IA2)	0.52	85	0	NA	1.25#
Interim Analysis (at PFS PA)	Agnostic	Agnostic	0.005	0.005	1.0#
Final Analysis (FA)	1	165 (assumed)	0.025	0.024927*	0.702858*

* Actual Haybittle Peto boundary will be calculated depending on the observed information fraction at OS IA.
# HR futility boundary family based on proof of concept criteria by Wiener et al 2020 and Saville et al 2011.

1.4.2 Primary Analysis for PFS and Interim Analysis for OS

The protocol effect size (HR) assumptions for PFS did not consider the potential impact of COVID-19 deaths when the study was initiated. In alignment with the FDA guidance[7] (Statistical Considerations for Clinical Trials During the COVID-19 Public Health Emergency, June 2020), in order to mitigate the potential impact on the PFS endpoint due to COVID-19, the follow-up will be extended and the primary PFS analysis will be conducted after sufficient PFS events are accrued considering the number of COVID-19 deaths before progressive disease. An OS interim efficacy analysis will be conducted at the time of the primary analysis for PFS when the superiority boundaries for PFS and ORR are crossed. Table 4.1, Table 4.2, and Figure 2 show the testing procedure and alpha spending for efficacy. A separate futility analysis for OS will be conducted at the time of the primary analysis for PFS, Table 3.2 and Figure 2.

1.4.3 Final Analysis for OS

The final analysis for OS will be conducted when 165 OS events are reached or 5 years after enrollment of the last subject, whichever is earlier. If the OS final analysis is significant, then IRR will be tested.

1.5 Overall Type-I Error Control

The one-sided overall type 1 error will be controlled at 0.025 by a closed testing procedure for testing the primary endpoint (PFS) and key secondary endpoints. The key secondary endpoints will be tested only if the primary endpoint (PFS) reaches statistical significance and will be ranked and tested at a 1-sided statistical significance of 0.025 sequentially in the hierarchical order listed in Table 4.1 and Table 4.2.

Table 4.1. Rank Order of the Endpoints and Analysis Timing

Rank Order	Endpoint	IA1 for PFS	IA2 for PFS	Primary Analysis for PFS	Final Analysis for OS
1	PFS	Futility	0.001	0.024	
2	ORR	No test	0.025 if PFS is SS at IA2	0.025 if PFS is not SS at IA2 but SS at PA.	
3	OS	No test	Futility	0.005 if PFS and ORR are SS at IA2 or PA.	0.024927* if PFS and ORR are SS at PA.
4	IRR	No test	No test	0.025 if PFS, ORR, and OS are all SS.	0.025 if PFS, ORR, and OS are all SS.

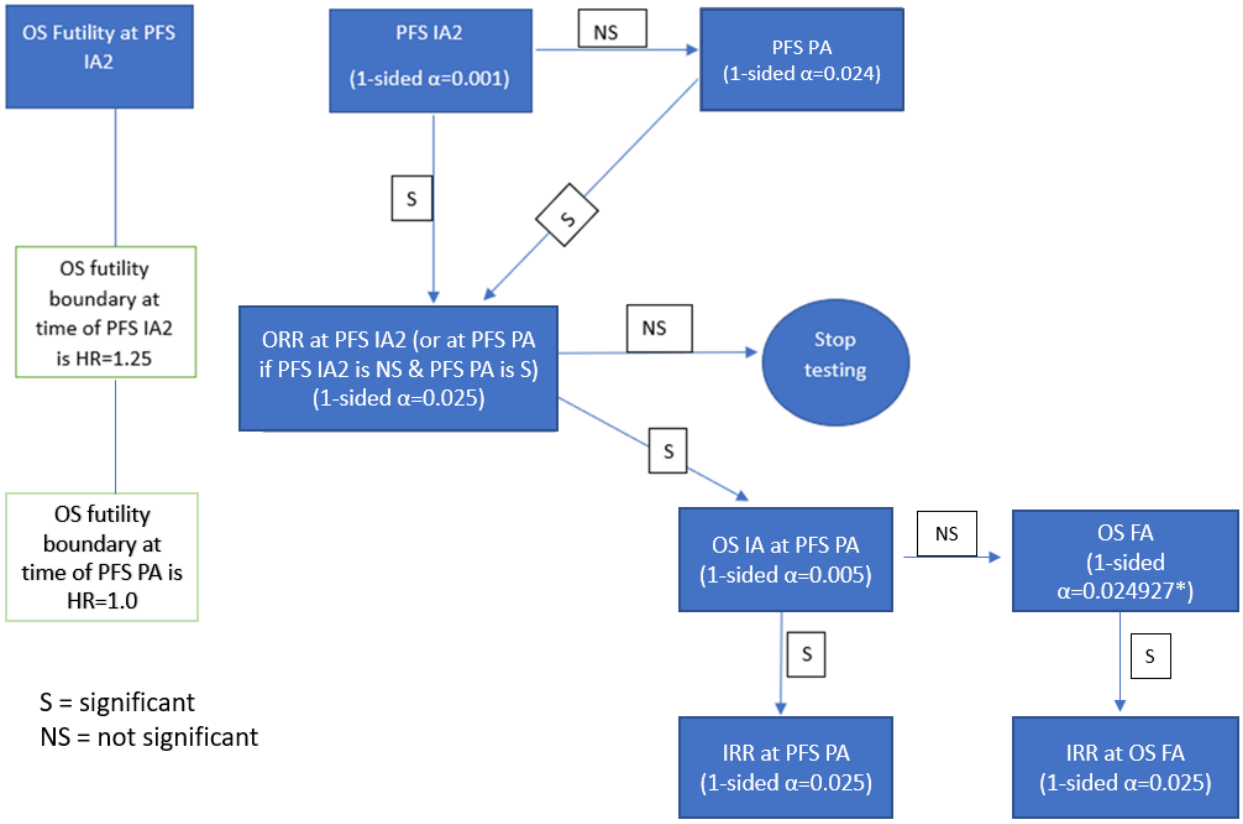
* Actual alpha will be calculated depending on the observed information fraction. SS statistically significant.

Table 4.2. One-Sided Alpha Spending by Testing Scenario

Scenario	PFS IA2	ORR at PFS IA2	PFS PA	ORR at PA	OS IA at PA	OS Final Analysis
1	SS at α=0.001	SS at α=0.025	No test - descriptive	No test - descriptive	Not SS at α=0.005	Test at α=0.024927*
2	SS at α=0.001	SS at α=0.025	No test – descriptive	No test – descriptive	SS at α=0.005	No test - descriptive
3	SS at α=0.001	Not SS at α=0.025	No test - descriptive	No test	No test	No test
4	Not SS at α=0.001	No test	SS at α=0.024	SS at α=0.025	Not SS at α=0.005	Test at α=0.024927*
5	Not SS at α=0.001	No test	SS at α=0.024	SS at α=0.025	SS at α=0.005	No test - descriptive
6	Not SS at α=0.001	No test	SS at α=0.024	Not SS at α=0.025	No test	No test
7	Not SS at α=0.001	No test	Not SS at α=0.024	No test	No test	No test

* Actual alpha will be calculated depending on the observed OS information fraction at PA. SS statistically significant.

Figure 2. Testing Procedure



* Actual alpha will be calculated depending on the observed OS information fraction at PA.

1.6 Blinding and Randomization Methods

1.6.1 Blinding Method

This is a double-blind Phase 3 study. Subjects, investigators, and the sponsor's study team members will remain blinded to treatment assignment until the primary analysis for PFS is completed. Examples of personnel who may be unblinded during the study are:

- The independent DMC, and the independent biostatistician and statistical programmers from an independent Statistical Support Group who are responsible for preparing interim tables, listings, and graphs for DMC review. Unblinding procedures and the control of the unblinded data are described in the DMC charter.
- Sponsor's representative responsible for pharmacokinetics testing and analysis.
- Sponsor safety representative to fulfill regulatory reporting requirements for suspected unexpected serious adverse reactions.

- In case of an urgent safety concern, site personnel and the sponsor may be unblinded, if treatment assignment information is needed to determine further actions to address the urgent safety concern (e.g., life-threatening event, medication error, such as an accidental overdose).

1.6.2 Randomization Method

Central randomization was implemented in this study. Randomization was used to minimize bias in the assignment of subjects to treatment groups, to increase the likelihood that known and unknown subject attributes (e.g., demographic and baseline characteristics) are evenly balanced across treatment groups, and to enhance the validity of statistical comparisons across treatment groups.

Subjects were randomized at a 3:1 ratio to receive either Treatment Arm A (ibrutinib/rituximab) or Treatment Arm B (placebo/rituximab) stratified by the following factors.

- Age group (60-69 vs. ≥ 70 years)
- FLIPI-1 score at Screening (low vs. intermediate/high)
- ECOG performance status score (0/1 vs. 2)
- Geographic region (North America versus Rest of World)

The interactive voice or web response system (IXRS) assigned a unique treatment code, which dictated the treatment assignment and matching study treatment kit for the subject. The requestor had to use his or her own user identification and personal identification number when contacting the IXRS and then provided the relevant subject details to uniquely identify the subject.

2 GENERAL ANALYSIS CONSIDERATION

Subjects will be analyzed and summarized by treatment as randomized for efficacy endpoints and by actual treatment received for safety analyses. For the exploratory endpoints, analyses will be carried out and results summarized only if the data provide clinically meaningful information.

2.1 Analysis Sets

Intent-to-Treat Population

The intent-to-treat (ITT) population consists of all subjects randomized. The ITT population will be the primary population for all efficacy analyses for the analysis unless specified otherwise.

Safety Population

The safety population consists of all subjects who received at least 1 dose of study treatment (rituximab, ibrutinib/placebo).

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. However, if the sample size does not warrant any statistical analyses, then descriptive summaries will be provided. In addition, the subgroup variables and the cutoff values are subject to change if warranted to better represent the data. Table 5 provides a list of subgroups.

Table 5. Subgroup Definition

Subgroup (by baseline characteristics)	Definition of Subgroup	Analysis Type
Age	< 65, ≥ 65 and 60-69, ≥ 70 and < 75, ≥ 75	E, S
Sex	Male, Female	E, S
Race	White, Non-White	E, S
Region	US, Non-US	E, S
ECOG as recorded on CRFs	0-1, 2	E
WHO criteria grade	1, 2, 3a	E
Ann Arbor stage	II, III, IV	E
FLIPI-1 score as recorded on CRFs	Low, intermediate/high	E
Number of nodal areas involved	≤ 4, > 4	E
Creatinine clearance	< 60, ≥ 60 mL/min	S
Liver function based on NCI criteria	Normal, Abnormal	S

ECOG: eastern cooperative oncology group, CRF: case report form, FLIPI: follicular lymphoma international prognostic index.
E: efficacy (PFS) S: safety (adverse events)

3 SUBJECT INFORMATION

This section provides a general outline of the subject information that will be included in the summary tables.

3.1 Subject Disposition

Subject enrollment will be summarized by geographic region, country and site. Subject disposition for each study drug and for study participation will be tabulated. Overall treatment duration and time on study will be summarized by treatment arm.

The disposition tables will include the following summaries by treatment and overall.

- Analysis populations (all subjects)
- Enrollment by region, country, and investigator (ITT population)
- Summary of randomization stratification per IXRS (ITT population)
- Study treatment disposition and discontinuation (ITT population)
- Disposition of study (ITT population).

The Kaplan-Meier estimates will be calculated to estimate the time on study using reversed censoring from the OS analysis.

3.2 Demographics and Baseline Characteristics

Subject demographics, baseline characteristics and disease characteristics will be summarized with descriptive statistics for the ITT population by treatment arm and overall. Categorical variables will be summarized with the number and percentage of subjects. Continuous variables will be summarized with descriptive statistics (number of non-missing observations, mean and standard deviation, median, minimum, and maximum).

3.3 Concomitant Medications

Medications will be coded to Anatomical Therapeutic Chemical (ATC) class and the preferred drug name (hereafter referred as "preferred name") per World Health Organization (WHO) drug dictionary.

Concomitant medications will be summarized by ATC class and preferred term (PT) for each treatment arm in the safety population. The summarization includes all the concomitant medications taken at any time while on study treatment (i.e., from the date of first dose through the date of last dose of the study treatment). Each subject will be counted once for each PT, and each ATC class. The following concomitant medications will be summarized separately.

- CYP3A inhibitors and inducers
- Anticoagulants and antiplatelet agents
- Blood supportive products and immunoglobulin
- Growth factors

3.4 Extent of Exposure to Study Treatment

Exposure to study treatment will be summarized by treatment arm for the safety population. Descriptive statistics for the following data will be provided for Ibrutinib/Placebo and Rituximab respectively.

Ibrutinib/Placebo: treatment duration (month), cumulative dose received (g), average daily dose (mg/day), relative dose intensity (%), number (%) of subjects with dose reduction due to adverse events (AEs), number (%) of subjects with dose adjustment due to CYP inhibitor.

Rituximab: treatment duration (month), number (%) of infusion received, cumulative dose received (g), average dose level per administration (mg/infusion), relative dose intensity (%), number (%) of subjects who had any missed dose or unintended dose level due to AE, number (%) of subjects with any infusion related reaction, number (%) of subjects with any infusion interruption.

3.5 Subsequent Anti-Cancer Treatment

Subsequent anti-cancer treatment will be summarized by treatment arm for the safety population.

3.6 Visit Impact Due to COVID-19

Visit impact due to the COVID-19 pandemic will be summarized by the primary reason for the impact to the visit and the type of visit recorded on the Visit Impact CRF page for each treatment arm. Information for treatment discontinuation and study exit during the COVID-19 pandemic will be listed.

4 ANALYSIS FOR EFFICACY ENDPOINTS

Analysis of endpoints will be conducted on the ITT population unless specified otherwise. To control the Type I error rate, secondary endpoints will be tested using the fixed sequence testing procedure if PFS shows superiority. The secondary endpoints will be ranked in the following hierarchical order: ORR, OS, infusion related reaction rate. For subgroup, sensitivity, and exploratory analyses, only the analyses that provide meaningful information will be presented in the clinical study report (CSR). Supplementary analyses will be conducted to address the impact of intercurrent events. Table 6 lists the efficacy endpoints and analysis methods.

The 4 stratification factors used in the randomization as recorded in the IXRS are age group (60-69 vs. ≥70 years), FLIPI-1 score (low vs. intermediate/high), ECOG performance status score (0/1 vs. 2) and Geographic region (North America versus Rest of World). In order to avoid having many small strata "Geographic Region" will not be included in the stratified efficacy analysis since it is not considered a prognostic factor. In addition, since age is part of FLIPI-1 score, correlated with ECOG performance status score, and forms several small strata, thus based on the clinical importance, age will not be included in the stratified efficacy analysis as one of the stratification factors. For the remaining factors, FLIPI and ECOG, small strata are combined to form bigger strata with ≥10 events and ≥40 patients in each stratum.

The investigators will perform the overall disease assessment based on the Revised Response Criteria for Malignant Lymphoma (Cheson 2014, Lugano Classification); referred to as Cheson 2014 or Lugano Classification Criteria. An Independent Review Committee (IRC) will be implemented to conduct the overall disease assessments based on the same criteria.

The overall disease assessment per the investigators will be used to conduct the primary efficacy analyses, and the overall disease assessment per IRC will be used to conduct sensitivity analyses.

Table 6. Definitions and Analyses for Endpoints

Endpoint	Definition	Analysis Method
Primary Endpoint		
PFS assessed by investigator	Time from the date of randomization to the date of the first documented disease progression (PD) or death due to any cause, whichever occurred first, prior to the initiation of subsequent anti-cancer therapy, regardless of premature discontinuation of study treatment. Stem cell transplant is not considered a subsequent anti-cancer therapy for PFS. For the evaluation of disease progression, an adequate post-baseline assessment is defined as an assessment where there is enough evidence to indicate the subject had or had not progressed based on Cheson 2014 (Lugano Classification Criteria). Subjects without baseline or post-baseline disease assessment will be censored at the date of randomization.	Primary Analysis: Subjects without a PFS event will be censored at the date of the last adequate post-baseline assessment showing no evidence of PD. Stratified log-rank test is the primary analysis comparing treatment effects. Cox proportional hazards model stratified by the pre-specified stratification factors will be used to estimate hazard ratio (HR) and its associated confidence interval (CI). In addition, Kaplan-Meier estimates and median PFS with its associated CI will be displayed. Brookmeyer-Crowley confidence interval based on the log-log-transformed Greenwood variance estimate will be calculated for median PFS (mPFS); KM estimates at selected landmark points will also be provided. Handling strategies to address intercurrent events in Section 1.2.1 will be implemented. Primary Analysis Censoring Rules A. Treatment policy strategy will be used to handle the intercurrent events of discontinuation of study treatment; i.e ignore the intercurrent events and use time to PD or death, whichever occurs first regardless of discontinuation of study treatment. B. Hypothetical strategy will be used to handle the intercurrent events of subjects who received anti-cancer treatment; i.e. subjects who initiated subsequent anti-cancer treatment will be censored at the last adequate post-baseline disease assessment showing no evidence of PD before the use of subsequent anti-cancer therapy.

		C. Composite variable strategy will be used to handle the intercurrent events of death due to COVID-19 without PD; i.e. consider COVID-19 deaths without PD a PFS event. D. Subjects who missed two or more consecutive overall disease assessments will be censored at the last adequate post-baseline disease assessment prior to the first missed disease assessment, regardless of whether or not a PFS event subsequently occurs. The algorithm to define missing 2 or more consecutive assessments is described in Appendix 6.1. This particular censoring rule will be applied to the primary analysis for US regulatory purposes only. The Sponsor will extend follow-up time to accrue additional PFS events to make up for the potential impact of COVID in accordance with FDA`s COVID-19 guidance (FDA guidance, June 2020). <u>Sensitivity/Supplementary Analyses:</u> 1. Non-stratified log-rank test and non-stratified Cox regression model. Same censoring convention as the primary analysis. 2. Change on the estimand's strategy for handling death due to COVID-19 without PD from composite variable strategy to hypothetical strategy; i.e. subjects who died due to COVID-19 without PD will be censored at the last adequate post-baseline disease assessment. 3. Use time to PD or death regardless of whether subjects missed any overall disease assessments. This will be applied to the primary analysis for Other Regions and is included as a sensitivity analysis for the US. 4. PFS analysis based on the ITT principle. Ignore the intercurrent events (including the use of subsequent anti-cancer therapy) and use time to PD or death regardless of whether subjects missed

Endpoint	Definition	Analysis Method
		any overall disease assessments. 5. PFS assessed by Independent Review Committee (IRC); conduct same analyses as specified for the primary and sensitivity/supplementary analyses for PFS assessed by investigator. Subgroup analysis: Hazard ratio and its 95% CI based on unstratified Cox regression model for each subgroup. A forest plot will be provided. Same censoring convention as the primary analysis.
Secondary Endpoints		
ORR assessed by investigator	The proportion of subjects achieving a best overall response of CR or PR based on Cheson 2014 (Lugano Classification Criteria) per investigator assessment at or prior to initiation of subsequent anti-cancer therapy. Subjects who did not have any post-baseline disease assessments or initiated subsequent anti-cancer therapy prior to initial CR/PR will be considered non-responders when summarizing ORR.	CMH chi-square test controlled for the pre-specified stratification factors, rate ratio and 95% CI. Sensitivity Analysis: ORR by IRC: same analysis as above. A forest plot will be provided for subgroups.
DOR	DOR is defined as the time from initial CR or PR to PD or death due to any cause, whichever is first reported, regardless of discontinuation of study treatment. If such event does not occur, then subjects will be censored at the last adequate disease assessment as required for PFS censoring.	Duration of response (DOR) by investigator: Kaplan-Meier estimates and associated CIs. Censoring rules for DOR will follow the censoring rules as the PFS primary analysis for US and Other Regions, respectively. Sensitivity Analyses: Same as the PFS sensitivity censoring rules for US and Other Regions, respectively. DOR by IRC: same analysis as above
OS	Time from the date of randomization to the date of death from any cause. Subjects who are not known to have died will be censored at the last known alive date.	Efficacy analysis is similar to the PFS primary analysis. Unstratified log-rank test, unstratified Cox regression model to estimate HR and its associated CI will be applied if the stratification factors selected for the PFS primary analysis do not warrant stratified analysis. Kaplan-Meier estimates, median OS, and associated CI. A forest plot will be provided for subgroups. Supplementary analysis: • censor subjects who died due to COVID-19 at the day prior to the death date.

Endpoint	Definition	Analysis Method
Rate of infusion related reaction	Proportion of subjects experiencing infusion related reactions (IRR) that start on the day of a rituximab infusion and are assessed as related or possibly related to rituximab.	Chi-square test, rate ratio and 95% CI. For each treatment, Clopper-Pearson method is used for 95% CI of the estimated rate.

Endpoint	Definition	Analysis Method
Exploratory Endpoints		
CR30	The proportion of subjects achieving a best overall response of CR per investigator assessment at or prior to 30 months and at or prior to initiation of subsequent anti-cancer therapy.	Chi-square test, rate ratio and 95% CI. For each treatment, Clopper-Pearson method is used for 95% CI of the estimated proportion.
MRD negativity rate	The proportion of follicular lymphoma subjects who achieve a CR and who reach MRD-negative disease status.	Chi-square test or Fisher's exact test depending on the sample size, rate ratio and 95% CI for each treatment (Clopper-Pearson method).
FACT-Lym total score and subscale scores	Change from baseline by visit regardless of initiation of subsequent anti-cancer therapy, proportion of subjects achieving clinically meaningful improvement or worsening on or prior to initiation of subsequent anti-cancer therapy, time to improvement or worsening on or prior to initiation of subsequent anti-cancer therapy. Clinically meaningful improvement is defined as an increase from baseline ≥ 7 points for FACT-Lym total score, (≥ 6 points for Trial Outcome Index, ≥ 3 points for lymphoma subscale score). Clinically meaningful worsening is defined as a reduction from baseline ≥ 5 points for lymphoma subscale score, (≥ 6 points for Trial Outcome Index.)	Descriptive summary statistics. The proportion of subjects achieving clinically meaningful improvement or worsening 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 will be analyzed using the Mixed-Effects Model Repeated Measure (MMRM). The model includes the continuous baseline scores, and categorical fixed effects of pre-specified stratification factors, treatment, time point and treatment-by-time point interaction. An unstructured variance covariance matrix will be used. 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. Time to event analyses will be conducted for the time to improvement/worsening endpoints. Subjects who have not met the definition of improvement/worsening will be censored at the last non-missing post-baseline PRO assessment prior to or on the date of initiating subsequent anti-cancer therapy.
TTP	TTP is defined as time from the date of randomization to the date of the first documented PD prior to the initiation of subsequent anti-cancer therapy, regardless of premature discontinuation of study treatment. Subjects without baseline or post-baseline disease assessment will be censored at the date of randomization. Subjects without PD will be censored at the last adequate disease assessment.	TTP by investigator: Kaplan-Meier estimates and associated CIs. <u>Censoring rules for TTP will follow the censoring rules as the PFS primary analysis for for US regulatory purposes.</u>

Endpoint	Definition	Analysis Method
DOCR	DOCR is defined as the time from initial CR to PD or death due to any cause, whichever is first reported, regardless of discontinuation of study treatment. DOCR will only be derived and analyzed for those subjects who achieved an overall response of CR. Subjects who are complete responders without PD or death will be censored at the last adequate disease assessment.	DOCR by investigator: Kaplan-Meier estimates and associated CIs. Censoring rules for DOCR will follow the censoring rules as the PFS primary analysis for for US regulatory purposes.

Note: CMH: Cochran-Mantel-Haenszel; CI: confidence interval; CR: complete response; HR: hazard ratio; ORR: overall response rate; OS: overall survival; PD: progressive disease; PFS: progression-free survival; PR: partial response; TTP: time to progression; DOCR: duration of complete response.

4.1 Exploratory Efficacy Analyses

Analyses of PFS, DOR, TTP, and DOCR for the following:

- subjects who discontinued ibrutinib for any reason other than PD vs. who did not discontinue ibrutinib
- subjects who discontinued ibrutinib for any reason other than PD after achieving BOR of SD or PR vs.CR
- subjects by ibrutinib treatment duration (< 24 months vs >24 months)
- subjects by ibrutinib treatment duration (< 24 months vs >24 months), excluding those who discontinued ibrutinib due to PD

FACT-Lym Total score change from baseline by visit will be presented for:

- subjects who discontinued ibrutinib for any reason other than PD vs. who did not discontinue ibrutinib
- subjects who discontinued ibrutinib for any reason other than PD after achieving BOR of SD or PR vs.CR

FACT-Lym Total score change from baseline by visit will be descriptively presented with the categorical effects including only subject comparison, visit, and subject comparison-by-visit interaction.

5 SAFETY ASSESSMENTS

Safety data will be summarized by treatment for the safety population unless otherwise indicated. Table 7 summarizes the safety analyses to be carried out. AEs will be coded in accordance with the Medical Dictionary for Regulatory Activities (MedDRA). Severity of AEs will be graded by the investigator according to the NCI Common Terminology Criteria for Adverse Events (NCI-CTCAE) v4.03.

The treatment-emergent period is defined as the period from the date of the first dose of study treatment up to 30 days after the date of the last dose of study treatment or the day before initiation of subsequent anti-cancer 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. TEAEs or deaths related to COVID-19 infection will be summarized or listed separately. 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. When central lab values are missing and local lab values are available, local lab values will be used. 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.

Table 7. Summary of Safety Assessments

Assessment Type	Definition	Analysis Methods
AE	TEAEs, COVID-19 TEAEs, SAEs, grade 3 or worse TEAEs, related TEAEs, TEAEs leading to treatment discontinuation, TEAEs leading to dose reduction, TEAEs leading to death, protocol- defined events of special interest and other safety observations.	Descriptive summary statistics and/or listings
Lab	Worst post-baseline toxicity grade for selected CTCAE gradable hematology and chemistry. Abnormalities in creatinine clearance, uric acid, and liver function.	Descriptive summary statistics and/or listings
Vital Signs and other Observations Related to Safety	Blood pressure, heart rate, body temperature, new or worsened eye-related symptoms	Descriptive summary statistics and/or listings

TEAE: treatment emergent adverse event; SAE: serious adverse event; CTCAE: Common Terminology Criteria for Adverse Events

5.1 Exposure-Adjusted Incidence Rate

Exposure-adjusted incide rate (EAIR) takes into account the patients' exposure time in quantifying the risk of an AE. For a given AE, EAIR is defined as the rate of the number of patients who experience at least one event over total time at risk for all subjects (Zhou, et. al.

2015; Jia, et. al. 2020). EAIR "adjusts for exposure" by accounting for the variability in patients time at risk as opposed to crude incidence rates, which divides the number of patients with events by the total number of patients. The time at risk is different for patients who have experienced the event and subjects who are event-free throughout the treatment-emergent period.

For patients with at least one event, the time at risk is the time from first dose date to the start date of the first event. The choice to restrict the time at risk to the onset of first event is due to EAIR's epidemiological origin and is based on the assumption that occurrences of multiple events are interdependent. Hence, for patients with events, the time at risk can be represented, in patient-years at risk, as:

(Start date of the first event first dose date + 1)/365.25 (equation 1)

For patients with no event, the time at risk is the time from the first dose date to the end of the treatment-emergent period which can be represented, in patient-years at risk, as:

(Last dose date + 30 days first dose date + 1)/365.25 (equation 2)

The *total patient-years at risk* is the sum of the patient-years at risk for all subjects. For patients with events, patient-years at risk is calculated using equation 1, while for patients with no event, patient-years at risk is calculated using equation 2.

The *100 patient-years at risk* is the total patient-years at risk divided per 100 patients and can be interpreted as the average of length time (in years) to observe the first onset of AE if 100 patients were treated for 1 year.

For a given AE, the EAIR is represented as the number of patients with events per 100 patient-years at risk:

*(Number of subjects with events/total patient-years at risk)*100,*
or, equivalently
Number of subjects with events/100 patient-years at risk

An equivalent interpretation of EAIR can be expressed as the "average number of patients experiencing the event of interest if 100 patients were treated for 1 year".

EAIRs may be summarized for all TEAEs, fatal TEAEs, and serious TEAEs.

5.2 Exploratory Safety Analyses

TEAE summaries will be presented by yearly intervals for both treatment arms using the prevalence of new and ongoing events during each ibrutinib/placebo treatment duration interval (> 0-1 years, > 1-2 years, > 2-3 years, > 3-4 years, > 4-5 years, > 5-6 years, > 6-7 years) and for the overall TEAE period. Specifically, the treatment duration interval > 0-1 years includes all subjects who received ibrutinib/placebo treatment up to 1 year; > 1-2 years includes all subjects who received ibrutinib/placebo treatment more than 1 year and up to 2 years, etc.

For the determination of yearly prevalence, newly occurring TEAEs in the respective treatment duration yearly-interval as well as TEAEs still ongoing from the previous treatment duration yearly-interval will be included. Ongoing events for a subject from previous time period are only included in the next time period when the subject is at risk (i.e. within 30 days of last ibrutinib/placebo dose date) at the start of the next time period.

The prevalence of TEAEs will be summarized across yearly-intervals by MedDRA SOC and PT for all TEAEs, grade 3 and higher TEAEs, SAEs and TEAEs related to ibrutinib/placebo.

In addition, all TEAEs by SOC and PT will be summarized across yearly intervals for the following:

- subjects who discontinued ibrutinib for any reason other than PD vs. who did not discontinue ibrutinib
- subjects who discontinued ibrutinib for any reason other than PD after achieving BOR of SD or PR vs.CR

TEAEs leading to death will be summarized by SOC and PT at the study-duration yearly interval when the subject died. Study duration is considered instead of treatment duration in order to include events that started prior to 30 days after ibrutinib/placebo discontinuation, and for which death may have occurred outside of that time period.

6 APPENDICES

6.1 Algorithm for Defining ≥2 Consecutive Missing Assessments

This algorithm is used for the PFS primary analysis for US regulatory submission and for the PFS sensitivity analyses for Other Regions.

This study has 2 planned imaging schedules: every 4 cycles (1 cycle 28 days) from Cycle 5 through Cycle 37, and every 6 cycles afterwards (see Table 8 below). Based on the change between the schedules, one transition of the imaging schedule (time period 2 in Table 8 below) is described as follows:

- When the adequate non-PD assessment occurs at Cycle 33, the next two subsequent assessments are planned at Cycle 37 (+4 cycles) and Cycle 43 (+6 cycles), respectively.

This algorithm will follow two steps in sequence:

1. Step 1 (Columns 2-4 in Table 8): Calculate the study day of the adequate non-PD assessment. Map the adequate non-PD assessment to a closest scheduled visit based on the study day using the midpoint time period window (see footnote of Table 8 for details). Once the assessment is mapped to a scheduled visit, the imaging interval for the next two subsequent consecutive visits can be decided based on the planned imaging schedule.
2. Step 2 (column 5 in Table 8): Calculate the maximum gap in days allowed for the next two subsequent visits based on the planned imaging interval plus the specified window (+/- 28 days) for each visit.

If the gap between the adequate non-PD assessment and the next adequate assessment or death exceeds the maximum gap, the subject has a ≥ 2 consecutive missing assessments and its PFS will be censored at the last adequate non-PD assessment before the missing interval. If there are 2 more missing intervals, the PFS will be censored at the last adequate assessment before the first missing interval. If the first missing interval occurs right after the randomization date, then PFS will be censored at the randomization date.

Table 8. Algorithm for defining ≥2 consecutive missing assessments

Time Period	Adequate Non-PD Assessment		Imaging Schedule for the Two Subsequent Consecutive Assessments after the Adequate Non-PD Date	Maximum gap[&] in Days Allowed between the Adequate Non-PD Date and Next Adequate Assessment or Death
	Scheduled Visit	Study Day Time Period Window[#]		
1	Cycle 5 to Cycle 29	Day 2 - Day 841	every 4 cycles (112 days)	280 = (112+28)*2
2	Cycle 33	Day 842 - Day 953	4 cycles (112 days) & 6 cycles (168 days)	336 = (112+28) + (168+28)
3	Cycle 37 and up	Day 954 and up	every 6 cycles (168 days)	392 = (168+28)*2

#: The visit window for each visit is based on the midpoint between two scheduled visits (i.e. target day +/ the half planned interval between two scheduled visits). The study day in time period 1 can be as early as one day after randomization (Day 2). If the adequate non PD assessment is the same as or earlier than the midpoint, it will be assigned to the earlier visit, and if the assessment is later than the midpoint, it will be assigned to the later visit. Study day is calculated in reference to the randomization date. 1 cycle 28 days.

&: A 28 day window is allowed to provide more flexibility regarding the timing of the data collection for disease assessment considering the COVID 19 situation and the indolence of first line follicular lymphoma

7 REFERENCES

1. Hwang I, Shih W, De Cani J. Group sequential designs using a family of type I error probability spending functions. Stat Med. 1990;9(12):1439-45.
2. Cheson BD, Fisher RI, Barrington SF, et al. Recommendations for Initial Evaluation, Staging, and Response Assessment of Hodgkin and Non-Hodgkin Lymphoma: The Lugano Classification. J Clin Oncol. 2014;32:3059-68.
3. Ghielmini M, Schmitz SFH, Cogliatti SB, et al. Prolonged treatment with rituximab in patients with follicular lymphoma significantly increases event-free survival and response duration compared with the standard weekly x 4 schedule. Blood. 2004;103:4416-4423.
4. Hainsworth JD, Litchy S, Burris HA, et al. Rituximab as a first-line and maintenance therapy for patients with indolent non-hodgkins lymphoma. J Clin Oncol. 2002;20:4261-7.
5. Fowler N, Nastoupil L, de Vos S, et al. The combination of ibrutinib and rituximab demonstrates activity in first-line follicular lymphoma. Br J Haematol. 2020;189(4):650-60.
6. Zucca E, Rondeau S, Vanazzi A, et al. Short regimen of rituximab plus lenalidomide in follicular lymphoma patients in need of first-line therapy. *Blood.* 2019;134(4):353-362.
7. FDA Guidance for Industry: Statistical Considerations for Clinical Trials During the COVID-19 Public Health Emergency, June 2020
8. Saville BR, Kim YS, Koch GG. Graphical displays for clarifying how allocation ratio affects total sample size for the two sample logrank test. Pharmaceut. Statist. 2011;10:40-44.
9. Wiener LE, Ivanova A, Koch GG. Methods for clarifying criteria for study continuation at interim analysis. Pharmaceutical Statistics. 2020;19:720-732.
10. Akazawa K, Nakamura T, Palesch Y. Power of logrank test and Cox regression model in clinical trials with heterogeneous samples. Stat Med. 1997;16(5):583-97.
11. Zhou, Y., Ke, C., Jiang, Q., Shahin, S., & Snapinn, S. (2015). Choosing Appropriate Metrics to Evaluate Adverse Events in Safety Evaluation. Therapeutic innovation & regulatory science, 49(3), 398 404
12. Jia, Q., Kuo, F., Wu, C., & Xu, P. (2020) Risk-based and Exposure-based Adjusted Safety Incidence Rates. In PharmaSUG 2020. https://www.pharmasug.org/proceedings/2020/SA/PharmaSUG-2020-SA-103.pdf

8 REVISION HISTORY

Table 9. SAP Version History Summary

Version Number	Date	Descriptions
1.0	08Dec2020	Original version
2.0	19Oct2021	Amendment summary: Below are the major changes made. 1. Update secondary endpoints and sample size determination per protocol amendment 2 dated March 10, 2021. 2. Modified the second interim analysis boundaries from Rho spending function to Lan-DeMets alpha spending function with O'Brien-Fleming parameter. 3. Update the alpha spending and testing procedure for secondary endpoints per protocol amendment 2. 4. Remove interim analysis for ORR. 5. Clarify the description regarding stratification factors in Part 1. 6. Add infusion related reaction rate to the sequential testing procedure and rank order for the secondary endpoints. 7. Clarify sections on interim and final analyses. 8. Add estimands and summary of estimand attributes for the primary and key secondary endpoints per ICH E9(R1). 9. Add the subgroup definition table for subgroup analysis. 10. Add descriptive summary of COVID-19 impact in disposition, safety, compliance; also add sensitivity and/or supplementary analysis for efficacy endpoints. 11. Add the description to handle COVID-19 deaths in the PFS primary and supplementary/sensitivity analyses. 12. Update the analyses for secondary endpoints. 13. Add analyses for PFS by IRC, ORR by IRC and DOR by IRC. 14. Add detailed analyses for exploratory endpoints such as CR30 and patient reported outcomes (PROs). 15. Add detailed definition of treatment emergent adverse events, and summary for treatment emergent adverse events and deaths related to COVID-19 infection.
3.0	29Apr2022	Amendment summary: 1. Add time to worsening for FACT-Lym scores. 2. Update censoring rule due to subsequent anti-cancer therapy for PFS 3. Revise sample size to be 445 as one subject was randomized twice 4. Add OS futility analysis at IA2 and modify boundaries for PFS analysis 5. Update extent of exposure variables. 6. Signatory for programming team changed 7. Added a figure of the testing procedure

		8. For the stratified efficacy analyses, stratification factors with small strata (<10 events) were removed. A pre-specified selection method and reference were given in the main text.
4.0	13Feb2023	Amendment summary: 1. Removed Part 2 language. 2. Update PFS primary censoring rule to include censoring for COVID-19 deaths without PD. 3. Added OS interim analysis at PFS primary analysis and modify OS final analysis timing. Added OS boundary values and table. 4. Clarified ECOG and FLIPI-1 source data for subgroup analysis. 5. Removed "Section 6 Changes in Protocol Planned Analysis" as changes in SAP v3 have been reconciled with Protocol Amendment 4. 6. Added algorithm for defining 2 or more missed assessments.
5.0	01Jun2023	Amendment summary: 1. Update PFS primary censoring rule to include censoring for subjects who missed two or more consecutive overall disease assessments prior to the PFS event. 2. Change the superiority boundary for the interim OS analysis from Lan-DeMets alpha spending function with O'Brien-Fleming boundary to Haybittle-Peto boundary.
6.0	01Mar2024	Amendment summary: 1. Update the futility analysis of PFS (IA1) and the futility analysis of OS at PFS IA2 by spending 0.00001 alpha to ensure strict control of family wise error rate. 2. Update the PFS primary censoring rules to not censor for COVID-19 deaths without PD. Accrue additional PFS events to make up for the potential impact of COVID in accordance with FDA`s COVID-19 guidance. 3. Update the PFS primary censoring rule, for US regulatory submission only, to include censoring for subjects who missed two or more consecutive overall disease assessments, regardless of whether or not a PFS event subsequently occurs. 4. Update the algorithm to defining ≥2 consecutive missing assessments to include subjects who missed two or more consecutive overall disease assessments, regardless of whether or not a PFS event subsequently occurs, and specifying a 28-day visit window. 5. Added "Section 5.1 Exposure-Adjusted Incidence Rate" 6. Added "Section 8.1 Changes to Planned Analyses in the Protocol"
v6.0 Addendum	17Apr2024	Exploratory efficacy and safety analyses were added to further assess the efficacy and cumulative safety of ibrutinib as administered in the protocol,
7.0	13May2024	Amendment summary: 1. Incorporate SAP v6 addendum

		a. Updated Section 1.2.3 by adding TTP and DOCR as exploratory endpoints b. Updated Table 4.1 by adding the definition and analyses method of TTP and DOCR c. Added "Section 4.1 Exploratory Effiacy Analyses" d. Added "Section 5.1 Exploratory Safety Analyses" 2. Remove the 0.00001 alpha allocation at the futility analysis of PFS (IA1) and the futility analysis of OS at PFS IA2. a. Update Table 3.1, Table 3.2, Table 4.1, Table 4.2, and Figure 2.

8.1 Changes to Planned Analyses in the Protocol

1. IRC will be implemented to do the overall disease assessment based on Cheson 2014 (Lugano Classification Criteria). A sensitivity/supplementary analyses will be conducted for the endpoints PFS, ORR and DoR using the IRC assessment.
2. The superiority boundary for the interim OS analysis was changed from O'Brien Fleming boundary to Haybittle-Peto boundary.
3. Per US FDA information request, PFS primary censoring rules was updated to include censoring of subjects with ≥ 2 consecutive missed assessments at the last adequate assessment before the missing interval, regardless of whether or not a PFS event subsequently occurs.