16.1.9 DOCUMENTATION OF STATISTICAL METHODS AND INTERIM ANALYSIS PLAN

STATISTICAL ANALYSIS PLAN

Study Protocol Number:	BGB-3111-212
Study Protocol Title:	An International, Phase 2, Open-Label, Randomized Study of BGB-3111 Combined with Obinutuzumab Compared with Obinutuzumab Monotherapy in Relapsed/Refractory Follicular Lymphoma
Date:	19 July, 2022
Version:	2.0

SIGNATURE PAGE

Author:

Signature
Date: _____

Approval:

Signature
Date: _____

Signature
Date: _____

Signature
Date: _____

TABLE OF CONTENTS

LIST OF ABBREVIATIONS AND DEFINITIONS OF TERMS

Abbreviation	Term
AEs	Adverse events
ADI	Actual dose intensity
AUC	Area under the plasma concentration-time curve
BID	Bis in die (twice a day)
BMI	Body mass index
BTK	Bruton tyrosine kinase
CI	Confidence interval
C_{max}	Maximum observed plasma concentration
CR	Complete response
CT	Computed tomography
DIPP	Data integrity protection plan
DOR	Duration of response
eCRF	Electronic case report form
EAIR	Exposure-Adjusted Incidence Rate
ECOG	Eastern Cooperative Oncology Group

EDC	Electronic data capture
FDA	Food and Drug Administration
FDG	[18F]fluorodeoxyglucose
FL	Follicular lymphoma
GHS/QoL	Global health status/Quality of life
ICR	Independent central review
ITT	Intent-to-Treat
MedDRA	Medical Dictionary for Regulatory Activities
NCI-CTCAE	National Cancer Institute Common Toxicity Criteria for Adverse Events
NHL	Non-Hodgkin's lymphoma
ORR	Overall response rate
OS	Overall survival
PD	Progressive disease
PFS	Progression-free survival
PK	Pharmacokinetic
PR	Partial response

PT	Preferred term
QD	Quaque die (once a day)
RDI	Relative dose intensity
SAEs	Serious adverse events
SMQ	Standardized MedDRA Query
sNDA	Supplemental new drug application
SOC	System organ class
TEAE	Treatment-emergent adverse event
WHO DD	World Health Organization Drug Dictionary

1 INTRODUCTION

The purpose of this statistical analysis plan (SAP) is to describe the procedures and the statistical methods that will be used to analyze and report the results for BGB-3111-212: An International, Phase 2, Open-Label, Randomized Study of BGB-3111 Combined with Obinutuzumab Compared with Obinutuzumab Monotherapy in Relapsed/Refractory Follicular Lymphoma.

The focus of this SAP is the planned Primary Analysis specified in this study protocol.

Details of the pharmacokinetic/pharmacogenomics analyses are not described in this SAP.

The SAP Version 2.0 is an amendment to the original SAP Version 1.0 (dated Nov 12, 2021) to clarify the addition of an Updated Analysis with a data cutoff at 12 months after randomization of the last patient as requested by the US Food and Drug Administration (FDA).

2 STUDY OVERVIEW

This is an international, phase 2, open-label, randomized, active-control study of BGB-3111 (zanubrutinib) plus obinutuzumab versus obinutuzumab monotherapy in 210 patients with relapsed or refractory follicular lymphoma. The primary efficacy endpoint is overall response rate (ORR) determined by independent central review (ICR). Disease response will be assessed per the Lugano Classification for Non-Hodgkin Lymphoma (NHL) (Cheson et al, 2014) - hereafter referred to as Lugano Classification for NHL.

Central randomization (2:1) will be used to assign patients to one of the following study drug treatments:

Arm A: zanubrutinib plus obinutuzumab

Arm B: obinutuzumab monotherapy

Randomization will be stratified by the number of prior lines of therapy (2 to 3 vs > 3), rituximab-refractory status (yes vs no), and geographic region (China vs ex-China). Treatment with zanubrutinib plus obinutuzumab and treatment with obinutuzumab monotherapy will be open label. Study treatment must commence within 5 days after randomization.

Each cycle consists of 28 days. Study drug treatments will be administered as follows, depending on cohort and treatment assignment:

- Zanubrutinib will be administered as two 80 mg capsules by mouth twice a day (160 mg twice a day) with or without food.
- Obinutuzumab will be administered as follows: 1,000 mg intravenously on Days 1, 8, and 15 of Cycle 1, then 1,000 mg on Day 1 of Cycles 2 to 6, then 1,000 mg every 8 weeks (at the discretion of the investigator, obinutuzumab may be administered as 100 mg on Day 1 and 900 mg on Day 2 of Cycle 1 instead of 1,000 mg on Day 1 of Cycle 1). Responding patients may continue to receive maintenance obinutuzumab every 8 weeks for an additional 24 months (eg, maximum total duration of obinutuzumab of approximately 30 months [maximum 20 doses]).
- At the discretion of the investigator, patients in Arm B will be eligible to receive crossover treatment with zanubrutinib plus obinutuzumab if they experience progressive disease (PD) or their disease does not respond to therapy with a complete response (CR) or partial response (PR) after 12 months. This must be confirmed by ICR. For patients who initiate crossover treatment with zanubrutinib plus obinutuzumab, safety, laboratory, and response evaluation assessments will continue to be performed per the Schedule of Assessments (Appendix 9 in the protocol).

Tumor assessments, including imaging studies, will be performed at screening, every 12 weeks from Cycle 1 Day 1 for 24 months, then every 24 weeks for 24 months, and then yearly until disease progression. All known sites of disease must be documented at screening and reassessed at each subsequent tumor evaluation. All patients must undergo PET-CT scan during screening. Patients whose disease is not [18F]fluorodeoxyglucose (FDG)-avid at screening will be followed by CT-based assessments alone. Patients whose disease is FDG-avid at screening, by either site assessment or the central imaging vendor, will be followed by an integration of PET-CT and CT-based assessments as follows:

- PET-CT scans are required at screening, end of Cycles 3, 6, and 12, and to confirm a result on CT scan (CR/PR or disease progression)
- CT scans with contrast are required at all other tumor response assessments

Patients receiving zanubrutinib should remain on study treatment until disease progression is confirmed by ICR. Patients receiving obinutuzumab should remain on study treatment until either

disease progression is confirmed by ICR or approximately 30 months of treatment with obinutuzumab (maximum 20 doses), whichever occurs first.

Assessments of safety will include AEs, SAEs, clinical laboratory tests, physical examinations, and vital signs. AEs will be graded for severity per National Cancer Institute Common Terminology Criteria for Adverse Events (NCI-CTCAE) version 4.03. An independent data monitoring committee will periodically monitor safety data.

3 STUDY OBJECTIVES

All primary and secondary objectives will compare zanubrutinib plus obinutuzumab versus obinutuzumab monotherapy.

3.1 PRIMARY OBJECTIVES:

- To evaluate efficacy, as measured by ORR determined by ICR

3.2 SECONDARY OBJECTIVES:

- To evaluate efficacy, as measured by the following:
    - ORR determined by investigator assessment
    - Duration of response (DOR) determined by ICR and by investigator assessment
    - Progression-free survival (PFS) determined by ICR and by investigator assessment
    - Overall survival (OS)
    - Rate of CR or complete metabolic response determined by ICR and by investigator assessment
    - Time to response determined by ICR and by investigator assessment
    - Patient-reported outcomes
- Safety and tolerability
- Pharmacokinetics (zanubrutinib plus obinutuzumab arm only)

3.3 EXPLORATORY OBJECTIVES:

- ORR in obinutuzumab arm after crossover to receive zanubrutinib plus obinutuzumab

4 STUDY ENDPOINTS

4.1 PRIMARY ENDPOINTS

- The primary endpoint is ORR determined by ICR using Lugano Classification for NHL. The ORR is defined as the proportion of patients who achieve either CR or PR as best overall response, defined as best response achieved during the entire follow-up period, until the data cutoff date, the start of a new anticancer therapy, or the crossover date for patients in Arm B who cross over to Arm A.

4.2 SECONDARY ENDPOINTS

- ORR determined by investigator assessment, defined similarly as ORR determined by ICR.
- DOR determined by ICR and by investigator assessment, defined as the time from the date that response criteria are first met to the date of first documentation of disease progression or death, whichever occurs first. For patients in Arm B who cross over to Arm A, the disease assessment after the crossover will not be included in the derivation.
- PFS determined by ICR and by investigator assessment, defined as the time from randomization to the date of first documentation of disease progression or death, whichever occurs first. For patients in Arm B who cross over to Arm A, the disease assessment after the crossover will not be included in the derivation.
- OS defined as the time from randomization to the date of death due to any reason.
- Rate of CR or complete metabolic rate determined by ICR and by investigator assessment, defined as the proportion of patients who achieve CR or complete metabolic rate as best overall response from randomization until the data cutoff date, the start of a new anticancer therapy, or the crossover date for patients in Arm B who cross over to Arm A.
- Time-to-response determined by ICR and by investigator assessment, defined as the time

from randomization to the time the response criteria are first met. For patients in Arm B who cross over to Arm A, the disease assessment after the crossover will not be included in the derivation.

- Patient-reported outcomes measured by EORTC QLQ-C30 and EQ-5D-5L questionnaires.
- Safety parameters, including AEs, SAEs, AEs of special interest, clinical laboratory tests, physical exams and vital signs.
- Pharmacokinetic (PK) parameters such as apparent clearance of the drug from plasma (CL/F) and AUC_{0-12}

4.3 EXPLORATORY ENDPOINT

- ORR in Arm B after crossover to Arm A by investigator assessment, defined as the proportion of patients who achieve a CR or PR from the crossover date until the data cutoff date, or the start of a new anticancer therapy.

5 SAMPLE SIZE CONSIDERATIONS

The sample size calculation is based on the comparison of the primary endpoint of ORR in the ITT Analysis Set. Assuming $ORR_A = 0.55$ and $ORR_B = 0.30$, 210 patients will be enrolled in a 2:1 ratio (140 patients in Arm A and 70 patients in Arm B) to provide a power of approximately 91% in testing ORR_A versus ORR_B using a normal approximation to binomial distribution with a 2-sided significance level of 0.05 with continuity correction. The sample size calculation is based on methods in Hulley et al, 2013 and Fleiss et al, 1980.

6 STATISTICAL METHODS

6.1 ANALYSIS SETS

Intent-to-Treat (ITT) Analysis Set

The ITT Analysis Set includes all patients randomized to a treatment arm. Patients will be grouped by the assigned treatment at randomization. The ITT Analysis Set will be used for efficacy analyses unless otherwise specified.

Safety Analysis Set

The Safety Analysis Set includes all patients in the ITT Analysis Set who received any dose of any study drug. Patients will be grouped by actual treatment received. The Safety Analysis Set will be used for all safety analyses unless otherwise specified.

Per-Protocol Analysis Set

The Per-Protocol Analysis Set includes all patients in the ITT Analysis Set who received any dose of any study drug and had no critical protocol deviations. The categories of critical protocol deviations are defined in Section 6.3.2.

PK Analysis Set

The PK Analysis Set includes zanubrutinib-treated patients in Arm A who have at least one post-baseline PK concentration measurement. The PK Analysis Set will be used for PK analyses.

6.2 DATA ANALYSIS GENERAL CONSIDERATIONS

6.2.1 Definitions and Computations

Study Treatment (study drug): Study drugs include zanubrutinib and/or obinutuzumab.

Study Day: Study Day will be calculated in reference to the date of first dose of either study drug (Study Day 1). For patients not dosed, randomization date will be used instead of the first dose date. For assessments conducted on or after Study Day 1, Study Day will be calculated as (assessment date – Study Day 1 + 1). For assessments conducted before Study Day 1, Study Day will be calculated as (assessment date – Study Day 1). There is no Study Day 0.

In the situation where the event date is partial or missing, the date will appear partial or missing in the listings. Study Day and any corresponding durations will be presented based on the imputations specified in the Appendix A.

Treatment duration: Treatment duration will be calculated as (date of last dose of study treatment – Study Day 1 + 1).

Baseline: Unless otherwise specified, a baseline value is defined as the last non-missing value collected before Study Day 1 or before randomization for patients not dosed. Note that assessments that occur on the Study Day 1 but prior to the time of the first dose can qualify to be a baseline value.

Post-baseline: A post-baseline value or assessment is defined as a value or assessment after Study Day 1 or after randomization for patients not dosed.

All calculations and analyses will be conducted using SAS version 9.3 or higher.

6.2.2 Handling of Missing Data

Missing data will not be imputed unless otherwise specified elsewhere in the SAP. Missing dates or partially missing dates will be imputed conservatively for adverse events (AEs) and prior/concomitant medications/procedures. Specific rules for handling missing or partially missing dates for diagnosis, progression/relapse to prior therapy, AEs, and prior/concomitant medications/procedures are provided in the Appendix A.

When summarizing categorical variables, patients with missing data are generally included in the denominator to calculate percentages unless otherwise specified. When needed, the category of "Missing" is created and the number of patients with missing data is presented.

When summarizing continuous variables, patients with missing data are not included in the calculations unless otherwise specified.

By-visit endpoints will be analyzed using observed data unless otherwise specified. For observed data analyses, missing data will not be imputed, and only the observed records will be included.

6.2.3 Adjustment for Covariates

Stratified analysis will be performed to adjust for important baseline covariates for the primary and key secondary endpoints. Details of the stratified analyses are provided in Section 6.4.

6.2.4 Multiplicity Adjustment

Not applicable. The only hypothesis test to be performed in this study is for the primary endpoint with the two-sided significance level of 0.05. All other p-values will be presented for descriptive purposes only.

6.2.5 Data Integrity

Before pre-specified statistical analysis begins, the integrity of the data should be reviewed to assure fit-for-purpose. The data set for analysis should be an accurate and complete representation of the patients' relevant outcomes from the clinical database. All essential data should be complete

and reviewed up to a pre-specified cutoff date. Critical consistency checks and appropriate source data verification should be completed according to the final data extraction plan.

Though the study is open-label, access to study data is controlled for specific sponsor study team members overseeing the conduct of the study or analyzing study data. The sponsor will not have access to aggregated efficacy or safety data summaries by treatment arm according to the study's data integrity protection plan (DIPP).

6.3 PATIENT CHARACTERISTICS

6.3.1 Patient Disposition

The number (%) of patients screened, randomized, treated, discontinued from treatment with reasons, remained on treatment, completed treatment, discontinued from study with reasons, remained on study, completed study, and the duration of study follow-up will be summarized in the ITT Analysis Set.

Study follow-up time is defined as the time from randomization to the date of death or the end of study (whichever occurs first) for the patients who discontinued from study, or the data cutoff date for patients still on study. Study follow-up time will be summarized descriptively.

6.3.2 Protocol Deviations

Patients with important protocol deviations will be identified and documented. Important protocol deviations will be summarized by category for all patients in the ITT Analysis Set. COVID-19-related protocol deviations will be summarized. A listing of important protocol deviations will also be provided.

Critical protocol deviations will also be identified and used to define the Per-Protocol Analysis Set. Critical protocol deviation criteria will be established before the database lock. The critical protocol deviations will include but are not limited to selected protocol deviations from the following categories:

- Patient randomized without satisfying study eligibility criteria
- Patient developed study drug withdrawal criteria but was not withdrawn
- Patient received prohibited concomitant treatment
- Patient received the wrong study treatment

6.3.3 Randomization Stratification Factors

The number of patients with each of the randomization stratification factors by IRT and EDC will be summarized by treatment arm. The randomization stratification factors include number of prior lines of therapy (2 to 3 vs > 3), rituximab-refractory status (yes vs no), and geographic region (China vs ex-China). In addition, stratification consistencies between IRC and EDC will be summarized.

6.3.4 Demographic and Other Baseline Characteristics

Demographic and baseline characteristics will be summarized using descriptive statistics in the ITT Analysis Set, including the following variables:

- Age (years) and age group (< 60 vs ≥ 60, < 65 vs ≥ 65 and < 75 vs ≥ 75)
- Gender
- Race and ethnicity
- Geographic region
- Weight and body mass index (BMI, kg/m^2)
- Eastern Cooperative Oncology Group (ECOG) performance status (0, 1, 2)

A listing of demographic and baseline characteristics will be also provided.

6.3.5 Disease History

The number (%) of patients reporting a history of disease and characteristics, as recorded on the eCRF, will be summarized in the ITT Analysis Set. Disease characteristics include the following variables:

- Time since initial diagnosis to first dose (months)
- Follicular Lymphoma International Prognostic Index (FLIPI) score at initial diagnosis and screening (low [0-1], intermediate [2], high [≥3])
- Ann Arbor stage at initial diagnosis and screening (I-II, III-IV)
- Age at initial diagnosis and screening (years) (< 60, ≥ 60)

- LDH at initial diagnosis and screening (≤ ULN, > ULN)
- Hemoglobin level at initial diagnosis and screening (< 120 g/L, ≥ 120 g/L)
- Number of nodal sites at initial diagnosis and screening (≤ 4, > 4)
- Bone marrow involvement at initial diagnosis and screening (absent, present, missing)
- Longest diameter of the largest involved node at initial diagnosis and screening (< 6 cm, ≥ 6 cm, unknown)
- Beta-2 microglobulin at initial diagnosis and screening (normal, raised, unknown, missing)
- Baseline bulky disease (≥ 5 cm, ≥ 7 cm, ≥ 10 cm)
- B symptoms (yes, no)
    - Unexplained fever of ≥ 101°F (yes, no)
    - Drenching night sweats (yes, no)
    - Unexplained > 10% weight loss in previous 6 months (yes, no)
- Baseline platelet count (10^9/L) and category (≤ 100 x10^9/L, > 100 x10^9/L)
- Baseline absolute neutrophil count (10^9/L) and category (≤ 1.5 x10^9/L, > 1.5 x10^9/L)

6.3.6 Medical History

Medical history will be coded using Medical Dictionary for Regulatory Activities (MedDRA) (version 24.0). The number (%) of patients reporting a history of any medical condition, as recorded on the eCRF, will be summarized by system organ class (SOC) and preferred term (PT).

6.3.7 Prior Anticancer Therapies and Surgeries

The following information related to prior therapy for lymphoma will be summarized by Anatomical Therapeutic Chemical (ATC) medication class Level 2 and World Health Organization Drug Dictionary (WHO DD) drug codes (version March 2021 or later) preferred name:

- Number of regimens of prior anticancer therapies

- Time from the end of the last line of prior anticancer therapy to first dose of study drug (months)
- Duration of last therapy (months)
- Best response to the last line of prior anticancer therapy (CR, PR, stable disease, PD, unknown)
- Rituximab-refractory status (yes, no, unknown)
- Refractory to the most recent line of therapy (yes, no, unknown)
- Progression-free and treatment-free for ≥ 12 months since last rituximab-containing regimen
- Progression of disease within 24 months of completion of the first line of therapy
- Progression of disease within 12 months of completion of the most recent line of therapy
- Progression of disease within 12 months of completion of the most recent line of therapy or refractory disease
- Prior anticancer radiotherapy
- Prior anticancer surgeries
- Prior stem cell transplantations

Refractoriness to the most recent line of therapy is defined as less than a PR. Rituximab-refractory status will be summarized based on the enrollment page in eCRF.

6.3.8 Prior and Concomitant Medications

Prior and concomitant medications will be coded using the WHO DD drug codes (version March 2021 or later) and will be further classified to the appropriate Anatomical Therapeutic Chemical (ATC) code.

Prior medications are defined as medications that started before the first dose date of study drug. Concomitant medications will be defined as medications that (1) started before the first dose of study treatment and were continuing at the time of the first dose of study treatment, or (2) started on or after the date of the first dose of study treatment up to 30 days after the last dose of

zanubrutinib or 90 days after the last dose of obinutuzumab or initiation of a new anticancer therapy. For the purpose of determining if a medication should be noted as a concomitant medication, the imputation rules stated in **Error! Reference source not found.** will be used.

The number (%) of patients reporting prior medications and concomitant medications will be summarized by ATC medication class Level 2 and WHO DD preferred name.

6.4 EFFICACY ANALYSIS

Statistical testing will be performed to compare the efficacy of zanubrutinib plus obinutuzumab (Arm A) and obinutuzumab (Arm B). All efficacy analyses will be performed using the ITT Analysis Set. For patients in Arm B who crossed over to Arm A, data before the crossover date will be analyzed for primary and secondary efficacy endpoints unless otherwise specified.

Stratified analyses will be based on the randomization stratification factors per IRT unless otherwise specified.

6.4.1 Primary Efficacy Endpoint

The null and alternative hypotheses for comparing ORR by ICR are set as follows:

H_0: $ORR_A = ORR_B$

H_a: $ORR_A > ORR_B$

where ORR_A is the ORR in Arm A (zanubrutinib plus obinutuzumab) and ORR_B is the ORR in Arm B (obinutuzumab monotherapy). The null hypothesis will be tested using the Cochran-Mantel-Haenszel method stratified by the randomization factors (number of prior lines of therapy [2 to 3 vs > 3], rituximab-refractory status [yes vs no], and geographic region [China vs ex-China]) per IRT, at the two-sided significance level of 0.05. If the null hypothesis can be rejected, it will be concluded that the superiority of Arm A over Arm B in ORR is demonstrated at the significance level. If the number of patients in a particular stratum is too small, this stratum may be combined with other strata for analysis.

The 95% confidence interval (CI) for the Mantel-Haenszel common risk difference (Mantel-Haenszel, 1959) will be constructed using a normal approximation and Sato's standard error (Sato 1989) stratified by the randomization factors (number of prior lines of therapy [2 to 3 vs > 3], rituximab-refractory status [yes vs no]), and geographic region [China vs ex-China]) per IRT.

A Clopper-Pearson 95% CI of ORR will be constructed for each arm.

Disease control rate by ICR, defined as the proportion of patients who achieved CR, PR, or stable disease, will be also analyzed using the same methods as in the analysis of ORR.

The data cutoff for the Primary Analysis will be 3 months after randomization of the last patient.

6.4.2 Secondary Efficacy Endpoints

6.4.2.1 ORR by Investigator Assessment

ORR per investigator assessment will be summarized and analyzed using the same methods used for ORR per ICR.

6.4.2.2 Progression-Free Survival

Progression-free survival (PFS) by ICR is defined as the time (in months) from the randomization date to the first documented disease progression or death due to any cause, whichever occurs first. The disease assessments after crossover for patients in Arm B who cross over to Arm A will not be included in the PFS analysis.

PFS will be right-censored for patients who met one of the following conditions: 1) alive without baseline or post-baseline tumor assessment; 2) start a new anticancer treatment before documented PD or death; 3) PD or death after missing ≥ 2 consecutive planned disease assessments; 4) alive without documented PD; 5) lost to follow-up without documented PD or death; 6) crossover from Arm B to Arm A. For such patients, the primary analysis of PFS will be right-censored according to the convention described in Table 1. The censoring rule is based on the FDA Guidance for Industry, 'Clinical Trial Endpoints for the Approval of Non-Small Cell Lung Cancer Drugs and Biologics' (2015, Table C1, https://www.fda.gov/regulatory-information/search-fda-guidance-documents/clinical-trial-endpoints-approval-non-small-cell-lung-cancer-drugs-and-biologics).

Table 1: Date of Progression or Censoring for Progression-free Survival

Situation	Date of Progression Event	Outcome
Death or PD between 2 planned disease assessments	Date of death or PD, whichever occurs first	Event
Death before the first disease assessment	Date of death	Event

Alive without baseline or post-baseline disease assessment	Date of randomization	Censored
Start new anticancer treatment before documented PD or death	Date of last disease assessment prior to the date of new anticancer treatment	Censored
PD or death after missing ≥ 2 consecutive planned disease assessments	Date of last disease assessment before death or PD; Use date of randomization if no assessment was performed before death or PD.	Censored
Alive without documented PD	Date of last disease assessment	Censored
Lost to follow-up* without documented PD or death	Date of last disease assessment	Censored
Crossover from Arm B to Arm A	Date of last disease assessment prior to the first zanubrutinib treatment after crossover	Censored

*This includes the consent withdrawal.

The hazard ratio (Arm A/Arm B) for PFS and its 95% CI will be estimated from a stratified Cox regression model stratified by the randomization factors (number of prior lines of therapy [2 to 3 vs > 3], rituximab-refractory status [yes vs no], and geographic region [China vs ex-China]).

The distribution of PFS, including median and other quartiles and PFS rate at selected timepoints such as 12, 18 and 24 months, will be estimated using the Kaplan-Meier method. The 95% CI for median and other quartiles of PFS will be generated by using the Brookmeyer method (Brookmeyer and Crowley 1982), whereas the 95% CI for PFS event-free rate at landmark times will be generated by using the Greenwood formula (Greenwood 1926). Duration of follow-up for PFS will be estimated by the reverse Kaplan-Meier method (Schemper and Smith 1996).

Kaplan-Meier curves for PFS will be presented for each arm. A listing of the PFS-related information (eg, the date of progression or censoring and the corresponding reasons) will also be provided.

A sensitivity analysis of PFS by ICR will also be provided using the Per-Protocol Analysis Set. To account for the impact of COVID-19, a sensitivity analysis of PFS by ICR will be performed by censoring deaths due to COVID-19.

PFS per investigator assessment will be performed using the same methods used for PFS per ICR.

6.4.2.3 Duration of Response

DOR by ICR is defined as the time from the date of the earliest PR or CR to the date of PD or death for any cause, whichever occurs earlier. DOR will be analyzed using the same methods as in the analysis of PFS with the exception that treatment arm comparisons will not be performed.

A sensitivity analysis of DOR by ICR will also be analyzed in the Per-Protocol Analysis Set. To account for the impact of COVID-19, a sensitivity analysis of DOR by ICR will be performed by censoring deaths due to COVID-19.

DOR per investigator assessment will be performed using the same methods used for DOR per ICR.

6.4.2.4 Overall Survival

OS is defined as the time from randomization to the date of death due to any cause. All patients, including those who crossed over from Arm B to Arm A and remained alive before data cutoff or discontinued the study (discontinued the study due to reasons other than death), will be censored at the last known date the patient is alive on or prior to data cutoff. OS will be analyzed using the same methods as in the analysis of PFS.

A sensitivity analysis will be performed where crossover from Arm B to Arm A will be treated as a censoring event. To account for the impact of COVID-19, a sensitivity analysis of OS will be performed by censoring deaths due to COVID-19.

6.4.2.5 Complete Response Rate

CR rate by ICR is defined as the proportion of patients who achieve CR or complete metabolic response as best overall response from randomization until the data cutoff date, the start of a new anticancer therapy, or the crossover date for patients in Arm B who cross over to Arm A. The CR rate will be compared between treatment groups using Fisher's exact test. A Clopper-Pearson 95% CI of CR rate will be presented for each arm.

CR rate per investigator assessment will be performed using the same methods used for CR rate per ICR.

6.4.2.6 Time to Response

Time to response by ICR for responders is defined as the time from the randomization date to the date of the earliest qualifying response until the data cutoff date, the start of a new anticancer therapy, or the crossover date for patients in Arm B who cross over to Arm A. Time to response will be summarized by descriptive statistics for each arm.

Time to response per investigator assessment will be performed using the same methods used for time to response per ICR.

6.4.2.7 Patient-Reported Outcomes

The scoring of the EORTC QLQ-C30 and EQ-5D-5L will follow their corresponding manuals (Fayers et al. 2001; EuroQol Group 1990; Herdman et al 2011).

Compliance rates will be provided for both EORTC QLQ-C30 and EQ-5D-5L, which is defined as ratio of the number of patients that completed questionnaires to the number of patients that expected to complete the questionnaires per each visit and arm from the ITT Analysis Set.

EORTC QLQ-C30

Descriptive Analysis:

The scores of the EORTC QLQ-C30 questionnaire will be summarized at each assessment timepoint for each treatment arm. Summaries will include: scores and changes from baseline in the 1 global health status scale, 5 functional scales (physical, role, emotional, cognitive, and social), 3 symptom scales (fatigue, pain, and nausea/vomiting), and 6 single items (dyspnea, insomnia, appetite loss, constipation, diarrhea, and financial difficulties).

The key PRO end points are global health status/QoL (GHS/QoL), physical function, role function, fatigue, pain, nausea/vomiting and diarrhea. The key clinical cycles are cycle 3 (Week 12) and cycle 6 (Week 24). The two cycles are clinically justifiable as there will be enough patients in both cycles to assess the short-term and long-term effects of treatments for comparison purposes.

Mixed Model Analysis:

In addition, the key PRO end points (GHS/QoL, physical function, role function, fatigue, pain, nausea/vomiting and diarrhea) will be used to assess the clinically meaningful differences between treatment arms using a restricted maximum likelihood-based linear mixed model for repeated

measures (MMRM) to account for missing data under Missing at Random (MAR) assumption (Mallinckrodt et al, 2008). The model will include the repeated measurement of the key PRO end points scale score at baseline, Week 12 and Week 24 as the dependent variable, and treatment, time (as categorical variable), treatment by time interaction, and randomization stratification factors (number of prior lines of therapy [2 to 3 vs > 3] and rituximab-refractory status [yes vs no] only, no geographic region) as covariates. The random patient effects will include patient random intercept on QLQ-C30 GHS/QoL score which is assumed to follow a normal distribution. A point estimate of the treatment difference between treatment arms, 95% CI, and the corresponding p-value will be provided.

EQ-5D-5L

The EQ-5D-5L comprises a descriptive domain and a Visual Analogue Scale (VAS) with the following 5 dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Each dimension has 5 levels: no problems, slight problems, moderate problems, severe problems, and extreme problems. The EQ VAS records the respondent's self-rated health on a 0 to 100 scale, with 100 labelled 'the best health you can imagine' and 0 'the worst health you can imagine'.

EQ-5D-5L scores of the descriptive domain will be summarized descriptively by treatment arm at each assessment timepoint, including the number and percentage of patients reporting each level of problem on each dimension of the EQ-5D-5L. Change from baseline in scores of VAS will be summarized descriptively.

6.4.3 Sensitivity Analysis

For the primary endpoint of ORR, the following sensitivity analyses may be performed:

- ORR using the Per-Protocol Analysis Set
- ORR excluding patients who died due to COVID-19 to account for the impact of COVID-19
- ORR using logistic regression adjusting stratification factors by IRT, progression of disease within 12 months of completion of the most recent line of therapy (yes vs no), age at baseline, and FLIPI risk category [0-1 vs 2 vs > 3] at baseline

The sensitivity analyses for the secondary endpoints are addressed in Section 6.4.2.

6.4.4 Subgroup Analyses

The primary and selected secondary endpoints will be summarized, as appropriate, in the following subgroups (if collected in both IRT and in the EDC, data in the EDC will be used to define the subgroups):

- Age group (< 65 vs ≥ 65 years; < 75 vs ≥ 75 years)
- Sex (male vs female)
- Geographic region (China vs ex-China)
- Prior lines of therapy (2-3 vs > 3)
- Baseline ECOG performance status (0 vs ≥ 1)
- Bulky disease (yes: any target lesion longest diameter ≥ 5 cm vs no)
- Bulky disease (yes: any target lesion longest diameter ≥ 7 cm vs no)
- Bulky disease (yes: any target lesion longest diameter ≥ 10 cm vs no)
- FLIPI risk category (low [0-1], intermediate [2], high [≥3])
- Rituximab-refractory status (refractory vs not refractory)
- Refractory status to the most recent line of therapy (refractory, not refractory, or unknown)
- Progression of disease within 24 months of starting the first line of therapy (yes, no, or unknown)
- Progression of disease within 24 months of starting the first line of chemoimmunotherapy (yes, no, or unknown)
- Progression of disease within 6 months of completion of the most recent line of therapy (yes, no, or unknown)
- Progression of disease within 12 months of completion of the most recent line of therapy (yes, no, or unknown)

The subgroup variables and the cutoff values are subject to change if warranted to better represent the data. Some subgroup analyses may not be done if majority patients are in one category.

6.4.5 Exploratory Efficacy Endpoints

For patients in Arm B cross over to Arm A, ORR by investigator assessment after crossover will be summarized. The ORR is defined as the proportion of patients who achieve either CR or PR from the crossover date until the data cutoff date or the start of a new anticancer therapy.

6.5 SAFETY ANALYSES

All safety analyses will be performed by treatment arms based on the Safety Analysis Set. Safety and tolerability will be assessed, where applicable, by incidence, severity, and change from baseline values for all relevant parameters including AEs, laboratory values, vital signs, and physical examination. For patients in Arm B who cross over to Arm A, data before the crossover date will be included for the safety analyses unless otherwise specified.

6.5.1 Extent of Exposure

The following measures of the extent of study drug exposure will be summarized:

- Duration of study treatment exposure (in months)
- Number (%) of treatment cycles for zanubrutinib and obinutuzumab respectively
- Total dose (mg) received for zanubrutinib and obinutuzumab, respectively
- Actual dose intensity (ADI, mg/day) and relative dose intensity (RDI, %) of zanubrutinib
- Number of infusions and relative infusion frequency (%) of obinutuzumab
- Number (%) of patients with dose reduced, interrupted, missed along with the corresponding reasons for zanubrutinib
- Number (%) of patients with dose interrupted along with the corresponding reasons for obinutuzumab

Duration of exposure in months is defined as (the last zanubrutinib/ obinutuzumab/ combination administration date [data cutoff date for treatment ongoing patients] – the first zanubrutinib/ obinutuzumab/ combination administration date + 1)/30.4375.

Number of treatment cycles is defined as duration of treatment (days) divided by 28.

The ADI of zanubrutinib is defined as the cumulative dose (mg) of zanubrutinib divided by the duration of zanubrutinib (day), which is defined as the last zanubrutinib administration date – the first zanubrutinib administration date + 1.

The RDI of zanubrutinib is defined as the ratio of ADI (mg/day) of zanubrutinib and the planned dose intensity (PDI, mg/day) of zanubrutinib, which is 320 mg/day.

The relative infusion frequency of obinutuzumab is defined as the ratio of the actual number of infusions administered and the planned number of infusions (1,000 mg on Day 1, 8, and 15 of Cycle 1, then 1,000 mg on Day 1 of Cycles 2 to 6, then 1,000 every 8 weeks). Infusions on Day 1 and 2 of Cycle 1 are counted as one infusion.

6.5.2 Adverse Events

AEs will be graded by the investigators using CTCAE v4.03. The AE verbatim descriptions (investigator terms from the eCRF) will be classified into standardized medical terminology using the Medical Dictionary for Regulatory Activities (MedDRA). AEs will be coded to the MedDRA (version 24.0) lower level term closest to the verbatim term. The linked MedDRA PT and primary SOC are also captured in the database.

A treatment-emergent adverse event (TEAE) is defined as an AE that had an onset date or increase in severity level on or after the first dose of study drug through 30 days after the last dose of zanubrutinib or 90 days after the last dose of obinutuzumab (permanent discontinuation of study drug) or prior to the initiation of new anticancer therapy, whichever is sooner. Worsening of an AE to Grade 5 after 30/90 days from the last dose of zanubrutinib/obinutuzumab (prior to the initiation of new anticancer therapy) will also be considered as TEAE. Summary tables will be based on TEAEs.

An AE overview table, including the number of patients with TEAEs, treatment-emergent serious AEs (SAEs), TEAEs with Grade 3 or above, TEAEs that led to death, TEAEs that led to treatment discontinuation, TEAEs that led to dose modification (reduction, interruption, delay), treatment-related TEAEs, TEAEs of special interest, Grade 3 or above TE AESIs, and serious TE AESIs will be provided.

Treatment-related AEs include those events considered by the investigator to be related, possibly or probably related to study drug or with a missing assessment of the causal relationship.

The incidence of TEAEs will be reported as the number (%) of patients with TEAEs by SOC, PT and the worst grade. A patient will be counted only once by the highest severity grade within an SOC and PT, even if the patient experienced more than 1 TEAE within a specific SOC and PT.

Summaries of TEAEs by SOC and PT will be presented for the following categories:

- Any TEAE
- Any Grade 3 or higher TEAE
- Any treatment-related TEAE
- Any treatment-related Grade 3 or higher TEAE
- Any serious TEAE
- Any treatment-related serious TEAE
- Any TEAE leading to dose reduction
- Any TEAE leading to dose interruption
- Any TEAE leading to dose delay (obinutuzumab only)
- Any TEAE leading to dose discontinuation
- Any TEAE leading to death
- Any treatment-related TEAE leading to death

Moreover, all the summaries above will also be provided by PT only and by SOC, PT and maximum severity.

Incidence of TEAEs of special interest by category and PT will be presented for the following categories:

- Any TEAE of special interest
- Any Grade 3 or higher TEAE of special interest
- Any treatment-related TEAE of special interest
- Any treatment-related Grade 3 or higher TEAE of special interest

- Any serious TEAE of special interest
- Any treatment-related serious TEAE of special interest
- Any TEAE of special interest leading to dose reduction
- Any TEAE of special interest leading to dose interruption
- Any TEAE of special interest leading to dose discontinuation

The categories and detailed search criteria for TEAEs of special interest are described in Appendix B.

Time to the first TEAE of special interest along with the cumulative event rate and monthly hazard rate will be provided.

Given the imbalance in treatment duration between the two arms, an exposure-adjusted incidence rate (EAIR) for the TEAE of special interest will be provided. EAIR is calculated as the number of patients experiencing the TEAE of special interest divided by the total time from Study Day 1 to the first event date or the last dose date if the first event occurred after the last dose date or if there is no event.

An overall summary of death and cause of death will be presented for the following categories:

- Total deaths
- Deaths within 30/90 days of last dose of study drug
- Deaths more than 30/90 days of last dose of study drug

Listings of all AEs, SAEs, AEs leading to dose reduction, AEs leading to dose interruption, AEs leading to dose delay, AEs leading to dose discontinuation, and all deaths will be provided.

6.5.3 Laboratory Values

All hematology, serum chemistry, and coagulation results for each patient will be presented in data listings. The baseline value, post-baseline value and change from baseline for all hematology and serum chemistry parameters will be summarized at each scheduled visit.

The laboratory parameters of interest for these summaries are:

Hematology	Serum Chemistry		Coagulation
Hemoglobin (decrease)	Alanine transaminase (ALT) (increase)	Albumin (decrease)	Activated partial thromboplastin time (aPTT) (increase)
Platelets (decrease)	Aspartate transaminase (AST) (increase)	Uric Acid (increase)	International Normalized Ratio (INR) (increase)
WBC (increase, decrease)	Alkaline Phosphatase (increase)	Sodium (increase, decrease)	
Absolute Neutrophil Count (ANC, decrease)	Total Bilirubin (increase)	Phosphorus (decrease)	
Absolute Lymphocyte Count (increase, decrease)	Creatinine (increase)	Potassium (increase, decrease)	
	Calcium (increase, decrease)	Magnesium (increase, decrease)	
	Glucose (increase, decrease)		

For hypocalcemia and hypercalcemia, serum calcium will be corrected using the formula:

> Corrected calcium = Serum calcium + 0.8 * (4 – serum albumin) where serum calcium is recorded in mg/dL and serum albumin is recorded in g/dL.

Descriptive summary statistics (n, mean, standard deviation, median, minimum, maximum for continuous variables; n [%] for categorical variables) for laboratory parameters and their changes from baseline will be summarized by visit.

Shift tables assessing the toxicity grade at baseline versus worst toxicity recorded post-baseline will be presented. A summary of the number (%) of patients with Grade 3 or higher toxicity will be provided for each laboratory parameter of interest. A listing of all Grade 3 or higher laboratory values will be provided. Box-whisker plots may be generated for parameters of interest.

Incidence of patients who met one or more of the Hy's law criteria will be summarized. A listing of patients that met one or more of the Hy's law criteria will be generated.

6.5.4 Vital Signs

Descriptive statistics for vital sign parameters (systolic and diastolic blood pressure, heart rate, temperature, and weight) and changes from baseline will be presented by visit and treatment group for all visits. A listing by patient and assessment timepoint will be generated.

6.5.5 Electrocardiograms (ECG)

The QTc-Fridericia (QTcF) interval values will be listed.

6.5.6 ECOG

ECOG performance status will be summarized at each visit. Shift tables assessing the ECOG performance status at baseline versus worst performance status post-baseline will be presented.

6.6 PK Analyses

Summary of plasma zanubrutinib concentrations at the scheduled time of collection (i.e. Cycle 1 Day 1 [2-hour postdose] and Cycle 2 Day 1 [predose and 2-hour postdose]) will be provided along with a listing of the PK sample collection times.

No other PK analyses are planned for the Primary Analysis and the Updated Analysis.

7 INTERIM ANALYSIS

No formal interim analyses are planned for this study.

8 CHANGES IN THE PLANNED ANALYSIS

This SAP provides more detail on the statistical methods than is provided in the protocol.

For the primary efficacy endpoint analysis, the null hypothesis was tested at the 2-sided significance level of 0.05, which gives the equivalent test result with 1-sided significance level of 0.025 as in the protocol.

An Updated Analysis with a data cutoff at 12 months after randomization of the last patient was added to provide additional follow-up for duration of response to satisfy a request communicated at the pre-sNDA meeting held with the US FDA on May 16, 2022.

9 REFERENCES

Brookmeyer R, Crowley J. A confidence interval for the median survival time. Biometrics. 1982;38:29-41.

Cheson, B.D., et al. Recommendations for Initial evaluation, staging, and response assessment of Hodgkin and non-Hodgkin lymphoma: the Lugano classification. Journal of Clinical Oncology, 2014; 32(27): 3059–3068.

EuroQol Group. EuroQol-a new facility for the measurement of health-related quality of life. Health Policy. 1990;16(3):199-208.

Fayers PM, Aaronson NK, Bjordal K, et al, on behalf of the EORTC Quality of Life Group. The EORTC QLQ-C30 Scoring Manual (3rd Edition). Published by: European Organisation for Research and Treatment of Cancer, Brussels. 2001.

Greenwood M. The Natural Duration of Cancer. Reports on Public health and medical patients. 1926;33:1-26.

Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res. 2011;20:1727-36.

Mallinckrodt CH, Lane PW, Schnell D, Peng Y, and Mancuso, JP. Recommendations for the primary analysis of continuous endpoints in longitudinal clinical trials. Drug Information Journal, 2008;42:303-319.

Mantel N, Haenszel W. Statistical Aspects of Analysis of Data from Retrospective Studies of Disease. Journal of the National Cancer Institute. 1959;22:719–748.

Sato T. On the Variance Estimator of the Mantel-Haenszel Risk Difference. Biometrics. 1989; 45:1323–1324.

APPENDIX A. IMPUTATION OF MISSING OR PARTIALLY MISSING DATES

In general, missing or partial dates will not be imputed. The following rules will apply for the specific analysis and summary purposes mentioned below only.

A.1 Prior/Concomitant Medications/Procedures

When the start date or end date of a medication is partially missing, the date will be imputed to determine whether the medication is prior or concomitant. The following rules will be applied to impute partial dates for medications:

If start date of a medication is missing, impute as follows:

- If both month and day are missing, then set to January 01
- If only day is missing, then set to the first of the month

If end date of a medication is missing, impute as follows:

- If both month and day are missing, then set to December 31
- If only day is missing, then set to last day of the month
- If start date or end date of a medication is completely missing, do not impute.

A.2 Adverse Events

The imputation rule for the safety analyses will be used to address the issues with partial dates. When the start date or end date of an AE is partially missing, the date will be imputed to determine whether the AE is treatment-emergent. When in doubt, the AE will be considered treatment-emergent by default. The following rules will be applied to impute partial dates for AEs:

If the start date of an AE is partially missing, impute as follows:

- If both month and day are missing, then the imputed day and month will be January 01 or the first dosing date if they have the same year, whichever is later.
- If only day is missing, then the imputed day will be the first day of the month or the first dosing date if they have the same month and year, whichever is later
- If the start date is completely missing, the imputed day will be the first dosing date as long as AE end date is not before the first dosing date.

If the end date of an AE is partially missing, impute as follows:

- If both month and day are missing, then set to December 31
- If only day is missing, then set to last day of the month
- If the end date is completely missing, do not impute.
- If the imputed AE end date > last known alive date or end of study date, then set to the last known alive date or end of study date, whichever occurs first.

A.3 Deaths

In case complete death dates are not recorded, impute as follows:

- If both month and day are missing, then the imputed month and day will be 01Jan or the last date of a patient known to be alive + 1, whichever is later.
- If only day is missing, the death will be assumed to be on the first day of the month or the last date of a patient known to be alive +1, whichever is later.

A.4 New Anticancer Therapy

If the start day of a new anticancer treatment, radiotherapy, or surgery is incomplete or missing, impute as follows:

- If only day is missing, then the imputed day will be the first day of the month; if the imputed anticancer treatment start date is prior to the study treatment end date, change the imputed anticancer treatment start date as 'study treatment end date'.
- No imputation will be performed for all other types of missing dates.

A.5 Date of Diagnosis or Date of Progression to Any Prior Therapy

If an initial diagnosis date or disease progression date to any prior therapy is missing, impute as follows:

- If both month and day are missing, then set to January 01.
- If only day is missing, then set to the first of the month.
- If a diagnosis date or progression date is completely missing, do not impute.

A.6 Prior Therapies, Radiotherapies, or Surgeries

If a prior therapy, radiotherapy, or surgery date is missing, impute as follows:

- If only day is missing, then set to the 15th of the month or the first study drug dose date - 1, whichever is earlier.
- No imputation will be performed for all other types of missing dates.

A.7 Date of Last Study Drug Administration

When the end date of a study drug administration is partially missing, the date will be imputed to calculate the extent of exposure and to define the duration of the TEAE. The following rules will be applied to impute the partial dates for the end date of a study drug administration:

- If both month and day are missing, then set to the start date of the corresponding cycle.
- If only day is missing, then the imputed day will be the first day of the month or the start date of the corresponding cycle if they have the same month and year, whichever is later.

APPENDIX B. ADVERSE EVENTS OF SPECIAL INTEREST CATEGORIES AND SEARCH CRITERIA

AESI Category	Search Criteria
Hemorrhage	Haemorrhage terms (excluding laboratory terms) (SMQ) Narrow
Major hemorrhage - Defined as serious or ≥ Grade 3 bleeding at any site, or central nervous system bleeding of any grade	Major haemorrhage: Subdural haematoma PT, Subdural haemorrhage PT All haemorrhage PTs if AE SOC is "Nervous system disorders" or Serious or ≥ Grade 3 haemorrhage PT if AE SOC is not "Nervous system disorders"
Atrial fibrillation and/or flutter	Atrial fibrillation PT, Atrial flutter PT
Hypertension	Hypertension (SMQ) Narrow
Second primary malignancies Skin cancers	Malignant tumours (SMQ) Narrow Subcategory - Skin malignant tumours (SMQ) narrow
Tumor lysis syndrome	Tumour lysis syndrome (SMQ) Narrow
Infection Opportunistic Infections	Infections: Infections and Infestations SOC Subcategory - Opportunistic infections: Opportunistic infections (SMQ) Narrow
Neutropenia	Neutropenia PT, Neutrophil count decreased PT, Febrile neutropenia PT, Agranulocytosis PT, Neutropenic infection PT, Neutropenic sepsis PT
Thrombocytopenia	Thrombocytopenia PT, Platelet count decreased PT
Anemia	Anaemia PT, Haemoglobin decreased PT