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 Rituximab, Cyclophosphamide, Doxorubicin, Vincristine, and Prednisone (R-CHOP) in Subjects With Newly Diagnosed Non-Germinal Center B-Cell Subtype of Diffuse Large B-Cell Lymphoma

Protocol Number: PCI-32765DBL3001**;** Phase 3

JNJ-54179060 (ibrutinib)

Issue/Report Date: 21 March 2018
Document No.: EDMS-ERI-57717339

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

TABLE OF CONTENTS

LIST OF IN-TEXT TABLES AND FIGURES

TABLES

FIGURES

ABBREVIATIONS

ABC	activated B cell-like subtype of DLBCL
AE	adverse event
ALC	absolute lymphocyte count
ALT	alanine aminotransferase
ANC	absolute neutrophil count
AST	aspartate aminotransferase
ATC	Anatomical Therapeutic Chemical
AUC	area under the curve
CI	confidence interval
CNS	central nervous system
CR	complete response
CRF	case report form
CT	computed tomography
DLBCL	diffuse large B-cell lymphoma
DMC	Data Monitoring Committee
DPS	Data Presentation Specifications
ECOG	Eastern Cooperative Oncology Group
eCRF	electronic case report form
EFS	event-free survival
EQ-5D-5L	EuroQol questionnaire
FACT-G	Functional Assessment of Chronic Illness Therapy-General
FACT-Lym	Functional Assessment of Cancer Therapy-Lymphoma
GEP	gene expression profiling
HR	hazard ratio
IHC	immunohistochemistry
IRC	Independent Review Committee
ITT	Intent-to-Treat population
LDH	lactate dehydrogenase
MedDRA	Medical Dictionary for Regulatory Activities
NCI-CTCAE	National Cancer Institute Common Terminology Criteria for Adverse Events
non-GCB	non-germinal center B cell-like subtype of DLBCL
OS	overall survival
PD	progressive disease
PET	positron emission tomography
PFS	progression-free survival
PR	partial response
PRO	patient-reported outcome(s)
R-CHOP	rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone
R-IPI	Revised International Prognostic Index
ROC	receiver operating characteristic
SAE	serious adverse event
SD	stable disease
SF	screen failure
TEAE	treatment-emergent adverse event
TTW	time to worsening
U.S.	United States
WHO	World Health Organization

1. INTRODUCTION

This clinical study is a part of a comprehensive ibrutinib clinical development plan to evaluate the safety and efficacy of ibrutinib for subjects with B-cell malignancies. This randomized, double-blind, placebo-controlled, multicenter, Phase 3 study is designed to evaluate the efficacy and safety of ibrutinib in combination with rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone (R-CHOP). The study will evaluate if ibrutinib in combination with R-CHOP will prolong event-free survival (EFS) in subjects with newly diagnosed non-germinal center B cell-like (non-GCB) subtype of diffuse large B-cell lymphoma (DLBCL) selected by immunohistochemistry (IHC) or newly diagnosed subjects with activated B cell-like (ABC) subtype of DLBCL identified by gene expression profiling (GEP) or both populations.

The purpose of this Statistical Analysis Plan is to present key elements, including definitions and statistical methods, for the planned analyses for the primary, secondary, and safety endpoints. Details for derived variables and data handling conventions will be added to the Data Handling section and the Data Presentation Specifications (DPS).

1.1. Study Objectives

The primary objective of this study is to evaluate if the addition of ibrutinib to R-CHOP prolongs EFS compared with R-CHOP alone in subjects with newly diagnosed non-GCB subtype of DLBCL or in subjects with newly diagnosed ABC subtype of DLBCL or in both populations.

The secondary objectives are to:

- Evaluate progression-free survival (PFS)
- Evaluate the complete response (CR) rate
- Evaluate overall survival (OS)
- Evaluate patient-reported lymphoma symptoms and concerns as measured by the Lym subscale of the Functional Assessment of Cancer Therapy-Lymphoma (FACT-Lym)
- Characterize the pharmacokinetics of ibrutinib and explore the potential relationships between ibrutinib metrics of exposure with relevant clinical or biomarker information
- Evaluate the safety of ibrutinib when combined with R-CHOP

The exploratory objectives are to:

- Evaluate patient-reported outcomes (PRO), related to well-being and general health status, utilizing the FACT-Lym and EuroQol (EQ-5D-5L)
- Explore the relationship between relevant biomarkers (eg, GEP, gene mutations) with clinical outcomes and mechanism of resistance

1.2. Study Design

This is a randomized, double-blind, placebo-controlled, multicenter, Phase 3 study to compare the efficacy and safety of ibrutinib in combination with R-CHOP versus R-CHOP alone in subjects with newly diagnosed non-GCB DLBCL or in subjects with newly diagnosed ABC subtype of DLBCL or in both populations.

Approximately 800 eligible subjects will be stratified by Revised International Prognostic Index (R-IPI; 1-2 vs. 3-5), region (United States [U.S.]/Western Europe vs. Rest of World), and number of pre-specified treatment cycles (6 cycles vs. 8 cycles), then randomized in a 1:1 ratio to receive either placebo+R-CHOP or 560 mg ibrutinib+R-CHOP. Prior to randomization, sites will pre-specify treatment of subjects with either 6 or 8 cycles. All subjects at individual sites must receive the same number of treatment cycles (6 or 8; no adjustment is permitted once pre-specified).

1.2.1. Revised International Prognostic Index

The R-IPI is calculated based upon the following IPI risk factors (1 point for each factor present):

- Age >60 years
- Eastern Cooperative Oncology Group (ECOG) performance status ≥2
- Elevated lactate dehydrogenase (LDH)
- ≥2 extranodal sites of disease
- DLBCL Stage III or IV

Table 1 lists the standard IPI and the R-IPI risk groups.

Table 1: International Prognostic Index (IPI) and Revised IPI Categories

Risk group	Number of IPI factors
Standard IPI	
Low	0, 1
Low-intermediate	2
High-intermediate	3
High 4, 5	4, 5
Revised IPI	
Very good[a]	0
Good	1, 2
Poor	3, 4, 5

IPI=international prognostic index
[a] The risk group of Very good (ie, R-IPI score of 0) does not meet the study eligibility criteria.

In this study, the R-IPI risk group will be used as 1 of the randomization stratification factors (ie, for eligible subjects with an R-IPI score ≥1). The stratification factors based upon R-IPI are

the 2 higher risk groups: Good (High, R-IPI score of 1 to 2) and Poor (Low, R-IPI score of 3 to 5).

1.2.2. Study Phase

The study will include a Pretreatment (Screening) Phase prior to randomization; an Active Treatment Phase, which includes visits at the start of each cycle and an End-of-Treatment visit; and a Posttreatment Follow-up Phase, which continues until death, loss to follow up, consent withdrawal, or study end, whichever occurs first.

- The Screening Phase will extend up to 30 days prior to randomization, including the date of randomization unless otherwise specified. All subjects must sign an informed consent form prior to the conduct of any study-related procedures.
- The Active Treatment Phase will extend from Day 1 Cycle 1 until study treatment discontinuation due to disease progression, initiation of subsequent antilymphoma therapy, unacceptable toxicity, withdrawn, or completion of study treatment.
  - During the Active Treatment Phase, all subjects will receive R-CHOP as background therapy for 6 or 8 cycles as pre-specified (21 days per cycle). Subjects who discontinue R-CHOP without disease progression prior to completing 6 or 8 cycles will continue study drug until 6 or 8 cycles (depending on the number of cycles pre-specified by each site) are completed or the end of the Active Treatment Phase, whichever occurs first.
  - During the Active Treatment Phase, disease response assessment will be performed at both Cycle 4 (computed tomography [CT] only) and End-of-Treatment visit/early withdrawal (CT and PET). Whole body PET scan is recommended but not mandated at baseline.
  - Disease response assessment criteria will be based on the Revised Response Criteria for Malignant Lymphoma (Cheson 2007). The progressive disease (PD) notification form (Study Event form) were to be faxed to the medical monitors within 24 hours of the PD declaration.
  - At the end of treatment, all subjects are required to fill out the End-of-Treatment visit form within 30 days after the last dose of study drug. If a subject requires subsequent systemic antilymphoma therapy between the last dose of study drug and the expected date of the End-of-Treatment visit, then the End-of-Treatment visit should be completed before initiation of subsequent systemic antilymphoma therapy.
- The Posttreatment Follow-up Phase will start after subjects complete the End-of-Treatment procedures. Study end is defined as when 50% of the randomized subjects have died; 5 years after the last subject is randomized; or the sponsor terminates the study, whichever occurs first. The Posttreatment Follow-up Phase includes 2 parts, the pre-PD follow-up phase and the post-PD survival follow-up phase.

1.2.3. Planned Clinical Cutoff

Three clinical cutoffs were planned. The first clinical cutoff is an interim analysis, which will occur when approximately 270 EFS events are observed. The second clinical cutoff is the final analysis of EFS, which will occur at 30 months after the 800th subject is randomized. Treatment assignment will be unblinded to the study team at this clinical cutoff for the final analysis of EFS and the analysis results will form the basis for the primary clinical study report. The last clinical cutoff will occur at the end of study. The main purpose of the last clinical cutoff is to update the long-term survival status. All available data prior to the time of a clinical cutoff will be included in each of the respective analyses.

Per the original study design, the number of events required for the interim analysis is 270 EFS events. As of 12 July 2017, 230 EFS events have been reached. Due to this lower than expected EFS event rate and the short time expected between the interim and final analysis, the sponsor has omitted the first clinical cutoff for the interim analysis. Omission of the interim analysis is documented in protocol Amendment INT-3 issued on 16 October 2017.

1.2.4. Disease Response Schedule and Method

The primary efficacy analysis of EFS will be based on investigator's assessment following the Cheson 2007 criteria (Cheson 2007). The investigator will evaluate sites of disease by CT scans of the neck, chest, abdomen, and pelvis with intravenous and oral contrast as indicated. Whole body PET scan is recommended but not mandated at baseline, but is required at the end of treatment. Magnetic resonance imaging may be used to evaluate sites of disease that cannot be adequately imaged using CT. Other sites of disease will be evaluated by radiological imaging, physical examination, or other procedures as necessary (to be performed throughout the study using the same method of assessment per subject).

An Independent Review Committee (IRC) will perform disease assessments for a random sample of subjects. If the results of EFS, based on the IRC's assessment, do not confirm the primary analysis results using the audit plan (Dodd method; Dodd 2011), then IRC will perform an independent assessment for all randomized subjects. The details are addressed in the IRC audit plan (Attachment 1).

At each site visit, subjects will be evaluated for toxicity. Safety evaluations will include adverse event (AE) monitoring, physical examinations, concomitant medication usage, and clinical laboratory parameters. Blood samples will be drawn for assessment of pharmacokinetic parameters.

1.2.5. Data Monitoring Committee

An independent Data Monitoring Committee (DMC) will be established to monitor data on an ongoing basis to ensure the safety of the subjects enrolled in this study and to evaluate the efficacy of the treatment at the time of interim analysis. The DMC is constituted according to regulatory agency guidelines. The DMC will meet periodically to review interim data. After the

review, the DMC will make recommendations regarding the conduct of the study. Further details regarding the responsibilities, authorities, and procedures will be provided in the DMC charter.

1.3. Statistical Hypotheses for Study Objectives

The primary objective of this study is to evaluate if the addition of ibrutinib to R-CHOP prolongs EFS compared with R-CHOP alone in subjects with newly diagnosed non-GCB subtype of DLBCL (intent-to-treat [ITT] population), subjects with newly diagnosed ABC subtype of DLBCL (ABC population), or in both populations. The statistical hypotheses are described as follows:

H_0: Null hypothesis, the EFS distributions of the experimental treatment group, $S_T(t)$, and the placebo group, $S_P(t)$, are not different, at all timepoints t for both the ITT population and ABC population:

$$S_T(t) = S_P(t), \text{ for all } t > 0 \qquad (1)$$

versus

H_1: Alternative hypothesis, the EFS distributions of the experimental treatment group, $S_T(t)$, are stochastically greater than that of the placebo group, $S_P(t)$, for the ITT population, and/or for the ABC population:

$ST(t) \geq SP(t)$, for all $t \geq 0$, with strict inequality for some t

These hypotheses will be tested using a stratified log-rank test as detailed in Section 5.

The primary objective for the original design was to evaluate if the addition of ibrutinib to R-CHOP prolongs EFS compared with R-CHOP alone in subjects with newly diagnosed non-GCB subtype of DLBCL only. The protocol was amended (protocol Amendment INT-3) to change the primary objective to "The addition of ibrutinib to R-CHOP prolongs EFS compared with R-CHOP alone in subjects with newly diagnosed non-GCB subtype of DLBCL selected by IHC, or subjects with newly diagnosed ABC subtype of DLBCL identified by GEP, or in both populations".

1.4. Sample Size Justification

This study is designed to evaluate the effect of treatment on EFS and is powered for this endpoint. DLBCL is an aggressive but potentially curable disease with 30% to 55% of patients achieving durable cure (Fisher 1993). For further insight into event estimation and the power calculation for this study, the study population was differentiated into 2 subgroups based on potential curability with treatment, ie, those who could achieve durable cure (curable population; those who achieve CR and are not expected to relapse irrespective of follow-up time) and those who are not expected to achieve durable cure (non-curable population). Simulation studies were conducted for estimating the event numbers, interim analysis timepoint, and power based on the following considerations:

- A 1:1 randomization ratio between 2 treatment arms.
- A total of 800 subjects to be enrolled (400 subjects per treatment arm).
- Assuming the cure rate for the control arm (placebo+R-CHOP) is 40% and the targeted cure rate of improvement is 10% for the active treatment arm ibrutinib+R-CHOP (ie, the cure rate for ibrutinib+R-CHOP is 50%). The median EFS is assumed to be 15 years for cured subjects.
- Among those subjects who are not cured, a targeted hazard ratio (HR) of 0.75 is assumed. This corresponds to a 4-month increase in median EFS for the active treatment arm (ibrutinib+R-CHOP) relative to the control arm (placebo+R-CHOP), assuming the median EFS for the control arm (placebo+R-CHOP) is 12 months.
- One interim analysis to be performed after approximately 270 EFS events have been observed, for superiority testing at the nominal significance level of 0.002 (1-sided).

With approximately 800 subjects (about 400 subjects per treatment arm) to be randomized in approximately 27 months (30 subjects per month) and with a study follow-up period of 30 months after the last subject is randomized, it is anticipated that approximately 419 EFS events will be observed and the study will have at least 90% power to show statistical significance at an overall alpha level of 0.025 (1-sided).

The nominal alpha level for the final analysis will be calculated using the cumulative alpha spending function of the power family (Jennison and Turnbull 2000) based on the actual number of events at the final analysis. Table 2 shows the calculated nominal significance level at the final EFS analysis assuming the final observed EFS events range from 350 to 450. Based on the cumulative alpha spending function with an interim analysis of 270 EFS events and a nominal significance level of 0.002 (1-sided), if the final observed number of EFS events is 420, the final nominal significance level will be 0.02463 (1-sided) to maintain the overall alpha of 0.025 (1-sided).

Table 2: Calculated Nominal Significance Level at the Final Analysis

EFS Events	Information Fraction	ρ	Nominal Significance Level (1-sided)
350	0.771	9.66	0.02489
360	0.750	8.78	0.02485
370	0.730	8.02	0.02482
380	0.711	7.37	0.02478
390	0.692	6.80	0.02474
400	0.675	6.55	0.02472
410	0.659	6.08	0.02468
420	0.643	5.66	0.02463
430	0.628	5.47	0.02461
440	0.614	5.11	0.02456
450	0.600	4.94	0.02453

ρ =index parameter for alpha spending function of the power family.

Simulation studies were also conducted based on the different assumptions of cure rate improvement and the risk reductions for those uncured subjects as detailed below:

- The cure rate is 40% in the control arm (placebo+R-CHOP).
- A 5% or 10% improvement in cure rate with ibrutinib+R-CHOP treatment.
- Median EFS is 15 years for both the control and active treatment arms in the curable population.
- For the non-curable population in the control arm, the median EFS is assumed to be 12 or 15 months.
- For the non-curable population in the active treatment arm, the median EFS is based on the HRs of 0.70, 0.72, 0.75, and 0.80 for the ibrutinib+R-CHOP arm relative to placebo+R-CHOP arm.
- Dropout rate is 5%.
- The nominal statistical significance level of 0.002 (1-sided) is used at the interim analysis. For the final analysis, a nominal significance level of 0.024 (1-sided) will be used. This will result in an overall significance level of less than or equal to 0.025 (1-sided) given the anticipated total number of events at the final analysis.

Based on 10,000 simulation runs, the results listed in Table 3 indicate that the study power will be maintained at least 80% with a 10% improvement in cure rate. For a 5% (from 40% to 45%) improvement in cure rate, HRs of 0.7 or less are needed to maintain an 80% power.

Table 3: Simulation Results Based on Different Cure Rate Improvement and Different HR Assumptions for Non-curable Population

	Non-curable Control arm Median EFS (months)	Non-curable TRT arm Median EFS (months)	Non-curable Population HR	Timing of IA (months) (when 270 events occur)	Number of Events at Final Analysis of EFS	Power (%) at Final Analysis
Cure rate Control, 40%; TRT, 50%	12	17	0.70	32.8	415	96
	12	16.7	0.72	32.6	417	95
	12	16	0.75	32.3	419	92
	12	15	0.80	31.9	423	88
	15	21	0.70	36.3	389	96
	15	20.8	0.72	36.1	391	95
	15	20	0.75	35.8	394	93
	15	18.8	0.80	35.3	398	87
Cure rate Control, 40%; TRT, 45%	12	17	0.70	31.9	428	82
	12	16.7	0.72	31.7	429	78
	12	16	0.75	31.5	432	70
	12	15	0.80	31.1	436	57
	15	21	0.70	35.4	400	84
	15	20.8	0.72	35.2	402	79
	15	20	0.75	34.8	405	72
	15	18.8	0.80	34.3	410	59

EFS=event-free survival; HR=hazard ratio; IA=interim analysis; TRT=active treatment.

In this study, the primary endpoint analysis measures a combination of improvement in the cure rate and improvement in the EFS interval among those subjects who are not cured. A statistically significant p-value for the log-rank test indicates that the addition of ibrutinib to R-CHOP improves the clinical outcome of the experimental group when both parameters, ie, difference in cure rate and difference in EFS of the non-cured population, are taken into account. A wide range of outcomes may result in a statistically significant difference between the groups. Table 4 presents a few examples of possible clinical outcomes and the corresponding probability that such outcomes will result in a statistically significant difference between the treatment arms.

Table 4: Examples of Possible Study Outcomes Based on Median EFS of 12 Months for Non-cured Subjects

Cure Rate		EFS HR for Non-cured Subpopulation	P (success) at IA (boundary p-value = 0.002)	Cumulative P (success) at Final Analysis (either p-value < 0.002 at IA or <0.024 at final analysis)
Control	Active Treated			
40%	50%	0.70	0.74	0.96
		0.75	0.61	0.92
		0.80	0.47	0.88
40%	45%	0.70	0.49	0.82
		0.75	0.33	0.70
		0.80	0.20	0.57

EFS=Event-free Survival; HR=Hazard Ration; IA=Interim Analysis; P=Probability.

As addressed in Section 1.2.3, the interim analysis was omitted due to the lower than expected EFS event rate. The simulation study, to examine the original study power, discussed in the above section will be not applicable per protocol Amendment INT-3. An updated power calculation is described in Section 1.4.2. The significant alpha level for the final analysis will be overall 2-sided 0.05 instead of 0.048. The primary hypothesis was also changed into simultaneous testing for both ITT and ABC populations. The overall significance level will be based on a 2-sided alpha of 0.05 for multiplicity adjustment shared between the ITT and ABC populations. Multiplicity adjustment details will be discussed in Section 1.4.1. The power estimation for the ABC population will be provided in Section 1.4.2. The sample size for ABC will depend on the GEP test results.

1.4.1. Multiplicity Adjustment for Simultaneous Testing

The primary analysis of EFS will be performed with data from both the ITT population and ABC population simultaneously using the method of Song and Chi (Song and Chi 2007), which utilizes a 2-stage testing procedure that maintains reasonable study power and also strongly controls the familywise Type I error rate.

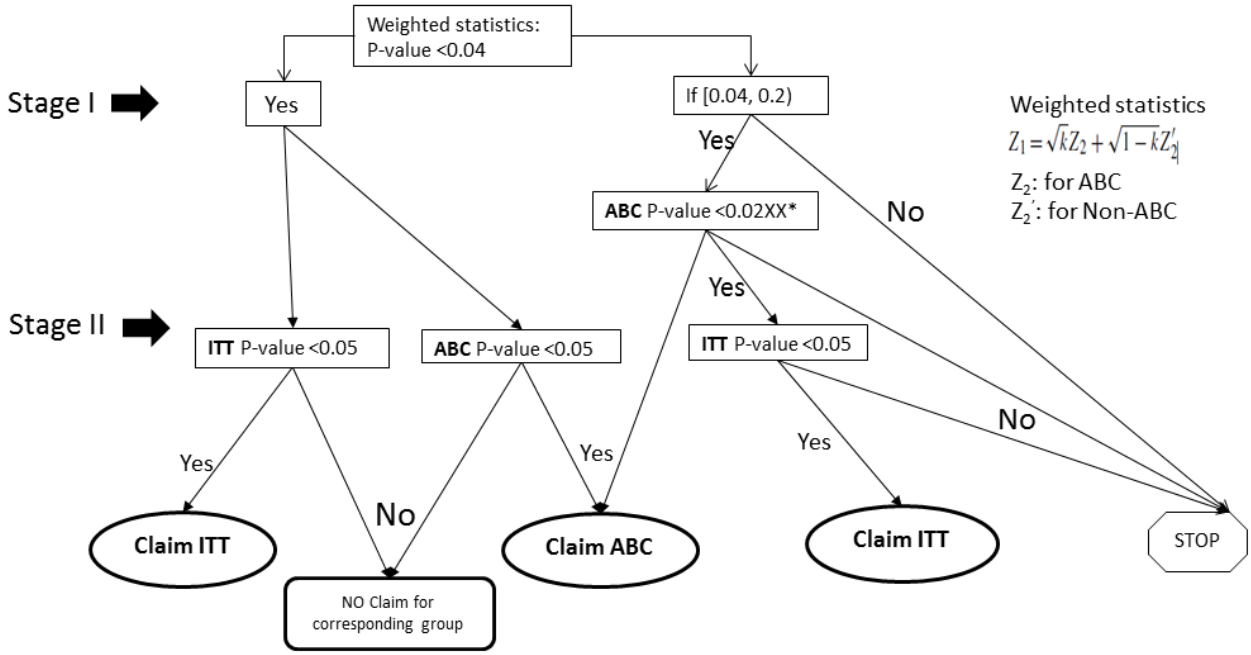
Song and Chi proposed a weighted test statistics Z_1, which combines the target population of ABC, represented by Z_2 and its complementary population, represented by Z_2'. k is the proportion of events in ABC population among total events in ITT population.

$$Z_1 = \sqrt{k}Z_2 + \sqrt{1-k}Z_2' \quad (2)$$

Figure 1 illustrates the Song and Chi 2-stage testing procedure. At Stage I, if the p-value associated with the weighted statistic Z_1 is less than 0.04, then proceed to Stage II for testing both the ITT (based on stratified log-rank test instead of weighted testing statistic Z_1) and the target subgroup (ABC population) at the alpha level of 0.05 separately. If the p-value for the weighted statistics Z_1 at Stage I is ≥0.04 and <0.2, then the ABC population will be tested at the corresponding significance level. The significance level for the ABC population is calculated to control the familywise error rate of 0.05 by incorporating the correlation between Z_1 and Z_2. If

significance is shown in the ABC population, then the ITT population can be re-tested at the significance level of 0.05 using a standard stratified log-rank test.

Figure 1: An Example of Song and Chi 2-Stage Testing Procedure



ABC=activated B cell-like DLBCL; ITT=intent-to-treat
*Actual significance level depends on the proportion of EFS events *k,* determined on blinded data before the signoff of this SAP. Please refer to Section 5.1.1 for the details.

Table 5 provides examples of calculated significance levels for the ABC population assuming different *k* - a proportion of EFS events in ABC population among total events in ITT population. For this study, we propose using a 2-sided alpha of 0.04 for the Stage I test (α1); the overall study population efficacy consistency control parameter assuming to be 2-sided alpha of 0.2. The actual significance levels for the ABC population are shown in tables 5 with *k* ranges from 0.6 to 0.8. The true significance level for ABC population will be based on the observed EFS events in ABC population vs. ITT population.

Table 5: Significance Alpha Level (2-sided) α_1 for the ABC Population with Different EFS Events Proportion.

K	α_1= 0.04
0.6	0.0255
0.7	0.0268
0.8	0.0300

α_1: Significance level for weighted test statistic Z_1; ABC=activated B cell-like DLBCL; EFS=event-free survival; *k*=proportion of EFS events in the ABC population among total events in ITT population.

The anticipated proportion of EFS events from the ABC subgroup is between 60% and 80%. This was based on the testing concordance (Table 6), using the 199 samples taken from Study DBL3001 (the selection scheme for those 199 subjects is in Attachment 2), the proportion of ABC is about 55% (57/104=54.8%). Subjects with the ABC subtype have an inferior efficacy profile than those with the GCB subtype (Lenz 2008). It is reasonable to assume more EFS events will come from patients with the ABC subtype than from those with the GCB or unclassified subtypes. Therefore, we assume that there will be at least 60% of the EFS events from the ABC population in our simulation studies.

Table 6: Concordance of IHC and GEP Data

		Nanostring GEP					
		ABC	UNC	GCB	Sample Not Evaluable	Total (N=199)	Evaluable Subjects (N=186)
[redacted] IHC[a]	non-GCB	57	22	25	6	110	104
	Randomized	45	18	21	5	89	84
	GCB	7	7	68	7	89	82

ABC=activated B cell-like subtype; UNC=unclassified; GCB=germinal center B cell-like subtype; GEP=gene expression profiling; IHC=immunohistochemistry; non-GCB=non-germinal center B cell-like subtype
[a] The data represents IHC results used for patient enrollment into Study DBL3001.

1.4.2. Power of Hypothesis Testing Based on Simulation Studies

DLBCL is a curable disease with multiple subtypes. It is a mixed population with cured and non-cured populations. Within cured and non-cured populations, DLBCL is further classified as ABC, GCB, and unclassified DLBCL subtypes by GEP. In the simulation setting that is used to estimate the study power based on the Song and Chi method (Song and Chi 2007), we assume that there are different treatment outcomes for the cured or uncured by GEP subtypes of ABC and non-ABC (GCB, and unclassified). Table 7 shows the simulation settings.

Table 7: Simulation Settings for Non-cured and Cured Population by GEP Subtype

		Treatment	Control
ABC	Cured	50%	40%
	Non-cured	HR=0.75	Median=18 months
non-ABC	Cured	60%	60%
	Non-cured	HR=0.8, 0.9, 1.0	Median=36 months

ABC=activated B cell-like DLBCL; HR=Hazard ratio
Note: For the cured population, the median is 180 months/15 years. *k*=0.6, 0.7, 0.8 for the proportion of ABC among the 800 subjects in the study.

Table 8 shows the simulation results on study power for both the ITT and ABC populations. An increased proportion of EFS events from the ABC subgroup is associated with increased study power. The reason to change the non-cured population median of 12 months to 18 and 36 months is due to the lower than expected EFS event accumulation.

Simulation results indicate that we will lose some power for the ITT population compared with the original design. However, we will have an opportunity to show statistical significance for the ABC subpopulation with reasonable probability. For example, with a proportion of events of k=0.6 and a HR of 0.8 for the non-cured and non-ABC population, the original study power for the ITT population is 0.702. After considering both the ITT and ABC populations based on Song and Chi method (Song and Chi 2007), the power is 0.693 for the ITT population and 0.722 for the ABC population. The overall study power, which is the probability that we will reach statistical significance for either ITT or ABC population, is 0.771.

Table 8: Simulation Results for Study Power

k (Prop of events in ABC)	HR (non-ABC, non-cured)	ABC	ITT	Either ITT or ABC	Original ITT
0.6	**0.8**	**0.722**	**0.693**	**0.771**	**0.702**
0.7	0.8	0.796	0.771	0.828	0.781
0.8	0.8	0.847	0.830	0.867	0.839
0.6	0.9	0.704	0.616	0.736	0.623
0.7	0.9	0.782	0.722	0.804	0.731
0.8	0.9	0.845	0.811	0.861	0.819
0.6	1.0	0.680	0.538	0.699	0.544
0.7	1.0	0.773	0.675	0.788	0.683
0.8	1.0	0.837	0.781	0.847	0.789

ABC=activated B cell-like DLBCL; HR=hazard ratio; ITT=intent-to-treat; *k*=proportion of EFS events in ABC vs. ITT population

1.5. Randomization and Blinding

Central randomization will be implemented in this study. Subjects will be randomly assigned in a 1:1 ratio to 1 of the 2 treatment arms based on a computer-generated randomization schedule prepared before the study and under the supervision of the sponsor. The randomization will be balanced by using randomly permuted blocks, and stratified by R-IPI score (1-2 vs. 3-5), region (U.S./Western Europe vs. Rest of World), and number of pre-specified treatment cycles (6 cycles vs. 8 cycles), then randomized in a 1:1 ratio to either Treatment Arm A (placebo+R-CHOP) or Treatment Arm B (ibrutinib+R-CHOP).

This is a double-blind study. Subjects, investigators, the IRC, and the sponsor's study team members will remain blinded to treatment assignment until the database has been locked for the clinical study report. Personnel who may be unblinded during the study are:

- The independent DMC, 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.

- Personnel performing blood serum concentration assays and analysis for pharmacokinetics.
- Unblinded sponsor safety representative and Ethics Committee for serious adverse event (SAE) reporting.

2. GENERAL ANALYSIS DEFINITIONS

2.1. Analysis Sets

The analysis populations are defined as:

1. Intent-to-Treat (ITT) population: all randomized subjects who are enrolled with the non-GCB DLBCL subtype by IHC. Subjects in this population will be analyzed according to the treatment to which they are randomized.
2. ABC population: ITT population who are identified as having the ABC subtype as determined by GEP (retrospectively determined using available clinical study formalin-fixed paraffin embedded tissue specimens). Subjects in this population will be analyzed according to the treatment to which they are randomized.
3. Per-protocol (PP) Population: all randomized subjects who undergo at least 1 postbaseline disease assessment and do not have major protocol violations as described below:
    a. did not meet all inclusion and exclusion criteria
    b. did not receive the treatment to which they were randomized
    c. had less than 75% of study treatment compliance

    Subjects in this analysis set will be analyzed according to the treatment to which they are randomized.
4. Biomarker population: all randomized subjects with biomarker data collected.
5. Pharmacokinetic-evaluable population: all randomized subjects who received at least 1 dose of study drug and had at least 1 pharmacokinetic sample obtained posttreatment.
6. Safety population: all randomized subjects who received at least 1 dose of study drug. Safety data will be analyzed according to the actual treatment received.

The ITT population will be used to summarize characteristics, efficacy, and PRO data and the safety population will be used to summarize the safety data, unless otherwise specified. The ABC population will be used to summarize the efficacy and/or other data as appropriate. The PP population will be used for sensitivity analysis of the primary endpoint EFS if there is a more than 10% difference between ITT and PP populations. The biomarker population will be used to summarize the biomarker data. Pharmacokinetic-evaluable population will be used to summarize the pharmacokinetic data.

2.2. Study Treatment and Study Medication

For the purpose of analysis, the term “study treatment” refers to all drugs, ibrutinib/placebo in combination with rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone (R-CHOP). The term “study medication or study drug” refers to ibrutinib/placebo.

2.3. Baseline Definitions or Conventions

Unless specified otherwise, the baseline value is defined as the last available (non-missing) value collected on or prior to the date of the first administration of study drugs.

2.4. Study Day and Visit Windows

The randomization day is considered to be Day 1. Study day is defined as the current date - Day 1 + 1.

The study date and corresponding study visit/cycle will be captured on each case report form (CRF). Visit windows will be created around the study day of each scheduled visit. They will be used to aggregate data values that are to be summarized by visit. Table 9 lists an example of the visit windows and the targeted study day. Details about the visit window will be specified in the DPS.

Table 9: Visit Windows

Scheduled Study Day	Visit Window	Scheduled Study Day	Visit Window
Cycle 1 Day 1[a]	1, 21	Cycle 5 Day 1 (Day 85)	85, 105
Cycle 2 Day 1 (Day 22)	22, 42	Cycle 6 Day 1 (Day 106)	106, 126
Cycle 3 Day 1 (Day 43)	43, 63	Cycle 7 Day 1 (Day 127)	127, 147
Cycle 4 Day 1 (Day 64)	64, 84	Cycle 8 Day 1 (Day 148)	148, 168

[a] Study Day 1 begins on the day of randomization.
Each cycle consists of 21 days.
Baseline value is defined as the last available value collected on or prior to the randomization day.
When there is a cycle delay, subsequent cycles will be shifted based on the above windows accordingly.

2.5. Imputation of Missing 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 and concomitant and subsequent therapies, date of progression/relapse on the last prior therapy, date of second PD, and date of initial diagnosis according to the following rules. Start date will be imputed before the end date.

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

The above imputations will be modified by the following rules:

- If such imputed date for prior therapies or initial diagnosis is on or after the randomization date, then randomization date - 1 day will be used. If such imputed date for subsequent therapies is before date of last dose, then date of last dose +1 day will be used.

- The imputed start date for subsequent therapies will be adjusted sequentially using the following steps:
  - If the imputed start date is before the treatment discontinuation date or (last dose date if no treatment discontinuation dates) but in the same year and month, then the treatment discontinuation date or last dose date if no treatment discontinuation date will be used.
  - If subsequent therapy end date is not missing and is before the imputed subsequent therapy start date, then the subsequent therapy end date will be used as the start date.
- If the imputed date is for a date of death and is before the last date that the subject is known to be alive, the latter date will be used.
- The imputed AE start date will be adjusted sequentially using the following steps:
  - If the imputed date is in the same year and month but before the first dose date, then the first dose date will be used, or if it is in the same year and month but after the last dose date + 30 days, then the last dose date + 30 days will be used.
  - If AE end date is not missing and the imputed AE start date is after the AE end date, then the AE end date will be used.
  - If the imputed AE start date is after date of death, then date of death will be used.
  - If the imputed AE start date is in the same month and year but after the 1st subsequent therapy start date, then 1st subsequent therapy start date will be used.
- If the imputed date is for an AE end date and is after the death date, then the death date will be used, or if the imputed AE end date is before the AE start date, then the AE start date will be used.
- The AE imputation rule will be used for concomitant medication.

2.6. Definitions of Subgroups

Table 10 provides definitions for each of the subgroups.

Table 10: Subgroups

Subgroup	Definition of Group	Analysis Type[a]
Gender	Male, Female	E, S, D
Race	Caucasian, non-Caucasian	E, S, D
Region	U.S./Western Europe, Rest of World	E, S, D
Age	<65, ≥65 years	E, S, D
Baseline ECOG	0-1, 2	E, S, D
Baseline R-IPI score	Good [1-2] (Low), Poor [3-5] (High)	E, D
Number of pre-specified treatment cycles	6 cycles, 8 cycles	E, D
Baseline DLBCL Stage of Disease	II, III-IV	E, D
Baseline lactate dehydrogenase level	≤ upper limit of local normal range, > upper limit of local normal range	E, D
Number of extranodal sites of disease at baseline	≤ 1, >1 extranodal sites	E, D
Hepatic Impairment	Normal vs (Mild or Moderate)[b]	S, D
Concomitant use of CYP3A inhibitor	Strength of CYP3A inhibitor (mild, moderate, strong, and other)	S
GEP subtype	GCB, ABC, Unclassified	E, D

[a] D=demographics; E=efficacy; S=safety. Analyses will be performed for each subgroup as appropriate.
ECOG=Eastern Cooperative Oncology Group
[b] National Cancer Institute Organ Dysfunction Working Group (NCI ODWG) Liver Function Classification.

2.7. Other General Definitions

- Subsequent antilymphoma therapy is not allowed prior to PD/relapse from CR except for either PET-positive or biopsy-proven residual disease after completion of at least 6 cycles of R-CHOP therapy. The start date of the first subsequent antilymphoma therapy will be entered in the electronic CRF (eCRF).
- A PET scan is required at the end of treatment. It is recommended but not mandated at baseline. The end of treatment includes the completion of the scheduled 6 or 8 cycles of R-CHOP treatment or early withdrawal.
- If an early withdrawal from treatment occurs after completion of 6 cycles of R-CHOP therapy, the initiation of subsequent antilymphoma therapy due to PET-positive or biopsy-proven residual disease will be counted as an EFS event. However, if the early withdrawal from treatment occurs before completion of 6 cycles of R-CHOP therapy, it will not be counted as an EFS event, even in the presence of PET-positive or biopsy-proven residual disease. Note that PET-positive residual disease is collected on an eCRF page.
- Regarding bone marrow biopsy, data reported on the 'Bone Marrow Biopsy and Aspirate' CRF will be used to derive bone marrow involvement. Subjects with a bone marrow aspirate or biopsy histology result of "positive" or histology negative but confirmed to be positive by IHC or flow cytometry is "yes"' at baseline based on any method of assessment will be considered to have bone marrow involved at baseline.

Disease progression collection after the initiation of subsequent antilymphoma therapy: after the initiation of subsequent antilymphoma therapy, the patient will continue to be followed to assess disease progression every 16 weeks in the first 24 months, then every 24 weeks up to PD/relapse from CR, withdrawal, death, or clinical cutoff for the primary analysis, whichever occurs first.

2.7.1. Treatment-Emergent Adverse Events

Treatment-emergent AEs are AEs that occur after the first dose of study treatment, and within 30 days following the last dose of study treatment; any AE that is considered study treatment-related regardless of the start date of the event; or any AE that is present at baseline but worsens in severity within 30 days following the last dose of study treatment or is subsequently considered study treatment-related by the investigator. Adverse events occurring after the initiation of the subsequent therapy will not be considered as the TEAE.

2.7.2. Year and Month

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

2.7.3. Age

Age in years is collected on an eCRF page.

2.7.4. Time from Initial Diagnosis to Randomization

Time from initial diagnosis to randomization in days will be calculated as (date of randomization - date of initial diagnosis) and the result will be rounded to the integer. Partially missing initial diagnosis dates will be imputed based on the rules provided in Section 2.5 of the SAP.

2.7.5. Date of Overall Response/Progressive Disease for Each Timepoint

When multiple or repetitive assessments for the timepoint are performed on different dates, then:

- For overall response:

The date is determined as the date of last assessment needed to document the response (CR, partial response [PR], or stable disease [SD])

- For PD (PD/relapse from CR):

The date is determined as the date of the earliest assessment used to document PD or relapse from CR. The PD date is collected on the eCRF page.

2.7.6. Date of Subsequent Systemic Antilymphoma Therapy Event

The subsequent systemic antilymphoma therapy event is defined as the initiation of systemic antilymphoma therapy for either PET-positive or biopsy-proven residual disease upon completion of at least 6 cycles of R-CHOP therapy. The date is determined as the date of the earliest disease assessment documenting the need for a subsequent systemic antilymphoma therapy event.

2.7.7. Date of Response/Best Response

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

2.7.8. Date of Progression

Date of first observed progression/relapse from CR.

2.7.9. Date of Subsequent Systemic Antilymphoma Therapy Event

Date of first subsequent systemic antilymphoma therapy event.

2.7.10. Date of EFS Event

Date of first observed EFS event: either progression, relapse from CR, subsequent systemic antilymphoma therapy event, or death.

2.7.11. Date of PFS Event

Date of first observed PFS event: either progression, relapse from CR, or death.

3. INTERIM ANALYSIS AND DATA MONITORING COMMITTEE REVIEW

3.1. Interim Analysis

In the original study design, 1 interim analysis was planned for the primary endpoint EFS, when approximately 270 EFS events had been observed. However, the planned interim analysis was omitted in protocol Amendment INT-3 due to the lower than expected event accumulation. The final analysis of the primary endpoint of EFS is calendar driven, more specifically, the clinical cutoff is 30 months after the 800th subject is randomized into the study.

3.2. Data Monitoring Committee

A DMC will be established to monitor data on an ongoing basis to ensure the safety of subjects enrolled in this study and to evaluate the efficacy of the treatment at the time of interim analysis. The committee will meet periodically to review interim data, assess the evidence of benefit or adverse effects of ibrutinib, and to monitor the conduct of the study. After the review, the DMC will make recommendations regarding the conduct of the study. The details regarding the DMC responsibilities, authorities, and procedures will be provided in a separate DMC charter.

At the interim analysis, the DMC could recommend stopping the study for efficacy, if the pre-specified stopping boundary was crossed. In addition to the planned interim analysis for efficacy, 4 safety review meetings were planned, which would occur approximately 1 month after 50, 250, 450, and 650 subjects have been randomized. The safety review was to focus on deaths, treatment discontinuations, SAEs, Grade ≥3 events, and AEs of special interest. Based on the results from these scheduled safety review meetings, the DMC chair requested an additional safety review meeting when the last subject was randomized. This additional meeting occurred on 29 January 2016. After reviewing the safety profiles for all 838 enrolled subjects, the DMC decided that there was no need to review any additional safety data. The next data review was

planned to occur at an interim analysis, when approximately 270 EFS events are observed. All deaths, treatment discontinuations, and SAEs will be reviewed by the sponsor's responsible physician on an ongoing basis to identify safety concerns and the DMC will be informed of any new potential signals.

As indicated earlier, the original planned interim analysis was omitted in protocol Amendment INT-3 due to the lower than expected EFS event rate for blinded data. The interim analysis-related statements are no longer applicable.

4. SUBJECT INFORMATION

All statistical analyses will be performed using statistical analysis system (SAS®). Analyses of disposition, demographic, baseline disease characteristics, and prior and concomitant therapy will be conducted on the ITT population. Analyses of treatment compliance and extent of exposure will be conducted on the Safety population. Analyses of background therapy (R-CHOP) 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, 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 and/or evaluable population, as appropriate.

4.1. Disposition Information

The number of subjects randomized, discontinuing and completing the study treatment will be reported for the ITT population. The reasons for discontinuation, as indicated by the investigators, will be summarized by the number of subjects reported. Those subjects who discontinue study treatment due to AEs will be further tabulated by relationship to study medication. The tabulation of the above information will be provided by treatment arms and combined.

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

4.2. Demographics and Baseline Characteristics

All demographic and baseline characteristics will be summarized for the ITT population.

Subject enrollment will be summarized by country and site.

Summary of baseline demographics, vital signs, disease characteristics, and laboratory tests graded by National Cancer Institute Common Terminology Criteria for Adverse Events

(NCI-CTCAE; the most recent version will be used), and stratification factors will be summarized by treatment arm and combined.

Two panels of baseline clinical laboratory tests will be summarized by frequency: hematology, and serum chemistry. Frequencies of NCI-CTCAE grade are to be provided for the following laboratory tests at baseline:

Hematology: hemoglobin, white blood cell, absolute neutrophil count (ANC), absolute lymphocyte count (ALC), and platelets.

Chemistry: sodium, potassium, creatinine, creatinine clearance, alanine aminotransferase (ALT), aspartate aminotransferase (AST), alkaline phosphatase, LDH, total bilirubin, albumin, and magnesium.

Note: Magnesium is to be evaluated on Day 1 of Cycles 1 and 2, and as clinically indicated.

4.3. Extent of Exposure

All extent-of-exposure summaries are to be presented for the safety population. The frequency of subjects who receive at least 1 postbaseline treatment cycle will be tabulated by the number of cycles. The number of treatment cycles will be summarized, along with time on study treatment and dose intensity for each study drug.

Summaries will be provided for total dose of each study drug administered by cycle and total cumulative dose. The number and percentage of subjects on treatment, and number and percentage of subjects with cycle delays will also be presented based on the CRF data collection. For each study drug, the following summaries are to be reported: number of subjects dosed and number of subjects with cycle delay, dose reduced, and dose discontinued.

The reasons for dose modification will be summarized for each of the following categories: cycle delay, dose reduction, and dose discontinuation.

For each type of dose modification mentioned above, AEs leading to change in dose will be summarized for each of the study drugs by treatment arms.

4.4. Protocol Eligibility and Major Protocol Deviations

Subjects with major protocol deviations will be listed by treatment arm. Protocol deviations will be based on clinical review, including, but not limited to the following: eligibility criteria, treatment compliance, subject safety, efficacy assessment deviations, and other(s) which might significantly impact safety/efficacy assessment. 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.

4.5. Prior and Concomitant Medications

For summarization purposes, medications will be coded to a generic term based on the World Health Organization (WHO) drug dictionary. Medications administered prior to the first dose of study medication will be considered prior medications. Concomitant therapies include those taken on or after first dose date of study treatment through 30 days after the last dose of study treatment. However, for CYP3A inhibitors or inducers, concomitant use is defined as administration on or after the first dose of study treatment through the last dose of study treatment.

The incidence of prior and concomitant medications will be summarized by Anatomical Therapeutic Chemical (ATC) class and drug generic term.

4.6. Medical History

Abnormal medical history findings reported by investigator will be coded using the Medical Dictionary for Regulatory Activities (MedDRA; the most recent version will be used) and summarized by system organ class and preferred term.

5. EFFICACY

5.1. Analysis Specifications

5.1.1. Level of Significance

In general, all tests will be performed at a 1-sided significance level of 0.025, unless otherwise specified. All interval estimations will be reported using 2-sided 95% confidence intervals (CIs).

Multiplicity adjustment for the primary endpoint EFS for both ITT population and ABC population is described in Section 1.4.1 using the Song and Chi method (Song and Chi 2007). Following Figure 1, if the weighted test statistics p-value at Stage I is <0.04 (2-sided), then the significance level will be 0.05 (2-sided) for both the ITT and ABC populations. If the p-value corresponding to the weighted test statistic at Stage I is ≥0.04 but <0.2 (2-sided), then the significance level for ABC population will be 0.02709, based on the actual proportion (176/247=0.7126) of the ABC EFS events among the total events in the ITT population. If the observed p-value for ABC population is < 0.02709, then the null hypothesis for ABC population will be rejected and we can re-test the ITT population using 0.05 (2-sided) significance level. The level of significance for secondary endpoints is specified in Section 5.3.5.

5.1.2. Data Handling Rules

The data handling rules will be outlined in the DPS.

5.1.3. General Analysis Considerations

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

Disease progression and response will be based on investigator assessment, using the Revised Response Criteria for Malignant Lymphoma (Cheson 2007). Detailed criteria for response categories can be found in the protocol. An IRC will perform an independent disease assessment for selected random samples based on the audit plan to determine disease progression and response, which will be evaluated as confirmation of the investigator tumor evaluation. The details will be described in audit plan (Attachment 1).

Key efficacy analysis will be performed based on both the ITT and the ABC populations as described in Section 2.1. The subgroup analyses (see the list of subgroups in Section 2.6) will be performed for both the primary endpoint EFS and the key secondary endpoints including PFS and OS.

5.2. Primary Efficacy Endpoint

5.2.1. Definition

The primary endpoint is EFS, which is defined as the duration from the date of randomization to the date of disease progression, relapse from CR as assessed by the investigator, initiation of systemic antilymphoma therapy for either PET-positive or biopsy-proven residual disease upon completion of at least 6 cycles of R-CHOP therapy, or death, whichever occurs first.

Subjects who have withdrawn from the study or who are lost to follow up will have EFS censored at the date of their last adequate disease assessment on or before the withdrawal date. For subjects who do not have an EFS event (ie, subjects who do not have PD/relapse from CR, a subsequent systemic antilymphoma therapy event, and are alive), as well as for those subjects with unknown EFS status or unknown survival status, as of the data cutoff date, EFS will be censored at the date of the subject's last adequate disease assessment. If there are no baseline or postbaseline disease assessments (and no death is observed), for a subject, EFS will be censored on the date of randomization. All subjects will continue to be followed until either confirmation of PD/relapse from CR, death, or clinical cutoff for the primary endpoint, whichever occurs first, for the final EFS analysis. The details of the censoring rules are listed in Table 11.

Table 11: Censoring Rules for Primary EFS Analysis

Situation	Outcome	Date	Event Description/ Censoring Reason
Progression/Relapse from CR[a] documented at a scheduled visit or between scheduled visits	EFS event[b]	PD date[a]	PD
Death before first documented PD	EFS event	Date of death	Death
PD or death, documented after clinical cutoff but prior to the database lock	EFS event	PD date or death date	PD or Death
Initiation of subsequent systemic antilymphoma therapy event due to residual disease[c] at a scheduled visit or between scheduled visits	EFS event	Earliest date of disease assessment documenting need for subsequent systemic antilymphoma therapy event[c]	New antilymphoma therapy due to residual disease[c]
No baseline tumor assessments	Censored	Date of randomization	No baseline disease assessment
No on-study tumor assessments, no death, and no withdrawal	Censored	Date of randomization	No death or on-study disease assessment
Withdrawal of consent to study participation or investigator/sponsor termination[e] before PD or death	Censored	Date of last adequate disease assessment[d]	Withdrew consent
Lost to follow up before documented PD or death	Censored	Date of last adequate disease assessment[d]	Lost to follow up

CR=complete response; EFS=event-free survival; PD=progressive disease; PET=positron emission tomography
[a] PD will indicate progression/relapse from CR; PD date is the earliest date of disease assessment documenting progression/relapse from CR.
[b] If more than 1 EFS event occurs for a subject, the event occurring first will be the subject's EFS event, and this date will determine the subject's time to EFS event.
[c] The initiation of subsequent systemic antilymphoma therapy event is a special case of initiation of subsequent systemic antilymphoma therapy, and is defined as the initiation of systemic antilymphoma therapy for either PET-positive or biopsy-proven residual disease upon completion of at least 6 cycles of R-CHOP therapy.
[d] Adequate disease assessment is defined as an assessment having sufficient evidence to correctly indicate that progression has or has not occurred. The Screening visit, the visit occurring at the end-of-treatment Cycle 4, the End-of-Treatment visit, and Pre-PD Posttreatment visits were designed to provide adequate disease assessment.
[e] Discontinuations/withdrawals mentioned in this table refer to those that apply to the entire study and not to those that solely apply to treatment.

5.2.2. Primary Analysis of EFS

The primary analysis will be based on the ITT and ABC populations following the Song and Chi method (Song and Chi 2007) discussed in Section 1.4.1. At Stage I, the stratified log-rank test will be used for both the ITT and ABC populations; the weight parameter of k will be the proportion of the number of EFS events observed in the ABC population (0.02709 based on data after CCO). The weighted test statistic Z_1 will be estimated using equation (2) in Section 1.4.1, which considers both the ABC and complementary ABC population (GCB, unclassified and unknown/missing). If the corresponding p-value (p_1) to the weighted test statistic Z_1 is less than 0.04 (2-sides), then test both the ITT and ABC populations using the significance level of 0.05 (2-sided) based on the stratified log-rank test. The stratification factors to be used in the analysis are R-IPI (1-2 vs. 3-5), region (U.S./Western Europe vs. Rest of World), and number of

pre-specified treatment cycles (6 cycles vs. 8 cycles). If p_1 is greater than or equal to 0.04 but less than 0.2, then test the ABC population using the significance level of 0.02709. If the ABC population is significant, then re-test the ITT population using the significance level of 0.05. (2-sided). The HR for ibrutinib+R-CHOP relative to placebo+R-CHOP and its associated 95% CI will be calculated based on the stratified Cox proportional hazards model, stratified by the stratification factors. The EFS distribution and median EFS with a 95% CI will be estimated using the Kaplan-Meier product-limit method for the ITT population.

In case there is strong evidence of crossing hazard, other test statistics such as Rényi statistics (Rényi 1953) may be used instead of log-rank test. In addition, the proportional hazards cure model will be used to estimate the cure rate for each treatment arm.

5.2.3. Sensitivity Analysis of EFS

Sensitivity analyses will be performed to evaluate the robustness of the analyses. Per protocol Amendment INT-3, the retrospective GEP test for all study screened subjects will be performed. Some of the screen failure patients due to GCB by IHC were proven to be ABC by GEP and should have been enrolled in the study. Imputation of the clinical data for those subjects who are GCB by IHC but ABC by GEP will be made to assess whether the results are consistent if the GEP test had been used for subject selection instead of IHC. In addition, approximately 10% of patients in the study may not have sufficient or unqualified samples for GEP testing. Sensitivity analyses will be performed after imputing the missing GEP results.

5.2.3.1. Imputing the Screen Failures due to GCB by IHC who are ABC by GEP

The original study design is to enroll subjects with the non-GCB subtype based on IHC. It is increasingly appreciated that DLBCL is heterogeneous in terms of morphology, genetics, and biologic behavior. Gene expression profiling and IHC are 2 of the assay platforms currently being used for determining cell of origin. Gene expression profiling platforms assign specimens into 1 of 3 diagnostic categories: GCB, ABC, or unclassified. Immunohistochemistry divides the population into 2 categories: GCB and non-GCB. The study protocol was amended to perform retrospective GEP testing for all screened study subjects. The sensitivity analysis will be performed by imputing the subjects with the ABC subtype that were classified as GCB by IHC. The hot deck imputation method will be used to impute the missing data.

For example, the missing A subjects (GCB by IHC and ABC by GEP and meeting all eligibility criteria) are allocated to the ibrutinib group and placebo group using a 1:1 randomization ratio. Randomly assign A x 50% subjects to ibrutinib group (A1), and A x 50% subjects to placebo group (A2).

- Given a percentage p, for each of the A1 x p subjects assigned to the ibrutinib group, sample the subjects based on the bootstrapping method with replacement from the ibrutinib group in the sampling pool. For each of the A x (1-p) subjects assigned to the ibrutinib group, sample the subjects with replacement from the placebo group in the sampling pool.

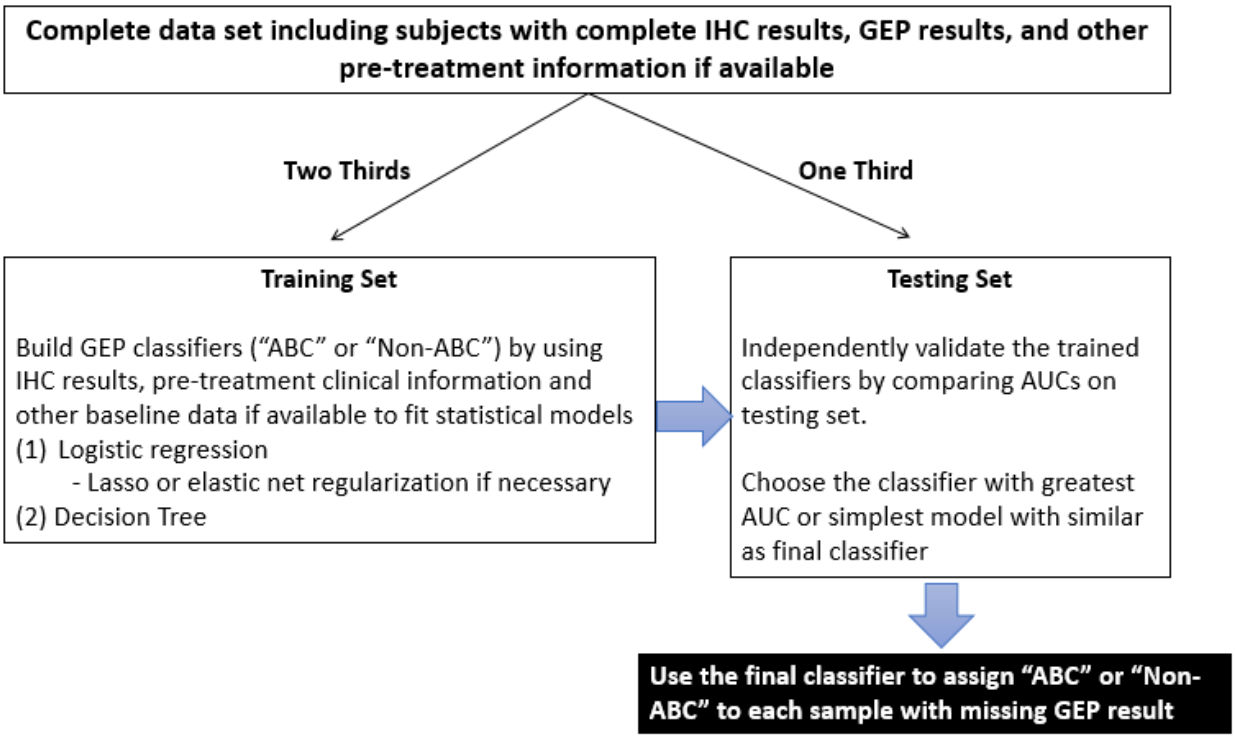
- For each of the A2 subjects assigned to the placebo group, sample the subjects with replacement from the placebo group in the sampling pool.
- Repeat above steps 1,000 times. Efficacy result will be estimated based on multiple imputation method through SAS MIANALYZE procedure.

The percentage p can be chosen as 100%, 75%, and 50%, respectively. The lower percentage of p is the lower estimated ibrutinib effect and favors the control arm.

5.2.3.2. Imputing Missing GEP Test Results

The GEP test will be retrospectively performed for all screened study subjects. It is expected that not every subject will have sufficient samples for the GEP test or will not be amenable to testing and so the GEP subtype will be undetermined. However, the missing GEP test results (~10%) can be reasonably assumed to be random. Therefore, the missing GEP results will be statistically imputed. A sensitivity analysis will be performed based on the imputed missing GEP results. The imputation method will follow the procedures illustrated in Figure 2. First, the complete dataset including subjects with known IHC results, GEP results, and other pretreatment or baseline information if available will be randomly split into 2 subsets: two thirds for the training set and one third for the independent testing set. The training set will be used to fit 2 classification models - logistics regression model and decision tree. Least absolute shrinkage and selection operator (lasso) and/or elastic net regularization method may be applied in variable selection in logistic regression should a great number of baseline covariates are available. Second, the predictive accuracy of the fitted models using the training set will be assessed using the testing set data. The predictive accuracy will be measured by the area under the receiver operating characteristic (ROC) curve. Generally, the model or GEP classifier with greatest area under the curve (AUC) estimated from testing dataset will be the final classifier. However, if the AUC of logistic regression model is close to but not the greatest, the consideration will be given to this model because it can be easily fitted, verified and interpreted. Missing GEP results will be imputed using the final classifier. The cutoff of the predicted probability of GEP ABC for the final classifier will be determined to maximize ABC concordance or sensitivity while Non-ABC concordance or specificity will be equal to or greater than 80%. This relatively high specificity will ensure reasonably high ABC predictive value, or positive predictive value (PPV).

Figure 2: Flow Chart for the Imputation of Missing GEP Sample



For the imputed ABC subjects, some of them may be already randomized in the study and some of them may be screen failures (SFs) due to the IHC GCB result. These SF cases do not have clinical data information because they were not enrolled in the study. Therefore, the hot deck method discussed in Section 5.2.3.1 will used to impute the postbaseline information for these SFs.

5.2.3.3. Sensitivity Analysis

Sensitivity analyses will include:

1) Stratified Wilcoxon test for the ITT population
2) Restricted Mean EFS Time difference (Uno 2014)
3) Subjects with EFS events observed after missing ≥2 consecutive scheduled disease assessment visits will be censored at the last adequate disease assessment before the 2 consecutive missing assessment visits
4) Censoring subjects who receive subsequent antilymphoma therapy and who did not meet the criteria for a subsequent antilymphoma therapy event at the date of last adequate disease assessment on or before subsequent antilymphoma therapy
5) Sensitivity analysis based on the final validated assay for the non-GCB subtype of DLBCL
6) Sensitivity analysis based on the imputed results for subjects who are ABC by GEP, but GCB by IHC will be performed as appropriate (data imputation will be based on Section 5.2.3.1)

7) Sensitivity analysis will be performed as appropriate for subjects who have a missing GEP sample, based on the imputed ABC subtype results (imputation is described in Section 5.2.3.2).

8) Sensitivity analysis based on the per-protocol population. If 90% or more subjects from ITT analysis set are included in PP analysis set then no analysis will be performed on PP analysis set.

5.2.3.4. Additional Covariate-adjusted Analysis of EFS

Additional exploratory analyses on EFS will be performed using a selected set of potential prognostic variables (obtained at or before baseline) as covariates in Cox regression models. Potential prognostic factors are as follows:

- Age (<65 vs. ≥65 years)
- Gender
- Race: (Caucasian vs. non-Caucasian)
- Revised International Prognostic Index (1-2 vs. 3-5)
- Region (U.S./Western Europe vs. Rest of World)
- Number of pre-specified treatment cycles (6 vs. 8 cycles)
- Baseline ECOG performance status (0-1 vs. 2)
- Baseline DLBCL stage of disease (II vs. III-IV)
- Baseline LDH level (≤ upper limit of local normal range vs. > upper limit of local normal range)
- Number of extranodal sites of disease at baseline (≤ 1 vs. >1 extranodal sites)
- GEP subtype, GCB vs. ABC vs. unclassified

Each factor will be assessed individually for prognostic values (1-sided p-value <0.05) using a univariate Cox model. Factors that are deemed to have prognostic value will be included as covariates in a multivariate Cox model to assess their significance in the presence of other factors. Selection methods (such as stepwise selection) will be used to identify the final set of prognostic factors (exit 1-sided p-value set to be 0.05). Treatment will then be added to this final model to assess the effect of treatment when adjusted for these prognostic factors.

5.2.3.5. Subgroup Analysis of EFS

Descriptive subgroup analysis will be performed for the selected potential prognostic variables (Section 5.2.3.4) to assess the internal consistency of the treatment benefit using a Forest plot. Hazard ratios between the 2 treatment arms within the subgroups and their 95% CIs will be calculated.

Additionally, exploratory subgroup analyses will be performed on EFS to determine the possible interaction of the subgroups with treatment using the Cox Proportional Hazard Model. If an

interaction results in statistical significance (1-sided p <0.10), then the nature of treatment comparisons within each subgroup stratum will be examined.

5.3. Secondary Efficacy Endpoints

The secondary endpoints, along with corresponding analysis methods, are defined in the following subsections. The treatment effect of ibrutinib+R-CHOP compared to placebo+R-CHOP will be tested with the corresponding statistics test. For the time to events endpoint, log-rank test will be used; for binary endpoint, the Cochran-Mantel-Haenszel Chi-square statistic will be used under the hierarchical test procedure, the comparison will be subject to the significance of the superiority test of ibrutinib over placebo.

5.3.1. Progression-free Survival

Progression-free survival is defined as the duration from the date of randomization to the date of progression, relapse from CR, or death, whichever occurs first.

Progression-free survival will be analyzed in a similar fashion as EFS. However, in contrast to EFS, the censoring rules for PFS will not be influenced by initiation of subsequent antilymphoma therapy, either as an event or otherwise. Sensitivity analysis, subgroup analysis, and the covariate-adjusted analysis will be performed as appropriate in a similar fashion as the primary endpoint EFS, except for the censoring rule; ie, PFS is not censored for subjects who received subsequent antilymphoma therapy.

The treatment effect of ibrutinib+R-CHOP compared to placebo+R-CHOP based on PFS will be tested with a stratified log-rank test adjusted for the randomization stratification factors.

The HR for ibrutinib+R-CHOP relative to placebo+R-CHOP and its associated 95% CI will be calculated based on the stratified Cox proportional hazards model, stratified by the stratification factors.

The PFS analysis will be performed for both the ITT and ABC populations.

5.3.2. Complete Response Rate

Complete response rate is defined as the proportion of subjects who have measurable disease at baseline and achieve a CR prior to the initiation of subsequent antilymphoma therapy.

All randomized subjects who have a valid baseline value will be included in this analysis. Subjects with missing post-randomization data are considered non-responders. Complete response rate will be summarized using descriptive statistics for categorical data by treatment arm.

The relative risk (ibrutinib+R-CHOP vs. placebo+R-CHOP) of CR will be reported along with the associated 95% CI. Statistical inference will be evaluated using the Cochran-Mantel-Haenszel Chi-square statistic, adjusted for the stratification factors R-IPI score, region, and number of pre-specified treatment cycles. Logistic regression analysis will also be performed to estimate an odds ratio and its associated 95% CI between the 2 treatment arms, adjusted for the stratification factors.

The CR analysis will be performed for both the ITT and ABC populations.

5.3.3. Overall Survival

Overall survival is defined as the duration from the date of randomization to the date of the subject's death. 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.

Overall survival will be analyzed using the stratified log-rank test for treatment comparison. The OS distribution and median OS with its 95% CI will be estimated using the Kaplan-Meier product-limit method. The HR for ibrutinib+R-CHOP relative to placebo+R-CHOP and its associated 95% CI will be calculated based on the stratified Cox proportional hazards model, incorporating the stratification factors.

The same subgroup analysis planned for EFS may be performed for OS if the number of events within each subgroup is sufficient.

The OS analysis will be performed for both the ITT and ABC populations.

In addition, exploratory analyses, using the Cox proportional hazards model with and without the covariates specified in Section 5.2.3.4, will be performed.

5.3.3.1. Time to Worsening on the Lym Subscale

Time to worsening (TTW) on the Lym subscale of the FACT-Lym is defined from the date of randomization to the start date of the worsening of patient symptoms. Worsening is defined by 2 criteria; a 3-point decrease and a 5-point decrease from baseline in patient symptoms (see Section 7). Time to worsening will be analyzed using the log-rank test.

The treatment effect of ibrutinib+R-CHOP compared to placebo+R-CHOP based on TTW will be tested with a stratified log-rank test adjusted for the randomization stratification factors. The HR for ibrutinib+R-CHOP relative to placebo+R-CHOP and its associated 95% CI will be calculated based on the stratified Cox proportional hazards model, stratified by the stratification factors.

5.3.4. Related Exploratory Analyses

5.3.4.1. Overall Response Rates

Overall response rate is defined as the proportion of subjects who achieve either a CR or PR as their best overall response as assessed by investigator. All responses of CR and PR after subsequent antilymphoma therapy will be excluded.

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

The order of overall response categories is: CR> PR> SD> PD> NE. The highest category of overall response is the best overall response.

Overall response rate will be estimated in the same manner as CR rate.

5.3.4.2. Duration of Response

Duration of response will be analyzed for subjects with a CR or PR according to the investigator and is defined as the interval between the date of initial documentation of a response and the date of first documented evidence of PD or death. The censoring rule for duration of response is the same as PFS.

Only subjects who achieve a CR or PR will be included in the analysis of duration of response. Duration of response will be summarized descriptively using the Kaplan-Meier method.

5.3.5. Multiplicity Adjustment for Secondary Endpoints

Inferential hypothesis testing for secondary endpoints will be conducted only if the primary analysis of EFS achieves statistical significance.

A sequential gate-keeping procedure will be used to test other secondary endpoints if the primary endpoint of EFS achieves statistical significance for both the ITT and ABC populations. Formal hypothesis testing will be made for the ITT population only in hierarchical order. More specifically, if PFS achieves statistical significance, then CR will be tested, If CR achieves statistical significance, then OS will be tested, followed by TTW in the Lym subscale of the FACT Lym. All statistical significance levels are defined at a 1-sided alpha of 0.025.

5.4. Other Exploratory Efficacy Endpoints

Additional exploratory efficacy endpoints:

- Evaluate the mean change from baseline in EQ-5D-5L score for each postbaseline assessment and/or the FACT-Lym score, as appropriate.
- Evaluate biomarkers associated with resistance to ibrutinib in subjects who progress on ibrutinib treatment compared to those who remain in CR.

6. SAFETY

All safety analyses will be based on the safety population. The baseline value for safety assessment is defined as the value collected at the time closest to, but prior to, the start of study drug administration.

Safety parameters to be evaluated are the incidence, intensity, and type of AEs, clinically significant changes in the subject's physical examination findings, vital signs measurements, and clinical laboratory results by treatment arm.

Safety analyses will be performed by the treatment actually received.

6.1. Adverse Events

6.1.1. All Adverse Events

The verbatim terms used in the CRF by investigators to identify AEs will be coded using the MedDRA (the latest version will be used). All reported AEs with onset date on or after the first dose of study drug up through 30 days after the last dose (ie, TEAEs) will be included in the analysis. For each AE, the percentage of subjects who experience at least 1 occurrence of the given event will be summarized by treatment arm.

Treatment-emergent AEs are AEs that occur after the first dose of study drug, through the Active Treatment Phase, and for 30 days following the last dose of study drug; any AE that is considered study drug-related regardless of the start date of the event; or any event that is present at baseline but worsens in severity or is subsequently considered study drug-related by the investigator.

Analysis of TEAEs includes:

- Incidence of AEs and SAEs by system organ class and preferred term
- Incidence of AEs and SAEs by severity
- Incidence of AEs and SAEs by relationship to study drug
- Incidence of AEs by subgroup (eg, age, sex, race, region, baseline ECOG, baseline hepatic impairment, and concomitant use of CYP3A inhibitor)

For each 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 (highest to lowest incidence within ibrutinib+R-CHOP arm). The same summary will be provided for study drug-related AEs (study drug-related includes possible, probable, or very likely related to study drug), SAEs, and drug-related SAEs, as well as AEs leading to treatment discontinuation and death.

6.1.2. Adverse Events of Special Interest

Major hemorrhage is an AE of special interest and is a subset of hemorrhagic events that are Grade ≥3, serious, or belonging to central nervous system (CNS) hemorrhage/hematoma.

Other safety observations include cytopenia, lymphocytosis, atrial fibrillation and atrial flutter, infections, other malignancies, hypertension, Leukostasis, Tumor lysis syndrome, Hypersensitivity, Eye disorders, Hepatic disorders, SCAR (severe cutaneous adverse reactions), interstitial lung disease, Cardiac Arrhythmias (excluding Atrial Fibrillation) and rituximab-related infusion reaction. Where applicable, the currently agreed-upon grouping criteria for ibrutinib TEAEs will be applied. Treatment-emergent AEs of special interest will be summarized similarly to TEAEs.

6.2. Deaths

The incidence of deaths within 30 days after last dose are to be reported, as well as all deaths for all subjects treated. Cause of death will be summarized as well in this table. In particular, frequencies of deaths due to study-treatment-related AEs will be reported.

6.3. Clinical Laboratory Tests

Laboratory data of hematology, coagulation, and serum chemistry up to 30 days after last dose or the End-of-Treatment visit date, whichever is later, will be reported in SI units. Coagulation parameters are activated partial thromboplastin time and international normalized ratio/prothrombin time.

Summary statistics (mean, standard deviation, median, and range) will be calculated for all numerical laboratory results in SI units, including change from baseline at each timepoint of assessment and change from baseline to the last value.

Graphical displays of over-time summaries will be presented for the following key laboratory parameters: ANC, ALC, hemoglobin, white blood cell count, platelet count, total bilirubin, creatinine, creatinine clearance, alkaline phosphatase, and electrolytes (sodium, potassium).

Shift tables from baseline to worst value on study (from treatment start to 30 days after the last dose or the End-of-Treatment visit date, whichever is later) will be produced for selected laboratory parameters, to include hemoglobin, white blood cell count, platelet count, ANC, ALC, AST, ALT, LDH, total bilirubin, creatinine, creatinine clearance, alkaline phosphatase, albumin, magnesium, and electrolytes (sodium, potassium). The worst toxicity grade during treatment will be tabulated.

6.3.1. Creatinine Clearance

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

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

For males, the factor is 1 instead of 0.85.

6.3.2. Analysis of Lymphocytosis

A descriptive summary of ALCs with and without lymphocytosis, time to lymphocytosis, and duration of lymphocytosis will be provided by treatment. Lymphocytosis is defined as ALCs increasing ≥50% from baseline and achieving level $\geq 5x10^9$/L. For subjects with lymphocytosis, resolution of lymphocytosis is defined as 1) a decrease of ALC value to the baseline level or lower, or 2) or an achievement of ALC value that is $<5x10^9$/L, whichever occurs first. Time to lymphocytosis is defined as the interval between the date of first dose and the first date that subject had lymphocytosis, presented in weeks. Duration of lymphocytosis will be derived from the first date that a subject had lymphocytosis until the event recovered for the first time, presented in weeks. Subjects who have not recovered will be censored at the last laboratory assessment.

6.4. Electrocardiogram

Abnormal electrocardiogram findings will be summarized.

6.5. Vital Signs and Physical Examination Findings

Abnormal physical examination findings will be tabulated by body system.

6.6. Other Safety Parameters

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

7. PATIENT-REPORTED OUTCOMES

Two PRO instruments, the FACT-Lym and EQ-5D-5L will be administered in this study. The following section describes PRO analyses that assess subject-perceived disease burden and symptoms. These will be assessed throughout the study. Patient-reported outcome assessments using the FACT-Lym and 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.

FACT-Lym consists of the Functional Assessment of Chronic Illness Therapy-General (FACT-G) and a lymphoma-specific additional concerns subscale (Lym) (Protocol Attachment 6). Responses to all items are rated on a 5-point scale ranging from 0 "not at all" to 4 "very much". The FACT-G consists of three 7-item subscales scored 0 to 28 (physical well-being, social well-being, and functional well-being) plus one 6-item subscale (emotional

well-being) scored 0 to 24. The recall period is the past 7 days. The lymphoma scale includes 15 items and scores range from 0 to 60. Two summary scores may also be calculated: the FACT-Lym total score (FACT-G plus Lym) and the FACT-Lym trial outcome index score (physical well-being + functional well-being + lymphoma). Higher scores represent better functional status and well-being for all subscales and summary scales. The subscale of most interest in this study will be the Lym subscale.

The EQ-5D-5L is a standardized instrument for use as a measure of health outcome. For the purpose of this study, the EQ-5D-5L will be used to generate weighted utility scores for use in cost effective analyses. The EQ-5D-5L is a 5-item questionnaire and a visual analogue scale ranging from 0 (worst imaginable health state) to 100 (best imaginable health state) (Protocol Attachment 7). The scores for the 5 separate questionnaires are categorical and should not be analyzed as cardinal numbers. However, the scores for the 5 dimensions are used to compute a single utility score ranging from zero (0.0) to 1 (1.0) representing the general health status of the individual. The United Kingdom weights will be used to generate patient utilities from the 5 dimensions of the EQ-5D-5L in this study. Data will be collected using electronic capture to enhance ease of collection, while flagging questions that may be missed accidently by the study subject.

PROs will be collected before any other visit procedure according to the protocol Time and Events schedule. The FACT-Lym will be performed until disease progression, death, or the clinical cutoff, whichever comes first. The EQ-5D-5L will be performed until death or study end. Following disease progression, sites should attempt to administer the EQ-5D-5L every 16 weeks for 2 times, unless death or study end occurs first. Specifically, the FACT-Lym and EQ-5D-5L questionnaires will be administered on Day 1 of every cycle during the Active Treatment Phase and at the End-of-Treatment visit. During the pre-PD follow-up period, they will be administered every 16 weeks during the first 2 years, and, after the first 2 years, every 24 weeks, until PD, clinical cutoff, or 5 years. During the Post-PD follow-up period, the EQ-5D-5L will be performed every 16 weeks until death, study end, or 32 weeks, whichever occurs first. 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. For each of the PRO scales, descriptive statistics (number of observations, mean, standard deviation, median, minimum, maximum) of scores at baseline and postbaseline assessments will be reported by treatment group. Specifically, for both instruments, descriptive statistics will be calculated for the raw data and the change from baseline at each timepoint of assessment as well as for the changes from previous assessment. Time to worsening analysis is specified in the secondary endpoints analysis Section 5.3.3.1. Time to worsening will be analyzed using the log-rank test.

For the FACT-Lym, the primary interest is the Lymphoma subscale, however other FACT subscales will also be examined, including:

- The 4 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: Physical and Functional domains of the FACT-G items plus the Lym Subscale score)
- The FACT-Lym total score (FACT-G plus FACT-Lym)

In the event of missing item(s) in a FACT-Lym assessment, the subscale scores will be computed according 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 lymphoma subscale of the FACT-Lym is defined as the interval from the date of randomization to the start date of the worsening and improvement, respectively. Worsening and improvement is defined by 2 criteria; 3-point or greater and 5-point or greater reduction and increase from baseline in the Lym lymphoma subscale (Carter 2008; Cella 2005). Death and missing data due to very ill as noted on the CRF will also be considered as worsening. Subjects who did not have a baseline and/or postbaseline assessment, or who did not meet the worsening/improvement criterion prior to PD are censored at the clinical cutoff date. Other subjects without worsening/improvement are censored at the last known date without events. Time to worsening and time to improvement in the lymphoma subscale will be estimated using Kaplan-Meier methods. The HR for ibrutinib relative to placebo and its associated 95% CI will be calculated based on the stratified Cox proportional hazards model by the stratification factors at randomization. The best and worst change of the lymphoma subscale, Trial Outcome Index, and total score from baseline to postbaseline may be explored using a waterfall plot. Exploratory analyses may be conducted to examine time to worsening or improvement according to clinical response status (ie, responder vs. non-responder) or by subject baseline characteristics of interest (ie, ECOG performance status).

For the FACT-Lym and EQ-5D-5L scores, a mixed effects model with repeated measures analysis will be conducted to estimate change from baseline at each timepoint between the 2 treatments. Subjects in the ITT population who have a baseline value and at least 1 post-randomization value will be 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 in weeks as a continuous variable, treatment-by-time interaction, and stratification factors as fixed effects.

REFERENCES

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 Am Soc Hematology; 7 December 2008; San Francisco, California.

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]. Am Soc Hematology. 2005;106:Abstract 750.

Cheson BD, Pfistner B, Juweid ME, et al. Revised response criteria for malignant lymphoma. J Clin Oncol. 2007;25(5):579-586. Epub 2007 Jan 22.

Dodd LE, Korn EL, Freidlin B, Gray R, Bhattacharya S. An audit strategy for progression free survival. Biometrics. 2011;67:1092-1099.

Fisher RI, Gaynor ER, Dahlberg S, et al. Comparison of a standard regimen (CHOP) with three intensive chemotherapy regimens for advanced non-Hodgkin's lymphoma. N Engl J Med. 1993;328(14):1002-1006.

Jennison C, Turnbull B. W. Group sequential methods with applications to clinical trials. Chapman & Hall/CRC.

Lenz G, Wright G, Dave SS, et al. Stromal Gene Signatures in Large-B-Cell Lymphomas. N Engl J Med. 2008;359:2313-2123.

Rényi, Alfréd. On the theory of order statistics. Acta Mathematica Hungarica. 1953;4:91-231.

Song Y, Chi GYH. A method for testing a prespecified subgroup in clinical trials. Statist Med. 2007;26:3535-3549.

Uno, H, Claggett, B, Tian, L, et al. Moving beyond the hazard ratio in quantifying the between-group difference in survival analysis. J Clin Oncol. 2014;32(22):2380-2385.