BeiGene

STATISTICAL ANALYSIS PLAN

Study Protocol Number:	BGB-3111-304
Study Protocol Title:	An international, Phase 3, Open-label, Randomized Study of BGB-3111 Compared with Bendamustine plus Rituximab in Patients with Previously Untreated Chronic Lymphocytic Leukemia or Small Lymphocytic Lymphoma
Date:	May 13, 2021
Version:	1.0

SIGNATURE PAGE

Author: [redacted]

[redacted]

Signature

Date: ____

Approval

[redacted]

Signature

Date: ___

[redacted]

Signature

Date: ___

[redacted]

Signature

Date: ____

TABLE OF CONTENTS

LIST OF ABBREVIATIONS AND DEFINITIONS OF TERMS

Abbreviation	Definition
AE	adverse event
ALT	alanine aminotransferase
ANC	absolute neutrophil count
AST	aspartate aminotransferase
B+R	bendamustine plus rituximab
BTK	Bruton tyrosine kinase
CBC	complete blood count
CFR	Code of Federal Regulations
CI	confidence interval
CLL	chronic lymphocytic leukemia
CR	complete response
CRi	complete response with incomplete bone marrow recovery
CT	computed tomography
CYP	cytochrome P450
DMC	data monitoring committee
ECG	electrocardiogram
ECOG PS	Eastern Cooperative Oncology Group Performance Status
eCRF	electronic case report form
EDC	electronic data capture system
EORTC	European Organisation for Research and Treatment of Cancer
FCR	fludarabine, cyclophosphamide, and rituximab
FDA	Food and Drug Administration
FISH	fluorescence in situ hybridization
GCP	Good Clinical Practice
HBcAb	hepatitis B core antibody
HBsAb	hepatitis B surface antibody
HBsAg	hepatitis B surface antigen
HBV	hepatitis B virus
HCV	hepatitis C virus
HR	hazard ratio
IEC	Independent Ethics Committee
IGHV	immunoglobulin variable region heavy chain
IND	Investigational New Drug
IRB	Institutional Review Board
IRC	Independent Review Committee
IRT	Interactive Response Technology
ITT	Intent-to-Treat population
iwCLL	International Workshop on Chronic Lymphocytic Leukemia
LNN	Lower limit of normal
MAR	Missing at random
MedDRA	Medical Dictionary for Regulatory Activities
MMRM	Mixed model for repeated measures

Abbreviation	Definition
MRI	magnetic resonance imaging
NCI-CTCAE	National Cancer Institute Common Terminology Criteria for Adverse Events
ORR	overall response rate
OS	overall survival
PD	Progressive disease
PFS	progression-free survival
PFS2	progression-free survival 2
PK	Pharmacokinetic
PR	partial response
PRO	patient-reported outcome
SAE	serious adverse event
SD	stable disease
SE	Standard error
SLL	small lymphocytic lymphoma
TEAEs	Treatment emergent adverse events
ULN	upper limit of normal
WHO	World Health Organization

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 results for BGB-3111-304. The focus of this SAP is for the planned interim analysis and the final analysis specified in the study protocol. The plan is written in accordance with protocol version 4.0 dated 10 Feb 2021.

The analysis details for Pharmacokinetic (PK) and Biomarker analyses are not described within this SAP. Separate analysis plans will be completed for these analyses and will be attached to the clinical study report.

2 STUDY OVERVIEW

This is an international (approximately 175 sites), phase 3, open-label, randomized study of zanubrutinib versus bendamustine plus rituximab (BR) in approximately 710 adult patients aged 18 and above with previously untreated chronic lymphocytic leukemia or small lymphocytic lymphoma (CLL/SLL). The study includes approximately 450 patients in cohort 1 and approximately additional 80 patients from Chinese sites in cohort 1a to support further analysis in Chinese population. Patients in both cohort 1 and cohort 1a should be without del17p as shown by central laboratory fluorescence in situ hybridization (FISH). The study also includes patients with del17p by central laboratory testing: approximately 100 patients in Cohort 2 and approximately 80 patients in Cohort 3. For patients in Cohort 3 with a central FISH test result other than del17p-positive CLL/SLL, those with a local laboratory test result documenting pathogenic TP53 variant may be eligible for enrollment.

Cohort 1a will be opened to enrollment in China when the Cohort 1 sample size has been reached. Cohort 2 (Arm C) will be closed to enrollment when the Arm C sample size (approximately 100 patients) has been reached. Cohort 3 (Arm D) will be open for enrollment in selected countries/sites after Arm C closes. Arm D will be closed to enrollment when the Arm D sample size has been reached.

Patients assigned to Cohort 1/1a will be randomized to receive either zanubrutinib monotherapy or B+R. Patients assigned to Cohort 2 (Arm C) will receive zanubrutinib monotherapy. Patients assigned to Cohort 3 (Arm D) will receive venetoclax + zanubrutinib. The primary efficacy endpoint is progression-free survival (PFS) in Cohort 1 determined by independent central review. Disease response will be assessed per the 2008 International Workshop on Chronic Lymphocytic Leukemia (iwCLL) guidelines (Hallek et al, 2008) with modification for treatment-related lymphocytosis (Hallek et al, 2012; Cheson et al, 2012) for patients with CLL, and per Lugano Classification for Non Hodgkin Lymphoma (NHL) (Cheson et al, 2014) – hereafter referred to as "Lugano Classification for NHL" – for patients with SLL. The modification for treatment-related lymphocytosis is important because treatment with Bruton tyrosine kinase (BTK) inhibitors may cause lymphocytosis due to a redistribution of leukemia cells from the lymphoid tissues to the blood. In such cases, increased blood lymphocytosis is not a sign of treatment failure or progressive disease. Other modifications to iwCLL 2008 guideline, in addition to treatment-related lymphocytosis, are included in the study protocol response criteria.

Cohort assignment will be based on the presence or absence of a specific DNA mutation associated with poor clinical outcomes and poor response to standard chemoimmunotherapy: Cohort 1/1a (without del17p), Cohort 2 (with del17p), and Cohort 3 (with del17p or pathogenic TP53 variant). It is expected that approximately 80-90% of patients will have CLL/SLL without

del17p (Moreno et al, 2019) and will be assigned to Cohort 1/1a.

Central randomization (1:1) will be used to assign patients in Cohort 1/1a to one of the following study drug treatments:

- Arm A: zanubrutinib
- Arm B: bendamustine + rituximab (B+R)

Randomization will be stratified by age (< 65 years vs ≥ 65 years), Binet stage (C vs A or B), immunoglobulin variable region heavy chain (IGHV) mutational status (mutated vs unmutated), and geographic region (North America vs Europe vs Asia-Pacific). Because Cohort 1a enrolls only patients from Chinese sites, geographic region will not be a randomization stratification factor for Cohort 1a.

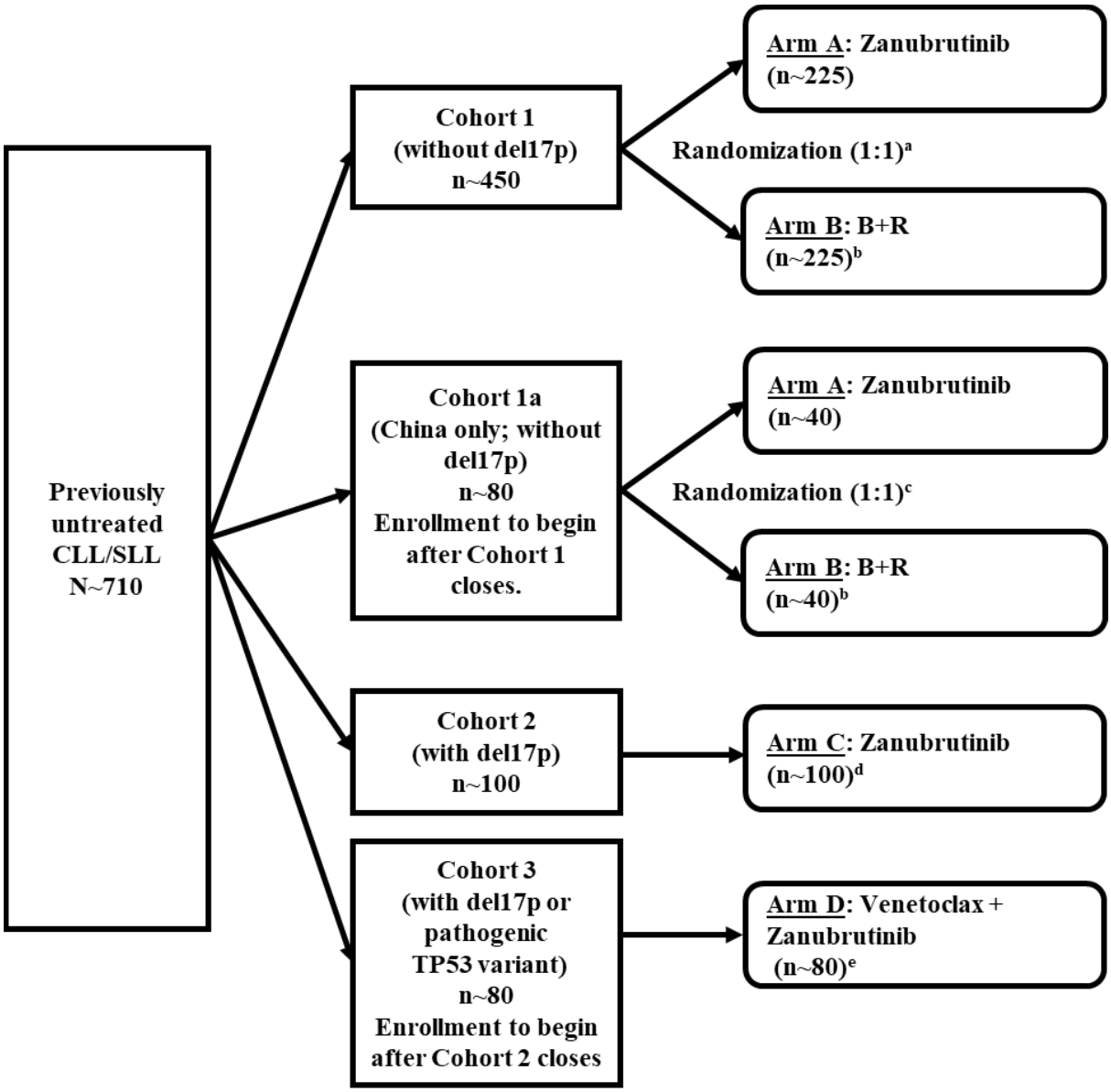
Patients in Cohort 2 (Arm C) will receive treatment with zanubrutinib. Patients in Cohort 3 (Arm D) will receive treatment with venetoclax + zanubrutinib.

Treatment with zanubrutinib, treatment with B+R, and treatment with venetoclax + zanubrutinib will be open label. Study treatment should be commenced within 5 days after randomization/treatment assignment. 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. Arm D only: zanubrutinib will be permanently discontinued in patients who experience unacceptable toxicity, disease progression, or who meet undetectable MRD4 requirements, unless otherwise agreed by the medical monitor to continue zanubrutinib.
- Bendamustine will be administered intravenously at a dose of 90 mg/m^2/day on the first 2 days of each cycle for 6 cycles.
- Rituximab will be administered intravenously at a dose of 375 mg/m^2 for cycle 1, and at a dose of 500 mg/m^2 for cycles 2 to 6.
- Venetoclax will be administered by mouth once daily with food, starting at Cycle 4 at 20 mg with dose escalation weekly up to 400 mg. Venetoclax will be permanently discontinued in patients who experience unacceptable toxicity, disease progression, or who meet undetectable MRD4 requirements or venetoclax will be continued until 24 cycles of venetoclax are completed, whichever comes first.

At investigator discretion, patients in Arm B of Cohort 1/1a may be eligible to receive crossover treatment with zanubrutinib at the time of disease progression confirmed by independent central review. To receive next-line therapy with zanubrutinib, a patient must meet the safety and laboratory requirements, and adhere to the safety, laboratory and efficacy assessments per zanubrutinib (Arms A and C) Schedule of Assessments. After crossover, disease response will only be evaluated by the investigator.

Study Schema



Abbreviations: B+R, bendamustine and rituximab; CLL, chronic lymphocytic leukemia; SLL, small lymphocytic lymphoma.

a. Randomization for Cohort 1 will be stratified by age (< 65 years vs ≥ 65 years), Binet stage (C vs A or B), IGHV mutational status (mutated vs unmutated), and geographic region (North America vs Europe vs Asia-Pacific).
b. Crossover for patients in Arm B to receive next-line zanubrutinib is allowed after disease progression is confirmed by independent central review.
c. The same stratification factors used for Cohort 1 will also be used for Cohort 1a, except for geographic region.
d. Cohort 2 (Arm C) will be closed to enrollment when the Arm C sample size (approximately 100 patients) has been reached.

e. Cohort 3 (Arm D) will be open for enrollment in selected countries/sites after Arm C closes. Arm D will be closed to enrollment when the Arm D sample size has been reached.

2.1 STUDY OBJECTIVES

All efficacy and safety objectives in Cohort 1 (patients without del17p) will compare zanubrutinib (BGB-3111) versus bendamustine plus rituximab.

2.1.1 Primary Objectives

- To compare efficacy between treatment groups in Cohort 1, as measured by progression-free survival determined by independent central review

2.1.2 Secondary Objectives

- To compare efficacy between Arms A and B in cohort 1, as measured by the following:
    - Overall response rate determined by independent central review and by investigator assessment
    - Overall survival
    - Duration of response determined by independent central review and by investigator assessment
    - Progression-free survival determined by investigator assessment
    - Patient-reported outcomes
- To compare efficacy between Arms A and B in pooled Cohort 1/1a patients from Chinese sites, as measured by the following:
    - Progression-free survival determined by independent central review and by investigator assessment
    - Overall response rate determined by independent central review and by investigator assessment
    - Duration of response determined by independent central review and by investigator assessment
- To evaluate efficacy in cohort 2 (patients with del[17p]) for Arm C, as measured by the following:
    - Overall response rate determined by independent central review and investigator review
    - Progression-free survival determined by independent central review and investigator review
    - Duration of response determined by independent central review and investigator review
- To evaluate efficacy in Cohort 3 (patients with del17p or pathogenic TP53 variant) for

Arm D, as measured by the following:

- Overall response rate determined by investigator review
- Progression-free survival determined by investigator review
- Duration of response determined by investigator review
- Assess undetectable minimal residual disease at $< 10^{-4}$ sensitivity (undetectable MRD4) at various timepoints in Arm D

- To compare safety between the treatment groups in Cohort 1
- To compare safety between the treatment groups in pooled Cohort 1/1a patients from Chinese sites
- To summarize safety in Cohort 2 (Arm C)
- To summarize safety in Cohort 3 (Arm D)
- To evaluate pharmacokinetics of zanubrutinib (Arms A and C)
- To evaluate pharmacokinetics of zanubrutinib and venetoclax (Arm D)

2.1.3 Exploratory Objectives

- To evaluate the following;
  - Progression-free survival 2 (for arms A, B and C) determined by investigator assessment
  - Candidate prognostic and predictive biomarkers and biomarkers of relapse
  - Overall survival in pooled Cohort 1/1a patients from Chinese sites
  - Patient-reported outcomes in pooled Cohort 1/1a patients from Chinese sites
  - Overall survival in Cohort 2
  - Patient-reported outcomes in cohort 2
  - Overall survival in Cohort 3
  - Patient-reported outcomes in Cohort 3
  - Time to recurrence of detectable minimum residual disease after discontinuation of zanubrutinib and/or venetoclax in Cohort 3

3 STUDY ENDPOINTS

3.1 PRIMARY ENDPOINTS

The primary endpoint is progression-free survival in Cohort 1 (patients without del17p) determined by independent central review using the modified iwCLL2008 guidelines with modification for treatment-related lymphocytosis and other modifications specified in the study protocol response criteria in patients with CLL and the Lugano Classification for NHL in patients

with SLL, and defined as the time from randomization to the date of first documentation of disease progression or death, whichever occurs first.

3.2 SECONDARY ENDPOINTS

- Overall response rate in Cohort 1 defined as the proportion of patients who achieve a complete response, complete response with incomplete bone marrow recovery, partial response, or partial response with lymphocytosis, determined by independent central review and by investigator assessment
- Overall survival in Cohort 1 defined as the time from randomization to the date of death due to any reason
- Duration of response in Cohort 1 determined by independent central review and by investigator assessment, using the iwCLL criteria with modification for treatment-related lymphocytosis (in patients with CLL) and the Lugano Classification for NHL (in patients with SLL), and defined as the time from the date that criteria for response (ie, PRL or better) are first met to the date that disease progression is objectively documented or death, whichever occurs first
- Progression-free survival in Cohort 1 determined by investigator assessment
- Patient-reported outcomes in Cohort 1 measured by the EQ-5D-5L and EORTC QLQ-C30 questionnaires
- Progression-free survival in pooled Cohort 1/1a patients from Chinese sites determined by independent central review and by investigator assessment
- Overall response rate in pooled Cohort 1/1a patients from Chinese sites determined by independent central review and by investigator assessment
- Duration of response in pooled Cohort 1/1a patients from Chinese sites determined by independent central review and by investigator assessment
- Overall response rate in Cohort 2 (patients with del17p), Arm C, determined by independent central review and by investigator assessment
- Progression-free survival in Cohort 2 (Arm C), determined by independent central review and investigator review
- Duration of response in Cohort 2 (Arm C), determined by independent central review and investigator review
- Overall response rate in Cohort 3 (patients with del17p or pathogenic TP53 variant), Arm D, determined by investigator review
- Progression-free survival in Cohort 3 (Arm D), determined by investigator review
- Duration of response in Cohort 3 (Arm D), determined by investigator review
- Cohort 3 (Arm D) only: undetectable MRD4 rate

- Safety parameters, including AEs, SAEs, clinical laboratory tests, physical examinations, and vital signs
- Pharmacokinetic parameters of zanubrutinib such as apparent clearance of the drug from plasma (CL/F) and AUC from time 0 to 12 hours postdose (AUC0-12) for Arms A, C, and D

3.3 EXPLORATORY ENDPOINTS

- Progression-free survival 2 (PFS2) for Arms A, B, and C, determined by investigator assessment, defined as the time from randomization to the date of progression on the next line of therapy subsequent to the study treatment.
- Clinical outcomes (eg, progression-free survival, overall response rate, duration of response, overall survival) correlated with baseline prognostic and predictive markers (eg, deletion 11q22-23, mutation status of IGHV, pathogenic TP53 variant, β-2 microglobulin level, deletion 13q14, trisomy 12)
- Overall survival in pooled Cohort 1/1a patients from Chinese sites
- Patient-reported outcomes in pooled Cohort 1/1a patients from Chinese sites
- Overall survival in Cohort 2 (Arm C)
- Patient-reported outcomes in Cohort 2 (Arm C), measured by the EQ-5D-5L and EORTC QLQ-C30 questionnaires
- Overall survival in Cohort 3 (Arm D)
- Patient-reported outcomes in Cohort 3, measured by the EQ-5D-5L and EORTC QLQ-C30 questionnaires
- Time to recurrence of detectable minimum residual disease after discontinuation of zanubrutinib and/or venetoclax in Cohort 3
- Medical resource utilization in Cohort 1/1a as assessed by the number of hospitalizations, length of hospital stay, and supportive care in patients
- Medical resource utilization in Cohort 2 as assessed by the number of hospitalizations, length of hospital stay, and supportive care in patients
- Medical resource utilization in Cohort 3 as assessed by the number of hospitalizations, length of hospital stay, and supportive care in patients
- Pharmacokinetic parameters of venetoclax such as apparent clearance of the drug from plasma (CL/F) and AUC from time 0 to 12 hours postdose (AUC_{0-12}) for Arm D

4 SAMPLE SIZE CONSIDERATIONS

The sample size calculation for Cohort 1 is based on the primary efficacy analysis of PFS by independent central review which compares arms A and B in cohort 1. Assuming the PFS HR (Arm A/Arm B) in cohort 1 is 0.58, 118 events are required to achieve 83.5% power at 2-sided

alpha of 0.05 to reject the null hypothesis when 1 interim analysis is planned after 73% of the target number of events at final analysis. If 450 patients are enrolled to Cohort 1 and randomized in a 1:1 ratio to arms A and B over a 25-month period (actual patient enrollment up to November 2018 and 28 patients per month enrollment rate after) and the hazard rate for drop-out of 0.0017/month, 118 PFS events are expected to be accumulated at 41 months from study start. This assumes a median PFS in arm B of 42 months (Woyach JA et al 2018) and that PFS follows exponential distribution. Approximately 710 patients will be enrolled, with 450 patients without the del17p mutation in Cohort 1 available for the primary efficacy analysis, approximately 80 additional patients from Chinese sites without the del17p mutation in Cohort 1a, approximately 100 patients with the del17p mutation in Cohort 2, and approximately 80 patients with the del17p mutation in Cohort 3. For patients in Cohort 3 with a central FISH test result other than del17p-positive CLL/SLL, those with a local laboratory test result documenting pathogenic TP53 variant may be eligible for enrollment.

Sample size selection for Cohort 1a was to accumulate enough PFS events among patients enrolled from Chinese sites to support more than 80% probability of demonstrating an HR <1 among patients enrolled from Chinese sites at the final analysis if the PFS HR based on the ITT Analysis Set crosses the prespecified statistical boundary at the final analysis. Sample size selections for Cohort 2 and 3 were driven by estimated patient availability.

5 STATISTICAL METHODS

5.1 ANALYSIS SETS

The Intent-to-Treat (ITT) Analysis Set includes all enrolled patients who are assigned a treatment group by the IRT system. The ITT Analysis Set will be the primary population for Cohort 1 and Cohort 1a efficacy analyses. Patients will be included in the treatment group originally assigned in the analysis.

The Safety Analysis Set includes all enrolled patients who received any dose of study drug. Patients will be included in the treatment group corresponding to the actual treatment received. The Safety Analysis Set will be used for all safety analyses.

The Per-Protocol Analysis Set includes all enrolled patients who received any dose of study medication and had no important protocol deviations. Criteria for exclusion from the Per-Protocol Population will be determined and documented before the database lock for the primary analysis.

The PK Analysis Set includes all zanubrutinib-treated patients for whom valid zanubrutinib PK concentration can be estimated.

The Analysis Set for Patients from Chinese Sites includes all patients from Chinese sites enrolled in Cohort 1 and Cohort 1a and randomized to a treatment group by the IRT system. The analysis results based on the Analysis Set for Patients from Chinese Sites will be used for China regulatory submission.

5.2 DATA ANALYSIS GENERAL CONSIDERATIONS

5.2.1 Definitions and Computations

Study day will be calculated in reference to the date of the first dose of study drug. For patients randomized/enrolled but did not receive any study drug, the reference will be the date of randomization/enrollment. For assessments conducted on or after the first dose of study drug, study day will be calculated as assessment date – date of the first dose of study drug + 1. For assessments conducted before the date of the first dose of study drug, study day is calculated as assessment date – date of the first dose of study drug.

In the situation where the event date is partial or missing, study day and any corresponding durations will be presented based on the imputations specified in Appendix 1.

The treatment-emergent period is defined as the period of time from the date and time of the first dose of study drug through 30 days for Arm A, C, and D patients and 90 days for Arm B patients after the last dose (permanent discontinuation of study drug) or initiation of new CLL/SLL related therapy. The treatment-emergent period will be used in the summaries of treatment-emergent adverse events (TEAEs).

The treatment duration will be calculated as date of the last dose of study drug – date of the first dose of study drug + 1.

Baseline: Unless otherwise specified, a baseline value related to CLL/SLL disease assessment is defined as the last non-missing value collected on or before the day of first study drug dose, or randomization date + 5 days if patients have never been dosed. Other baseline value, such as demographics, CLL/SLL treatment history, and pertinent medical history etc. is defined as last available value before randomization date.

All calculations and analyses will be conducted using SAS version 9.2 or higher.

5.2.2 Conventions

Unless otherwise specified, the following conventions will be applied to all analyses:

- 1 year = 365.25 days. Number of years is calculated as (days/365.25) rounded up to 1 significant digit.
- 1 month = 30.4375 days. Number of months is calculated as (days/30.4375) rounded up to 1 significant digit.
- Age will be calculated as the integer part of (date of informed consent – date of birth + 1)/365.25
- P-values will be rounded to 4 decimal places. P-values that round to 0.0000 will be presented as '< 0.0001' and p-values that round to 1.000 will be presented as '> 0.9999'.
- Time-to-event or duration of event endpoints will be based on the first/last date that a factor confirming the event was obtained rather than the associated visit date.
- Missing efficacy or safety data will not be imputed unless otherwise specified.

- For laboratory results collected as < or >, the numeric values will be used in analyses.
- For by-visit observed data analyses, percentages will be calculated based on the number of patients with non-missing data as the denominator.
- For continuous endpoints, summary statistics will include n, mean, standard deviation, median, first and third quartiles and range (minimum and maximum).

5.2.3 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 and prior/concomitant medications/procedures as provided in Appendix 1. Missing data for the health-related quality-of-life (HRQoL) data will be handled according to each PRO instrument manual (Fayer & Machin, 2000 in The EORTC QLQ-C30 (Third Edition), 2001; https://euroqol.org/publications/user-guides/).

When summarizing categorical variables, subjects 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, subjects with missing data are not included in calculations. No imputations are made.

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.

If the start day of a subsequent anti-cancer therapy is missing, it will be assumed to be the first day of the month.

If only the day of death date is missing, the death will be assumed to be on the first day of the month if the last known alive date is earlier. If the last known alive date is later than the first day of the month, then the death date will be assumed to be the last known alive date plus 1 day.

No imputation of AE grades will be performed. TEAEs with missing CTCAE grade will only be summarized in the all-grades column.

If the assessment of the relationship of an AE to study treatments is missing, then the AE is assumed to be related to the study treatments in the safety analysis, but no imputation should be done at the data level.

Imputation of PRO scales with missing items is described in Appendix 1.

5.2.4 Adjustments for Covariates

No adjustments for covariates are planned for primary and secondary analyses in the study except selected randomization stratification factors used in the stratified analyses and mixed model analyses. Baseline factors may be used in the model as covariates as supportive analyses for endpoints.

5.2.5 Multiple Comparisons/Multiplicity

Multiplicity adjustments will be made in the primary and secondary efficacy endpoint testing. Details on multiplicity adjustments are described in Section 5.4.1.3 PFS Testing Procedure and Section 5.4.2.9 Secondary Endpoint Testing Procedures for the secondary endpoints.

5.2.6 Data Integrity

Before pre-specified interim or final 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 subjects' relevant outcomes from the clinical database. All data should be complete and reviewed up to a pre-specified cutoff date. Consistency checks and appropriate source data verification should be complete.

In this open-label study, neither the subjects nor the investigators are blinded to treatment. However, access to efficacy data is controlled thus the Sponsor's staff overseeing the conduct of the study or analyzing/summarizing data do not have access to aggregated efficacy summary by treatment arm prior to study database lock. Details of the masking method are described in the study Data Integrity Protection Plan.

Assessment for progression and disease response is performed centrally by the IRC, whose members are blinded to the study treatment and without access to patient absolute lymphocytes count and study drug hold data in an initial review. The IRC data flow and workflow are described in the IRC charter.

5.3 SUBJECT CHARACTERISTICS

5.3.1 Subject Disposition

The number (percentage) of subjects screened, randomized, treated, discontinued from study drugs and discontinued from the study will be summarized. The primary reason for end of treatment (study drug discontinuation) and end of study will be summarized by categories.

Survival status (alive, death, or lost to follow-up) at the data cutoff date will be summarized using the data from the long-term follow-up.

5.3.2 Protocol Deviations

Important protocol deviation criteria will be established before the database lock. Important protocol deviations will be summarized for all randomized patients. They will also be listed by each category. The following categories will be considered when evaluating a deviation as an important protocol deviation:

- Enrollment of a patient into the study even though that patient did not meet all protocol-specified eligibility criteria
- Failure to withdraw a patient from study treatment when that patient met protocol criteria requiring withdrawal from study treatment

- Administration of the incorrect study treatment to a study patient
- Administration of a medication listed as prohibited according to the protocol
- Assignment of patient to the incorrect study cohort based on the study cohort selection assay
- Prior to obtaining informed consent, performance of a study-specific procedure that was not considered a standard or typical procedure for the disease under study

5.3.3 Demographic and Other Baseline Characteristics

Demographic and other baseline characteristics such as disease history will be summarized using descriptive statistics by treatment arm in the ITT analysis set and the analysis set for patients from Chinese sites, and cohort 2 and 3 safety analysis set. Variables include e.g. age, gender, ethnicity, race, geographic region, height, weight, body surface area, cancer type (CLL vs. SLL), Eastern Cooperative Oncology Group (ECOG) performance status, Binet stage, IGHV mutational status, time since diagnosis, sites of disease, presence or absence of disease-related constitutional symptoms, lymphadenopathy, hepatomegaly, splenomegaly, blood lymphocytes, marrow, platelet count, hemoglobin, neutrophils, serum immunoglobulins, coagulation, del(17p), and other selected hematology and chemistry lab measures and cytogenetic abnormalities.

5.3.4 Prior CLL/SLL related Drug Therapies and Surgeries

Because patients with previous systemic treatment for CLL/SLL are excluded from the study, prior CLL/SLL related drug therapy is not expected for the study population. A listing of prior CLL/SLL therapies will be provided.

5.3.5 Prior and Concomitant Medication and Therapy

Prior and concomitant medications will be coded using the World Health Organization Drug Dictionary (WHO DD) drug codes and will be further classified to the appropriate Anatomical Therapeutic Chemical (ATC) code.

The number (percentage) of subjects reporting prior and concomitant medications will be summarized by ATC medication class and WHO DD preferred term in the safety population. Prior medications are defined as medications that started before the first dose date. Concomitant medications will be defined as medications that (1) started before the first dose of study drug and were continuing at the time of the first dose of study drug, or (2) started on or after the date of the first dose of study drug up to 30 days for Arm A, C, and D patients and 90 days for Arm B patients after the patient's last dose or initiation of a new CLL/SLL related therapy, whichever is earlier.

5.3.6 Medical History

Medical History will be coded using MedDRA (version 22.0 or newer). The number (percentage) of subjects reporting a history of any medical condition, as recorded on the CRF, will be summarized by system organ class and preferred term in the safety population.

5.4 EFFICACY ANALYSIS

Analysis of efficacy endpoints comparing Arm A and B will be conducted in the cohort 1 and 1a ITT Analysis Set and the Analysis Set for Patients from Chinese Sites, unless otherwise specified. The following three randomization stratification factors will be used for the stratified analysis/test: age (< 65 years vs ≥ 65 years), Binet stage (C vs A or B), and IGHV mutational status (mutated vs unmutated) per IRT. To reflect the randomization process and maintain the integrity of randomization, all stratified tests will be based on randomization stratification factors as recorded in the IRT. Note geographic region variable, although a stratification factor in the randomization, is not included as a stratification factor in the stratified analysis/test. Efron (1977) method for tie handling will be used in time-to-event efficacy analysis.

5.4.1 Primary Efficacy Analyses

5.4.1.1 Primary Analysis of Progression-Free Survival in Cohort 1

PFS is defined as the time from randomization to the earlier of disease progression or death due to any cause:

PFS = (Disease Progression/Death Date – Randomization Date +1) / 30.4375

For purposes of calculating PFS, the start date for progressive disease is the date at which progression was first observed.

The response and disease progression will be centrally reviewed by IRC, which will be the primary source for the PFS analysis. Criteria for PD and response categories, as well as the process and convention of the IRC, are prospectively detailed in the IRC charter.

The duration of PFS will be right-censored for patients who met 1 of the following conditions: 1) no baseline disease assessments; 2) starting a new CLL/SLL related therapy before documentation of disease progression or death; 3) death or disease progression immediately after two or more missed consecutive disease assessments or; 4) alive without documentation of disease progression before the data cutoff date. For such patients, the primary analysis of PFS will be right-censored according to the convention described in Table 1. These conventions are based on December 2018 FDA Guidance for Industry, 'Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics,' and December 2012 EMA Appendix 1 to the Guideline on the Evaluation of Anticancer Medicinal Products in Man, 'Methodological Consideration for using Progressive-free Survival (PFS) or Disease-free Survival (DFS) in Confirmatory Trials.'

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

Situation	Date of Progression or Censoring	Outcome
No baseline disease assessments	Date of randomization	Censored

New CLL/SLL related treatment started before documentation of PD or death	Date of last disease assessment prior to start of a new CLL/SLL related treatment	Censored[a]
Death or PD immediately after two or more missed consecutive disease assessments	Date of last disease assessment with documented non-progression[b]	Censored
Alive and without PD documentation	Date of last disease assessment	Censored
Death or PD between planned disease assessments	Date of death or first disease assessment showing PD, whichever occurs first	Progressed
Death before first disease assessment	Date of death	Progressed

[a] Patient with a confirmed PD within 30 days or death within 90 days of start of next line therapy will not be censored. Censoring the event from a patient at a date close to patient PD or death date will fall into informative censoring thus biases the treatment effect. Further, counting the event is unlikely to overestimate the overall PFS.
[b] For investigator disease assessment, "non-progression" includes any response assessment code other than "PD" or "Not Done."

The primary analysis of PFS will include all randomized patients (ITT analysis set) in cohort 1. The distribution of PFS, including median PFS and PFS rate at selected timepoints such as 12, 24 and 36 months, will be summarized descriptively using the Kaplan-Meier method for each arm. The 95% confidence interval for median and other quartiles of PFS will be generated by using Brookmeyer method, whereas the 95% confidence interval for PFS rate at selected timepoints will be generated by using Greenwood formula. The primary inferential comparison of PFS

H_0: Hazard ratio (HR) (Arm A/Arm B) = 1

H_a: HR (Arm A/Arm B) < 1

between treatment groups will use the log-rank test stratified by age (< 65 years vs ≥ 65 years), Binet stage (C vs A or B), and IGHV mutational status (mutated vs unmutated) per IRT. The HR will be estimated using a stratified Cox proportional hazard model. Duration of follow-up for PFS will be estimated by reverse Kaplan-Meier method (Schemper and Smith 1996).

For the primary analysis of PFS, the point estimate of the hazard ratio and its 95% CI will be computed based on fixed design procedures using a Cox model. The adequacy of the proportional hazard assumption will be evaluated by examining Schoenfeld residual plot and Kaplan-Meier plot. If strong evidence of non-proportionality of the treatment effect is observed, the time axis will be partitioned using the time points suggested by the residual plot and a piecewise Cox model will be fitted as a sensitivity analysis.

5.4.1.2 Sensitivity Analysis of PFS in Cohort 1

In addition to the primary analysis of PFS based on IRC assessed disease outcomes, the robustness of PFS result will be assessed using additional sensitivity analyses, all of which will be based on the ITT analysis set unless otherwise specified, including the following:

5.4.1.2.1 Unstratified Analysis

In addition to the stratified primary analysis of PFS, the unstratified log-rank test will be performed to compare the PFS between treatment groups. The unstratified Cox proportional hazard regression model will be used to obtain the unstratified estimate of hazard ratio.

5.4.1.2.2 PFS Analysis Based on the Per-Protocol Analysis Set

In this PFS analysis, the Per-Protocol Analysis Set instead of the ITT analysis set will be used as the analysis population. The analysis method will be the same as that for the primary PFS analysis.

5.4.1.2.3 Initiation of Non-Protocol CLL/SLL Related Therapy Treated as a PFS Event

In this sensitivity analysis, initiation of non-protocol CLL/SLL related therapy will be treated as a PFS event whereby PFS is broadly defined as duration from randomization to documented disease progression, initiation of non-protocol CLL/SLL related therapy, or death, whichever occurs earlier. The data censoring rules are the same as those for the primary analysis of PFS except that the use of non-protocol CLL/SLL therapy will be treated as an event rather than a mechanism for censoring. The analysis method will be the same as that for the primary PFS analysis.

5.4.1.2.4 Initiation of Non-Protocol CLL/SLL Related Therapy Treated as neither a PFS Event nor a Censoring Event

In this sensitivity analysis, the use of non-protocol CLL/SLL related therapy will be ignored. The data censoring rules are the same as those for the primary analysis of PFS except that the initiation of non-protocol CLL/SLL related therapy will be excluded as a mechanism for censoring. The analysis method will be the same as that for the primary PFS analysis.

5.4.1.2.5 Death or Disease Progression Immediately After Two or More Missed Consecutive Disease Assessments as a PFS Event

In this sensitivity analysis, death or disease progression immediately after two or more missed consecutive disease assessments will be treated as a PFS event. The analysis method will be the same as that for the primary PFS analysis.

5.4.1.2.6 PFS Analysis under a Non-Proportional Hazard Function

If there is a substantial deviation from the proportional hazard assumption, a piecewise Cox model will be fitted to model the non-proportional hazard. The time-axis will be partitioned using time points suggested by Schoenfeld residual plot and a hazard will be estimated for each

interval. A piecewise weighted log-rank test with weights proportional to the piecewise log hazard ratios will be performed to test the treatment difference between the two arms. Further, methods adjusting for delayed treatment effect (Xu et al 2017) will be applied to analyze the data if appropriate.

5.4.1.2.7 PFS Analysis based on all Patients Randomized to Cohort 1 and Cohort 1a

In this sensitivity analysis, the primary PFS analysis will be repeated using all patients randomized to Cohort 1 and Cohort 1a.

5.4.1.2.8 Hospitalization due to COVID-19 Treated as a Censoring Event

In this sensitivity analysis, patient hospitalization due to COVID-19 will be treated as a censoring event. The data censoring rules are the same as those for the primary analysis of PFS except that the hospitalization due to Covid-19 will be added as a mechanism for censoring. The analysis method will be the same as that for the primary PFS analysis.

5.4.1.2.9 PFS Analysis based on Interval Censoring

In this sensitivity analysis, PD event dates will be assumed as interval censored, i.e. occurred between date of disease assessment right before PD and date of disease assessment with detected PD. A non-parametric method will be used to compare the 2 arms with the interval censored data (Huang 2008).

5.4.1.3 PFS Testing Procedure in Cohort 1

Up to 2 analyses of PFS are planned: an interim analysis and the final analysis. The outcomes determined by the IRC will serve as the primary data source for the primary analysis of PFS. Response and disease progression will be determined using the International Workshop on Chronic Lymphocytic Leukemia (iwCLL) guidelines with modification for treatment-related lymphocytosis for CLL and the modified Lugano classification for Non-Hodgkin Lymphoma for SLL. The IRC will centrally review the disease related tests and assessments to evaluate disease progression and response without knowledge of the randomization assignments, according to the pre-specified charter.

The monitoring boundary for early stopping in PFS will be determined using O'Brien-Fleming alpha spending function (Lan and DeMets 1983) for efficacy and Haybittle-Peto method (Haybittle 1971; Peto et al 1976) for futility so that the overall Type I error is less than or equal to 0.025 (1-sided). The interim analysis will be performed when approximately 86 events (73% of the target number of events at final analysis) from Arms A and B in Cohort 1 are observed. It is estimated that it will take approximately 33 months to observe 86 events under the assumptions described in Section 4. The futility will be non-binding.

Information is based on number of events. Monitoring boundaries will be calculated for the interim analysis based on the actual number of PFS events observed up to the data cutoff of the interim analysis. Deviation from the scheduled interim analyses will not affect overall Type I

error. Primary inference will be based on the stratified log-rank test and nominal p-value boundary will be used at both interim and final analyses. The final analysis of PFS will take place after 118 events observed in Cohort 1, which is estimated as approximately 41 months from study start.

Table 2 is an example scenario for the analyses of PFS. The boundaries will be adjusted based on actual number of events observed at the interim analysis. Nominal p-value boundary will be used for primary inference at both interim and final analyses. If the DMC determines that the observed p-value at the interim analysis is less than or equal to the crossing boundary (nominal significance level) at the analysis and recommends to stop the study for efficacy, the sponsor would evaluate the overall benefit-risk profile of using BGB-3111 vs. BR as a first-line treatment for CLL/SLL patients and consider stopping the trial.

Table 2: Stopping Boundaries for Interim and Final Analyses of Progression-Free Survival

	Time (months)	**# PFS events**	**Nominal 1-sided p-value (Z score) for efficacy**	**Nominal 1-sided p-value (Z score) for interim futility**
Interim analysis	33	86	< 0.009 (> 2.38)	> 0.313 (< 0.489)
Final analysis	41	118	< 0.022 (> 2.007)	-

Abbreviations: HR, hazard ratio; PFS, progression-free survival.

The inferential comparisons for the secondary endpoints will also be performed if a stopping boundary for PFS is met at any of the analyses.

5.4.2 Secondary Efficacy Analyses

5.4.2.1 Overall Response Rate (ORR) in Cohort 1

ORR will be estimated as the crude proportion of patients in each treatment group who achieve PR (including PR-L) or higher. Associated 95% Clopper-Pearson CI will be calculated by treatment group. The odds ratio (and 95% CI), which will be provided as a measure of the relative treatment effect, will be estimated using the stratified Cochran-Mantel-Haenszel method. The BR arm will serve as the reference treatment group in the calculations of the odds ratio. Given the high level of ORR (95%) observed in the BR arm patients in CLL10 study (Eichhorst et al 2016), the comparison between treatment groups for the ORR endpoint in cohort 1 will be descriptive.

Best overall response is defined as the best response recorded from randomization until the earliest date of data cut or start of new CLL/SLL related treatment or disease progression. Patients with no post-baseline response assessment (due to any reason) will be considered as non-responders. The proportion of each of the best response categories (CR, CRi, nPR, PR, partial response with lymphocytosis, SD, and progressive disease) will also be presented by treatment group.

5.4.2.2 Overall Survival (OS) in Cohort 1

OS is defined as the time from randomization to the date of death (whatever the cause):

OS = (Death Date – Randomization Date + 1) / 30.4375

Patients who are alive or lost to follow-up as of the data analysis cutoff date will be right-censored at the patient's date last known to be alive. The distribution of OS, including quartiles, will be summarized descriptively using the Kaplan-Meier method per each arm. Median follow-up for OS will be estimated according to the Kaplan-Meier estimate of potential follow-up also termed "reverse Kaplan-Meier" (Schemper and Smith 1996). The inferential comparison of OS

H_0: Hazard ratio (HR) (Arm A/Arm B) = 1

H_a: HR (Arm A/Arm B) < 1

between treatment groups will use the log-rank test stratified by age (< 65 years vs ≥ 65 years), Binet stage (C vs A or B), and IGHV mutational status (mutated vs unmutated) per IRT. The HR for BGB-3111 arm over the BR arm will be estimated using a stratified Cox proportional hazards model.

The survival rate at selected landmark times (e.g., 1 year, 2 years, and 3 years from randomization) will be estimated for each treatment group by the corresponding Kaplan-Meier estimate with its 95% confidence interval using Greenwood formula.

The final analysis of OS will be performed at the end of the study, approximately 5 years after first patient randomized. Two interim analyses of OS are planned at the time of the interim and final analysis of PFS. Given a 3-year 92% survival rate observed in the BR arm patients in the CLL10 study (Eichhorst et al 2016), the planned interim OS analyses are not expected to have enough power to show statistical difference between the two arms. Therefore, a one-sided 0.00005 alpha will be set for each of the two planned interim analyses.

The OS result will be assessed using additional sensitivity analyses, all of which will be based on the ITT analysis set, including the followings:

5.4.2.2.1 On-Treatment Analysis: Initiation of BGB-3111 in BR Arm Patients Treated as a Censoring Event

In this sensitivity analysis, initiation of BGB-3111 in BR arm patients will be treated as a censoring event. The analysis method remains the same as that for the primary analysis of OS.

5.4.2.2.2 Estimation Based on Inverse Probability of Censoring Weights (IPCW) Method (Robin and Finkelstein 2000)

In this analysis, BR arm patients crossed over to receive any BGB-3111 will be artificially censored at the time of switch, and remaining BR arm patients will be weighted based upon covariate values and a model of the probability of being censored. This allows patients who have not been artificially censored to be weighted in order to reflect their similarities to patients who

have been censored in an attempt to remove the selection bias caused by the censoring – patients who did not crossover and have similar characteristics to subjects who did cross-over receive higher weights. The IPCW version of Kaplan-Meier estimator, log-rank test, and Cox partial likelihood of the HR will be used for the OS analysis. The IPCW method will be only considered when more than 20% BR arm patients crossed over to receive BGB-3111.

5.4.2.2.3 Estimation Based on Iterative Parameter Estimation (IPE) Algorithm (Branson and Whitehead 2002)

The IPE procedure is an extension of rank preserving structural failure time model (RPSFTM). It uses parametric methods and a counterfactual framework to estimate the causal effect of the BGB-3111 treatment. In this analysis, a parametric accelerated failure time model is fitted to the original unadjusted ITT data to obtain an initial estimate of the treatment effect. The failure times of BR arm patients who received any BGB-3111 are then re-estimated using the model, and this iterative procedure continues until the new estimate is very close to the previous estimate, i.e. "converged." Similar to the analysis based on IPCW method, this analysis will be only considered when more than 20% BR arm patients crossed over to use BGB-3111.

5.4.2.3 Duration of response in cohort 1

The duration of overall response will be calculated for patients who achieve CR, CRi, nPR, PR, or PR-L. For such patients, the duration of overall response is defined as the number of days from the start of CR, CRi, nPR, PR, or PR-L (whichever response is achieved first), until the first date that progressive disease is objectively documented or death due to any cause.

DOR will be right censored based on the censoring conventions defined previously for primary PFS analysis. The distribution of DOR will be summarized descriptively using the Kaplan-Meier method. This analysis will include only the patients with best overall response of PR-L or better. Median follow-up for DOR will be estimated according to the Kaplan-Meier estimate of potential follow-up also termed "reverse Kaplan-Meier" (Schemper and Smith 1996)

5.4.2.4 Progression-free survival by investigator assessment in cohort 1

PFS by investigator assessment will be used as the analysis endpoint. The analysis will be performed using the same method as that for the primary PFS analysis based on the IRC assessed outcomes.

5.4.2.5 Patient-Reported Outcomes in cohort 1

The goal of the analyses is to compare the effects of BGB-3111 vs BR treatment based on the patients reported outcomes (PRO) using EORTC QLQ-C30 and EQ-5D-5L. Algorithms for calculating the scales are described in Appendix 5.

The scores of EORTC QLQ-C30 questionnaire will be summarized for each assessment timepoint. Summaries will include: scores and mean changes from baseline in the global health status/QoL scale and the five functional scales (Physical Functioning, Role Function, Emotional Functioning, Cognitive Functioning, and Social Functioning), three Symptom scales (fatigue, pain, and nausea

and vomiting), and six single items (dyspnea, insomnia, appetite loss, constipation, diarrhea, and financial difficulties).

The percentage of patients with a clinically meaningful change from baseline in 'global health status/QoL' will be summarized as "improved", "stable" or "worsened" and compared between arms A and B. Patients will be defined as:

- 'improved' if they have a score change of >+1 response category improvement (corresponding to a 16.7 point increase in derived scales, see details in Appendix 5),
- 'stable' if score change within ±1 response category maintenance,
- 'worsened' if score change >-1.

The scale will be compared between treatment groups using a restricted maximum likelihood-based mixed model for repeated measures (MMRM) to account for missing data under Missing at Random (MAR) assumption (Mallinckrodt et al, 2008). The analysis will be based on the ITT analysis set. The dependent variable of this model will be the QLQ-C30 Global Health Status/QoL score measured along with disease assessments at baseline, week 12 and week 24. The model will include treatment (pre-treatment, Zanubrutinib vs. BR arm), time (as categorical variable), treatment by time interaction, as well as the three randomization stratification factors, as fixed effects. The random patient effects will include patient random intercept on QLQ-C30 Global Health Status/QoL score which is assumed to follow a normal distribution. A point estimate of the treatment difference between the BGB-3111 and BR arms, the corresponding p-value and 95% confidence interval will be provided and used as the primary inference for the QLQ-C30 Global Health Status/QoL endpoint. A supportive analysis will also include week 36 and week 48 data in the analysis. Another supportive analysis will be based on patients completed 6 cycles of treatment. The QLQ-C30 Global Health Status/QoL measured at week 36 and week 48 will be used as the dependent variable in the MMRM model, whereas baseline score, treatment, along with the three randomization stratification factors will be used as the independent variables. Additional independent variables could be included in the MMRM model for the supportive analysis, such as β2 – microglobulin, bulky disease, ECOG status, and certain molecular abnormalities etc. by using model selection strategies. An exploratory analysis will include all ITT patients and their QLQ-C30 Global Health Status/QoL scores collected in the MMRM model.

Analyses of selected subscales of QLQ-C30 will also be performed. The following subscales of the symptom and functional scales will be analyzed using the same approach as used for the primary analysis of QLQ-C30 Global Health Status/QoL score:

- QLQ-C30 Fatigue
- QLQ-C30 Nausea/Vomiting
- QLQ-C30 Pain
- QLQ-C30 Diarrhoea
- QLQ-C30 Physical Functioning
- QLQ-C30 Role Functioning

The EQ-5D-5L comprises a descriptive system and an EQ Visual Analogue Scale (EQ 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 overall score, visual analogue scale, and their changes from baseline will be summarized descriptively.

5.4.2.6 Analysis in Cohort 1a ITT Analysis Set

PFS by independent central review and by investigator assessment will be summarized for Arms A and B in the Cohort 1a ITT Analysis Set. PFS HR between the 2 arms and its 2-sided 95% CI will be estimated from a stratified Cox regression model using randomization stratification factors (age [< 65 years vs ≥ 65 years], Binet stage [C vs A or B], and IGHV mutational status [mutated vs unmutated).

ORR and duration of response will be summarized for each treatment group.

5.4.2.7 Analysis for Patients from Chinese Sites

The analysis proposed in this section will be used for China regulatory submission. PFS by independent central review and by investigator assessment will be summarized for Arms A and B in the Analysis Set for Patients from Chinese Sites. PFS HR between the 2 arms and its 2-sided 95% CI will be estimated from a stratified Cox regression model using randomization stratification factors (age [< 65 years vs ≥ 65 years], Binet stage [C vs A or B], and IGHV mutational status [mutated vs unmutated) in the Analysis Set for Patients from Chinese Sites.

ORR and duration of response will be summarized for each treatment group.

If the interim cohort 1 PFS analysis result crosses the pre-specified boundary, descriptive analysis will be performed on the same datacut and patients from Chinese sites will be continued to be followed until 10 aggregate PFS events are accumulated and the PFS analysis in the Analysis Set for Chinese sites will be performed at that time. If the interim cohort 1 PFS analysis result does not cross the pre-specified boundary but the final PFS analysis result is statistically significant, the analysis for patients enrolled from Chinese sites will be performed concurrently with the final cohort 1 PFS analysis regardless of the number of PFS events accrued from patients enrolled in Chinese sites.

5.4.2.8 Cohort 2

PFS, ORR, duration of response and OS of cohort 2 (arm C) will be summarized descriptively. Independent central review data will be used for PFS, ORR and duration of response. The Kaplan-Meier method will be used to summarize the distribution of PFS, duration of response,

and OS including quartiles and event-free rates at selected timepoints. An estimate of ORR with 95% Clopper-Pearson CI will be generated.

5.4.2.9 Cohort 3

PFS, ORR, and duration of response of Cohort 3 (Arm D) will be summarized descriptively. Investigator assessed response data will be used for PFS, ORR and duration of response. The Kaplan-Meier method will be used to summarize the distribution of PFS and duration of response including quartiles and event-free rates at selected timepoints. An estimate of ORR with 95% Clopper-Pearson CI will be generated. Undetectable MRD4 rate will be estimated as the crude proportion of patients in Arm D who achieve undetectable MRD4 at certain time points, e.g. week 64 and 112 in the peripheral blood. Associated 95% Clopper-Pearson CI will be provided.

5.4.2.10 Secondary Endpoint Testing Procedures

The inferential tests associated with the interim and final analyses of PFS in cohort 1 (primary efficacy endpoint) will be assessed against an overall 1-sided significance level of 0.025 as described in Section 5.4.1.3. Study-wide type-I error will be controlled at the level 0.025 for the testing of the primary endpoint and one secondary endpoint OS in cohort 1. All other inferences will be descriptive without multiplicity adjustment. OS is tested only if the primary endpoint, PFS, is significant.

The significance level for the OS analysis at the interim and final PFS analysis will be 0.00005, and the final OS analysis will be assigned an one-sided alpha of 0.0249. The secondary endpoint testing procedure is closed testing procedures and preserves the family-wise error rate at 0.025 in the strong sense.

5.4.3 Subgroup Analyses in cohort 1

To determine whether the effect of the BGB-3111 regimen is consistent across various subgroups, the estimate of the hazard ratio (or odds ratio) for treatment group (with 95% CI) will be provided for the primary and selected secondary efficacy endpoints. Unstratified analysis will be performed within each subgroup. The example of baseline variables for the subgroup analysis are below:

- Age (< 65 years vs ≥ 65 years) and (< 65 years vs 65-75 years vs ≥ 75 years)
- Sex (male vs female)
- Race (White vs Black vs Asian vs all others)
- Geographic region (North America vs Europe vs Asian Pacific)
- Cancer type (CLL vs SLL)
- Binet stage (C vs A or B)
- ECOG (0 vs ≥ 1)

- Bulky disease (LDi < 5 cm vs ≥ 5 cm and LDi < 10 cm vs ≥ 10 cm)
- IGHV mutational status (mutated vs unmutated)
- Elevated LDH at baseline (No (≤ ULN) vs Yes (> ULN) per central lab)
- Cytopenias at baseline (Yes vs No)
- Chromosome 11q deletion (Yes vs No)
- Del(13q) (Yes vs No)
- Complex karyotype (< 3 vs ≥ 3 abnormalities) and (< 5 vs ≥ 5 abnormalities)
- Trisomy 12 (Yes vs No)
- TP53 mutation (Yes vs No)
- Serum β2 – microglobulin (≤ 3.5 mg/L vs > 3.5 mg/L)

The subgroup variables and the cutoff values are subject to change if warranted to better represent the data.

5.4.4 Exploratory Efficacy Analyses

OS of Cohort 2 (Arm C) and Cohort 3 (Arm D) will be summarized (separately) using descriptive statistics. The Kaplan-Meier method will be used to summarize OS including quartiles and event free rates at selected timepoints. OS for the Analysis Set for Patients from Chinese Sites will be analyzed using the same methods as described in the "Overall Survival" section.

PFS2 will be analyzed using the same methods used for other time to event endpoints such as PFS or OS. If the date of PD on second line therapy is missing, the end date for second line treatment will be used. PFS2 will be compared between the 2 treatment groups.

Cox and/or logistic regression models, as well as descriptive comparisons, may be used to explore the association between prognostic and predictive biomarkers and clinical outcomes.

HRQoL for the Analysis Set for Patients from Chinese Sites will be analyzed using the same method as described in the "HRQoL Analysis" section. Patient-reported outcomes for Cohort 2 and 3 will be summarized descriptively.

For cohort 3 patients discontinue either venetoclax or zanubrutinib or both due to achievement of CR/uMRD, patient time to recurrence of detectable minimum residual disease after discontinuation will be summarized descriptively using Kaplan-Meier method.

5.5 SAFETY ANALYSES

Safety will be assessed by monitoring and recording of all AEs graded by NCI-CTCAE v4.03. Laboratory values (CBC, serum chemistry, serum immunoglobulins, and coagulation), vital

signs, physical examinations and ECG findings will also be used in assessing safety. Descriptive statistics (e.g., n, mean, standard deviation, median, Q1, Q3, minimum, maximum for continuous variables; n [%] for categorical variables) will be used to analyze all safety data by treatment group in Cohort 1, Cohort 1a and Cohort 2, as well as by combining Cohort 1 Arm A with Arm C. Safety will be summarized separately for Arm D. Safety analysis will also be performed by treatment group for all patients enrolled from Chinese sites in Cohort 1 and Cohort 1a, which will be used for China regulatory submission.

5.5.1 Extent of Exposure

Extent of exposure to study drug will be summarized descriptively as the number of cycles received (number and percentage of patients), duration of exposure (date of last study drug administration-date of first study drug administration+1)/30.4375 (months), cumulative total dose received per patient (mg) and relative dose intensity (%).

For each patient, the relative dose intensity (actual vs planned) of each treatment will be calculated.

- For zanubrutinib, the planned dose intensity is 320 mg /day and the actual dose intensity is calculated as total doses taken divided by the duration of the treatment, i.e. last dose of zanubrutinib – first dose +1.

- For bendamustine, the planned dose is 180 mg/m2 per cycle and the total dose (mg/m2) received / number of cycles received will be calculated.

- For rituximab, the planned dose is 375 mg/m2 (first cycle) + 500 mg/m2 (subsequent cycles) and the total dose (mg/m2) received / number of cycle received will be calculated

The average relative dose intensity for each treatment group will be calculated by summarizing the relative dose intensity for individual patients. These results will be provided to determine the presence of any major differences between the treatment groups for the planned vs actual dose and schedule.

The number (and percentage) of patients with dose reduction, infusion interruption (Arm B patients only), and dose missed (Arm A, C and D patients only) will be summarized with the respective reasons. Frequency of dose modifications will be summarized by categories.

Patient data listings will be provided for all dosing records and for calculated summary statistics.

5.5.2 Adverse Events

AEs will be graded by the investigators using CTCAE v4.03. Neutropenia, thrombocytopenia, and anemia will also be graded based on the Grading Scale for Hematologic Toxicity in CLL Studies (Appendix 4) for CLL patients. The AE verbatim descriptions (investigator terms from the CRF) will be classified into standardized medical terminology using the Medical Dictionary for Regulatory Activities (MedDRA). Adverse events will be coded to the MedDRA Version 20.0 or

higher lower level term closest to the verbatim term. The linked MedDRA preferred term (PT) and primary system organ class (SOC) are also captured in the database.

A treatment-emergent AE is defined as an AE that had an onset date or was worsening in severity from baseline (pretreatment) on or after the first dose of study drug up to 30 days for Arm A, C, and D patients and 90 days for Arm B patients following study drug discontinuation or the start of new CLL/SLL related therapy, whichever comes first. Worsening of an event to Grade 5 after the reporting period should also be defined as TEAE. Missing and partially missing AE start dates will be imputed according to the specifications described in Appendix 1. Two sets of summary tables will be provided: TEAEs as defined above and TEAEs plus AEs/SAEs reported during the post-treatment follow-up phase (from the day after the last dose of study drug to the date of patient PD). All AEs, treatment emergent or otherwise, will be presented in subject data listings.

The incidence of TEAEs and TEAEs plus post-treatment phase AEs will be reported as the number (and percentage) of patients with TEAEs plus post-treatment phase AEs by system organ class and preferred term. A patient will be counted only once by the highest severity grade according to CTCAE v4.03 within a system organ class and preferred term, even if the patient experienced more than 1 TEAEs or post-treatment phase AEs within a specific system organ class and preferred term. The number (percentage) of patients with TEAEs plus post-treatment phase AEs will also be summarized by relationship to the study drug. Treatment-related AEs include those events considered by the investigator to be related to study drug or with missing assessment of the causal relationship. For patients in Arm B, treatment-related AEs include AEs considered to be related to either Bendamustine and/or Rituximab.

An overview table, including the incidence of and the number of subjects with TEAEs and TEAEs plus post-treatment phase AEs, serious adverse events (SAEs), treatment-related TEAEs plus post-treatment phase AEs, TEAEs plus post treatment phase AEs with grade 3 or above, treatment-related SAEs, TEAEs plus post-treatment phase AEs that led to death, and TEAEs that led to treatment discontinuation, dose reduction , dose interruption (Arm A and C patients only), dose delay/held (Arm B patients only) or infusion interruption (Arm B patients only) will be provided.

Tabular summaries of the following AE will also be provided:

- TEAEs plus post-treatment phase AEs
    - by system organ class
    - by preferred term in decreasing frequency
    - by AEs of special interest (AEIs)
- Serious adverse events
    - by system organ class
    - by preferred term in decreasing frequency

  - by AEs of special interest (AEIs)
- Treatment-related TEAEs plus post-treatment phase AEs
- TEAEs plus post-treatment phase AEs with grade 3 or above
- Treatment-related SAEs
- TEAEs resulting in discontinuation, reduction, interruption, dose delay/held or infusion interruption of protocol therapy
- Deaths within 30 days of the last administration of protocol therapy

Incidence and time to diarrhea (≥ grade 3), severe bleeding (defined as ≥ grade 3 or serious bleeding of any site or central nervous system bleeding of any grade), and atrial fibrillation (both new onset and exacerbation of existing atrial fibrillation) will also be summarized.

Patient data listings of all AEs, SAEs, treatment-related AEs, grade 3 or above AEs, AEs that led to death and AEs that led to treatment discontinuation will be provided.

5.5.2.1 Exposure-Adjusted Incidence Rates (EAIR)

Given the imbalance in treatment duration between the two arms, an exposure adjusted analysis is also planned to analyze AEIs (AE of interest). The analysis restricts on the occurrence of the first event per patient and ignores the existence of later (multiple) events as these cannot be assumed to occur independently of previous events.

The incidence rate for a patient is derived from the duration of treatment exposure of that patient. A patient's duration of exposure is given either 1) by the time when the event has occurred (non-censored data), or 2) by the total duration of treatment in case the patient does not show the adverse event of interest (censored data). Depending on whether a patient has an adverse event or not, the duration of exposure enters the denominator in its non-censored or censored form, respectively.

The average EAIR per AEI considers the first event per patient per AEI only, and the corresponding exposure time in the denominator:

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

Whereby $TEAE_{AEI,i}$ represents the first TEAE among all AEI TEAEs of patient i and t_i as time when the TEAE occurs (non-censored data) or total duration of treatment if no event occurs (censored data).

5.5.3 Laboratory Values

CBC, serum chemistry, serum immunoglobulin, and coagulation values will be evaluated for each laboratory parameter by treatment group. Abnormal laboratory values will be flagged and identified as those outside (above or below) the normal range. Reference (normal) ranges for

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

Laboratory parameters that are graded in NCI-CTCAE (v.4.03) will be summarized by shifts from baseline CTCAE grades to maximum post-baseline grades. In the summary of laboratory parameters by CTCAE grade, parameters with CTCAE grading in both high and low directions (eg, calcium, glucose, magnesium, potassium, sodium) will be summarized separately. Number (percentage) of patients with abnormal postbaseline laboratory values will be summarized.

The incidence of grade 3 and 4 hematological toxicities (including neutropenia, thrombocytopenia, and anemia, which is graded based on the Grading Scale for Hematologic Toxicity in CLL Studies (Appendix 4)) will be provided for all treatment group by cycles.

5.5.4 Vital Signs

Descriptive statistics for vital sign parameters (systolic and diastolic blood pressure [BP], pulse rate, temperature, and weight) and changes from baseline will be presented by visit and treatment group for all visits. Vital signs will be listed by patient and visit.

5.5.5 Electrocardiograms (ECG)

ECG assessments will be performed at the baseline. Descriptive statistics for ECG parameters will be presented. Overall interpretation of ECG will be summarized. A listing of ECG assessments, including unscheduled post-baseline assessments will be provided.

5.5.6 ECOG

A shift table from baseline to worst post-baseline in ECOG performance score will be generated. ECOG scores will be summarized by visit.

5.6 PHARMACOKINETIC ANALYSES

Plasma BGB-3111 concentrations will be summarized by scheduled time of collection. A population PK analysis may be performed to include plasma concentrations of BGB-3111 from this trial in an existing model. PK parameters such as CL/F and AUC0-12 may be derived from the population PK analysis if supported by data.

An exposure-response (efficacy or safety endpoints) analysis may be performed if supported by data. The results from the population PK and exposure-response analyses may be reported separately from the Clinical Study Report (CSR).

The detail analysis plan for PK will be described in a separate document.

5.7 OTHER ANALYSES

Exploratory analyses may be conducted as appropriate. Any exploratory analyses that are performed will be appropriately titled/labeled as exploratory and will be clearly distinguished from planned analyses when results are reported in the Clinical Study Report.

6 INTERIM ANALYSES

An independent DMC is convened for this study and acts in an advisory capacity to the Sponsor with respect to safeguarding the interests of study patients, assessing interim safety and efficacy data, and for monitoring the overall conduct of the study. The DMC meets approximately every 6 months after an initial meeting.

The objective of the interim analysis is to monitor for differences between treatment arms for evidence of substantial benefit or futility in the BTK-3111 arm. There will be 1 interim analysis of PFS by independent central review in Cohort 1. O'Brien-Fleming boundary approximated by Lan-DeMets spending function will be implemented for efficacy and Haybittle-Peto method for futility. This analysis is scheduled to occur after approximately 73% of the targeted total PFS events from Arms A and B in Cohort 1 are reported, which are anticipated to occur approximately 33 months after the first patient is randomized, under the assumptions used in Section 4. The futility will be non-binding.

7 CHANGES IN THE PLANNED ANALYSIS

None.

8 REFERENCES

Cheson BD, et al. Recommendations for initial evaluation, staging, and response assessment of Hodgkin and Non-Hodgkin lymphoma: The Lugano Classification. Journal of Clinical Oncology, 32: 3059-68, 2014.

Cheson BD, et al. Novel targeted agents and the need to refine clinical end points in chronic lymphocytic leukemia. Journal of Clinical Oncology, 30: 2820-22, 2012.

Common Terminology Criteria for Adverse Events, Version 4.03. Cancer Therapy Evaluation Program.14 June 2010.

Efron B. The Efficiency of Cox's Likelihood Function for Censored Data. Journal of the American Statistical Association, 72:359, 557-565, 1977. DOI: 10.1080/01621459.1977.10480613

Eichhorst B, Fink AM, and Bohlo J et al. First-line chemoimmunotherapy with bendamustine and rituximab versus fludarabine, cyclophosphamide, and rituximab in patients with advanced chronic lymphocytic leukaemia (CLL10): an international, open-label, randomised, phase 3, non-inferiority trial. Lancet Oncology, 17: 928-42, 2016.

European Medicines Agency: Appendix 1 to the guideline on the evaluation of anticancer medicinal products in man: methodological consideration for using progression-free survival (PFS) or disease-free survival (DFS) in confirmatory trials. 13 December 2012.

Hallek M, et al. Clarification of IWCLL criteria for a partial response to therapy. 2013; e letter, published November 13, 2013. http://www.bloodjournal.org/content/111/12/5446/tab-e-letters#clarification-of-iwcll-criteria-for-a-partial-response-to-therapy

Hallek M, et al. Response assessment in chronic lymphocytic leukemia treated with novel agents causing an increase of peripheral blood lymphocytes. 2012; e letter, published June 4, 2012. http://www.bloodjournal.org/content/early/2008/01/23/blood-2007-06-093906/tab-e-letters?sso-checked=true?sso-checked=true#response-assessment-in-chronic-lymphocytic-leukemia-treated-with-novel-agents-causing-an-increase-of-peripheral-blood-lymphocytes

Hallek M, et al. Guidelines for the diagnosis and treatment of chronic lymphocytic leukemia: a report from the International Workshop on Chronic Lymphocytic Leukemia updating the National Cancer Institute-Working Group 1996 guidelines. Blood 12: 5446-56, 2008.

Hallek M, Cheson BD, Catovsky D, et al. iwCLL guidelines for diagnosis, indications for treatment, response assessment, and supportive management of CLL. Blood, 131(25):2745-2760, 2018.

Haybittle, JL. Repeated assessment of results in clinical trials of cancer treatment. British Journal of Radiology, 44:793–797, 1971.

Huang J, Lee C, and Yu, Q. A Generalized Log-Rank Test for Interval-Censored Failure Time Data via Multiple Imputation, Statistics in Medicine, 27, 3217–3226, 2008.

Lan K and DeMets D. Discrete sequential boundaries for clinical trials. Biometrika 70: 659-663, 1983.

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, Vol. 42, pp.303-319, 2008.

Peto R, Pike MC, Armitage P, Breslow NE, Cox DR, Howard SV, Mantel N, McPherson K, Peto J, and Smith, PG. Design and analysis of randomized clinical trials requiring prolonged observation of each patient: I. Introduction and Design. British Journal of Cancer 34:585–612, 1976.

Schemper M, Smith TL. A note on quantifying follow-up in studies of failure time. Controlled Clinical Trials 17:343–346, 1996.

US Food & Drug Administration: Guidance for Industry. Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics. December 2018. www.fda.gov/cder/guidance/7478fnl.htm

Xu, Z., Zhen, B., Park, Y., and Zhu, B. Designing therapeutic cancer vaccine trials with delayed treatment effect. Statistics in Medicine 36: 592–605, 2017.

9 APPENDIX

9.1 APPENDIX 1 IMPUTATION OF MISSING/PARTIALLY MISSING DATES

Missing data will not be imputed unless otherwise specified. The following rules will be applied for the specific analysis and summary purposes mentioned below only.

9.1.1 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 adverse event is partially missing, the date will be imputed to determine whether the adverse event is treatment-emergent. When in doubt, the adverse event will be considered treatment emergent by default. The following rules will be applied to impute partial dates for adverse events.

If start date of an adverse event is partially missing, impute as follows:

- If both month and day are missing and year = year of treatment start date, then set to treatment start date
- If both month and day are missing and year ≠ year of treatment start date, then set to January 01
- If day is missing and month and year = month and year of treatment start date, then set to treatment start date
- If day is missing and month and year ≠ month and year of treatment start date, then set to first day of the month
- If start date is completely missing, set to treatment start date as long as adverse event end date is not before treatment start date

If end date of an adverse event 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 end date is completely missing, do not impute

9.1.2 Prior/Concomitant Medications/Procedures

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

If start date of a medication or procedure is partially 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 or procedure 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 start date or end date of a medication or procedure is completely missing, do not impute.

9.1.3 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 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 subject known to be alive + 1, whichever is later.

9.1.4 New CLL/SLL Related Therapy

If the start day of a subsequent anti-cancer therapy is incomplete or missing, impute as follows:

- If both month and day are missing, then the imputed month and day will be 01Jan or the last day of the month for the last adequate disease assessment if they have the same year.
- If only day is missing, then the imputed day will be the first day of the month.

9.1.5 Diagnosis

If a diagnosis date is partially missing, impute as follows:

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

If a diagnosis date is completely missing, do not impute.

9.1.6 HRQoL Scales

Have at least half of the items from the scale been answered?

- If Yes, use all the items that were completed, and apply standard equations for calculating the scale scores; ignore any items with missing values when making the calculations.

- If No, set scale score to missing

For single-item measures, set score to missing

9.2 APPENDIX 2 CLL RESPONSE DEFINITIONS

(From Modified IWCLL guidelines Hallek et al 2008 and Cheson et al 2012)

Parameter	Complete Response[e]	Partial Response[g]	Partial Response with Lymphocytosis[i]	Progressive Disease[j]
Group A				
Lymphadenopathy[a]	None > 1.5 cm	Decrease ≥ 50% from BL	Decrease ≥ 50% from BL	Increase ≥ 50% from nadir or new lesion
Hepatomegaly[b]	None	Decrease ≥ 50% from BL	Decrease ≥ 50% from BL	Increase ≥ 50% from nadir or new hepatomegaly when none at BL
Splenomegaly[c]	None	Decrease ≥ 50% from BL	Decrease ≥ 50% from BL	Increase ≥ 50% from nadir or new splenomegaly when none at BL
Blood lymphocytes	< 4000/µL	< 5000/µL OR decrease ≥ 50% from BL	Decrease < 50% or increase from BL	Progression based on increasing ALC alone Increase ≥ 50% from nadir and ALC ≥ 5000/µL[k]
Marrow[d]	< 30% lymphocytes and no B-lymphoid nodules. See footnote f for definition of CRi.	50% reduction in marrow infiltrate, or B-lymphoid nodules[h]	50% reduction in marrow infiltrate, or B-lymphoid nodules	Not Applicable
Group B				
Platelet count	> 100,000/µL	> 100,000/µL or increase ≥ 50% over BL	> 100,000/µL or increase ≥ 50% over BL	Decrease of ≥ 50% from BL secondary to CLL[l]
Hemoglobin	> 11.0 g/dL	> 11 g/dL or increase ≥ 50% over BL	> 11 g/dL or increase ≥ 50% over BL	Decrease of > 2 g/dL from BL secondary to CLL[l]

Parameter	Complete Response[e]	Partial Response[g]	Partial Response with Lymphocytosis[i]	Progressive Disease[j]
Neutrophils[d]	> 1500/µL	> 1,500/µL or increase > 50% over BL	> 1,500/µL or increase > 50% over BL	Not Applicable

Abbreviations: BL, baseline (The most recent data/value prior to first dose of study drug); CBC, complete blood count; CLL, chronic lymphocytic leukemia; CR, complete response; CRi, CR with incomplete hematopoietic recovery; CT, computed tomography; eCRF, electronic case report form; PR, partial response.
NOTE: A Best OR of CR, CRi, nPR, PR, or PR-L should be confirmed no earlier than 8 weeks from the time a response of CR, CRi, nPR, PR, or PR-L is first suspected (SD does not require confirmation). At a minimum, a physical exam and complete blood count with differential will be needed to confirm the Best OR.
NOTE: Group A criteria define the tumor load, Group B criteria define the function of the hematopoietic system (or marrow).

a. For CR, all CLL-related lesions must resolve. For determination of other categories, the percent change is calculated from the sum of the products of multiple lymph nodes (as evaluated by CT scans, or by physical examination). NOTE: A single lymph node increase from nadir of ≥ 50% in the longest diameter, if unequivocal, will confirm disease progression. NOTE: Other clinical factors that could result in temporary lymphadenopathy (ie, infection) should be ruled out in order to confirm progression.
b. Percent change is relative to only the abnormally enlarged portion of the organ documented at baseline. No firmly established international consensus of the size of a normal liver is available (Hallek et al, 2018); therefore, liver size should be evaluated by imaging and manual palpation and the size of the abnormal portion is to be documented at baseline and subsequent post-baseline response evaluations. Disease progression requires a ≥ 50% increase in the abnormal portion from nadir (if an abnormal portion exists at nadir) and a minimum absolute increase in the abnormal portion by at least 2 cm. NOTE: Other clinical factors that could result in temporary hepatomegaly (ie, infection) should be ruled out in order to confirm progression. NOTE: If the liver is barely palpable but still considered enlarged on physical examination, enter "0.5 cm" as the measurement for "centimeters below right costal margin" on the corresponding eCRF page.
c. Normal spleen length is ≤13 cm. Percent change is relative to only the abnormally enlarged portion of the organ documented at baseline. For example, a patient with a total craniocaudal spleen length at baseline of 18 cm has a 5 cm abnormal spleen length: 18 cm – 13 cm (normal spleen size)= 5 cm abnormal portion. If the total length of the spleen at Week 12 is 15 cm, then the % change from baseline is 3/5= 60% decrease, which meets the splenomegaly "A" parameter definition for partial response. Disease progression requires a ≥ 50% increase in the abnormal portion from nadir (if an abnormal portion exists at nadir) and a minimum absolute increase in the abnormal portion by at least 2 cm. NOTE: Other clinical factors that could result in temporary splenomegaly (ie, infection) should be ruled out in order to confirm progression. NOTE: If the spleen is barely palpable but still considered enlarged on physical examination, enter "0.5 cm" as the measurement for "centimeters below left costal margin" on the corresponding eCRF page.
d. These parameters are irrelevant for some response categories.
e. CR - Complete response: All CLL-related lesions must resolve. In addition, all the criteria must be met, including CBC data confirming CR without need for exogenous growth factors within the prior 14 days, and patients must lack disease-related constitutional symptoms.
f. CRi - Complete response with incomplete hematopoietic recovery: all the criteria met for complete response except the patient has at least 1 peripheral blood cytopenia (anemia, thrombocytopenia, and/or neutropenia).
g. PR - Partial response: Only those Group A parameters which were previously abnormal at BL can be used to evaluate PR response. If 2 or more Group A parameters are abnormal at BL, then at least 2 of them must meet PR criteria. If only one Group A parameter is abnormal at BL, then that parameter must meet PR criteria. In both cases, at least 1 Group B parameter must also meet PR criteria.
h. nPR - Nodular partial response: all the criteria met for complete response except for the presence of lymphoid nodules in the bone marrow

i. PRL - Partial response with lymphocytosis: PR criteria are otherwise met while blood lymphocytes are only decreased < 50% from BL or are or increased from BL. NOTE: PRL may only be assessed for patients treated with zanubrutinib and would override disease progression by ALC alone only for patients on zanubrutinib. NOTE**:** BTK inhibition may cause lymphocytosis due to a redistribution of leukemia cells from the lymphoid tissues to the blood. In such cases, increased blood lymphocytosis is not a sign of treatment failure or progressive disease.
j. PD - Progressive disease: at least 1 of the above progressive disease criteria must be met. Transformation to a more aggressive histology evidenced by biopsy is also considered as progressive disease.
k. For assessment of PD by ALC alone:
   - During treatment: the increase in ALC should be assessed against the baseline lymphocyte count and not cycle-by-cycle lymphocyte counts, which may not be stable.
   - After treatment: the increase should be assessed against the lowest lymphocyte count assessed at the first follow-up visit after the end of 6 cycles of therapy.
   - For patients who did not receive 6 cycles of treatment: compare ongoing ALC against the lowest lymphocyte count assessed starting from C7D1.

   For assessment of PD by new lesion, one of the following criteria must be met:
   - A new node > 1.50 cm in any axis
   - A new extranodal site > 1.0 cm in any axis OR if ≤ 1.0 cm in any axis, its presence must be unequivocal and must be attributable to CLL/SLL
   - Assessable disease of any size unequivocally attributable to lymphoma

l. After treatment, the progression of peripheral cytopenia(s), as documented by a decrease of hemoglobin levels by > 20 g/L (2 g/dL) from baseline or to < 100 g/L (10 g/dL), or by a decrease of platelet counts by ≥ 50% from baseline or to < 100 x 10^9/L (100,000/μL), which occurs at least 3 months after treatment, defines disease progression, if the bone marrow biopsy demonstrates an infiltrate of clonal CLL cells.

Other clinical factors that could result in the temporary appearance of a new lesion (ie, infection) should be ruled out in order to confirm progression.

SD - Stable disease: is absence of progressive disease and failure to achieve at least a PR or PRL.

For cases of Indeterminate response due to zanubrutinib dosing hold: see below.

Temporary withholding of study drug (eg, for drug-related toxicity, surgery, or intercurrent illness) can cause a transient worsening of disease and/or of constitutional symptoms. In such circumstances, and if medically appropriate, patients may resume therapy and relevant clinical, laboratory, and/or radiologic assessments should be performed to document whether tumor control can be maintained or whether actual disease progression has occurred.

Isolated increase in lymph nodes and/or splenomegaly (defined as vertical spleen length > 13 cm) during periods of zanubrutinib hold will not be considered as progressive disease unless confirmed by a repeat imaging studies at least 6 weeks after restarting study drug administration. The response category “indeterminate due to zanubrutinib hold” should be selected for such instances. Following the repeat imaging 6 weeks after restarting study drug, response should be in comparison to the imaging at baseline.

9.3 APPENDIX 3 MODIFIED LUGANO CLASSIFICATION FOR NON-HODGKIN LYMPHOMA

Response and Site	CT-Based Response (Patients Without PET-Avid Disease at Screening)*
Complete Lymph nodes and extralymphatic sites	Complete radiologic response (all of the following): • Target nodes/nodal masses must regress to ≤ 1.5 cm in longest transverse diameter of a lesion • No extralymphatic sites of disease
Non-measured lesion	Absent
Organ enlargement	Regress to normal
New lesions	None
Bone marrow	Normal by morphology, if indeterminate, immunohistochemistry negative
Partial Lymph nodes and extralymphatic sites	Partial remission (all of the following): • ≥ 50% decrease in sum of the product of the perpendicular diameters for multiple lesions of up to 6 target measurable nodes and extranodal sites • When a lesion is too small to measure on CT, assign 5 mm x 5 mm as the default value • When no longer visible, 0 x 0 mm • For a node > 5 mm x 5 mm, but smaller than normal, use actual measurement for calculation
Non-measured lesions	Absent/normal, regressed, but no increase
Organ enlargement	Spleen must have regressed by > 50% in length beyond normal
New lesions	None
Bone marrow	Not applicable
No response or stable disease Target nodes/nodal masses, extranodal lesions	Stable disease < 50% decrease from baseline in sum of the product of the perpendicular diameters for multiple lesions of up to 6 dominant, measurable nodes and extranodal sites; no criteria for progressive disease are met
Non-measured lesions	No increase consistent with progression
Organ enlargement	No increase consistent with progression
New lesions	None
Bone marrow	Not applicable
Progressive disease** Individual target nodes/nodal masses	Progressive disease requires at least 1 of the following cross product of the longest transverse diameter of a lesion and perpendicular diameter progression: An individual node/lesion must be abnormal with: • longest transverse diameter of a lesion > 1.5 cm and • Increase by ≥ 50% from cross product of the longest transverse diameter of a lesion and perpendicular diameter nadir and • An increase in longest transverse diameter of a lesion or shortest axis perpendicular to the longest transverse diameter of a lesion from nadir

Response and Site	CT-Based Response (Patients Without PET-Avid Disease at Screening)*
	○ 0.5 cm for lesions ≤ 2 cm ○ 1.0 cm for lesions > 2 cm • In the setting of splenomegaly, the splenic length must increase by > 50% of the extent of its prior increase beyond baseline (eg, a 15-cm spleen must increase to > 16 cm). If no prior splenomegaly, must increase by at least 2 cm from baseline • New or recurrent splenomegaly
Non-measured lesions	New or clear progression of pre-existing non-measured lesions
New lesions	• Regrowth of previously resolved lesions • A new node > 1.5 cm in any axis • A new extranodal site > 1.0 cm in any axis; if ≤ 1.0 cm in any axis, its presence must be unequivocal and must be attributable to lymphoma • Assessable disease of any size unequivocally attributable to lymphoma
Bone marrow***	New or recurrent involvement

Abbreviations: CT, computed tomography; FDG, [18F]fluorodeoxyglucose; MRI, magnetic resonance imaging; PET, positron emission tomography.
Modified from Cheson et al, 2014.
*Normal spleen length is ≤13 cm. NOTE: If the spleen is barely palpable but still considered enlarged on physical examination, enter "0.5 cm" as the measurement for "centimeters below left costal margin" on the corresponding eCRF page.
**Progressive disease must be confirmed by repeat imaging no sooner than 4 weeks from the first imaging that show possible progression to rule out pseudo-progression. Patients may continue study treatment while they wait for the confirmation imaging. An exception to the confirmatory scan requirement exists if one or more of the following clear, clinical signs of progression is present and the applicable reviewer believes the progression is unequivocal:

- Clinical symptoms and signs of disease progression (including worsening laboratory values, biopsy-proven disease transformation, etc.)
- Rapid progression of disease or of progressive tumor at critical anatomical sites (eg, spinal cord compression) that, in the opinion of the reviewer, is unequivocal

***After treatment, the progression of peripheral cytopenia, as documented by a decrease of hemoglobin levels by > 20 g/L (2 g/dL) from baseline or to < 100 g/L (10 g/dL), or by a decrease of platelet counts by ≥ 50% from baseline or to < 100 x 10^9/L (100,000/μL), which occurs at least 3 months after treatment, defines disease progression, if the bone marrow biopsy demonstrates an infiltrate of SLL cells.

NOTE: Temporary withholding of study drug (eg, for drug-related toxicity, surgery, or intercurrent illness) can cause a transient worsening of disease and/or of constitutional symptoms. In such circumstances, and if medically appropriate, patients may resume therapy, and relevant clinical, laboratory, and/or radiologic assessments should be performed to document whether tumor control can be maintained or whether actual disease progression has occurred.

BTK inhibition may cause lymphocytosis due to a redistribution of leukemia cells from the lymphoid tissues to the blood. In such cases, increased blood lymphocytosis is not a sign of treatment failure or progressive disease. The opposite may occur during periods of temporary holds of BTK inhibitors (due to adverse events or other reasons), and leukemia cells may redistribute from the blood to lymphoid tissue; this also is not a sign of treatment failure or progressive disease.

Isolated increase in lymph nodes and/or splenomegaly (defined as vertical spleen length > 13 cm) during periods of zanubrutinib hold will not be considered as progressive disease unless confirmed by a repeat imaging study at least 6 weeks after restarting study drug administration. The response category "indeterminate due to zanubrutinib hold" should be selected for such instances. Following the repeat imaging 6 weeks after restarting study drug, response should be in comparison to the imaging at baseline.

9.4 APPENDIX 4 GRADING SCALE FOR HEMATOLOGIC TOXICITY IN CLL STUDIES

Hematologic Grading Scheme

Grade[1]	Decrease in platelets[2] or Hgb[3] (nadir) from pretreatment value	Absolute neutrophil count/μL[4] (nadir)
0	No change to 10%	≥ 2000
1	11%-24%	≥ 1500 and < 2000
2	25%-49%	≥ 1000 and < 1500
3	50%-74%	≥ 500 and < 1000
4	≥ 75%	< 500

Abbreviation: HgB, hemoglobin.

Source: Hallek M, Cheson BD, Catovsky D, et al. Guidelines for the diagnosis and treatment of chronic lymphocytic leukaemia: a report from the International Workshop on Chronic Lymphocytic Leukemia updating the National Cancer Institute - Working Group 1996 guidelines. Blood. 2008;111(12):5446-56

1. Grades: 1, mild; 2, moderate; 3, severe; 4, life-threatening; 5, fatal. Death occurring as a result of toxicity at any level of decrease from pretreatment will be reported as Grade 5.
2. Platelet counts must be below normal levels for Grades 1 to 4. If, at any level of decrease, the platelet count is $< 20 \times 10^9$/L (20,000/μL), this will be considered Grade 4 toxicity, unless a severe or life-threatening decrease in the initial platelet count (eg, $< 20 \times 10^9$/L [20,000/μL]) was present pretreatment, in which case the patient is not evaluable for toxicity referable to platelet counts.
3. Hemoglobin (Hgb) levels must be below normal levels for Grades 1 to 4. Baseline and subsequent Hgb determinations must be performed before any given transfusions. The use of erythropoietin is irrelevant for the grading of toxicity, but should be documented.
4. If the ANC reaches $< 1 \times 10^9$/L (1000/μL), it should be judged to be Grade 3 toxicity. Other decreases in the WBC, or in circulating neutrophils, are not to be considered because a decrease in the WBC is a desired therapeutic endpoint. A gradual decrease in granulocytes is not a reliable index in CLL for stepwise grading of toxicity. If the ANC was $< 1 \times 10^9$/L (1000/μL) before therapy, the patient is not evaluable for toxicity referable to the ANC. The use of growth factors such as granulocyte colony-stimulating factor (G-CSF) is not relevant to the grading of toxicity, but should be documented.

9.5 APPENDIX 5 HEALTH RELATED QUALITY OF LIFE SCORING PROCESS

9.5.1 EORTC QLQ-C30 V3.0

The principle for scoring applies to all scales/scores. Raw scores are calculated as the average of the items that contribute to the scale.

A linear transformation to standardize the raw scores is utilized, so that the scores are ranged from 0 to 100. Increases in scores for functional domains (e.g., physical, role, social, emotional, etc.) are improvements while increases in scores for symptoms (e.g., fatigue, vomiting and nausea, diarrhea, pain, etc.) are deteriorations.

Missing Items

If at least half of the items for a scale are answered, then all the completed items are used to calculate the score. Otherwise, the scale score is set to missing.

In practical terms, if items I_1, I_2, ... I_n are included in a scale, the procedure is as follows:

Raw Score

For all scores, the raw score (RS), is the mean of the component items

RS = $(I_1+I_2+\ldots+I_n)$/n

Derived Scale

The derived scales are obtained from the raw scores as defined in the EORTC manual. The derived scales have a more intuitive interpretation: larger function scale or global health status.

QoL are improvements while larger symptom scales (e.g., pain, nausea, etc.) are deteriorations.

The derivation formulas are as follows.

Linear transformation
Apply the linear transformation to 0-100 to obtain the score S.

Functional scales: $S = \left\{1 - \frac{(RS-1)}{range}\right\} \times 100$

Symptom scales / items: $S = \{(RS-1)/range\} \times 100$

Global health status / QoL: $S = \{(RS-1)/range\} \times 100$

Scales of QLQ-C30

	Scale	Number of items	Item range	Item Numbers

Global health status/ QoL Global health status/QOL	QL2	2	6	29,30
Functional Scales				
Physical functioning	PF2	5	3	1, 2, 3, 4, 5
Role functioning	RF2	2	3	6, 7
Emotional functioning	EF	4	3	21, 22, 23, 24
Cognitive functioning	CF	2	3	20, 25
Social functioning	SF	2	3	26, 27
Symptom Scales/ items				
Fatigue	FA	3	3	10, 12, 18
Nausea and vomiting	NV	2	3	14, 15
Pain	PA	2	3	9, 19
Dyspnoea	DY	1	3	8
Insomnia	SL	1	3	11
Appetite loss	AP	1	3	13
Constipation	CO	1	3	16
Diarrhoea	DI	1	3	17
Financial Difficulties	FI	1	3	28

9.5.2 EQ5D-5L

For the 5 level Dimensions, scores on a 5-point likert scale of 1 to 5, with level 1 indicating no problem and level 5 indicating extreme problems. The Health State is defined by combining one level from each dimension.

EQ visual analogue scale (VAS) includes a scale of 0 to 100 with higher scores indicating better health status.

Scales of EQ-5D-5L

MOBILITY

I have no problems in walking about □

I have slight problems in walking about □

I have moderate problems in walking about □

I have severe problems in walking about □

I am unable to walk about □

SELF-CARE

I have no problems washing or dressing myself □

I have slight problems washing or dressing myself □

I have moderate problems washing or dressing myself □

I have severe problems washing or dressing myself □

I am unable to wash or dress myself □

USUAL ACTIVITIES (e.g. work, study, housework, family or leisure activities)

I have no problems doing my usual activities □

I have slight problems doing my usual activities □

I have moderate problems doing my usual activities □

I have severe problems doing my usual activities □

I am unable to do my usual activities □

PAIN / DISCOMFORT

I have no pain or discomfort □

I have slight pain or discomfort □

I have moderate pain or discomfort □

I have severe pain or discomfort □

I have extreme pain or discomfort □

ANXIETY / DEPRESSION

I am not anxious or depressed □

I am slightly anxious or depressed □

I am moderately anxious or depressed □

I am severely anxious or depressed □

I am extremely anxious or depressed □

- We would like to know how good or bad your health is TODAY.
- This scale is numbered from 0 to 100.
- 100 means the best health you can imagine.
  0 means the worst health you can imagine.
- Mark an X on the scale to indicate how your health is TODAY.
- Now, please write the number you marked on the scale in the box below.

YOUR HEALTH TODAY = ☐