STATISTICAL ANALYSIS PLAN

Study Protocol Number:	**BGB-3111-AU-003**
Study Protocol Title:	**A Phase I, Open-Label, Multiple-Dose, Dose Escalation and Expansion Study to Investigate the Safety and Pharmacokinetics of the BTK Inhibitor BGB-3111 in Subjects with B-Cell Lymphoid Malignancies**
Date:	19 October 2018
Version:	1

SIGNATURE PAGE

[redacted]

Signature

Date: _____

[redacted]

[redacted]

Signature

Date: _____

[redacted]

Approval

[redacted]

Signature

Date: _____

[redacted]

[redacted]

Signature

Date: _____

[redacted]

TABLE OF CONTENTS

LIST OF ABBREVIATIONS AND DEFINITIONS OF TERMS

Abbreviation	Term
AEs	Adverse events
ALT	Alanine aminotransferase
aPTT	Activated partial thromboplastin time
AST	Aspartate aminotransferase
ATC	Anatomical therapeutic chemical
AUC	Area under the plasma concentration-time curve
BID	Twice a day
BMI	Body mass index
BOR	Best of response
BTK	Bruton tyrosine kinase
CLL	Chronic lymphocytic leukemia
C_{max}	Maximum observed plasma concentration
CR	Complete response
CSR	Clinical study report
CT	Computed tomography
DBP	Diastolic blood pressure
DLBCL	Diffuse large B-cell lymphoma
DLT	Dose-limiting toxicity
DOR	Duration of response
eCRF	Electronic case report form

ECG	Electrocardiogram
ECOG	Eastern cooperative oncology group
EDC	Electronic data capture
FL	Follicular lymphoma
GCB	Germinal center B-cell-like
HCL	Hairy cell leukemia
INR	International normalized ratio
IPSS	International prognostic scoring system
IRC	Independent review charter
MCL	Mantle cell lymphoma
MALT	Mucosa associated lymphoid tissue
MedDRA	Medical Dictionary for Regulatory Activities
MRD	Minimal residual disease
MRR	Major response rate
MTD	Maximal tolerated dose
MZL	Marginal zone lymphoma
NCI-CTCAE	National Cancer Institute Common Toxicity Criteria for Adverse Events
ORR	Overall response rate
OS	Overall survival
PBMCs	Peripheral blood mononuclear cells
PD	Disease progression
PFS	Progression-free survival

PK	Pharmacokinetic
RP2D	Recommended phase 2 dose
PT	Preferred term
RT	Richter's transformation
Q1, Q3	First quartile, third quartile
QD	Once daily
QT	Electrocardiographic interval
SAEs	Serious adverse events
SAP	Statistical analysis plan
SBP	Systolic blood pressure
SLL	Small lymphocytic lymphoma
SMC	Safety monitoring committee
SOC	System organ class
SPD	Sum of products of diameters
SD	Standard deviation
$t_{1/2}$	Terminal half-life
TEAE	Treatment-emergent adverse event
t_{max}	Time to maximum observed plasma concentration
TTR	Time to response
V_d/F	Apparent volume of distribution
VGPR	Very good partial response
WHO-DD	World Health Organization Drug Dictionary

WM	Waldenström's macroglobulinemia

1 INTRODUCTION

This statistical analysis plan (SAP) describes the detailed plan for analysis of data in evaluation of safety and efficacy for the BGB-3111-AU-003. This document is based on the protocol Version 8 dated 25-SEP-2018. The analysis plan for pharmacogenomics and biomarker analyses are not included in this SAP.

2 STUDY OVERVIEW

This is a multicenter, phase 1, open-label, multiple-dose, dose escalation, first-in-human study with dose expansion cohorts used to further evaluate the efficacy and safety of the recommended Phase 2 dose (RP2D). The study is conducted in two sequential parts: Dose Escalation (Part 1), followed by Expansion (Part 2), and consists of a screening period, treatment until disease progression, intolerance or death, withdrawal of consent, loss to follow-up, or study termination by sponsor, and follow-up.

Part 1: Dose Escalation

This part of the study followed a modified 3+3 dose escalation scheme. The doses and schedules of interest for this part of the study were:

- 40 mg once a day (QD)
- 80 mg QD
- 160 mg QD
- 160 mg twice a day (BID)
- 320 mg QD

At least 3 patients with B-cell lymphoid malignancies were to be enrolled into each cohort. Additional patient(s), up to a maximum of 6 patients in total, were to be enrolled if more than 3 had been screened and were eligible for the cohort. The dose-limiting toxicity (DLT) assessment and dose-escalation scheme followed the standard 3+3 dose escalation design. For example, 3 additional patients were to be enrolled if a DLT is observed in 1 of 3 patients; 2 additional patients were to be enrolled if a DLT is observed in 1 of 4 patients; and 1 additional patient was to be enrolled if a DLT is observed in 1 of 5 patients. No additional patients are required if a DLT is observed in 1 of 6 patients. The period for DLT assessment was 21 days from first dose of zanubrutinib.

Continuous safety evaluation was performed by the sponsor, the coordinating investigator, and investigators. A Safety Monitoring Committee (SMC) determined dose levels to be administered and dose regimen during dose escalation and utilized the data available from the previous dose levels in their decision making.

In the event that a maximum tolerated dose (MTD) was not identified due to paucity of DLTs, the expansion dose and schedule was to be based on pharmacokinetic (PK) and pharmacodynamic (PD) studies of Bruton tyrosine kinase (BTK) inhibition in peripheral blood mononuclear cells (PBMCs), safety, tolerability and preliminary efficacy.

Approximately 25 patients were to be enrolled and treated during Part 1. At the time of development of this document, enrollment and dose escalation of Part 1 has completed. No DLT was identified. The doses and schedules of interest for Part 2 were determined as160 mg BID and 320 mg QD.

Part 2: Expansion

This part of the study includes cohorts of patients with specific types of B-cell malignancies. The disease cohorts for Part 2 are defined as 2a, 2b, 2c, 2d, 2e, 2f, 2g, 2h, 2i, 2j, 2k, 2l, and 2m. Approximately 380 patients will be enrolled and treated in Part 2.

The number of patients to be enrolled in each of the disease cohorts of Part 2 is described by diagnosis, sub-diagnosis, refractory status (relapsed or refractory [R/R] or treatment-naïve [TN]), dose, and schedule in Table 1 below.

The diagnoses and sub-diagnoses that are to be investigated in Part 2 and summarized in Table 1 are:

- CLL/SLL Chronic lymphocytic leukemia/small lymphocytic lymphoma
- MALT Mucosa associated lymphoid tissue
- RT Richter's Transformation
- WM Waldenström's macroglobulinemia
- FL Follicular lymphoma
- MCL Mantle cell lymphoma
- MZL Marginal zone lymphoma
- HCL Hairy cell leukemia
- DLBCL Diffuse large B-cell lymphoma
- GCB Germinal center B-cell like (sub-diagnosis of DLBCL)
- Non-GCB non-Germinal center B-cell like (sub-diagnosis of DLBCL)

This study will be considered complete once all patients have either manifested disease progression, ceased zanubrutinib due to intolerance, death, or withdrawal from the study.

Table 1: Planned sample size in Part 2 by diagnosis, sub-diagnosis, prior treatment status, starting dose, and schedule

		Diagnosis (sub-diagnosis)									Refractory status		Dose/schedule	
Part 2 Cohort	**n**	**CLL/SLL**	**WM**	**MCL**	**FL**	**MZL**	**DLBCL**	**HCL**	**MALT**	**RT**	**R/R**	**TN**	**320 mg QD**	**160 mg BID**
2a	40			x	x	X	GCB subtype				x		alternate	alternate
2b	40						non-GCB subtype				x			x
2c	70	x									x			x
2d	20		x								x			x
2e	20	x									x		x	x
2f	50		x								x	x	x	x
2g	20			x							x		x	x
2h	20	x										x	x	x
2i	20			>65yr, CIRS>=6								x	x	x
2j	10							x			x		x	x
2k	40				x	X			x		x			x
2l	15									X				x
2m	15	x	x	x	x	X	x	x	x	X	x			x
TOTAL	**380**													

CIRS: cumulative illness rating scale, a measure of comorbidity; QD: once a day; BID: twice a day

Cohort 2m includes patients otherwise eligible for cohorts 2a to 2l who failed to achieve a major response (partial response [PR] or better) after at least 6 months, had disease progression on prior BTK-inhibitor therapy (ibrutinib, acalabrutinib, zanubrutinib, or other BTK-inhibitor therapy), or discontinued BTK-inhibitor therapy due to an AE.

3 STUDY OBJECTIVES

3.1 PART 1

Primary Objectives

- To determine the safety and tolerability of zanubrutinib in patients with B-cell malignancies.
- To determine RP2D and regimen of zanubrutinib when given continuously orally.

Secondary Objectives

- To characterize the PK of zanubrutinib after drug administration.
- To determine the extent of BTK inhibition in PBMCs after treatment with zanubrutinib.
- To describe the preliminary antitumor activity of zanubrutinib.

Exploratory Objectives

[redacted]

3.2 PART 2

Primary Objectives

- To further assess the safety and tolerability of zanubrutinib, administered orally either QD or BID, in patients with specified B-cell malignancies.

Secondary Objectives

- To assess the preliminary antitumor activity of zanubrutinib at RP2D(s) in patients with specific B-cell malignancies.
- To further characterize the PK profile of zanubrutinib.
- To determine the extent of BTK inhibition in PBMCs after treatment with zanubrutinib.

Exploratory Objectives

- [redacted]
- [redacted]
- [redacted]
- [redacted]

4 STUDY ENDPOINTS

4.1 PART 1

Primary Endpoints

- Adverse events (AEs), serious adverse events (SAEs) per the National Cancer Institute Common Toxicity Criteria for Adverse Events (NCI-CTCAE) Version 4.03, physical examination, and laboratory measurements, used to assess safety.
- PK, BTK inhibition in PBMCs, safety and tolerability and preliminary efficacy, used to determine the RP2D and regimen.

Secondary Endpoints

- For a single dose profile: area under the plasma concentration-time curve from zero to the last measurable concentration (AUC_{last}), area under the plasma concentration-time curve from zero to infinity ($AUC_{0\text{-}inf}$), maximum observed plasma concentration (C_{max}), time to maximum observed plasma concentration (t_{max}), terminal half-life ($t_{1/2}$), apparent clearance (CL/F), and apparent volume of distribution (V_z/F).
- After steady-state (ss): $AUC_{last,ss}$, $C_{max,ss}$, and $t_{max,ss}$.
- Overall response rate (ORR), complete response rate (CRR), partial response rate (PRR), minimal residual disease (MRD) clearance rate, progression-free survival (PFS), overall survival (OS) and duration of response (DOR).
- BTK inhibition activity of zanubrutinib in PBMCs via BTK occupancy assay

Exploratory Endpoints

4.2 PART 2

Primary Endpoints

- AEs, SAEs per the NCI-CTCAE Version 4.03 (or higher), physical examination, and laboratory measurements.

Secondary Endpoints

- ORR, CRR, PRR, MRD clearance rate, PFS, OS and DOR.
- PK parameters as described for Part 1, including AUC_{last}, $AUC_{0\text{-}inf}$, C_{max}, t_{max}, $t_{1/2}$, CL/F, V_z/F, $AUC_{last,ss}$, $C_{max,ss}$, and $t_{max,ss}$.
- BTK inhibition activity of zanubrutinib in PBMCs via BTK occupancy assay

Exploratory Endpoints

[redacted]

4.3 EFFICACY ENDPOINTS FOR POOLED PART1 AND PART2 BY DISEASE TYPE

In addition to the endpoints defined separately for Part 1 and Part 2 per protocol in Section 4.1 and 4.2, the following efficacy endpoints are defined for pooled Part 1 and Part 2 cohorts for each disease type.

WM

Response of WM patients will have 2 types of assessment by Independent Review Committee (IRC) and 1 type of assessment by investigator, respectively:

- overall IgM assessment by IRC;
- overall combined assessment by IRC;
- overall combined assessment by investigator

All efficacy endpoints will be analyzed in the WM efficacy evaluable set as defined in Section 6.1.

The difference between these 2 types of assessment are summarized in Appendix A.

Primary efficacy endpoint

- Rate of VGPR or CR per overall IgM assessment by IRC

Secondary efficacy endpoints

- Rate of VGPR or CR per overall combined assessment by IRC
- Major Response Rate (MRR, PR or better) per overall IgM assessment by IRC
- MRR per overall combined assessment by IRC
- Overall Response Rate (ORR, MR or better) per overall IgM assessment by IRC
- ORR per overall combined assessment by IRC
- Duration of VGPR or CR, Duration of major response (PR or better), Duration of overall response (MR or better) per overall IgM assessment by IRC,
- Time to VGPR or CR, Time to major response (PR or better), Time to overall response (MR or better) per overall IgM assessment by IRC
- PFS per overall IgM assessment by IRC
- OS

Exploratory efficacy endpoints

- Rate of VGPR or CR, ORR, MRR per overall combined assessment by investigator
- Duration of VGPR or CR, Duration of major response (PR or better), Duration of overall response (MR or better) per overall combined assessment by investigator,
- Time to VGPR or CR, Time to major response (PR or better), Time to overall response (MR or better) per overall combined assessment by investigator
- PFS per overall combined assessment by investigator
- PFS per overall combined assessment by IRC
- Change of IgM level from baseline
- Tumor involvement in bone marrow
- Change of lymph node size from baseline
- Change of spleen size from baseline
- Change of hemoglobin from baseline

CLL/SLL

Primary efficacy endpoint for CLL/SLL patient is ORR defined as rate of PR-L or better. Secondary efficacy endpoints include rate of PR or better, Complete Response Rate, MRD clearance rate, PFS, OS, TTR and DOR.

Response criteria for patients with CLL/SLL are based on Halleck (2008) and Cheson (2012) for CLL and Modified Lugano Classification for Non-Hodgkin Lymphoma for SLL (Cheson 2014).

MCL

Primary efficacy endpoint for MCL patients is Overall Response Rate. Secondary efficacy endpoints include Complete Response Rate, PFS, OS, TTR and DOR. Response assessment will be based on modified Lugano classification for NHL (Cheson 2014). CT-based responses will also be evaluated. Primary analysis of efficacy endpoints will be based on IRC assessment.

Other B cell Lymphoid Malignancies

Primary efficacy endpoint for other B cell Lymphoid patients is Overall Response Rate. Secondary efficacy endpoints include Complete Response Rate, PFS, OS, TTR and DOR. Response assessment will be based on modified Lugano classification for NHL (Cheson 2014).

5 SAMPLE SIZE CONSIDERATIONS

Part 1

The number of dose levels examined and the emerging zanubrutinib toxicities will determine the sample size. It is anticipated that approximately 25 patients in Part 1 will be required to establish the selected dose and schedule of zanubrutinib when administered as a single agent.

Part 2

Approximately 380 patients will enroll in the disease cohorts in Part 2. In general, the sample size for individual disease cohorts was based on obtaining rigorous descriptions of the safety profile and estimates of the response rates for zanubrutinib in specific B-cell malignancies that have sufficient precision.

For example, initial data suggested a response rate of 30% for Part 2b (non-GCB DLBCL), at least 90% for Part 2c (relapsed or refractory CLL/SLL), and 80% for Part 2d/2f (relapsed or refractory WM). With 40 patients in Part 2b, the lower bound of a 95% confidence interval would be 17% if the observed response rate was 30%. With a total of 70 patients in Part 2c, the lower bound of 95% confidence interval would be 80% if the observed response rate was 90%. With a total of 50 patients in Part 2f, the lower bound of 95% confidence interval was 66% if the observed response rate was 80%.

6 STATISTICAL METHODS

6.1 ANALYSIS SETS

The Safety Set includes all patients who received any dose of zanubrutinib. It will be the primary analysis set of all safety analysis and of the final efficacy analysis except for WM patients.

The Efficacy Evaluable Set for WM is defined as all WM patients in the safety set with baseline IgM (or M-protein) >= 5 g/L and with no prior exposure to a BTK inhibitor.

The DLT Evaluable Set includes patients who had DLT or received at least 21 days of zanubrutinib in Part 1.

The PK Set includes all patients who have at least one PK sample collected (have at least one post-dose PK concentration) according to the protocol and laboratory manual.

The Per-protocol (PP) Set for non-WM disease groups includes patients who received any dose of zanubrutinib and had no important protocol deviations which might impact the primary efficacy and safety analysis per International Conference on Harmonization (ICH) of Technical Requirements for Registration of Pharmaceuticals for Human Use: Structure and Content of Clinical Study Reports E3 guideline. Criteria for exclusion from the PP will be determined and documented before the database lock. This set may be used as a sensitivity analysis of the efficacy endpoints as necessary.

The Per-protocol (PP) for WM efficacy evaluable set includes WM efficacy evaluable patients without important protocol deviations.

A summary of analysis sets will provide the number and percentage of patients in each analysis set.

In addition to the above analysis sets defined for the final analysis, efficacy analyses performed prior to the final analysis might exclude the patients who have started first dose less than 12 weeks before the data cutoff date.

6.2 DATA ANALYSIS GENERAL CONSIDERATIONS

Limited safety and efficacy analysis will be performed by Part 1 and Part 2 separately. The focus of the Part 1 analysis is to present dose effects and justify the RP2D. Part 2 will be analyzed separately for safety, but not for efficacy since it is close to the pooled parts 1 and 2 set (380, 94%, patients planned for Part 2 out of a total 405 patients). Complete safety and efficacy analyses will be performed for the pooled Parts 1 and 2 set only.

For Part 1, safety and efficacy will be summarized by the following dose level/schedule groups:

- 40 mg QD
- 80 mg QD
- 160 mg QD
- 160 mg BID
- 320 mg QD
- Part 1 Overall

For pooled Parts 1 and 2 set, safety will be summarized by the following dose level/schedule groups:

- 160 mg BID

- 320 mg QD
- Parts 1 and 2 Overall

For pooled Parts 1 and 2 set, efficacy and safety will be summarized by the following disease types also. Some disease types might be combined for additional analysis:

- CLL/SLL
- WM
- MALT
- RT
- MCL
- FL
- MZL
- GCB DLBCL
- Non-GCB DLBCL
- HCL
- Overall

A patient will be included in the dose level/schedule group per the first dose received unless otherwise specified.

Statistical programming and analyses will be performed using SAS® (SAS Institute Inc., Cary, NC, USA) version 9.3 or above.

6.2.1 Definitions and Computations

Study treatment (study drug) for this study is zanubrutinib.

Study day: Study day will be calculated relative to the date of the first dose of study treatment. For assessments conducted on or after the date of the first dose of study treatment, study day will be calculated as (assessment date – date of first dose of study treatment + 1). For assessments conducted before the date of the first dose of study treatment, study day is calculated as (assessment date – date of first dose of study treatment). There is no study day 0.

In the situation where the event date is partial or missing, the date will appear partial or missing in the listings; Study day and any corresponding durations will be presented based on the imputations specified in Analysis Details Specification document.

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

Baseline: Unless otherwise specified, a baseline value is defined as the last non-missing value collected before the first dose of study treatment.

.

6.2.2 Handling of Missing Data

Missing data will not be imputed unless otherwise specified. Missing dates or partially missing dates will be imputed conservatively for adverse events and prior/concomitant medications/procedures as provided in Analysis Details Specification document for analyses that require the date values.

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

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

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 treatment in the safety analysis summary. No imputation will be done in the AE listings.

6.2.3 Adjustment for Covariates

With possible exception of exploratory analyses, no adjustments for covariates are planned for this study. Baseline factors may be included in the models used for these exploratory analyses.

6.2.4 Multiplicity Adjustment

No formal hypothesis testing is planned for this study. Adjustment for multiplicity is not planned.

6.2.5 Data Integrity

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

and appropriate source data verification should be completed according to the final data extraction plan.

6.3 PATIENT CHARACTERISTICS

6.3.1 Patient Disposition

The following patient disposition information will be summarized:

- Number of patients enrolled
- Number of treated patients (Safety set)
- Number (%) of treated patients who discontinued treatment
- Reason for treatment discontinuation
- Number (%) of treated patients who discontinued study
- Reason for study discontinuation

The following patient disposition information will be summarized by country/site.

- Number of treated patients
- Number of treated patients who discontinued treatment
- Number of treated patients who discontinued study

The number (%) of patients still receiving treatment and still in the study at data cut-off will also be summarized.

Study follow-up time will be defined as the time from first dose date to the death date or end of study date for patients discontinued from study (whichever occurs first), or the database cutoff date for ongoing patients. Study follow-up time will be summarized by median and range.

6.3.2 Protocol Deviations

Important protocol deviation criteria will be established and patients with important protocol deviations will be identified and documented before the database lock. Important protocol deviations will be summarized by deviation category and listing will also be provided.

6.3.3 Demographic and Other Baseline Characteristics

Demographic and baseline characteristics including the following will be summarized using descriptive statistics:

- Age (years) and age categorized (years) as <65, ≥65 and <75, and ≥75

- Gender
- Race
- Ethnicity
- Geographic region and country
- Eastern Cooperative Oncology Group (ECOG) performance status (categorical)
- Height (cm) and weight (kg)
- Body mass index (BMI in kg/m^2)

6.3.4 Disease Characteristics

Baseline disease characteristics including followings will be summarized by descriptive statistics:

- B-cell lymphoid malignancies disease type
- Time since initial diagnosis
- Time since most recent progression
- Stage of disease at initial diagnosis
- Stage of disease at study entry
- For WM: Genotype status for MYD88 (Leu265Pro mutation, wild type, unknown), CXCR4 (WHIM mutation, wild-type, unknown), extramedullary disease (yes vs no), International Prognostic Scoring System risk status (IPSS; low, intermediate, high), hemoglobin, beta 2 microglobulin
- For CLL: Genotype status for del(17p), p53 mutation, del(11q), del(13q), IgVH mutation
- For MCL: International prognostic index (MIPI) (low, intermediate, high), blastoid variant, LDH
- Bulky disease
- Disease-related B-symptoms
- Serum immunoglobulin (IgM, IgA and IgG)
- Presence of lymphadenopathy, hepatomegaly, splenomegaly
- Bone marrow tumor cell infiltration

Stage of disease is defined by criteria that are specific to the disease subtypes. WM IPSS and is defined in Appendix B.

6.3.5 Medical History

Medical history will be coded using Medical Dictionary for Regulatory Activities (MedDRA) (version 20.0 or higher). The number (%) of patients reporting a history of any medical condition, as recorded on the CRF, will be summarized by system organ class (SOC) and preferred term (PT). A listing of medical history will be provided.

6.3.6 Prior Anti-Cancer Therapies and Surgeries

Prior therapy for lymphoma including following information will be summarized and listed:

- Treatment naïve vs relapsed/refractory (including refractory status)
- Number of prior treatment regimens
- Time (months) from the end of last treatment regimen to study treatment start
- Prior transplant
- Prior lymphoma-related surgery
- Prior radiotherapy

The therapies and surgeries with the same sequence/regimen number are counted as one prior therapy/surgery.

6.3.7 Prior and Concomitant Medications

Prior and concomitant medications will be coded using the World Health Organization Drug Dictionary (WHO DD) drug codes version March 2017 or later and will be further classified to the appropriate Anatomical Therapeutic Chemical (ATC) code.

Prior medications are defined as medications that started before the first dose date. Concomitant medications will be defined as medications that (1) started before the first dose of study treatment and were continuing at the time of the first dose of study treatment, or (2) started on or after the date of the first dose of study treatment up to 30 days after the patient's last dose or initiation of a new anti-cancer therapy. For the purpose of determining if a medication should be noted as a concomitant medication, the imputation rules stated in the Analysis Details Specification document will be used.

The number (%) of patients reporting prior medications and the number (%) of patients reporting concomitant medications will be summarized by ATC medication class Level 2 and WHO DD preferred name in the safety set. A listing of prior and concomitant medications will be provided.

6.4 EFFICACY ANALYSIS

The primary summaries of the efficacy data will be based on the Safety set in the final analysis except for WM. The primary summaries of the efficacy data for WM will be based on the WM Efficacy Evaluable Set, in which patients with baseline IgM (or M-protein) < 5g/L will be excluded. The efficacy endpoints for pooled Part 1 and Part 2 by disease type are listed in Section 4.3. Sections 6.4.1 to 6.4.5 describes the analysis methods for the efficacy endpoints common in all disease types. Sections 6.4.6 to 6.4.8 describe the disease type specific efficacy analysis.

6.4.1 Best overall response and response rate endpoints

A patient's best overall response (BOR) is the best response recorded throughout the study (prior to data cut-off). Responses recorded after initiation of new anti-cancer treatment will not be considered for best overall response. Response rate is a crude proportion of patients with best overall responses in corresponding response categories.

Patients without postbaseline disease assessment will be considered as non-responders. The 2-sided 95% exact binomial confidence interval will be calculated using the Clopper-Pearson method. Patients' best overall response (number and percentage) will be tabulated.

6.4.2 PFS

PFS is defined as the time (in months) from the first dose date of study drug to the date of the earliest occurrence of PD or death (due to any cause), whichever occurs first.

PFS will be right-censored for patients who met one of the following conditions: 1) no baseline disease assessments; 2) starting a new anti-cancer therapy before PD or death; 3) death or PD immediately after more than 6 months since last disease assessment (or 12 months if a patient is on the response assessment schedule of every 24 weeks); and 4) alive without documentation of disease progression. For such patients, the primary analysis of PFS will be right-censored according to the convention described in Table 3 which is based on the May 2007 FDA Guidance for Industry, 'Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics' (www.fda.gov/cder/guidance/7478fnl.htm).

Only those procedures with valid assessment results will be used in determination of PFS.

Table 3: Date of Progression or Censoring for Progression-free Survival

Situation	Date of Progression Event	Outcome
1. Death or PD between planned disease assessments	Date of death or first disease assessment showing PD, whichever occurs first	Event
2. Death before first PD assessment or between adequate assessment	Date of death	Event
3. No baseline disease assessments	Date of first dose	Censored
4. New anti-cancer treatment started before PD or death	Date of last disease assessment without PD prior to start of a new anti-cancer treatment	Censored
5. Death or PD more than 6 months[1] after last disease assessment	Date of last disease assessment that is before death or PD	Censored
6. Alive and without PD	Date of last disease assessment	Censored

[1] Or 12 months if a patient is on the assessment schedule of every 24 weeks.

The distribution of PFS, including median and PFS rate at selected timepoints such as 6, 9, and 12 months, will be estimated using the Kaplan-Meier method. The 95% confidence interval for median and other quartiles of PFS will be generated by using Brookmeyer method (Brookmeyer and Crowley 1982), whereas the 95% confidence interval for PFS rate at landmark times will be generated by using Greenwood formula (Greenwood 1926). Duration of follow-up for PFS will be estimated by reverse Kaplan-Meier method (Schemper and Smith 1996). Kaplan-Meier curves for PFS will be also generated.

A listing will be provided for the information of patient PFS, date of progression or censoring, and reason.

6.4.3 DOR

Duration of response (DOR) for responders (those who satisfy the criteria for ORR, CRR, VGPR and higher rate, and MRR respectively) is defined as time (in months) from the date of the earliest qualifying response to the date of PD or death for any cause (whichever occurs earlier). The analysis methods, including censoring rules, will be same as those for PFS.

6.4.4 Time to Response

Time to response (TTR) for responders is defined as time (in months) from the first dose of the study drug to the date of the earliest qualifying response. TTR will be summarized by sample statistics such as sample mean, median and range.

6.4.5 Overall Survival

Overall survival (OS) is defined as the time (in months) from the date of the first dose to death due to any cause. Patients who remained alive before data cutoff or discontinuation of the study (discontinued study due to reasons other than "Death") will be censored at the last date the patient was known to be alive on or prior to data cutoff. The analysis methods for OS will be same as those for PFS.

6.4.6 WM

6.4.6.1 Primary Efficacy Endpoint

Rate of VGPR or CR per overall IgM assessment by IRC is defined as proportion of patients who reached best response of VGPR or CR among the WM efficacy set. Rate of VGPR or CR will be analyzed per the methods described in section 6.4.1.

6.4.6.2 Secondary efficacy endpoints

The planned analysis for secondary endpoints is provided in Table 4.

Table 4: Secondary Endpoints and Associated Analysis Methods

Secondary efficacy endpoints		
Rate of VGPR or CR per overall combined assessment by IRC	Proportion of patients who reached best response of VGPR or CR per overall combined assessment by IRC among the WM efficacy set	Estimated from the crude proportion in the WM efficacy set with 2-sided 95% exact binomial confidence interval using the Clopper-Pearson method
Major Response Rate (PR or better) per overall IgM assessment by IRC	Proportion of patients who reached best response of PR or better per overall IgM assessment by IRC among the WM efficacy set	
Rate of VGPR or CR per overall combined assessment by IRC	Proportion of patients who reached best response of VGPR or CR per overall combined assessment by IRC among the WM efficacy set	

Major Response Rate (PR or better) per overall IgM assessment by IRC	Proportion of patients who reached best response of PR or better per overall IgM assessment by IRC among the WM efficacy set	
Major Response Rate per overall combined assessment by IRC	Proportion of patients who reached best response of PR or better per overall combined assessment by IRC among the WM efficacy set	
Overall Response Rate (MR or better) per overall IgM assessment by IRC	Proportion of patients who reached best response of MR or better per overall IgM assessment by IRC among the WM efficacy set	
Overall Response Rate per overall combined assessment by IRC	Proportion of patients who reached best response of MR or better per overall combined assessment by IRC among the WM efficacy set	
Duration of VGPR or CR per overall IgM assessment by IRC	Time from the date of the earliest VGPR or CR and the date of PD or death for any cause (whichever occurs earlier) per overall IgM assessment by IRC.	Median and event free rate at landmark time points estimated by Kaplan-Meier method (Section 6.4.3). Censoring rule follows Table 3. Only responders are included in the analysis.
Duration of Major Response per overall IgM assessment IRC	Time from the date of the earliest PR or better to the date of PD or death for any cause per overall IgM assessment by IRC	
Duration of Overall Response per overall IgM assessment by IRC	Time from the date of the earliest MR or better to the date of PD or death for any cause (whichever occurs earlier) per overall IgM assessment by IRC	

Time to VGPR or CR per overall IgM assessment by IRC	Time from the first dose of the study drug to the date of the earliest VGPR or CR per overall IgM assessment by IRC	Sample mean, median and range. Only responders are included in the analysis. (Section 6.4.4)
Time to Major Response per overall IgM assessment by IRC	Time from the first dose of the study drug to the date of the earliest PR or better per overall IgM assessment by IRC	
Time to Overall Response per overall IgM assessment by IRC	Time from the first dose of the study drug to the date of the earliest MR or better per overall IgM assessment by IRC	
PFS per overall IgM assessment by IRC	Time from the first dose date of study drug to the date of the earliest PD or death (due to any cause), whichever occurs first, per overall IgM assessment by IRC	Median and event free rate at landmark time points estimated by Kaplan-Meier method. (Section 6.4.2)
OS	Time from the first dose date of study drug to the date if death (due to any cause)	Median and event free rate at landmark time points estimated by Kaplan-Meier method. (Section 6.4.5)

6.4.6.3 Exploratory efficacy endpoints

Rate of VGPR or CR, Overall Response Rate, Major Response Rate per overall combined assessment by investigator will be analyzed using same method as that for Rate of VGPR or CR, Overall Response Rate, Major Response Rate per overall IgM assessment by IRC.

Duration of VGPR or CR, Duration of Major Response (PR or better), Duration of Overall Response (MR or better) per overall combined assessment by investigator, will be analyzed using same method as that for Duration of VGPR or CR, Duration of Major Response, Duration of Overall Response per overall IgM assessment by IRC.

Time to VGPR or CR, Time to Major Response, Time to overall response per overall combined assessment by investigator will be analyzed using the same method as that for Time to VGPR or CR, Time to Major Response, Time to for overall response per overall IgM assessment by IRC.

PFS per overall combined assessment by investigator and PFS per overall combined assessment by IRC will be analyzed using the same method as that for PFS per overall IgM assessment by IRC.

Change of IgM level from baseline will be summarized using descriptive statistics such as mean, median and range by visit and using a box-whisker plot. The summary of maximum reduction per patient will be also provided using descriptive statistics and using a waterfall plot. If the baseline IgM is not applicable , the available M-Protein will be used throughout the study and summarized together with IgM values for this endpoint.

Tumor involvement in bone marrow will be summarized by maximum reduction in cellularity for patients with baseline bone marrow involvement. Summary of maximum reduction per patient will be provided using descriptive statistics and using a waterfall plot.

Change of lymph node size from baseline will be summarized for patients with baseline lymphadenopathy by maximum reduction in SPD. Summary of maximum reduction per patient will be provided using descriptive statistics and using a waterfall plot. Lesions with diameter too small to measure or not reported or less than 0.5 cm will be imputed as 0.5 cm in the SPD calculation. Only nodal target lesions will be included in the SPD calculation for this endpoint. The measurements by IRC's radiology assessment will be used in the calculation.

Change of spleen size from baseline will be summarized for patients with baseline splenomegaly by maximum reduction in cranial to caudal length of spleen. Summary of maximum reduction per patient will be provided using descriptive statistics and using a waterfall plot. The measurements by IRC's radiology assessment will be used in the calculation.

Change of hemoglobin from baseline will be summarized by visit using descriptive statistics and using a box-whisker plot. The analysis will be repeated for the overall WM efficacy evaluable set and for the patients in the set with the baseline hemoglobin level <110 g/L. Summary of maximum increase per patient will be also provided using descriptive statistics and using a waterfall plot.

6.4.6.4 Subgroup Analysis

Primary and selected secondary efficacy endpoints will be summarized in the subgroups including below, as appropriate (i.e. when there is sufficient number of patients in the subgroup, otherwise relevant subgroups may be combined): sex, age group (<65 vs. ≥65 and >75 vs. ≥75), number of prior regimens (treatment naïve vs relapsed/refractory and 0 vs 1-2 vs >=2), ECOG-PS (0 vs. ≥1), genotype (MYD88 L265P mutation/CXCR4 wild-type vs MYD88 L265P mutation/CXCR4 WHIM vs MYD88 wild-type), baseline IgM level (<40 g/L vs >=40 g/L), baseline beta 2 microglobulin level (<3 mg/L, >=3 mg/L), baseline platelet count (≤ 100 x 10^9/L vs > 100 x 10^9/L), baseline hemoglobin (≤ 11.0 g/dL vs >11.0 g/dL), IPSS (low, intermediate, high), baseline extramedullary disease (yes vs no).

6.4.7 MCL

The primary analysis of the efficacy endpoints described in Sections 6.4.1-6.4.4 will be by IRC. Primary and selected secondary efficacy endpoints will be summarized in the subgroups including below, as appropriate (i.e. when there is sufficient number of patients in the subgroup, otherwise relevant subgroups may be combined): sex, age group (<65 vs. ≥65 and >75 vs. ≥75), number of prior regimens (treatment naïve vs relapsed/refractory and 0 vs 1-2 vs >=2), ECOG-PS (0 vs. ≥1), stage of disease at study entry (I-III vs IV), MIPI (low, intermediate, high), blastic variant (yes vs no), bulky disease (>=10cm vs <10cm, >=5cm vs <5cm).

6.4.8 CLL

MRD clearance rate is defined as the proportion of patients with negative MRD status among all CLL patients and will be estimated with 95% exact binomial confidence interval. The summary will be provided separately for the assessment of MRD in the bone marrow and in the peripheral blood. Listing for MRD results will be provided.

6.5 SAFETY ANALYSES

All safety analyses will be provided for the Safety set.

6.5.1 Extent of Exposure

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

- Starting dose and schedule
- Duration of zanubrutinib treatment (in weeks) and categorical distribution
- Cumulative dose of zanubrutinib (mg) administered

- Actual dose intensity of zanubrutinib (mg/day) administered
- Relative dose intensity of zanubrutinib
- Number (%) of patients with dose modifications (reduced/changed or missed/held) for zanubrutinib
- Number (%) of patients with zanubrutinib dose reduced/changed and number of reductions/changes
- Reasons for zanubrutinib dose reduction/change
- Reasons for zanubrutinib dose missing/holds

The relative dose intensity (RDI) is defined as the ratio of the actual dose intensity to the planned dose intensity in percentage. Planned dose intensity for Part 2 patients is 320 mg (per day). Planned dose intensity of Part 1 patients is based on the starting dose for each patient.The actual dose intensity is defined as the actual cumulative dose (mg) taken based on total dose per day divided by the duration of treatment (days).

6.5.2 Adverse Events

A TEAE is an AE with an onset time or increase in severity level on or after the first dose of study drug and within 30 days after the last dose of study drug or prior to the initiation of new anti-cancer therapy, whichever is sooner.

A treatment-related adverse event is an adverse event that is assessed by the investigator as related to zanubrutinib.

The AE verbatim descriptions (investigator terms from the CRF) will be classified into standardized medical terminology using the MedDRA. Adverse events will be coded to the MedDRA Version 20.0 lower level term closest to the verbatim term. The linked MedDRA PT and primary SOC are also captured in the database. AEs will be graded by the investigators using CTCAE v4.03.

TEAE will be summarized based on the number (%) of patients experiencing events by MedDRA SOC and PT. A patient reporting the same AE more than once will be counted only once when calculating incidence 1) within a given SOC, and 2) within a given SOC and PT combination. For such cases, the maximum CTCAE toxicity grade and strongest causal relationship to zanubrutinib for the event will be used in the incidence calculations.

An overall summary of TEAEs will include the number (%) of patients:

- With at least one TEAE
- With at least one treatment-related TEAE
- With at least one grade 3 or higher TEAE
- With at least one treatment-related grade 3 or higher TEAE
- With at least one fatal TEAE
- With at least one treatment-related fatal TEAE
- With at least one serious TEAE
- With at least one treatment-related serious TEAE
- With at least one TEAE leading to treatment discontinuation
- With at least one treatment-related TEAE leading to treatment discontinuation
- With at least one TEAE leading to dose reduced
- With at least one treatment-related TEAE leading to dose reduced
- With at least one TEAE leading to dose held
- With at least one treatment-related TEAE leading to dose held

Summaries of all above TEAEs will be provided by SOC and PT.

Summaries of all TEAEs, SAE, grade 3 or higher TEAEs, treatment-related TEAEs and TEAE leading to treatment discontinuation will also be provided by PT only.

All TEAE, SAE, grade 3 or higher TEAEs, treatment-related TEAEs and TEAEs leading to treatment discontinuation will also be summarized by SOC, PT, and maximum severity.

AE of special interest (AESI) will be defined and summarized by AESI category name and PT for each of following groups:

- All AESIs
- AESIs grade 3 or higher
- Serious AESIs
- AESIs by maximum severity

Time to first AESI along with cumulative event rates at milestone timepoints as well as patient incidence in 3-month intervals over time may be presented for selective AESIs.

The following AEs of hepatic events will be summarized by PT:

- All AEs of hepatic events
- AEs of hepatic events grade 3 or higher
- Serious AEs of hepatic events
- AEs of hepatic events by maximum severity

An exploratory analysis of the correlation between the incidence of an AE of hemorrhage and the prior use of anti-platelet and anti-coagulant medications will be performed.

A summary of the number of deaths and the cause of death, classified by deaths within 30 days of last dose of study drug and deaths more than 30 days after the last dose, will be provided.

Listings of deaths, all AEs, serious AEs, AEs leading to death, AEs leading to dose modification or discontinuation of zanubrutinib in the Safety set, and AEs classified as potential DLTs (i.e. TEAE started within 21 days after first dose and met at least one of the criteria in Protocol Section 4.1.1.1 without considering data not collected in the AE CRF) in the DLT Evaluable set, will be provided.

6.5.3 Laboratory Values

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

Selected laboratory test results will be assigned toxicity grades using the IWCLL 2008 Grading Scale (Hallek et al 2008) for hematological toxicity in CLL/SLL patients and NCI CTCAE 4.03 for hematological toxicities in non-CLL/SLL patients and non-hematological toxicities of all patients. . These laboratory parameters of interest are:

Hematology	**Serum Chemistry**		**Coagulation**
Hemoglobin (decrease)	Alanine transaminase (ALT) (increase)	Albumin (decrease)	Activated partial thromboplastin time (aPTT) (increase)
Platelets (decrease)	Aspartate transaminase (AST) (increase)	Uric Acid (increase)	International Normalized Ratio (INR) (increase)

WBC (increase, decrease)	Alkaline Phosphatase (increase)	Sodium (increase, decrease)
Absolute Neutrophil Count (ANC, decrease)	Total Bilirubin (increase)	Phosphorus (decrease)
Absolute Lymphocyte Count (increase, decrease)	Creatinine (increase)	Potassium (increase, decrease)
	Calcium (increase, decrease)	Magnesium (increase, decrease)
	Glucose (increase, decrease)	

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

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

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

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

6.5.4 Vital Signs

Actual value and change from baseline for all vital signs and weight will be summarized at each scheduled visit. For the Week 1 Day 1 and Week 2 Day 1 visits (i.e. PK sample taken date), the actual value and change from pre-dose to post-dose for all vital signs will be summarized by timepoint of the same day.

Shift tables assessing the toxicity grade at baseline versus worst toxicity recorded post-baseline will be presented. Box-whisker plots will be generated for actual value and change from baseline for systolic and diastolic blood pressure.

6.5.5 Physical Examination

Physical examination results will be listed without summary.

6.5.6 Electrocardiograms (ECG)

Actual value and change from baseline for the PR, QT, QTc-Fridericia, QTc-Bazzett, VR, and QRSD intervals will be summarized. If triplicate readings are recorded (e.g. Screening), the average of the readings for the visit will be used for the summary. .

The number and percentage of patients satisfying the following QT and QTcF conditions at any time post-baseline will be summarized:

- > 450, > 480, or > 500 msec
- ≤ 30 msec increase from baseline, > 30 and ≤ 60 msec increase from baseline, or > 60 msec increase from baseline

6.5.7 ECOG

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

6.6 PHARMACOKINETIC ANALYSES

The population of interest for the pharmacokinetic (PK) analyses will include all patients who have at least one PK sample collected according to the protocol and laboratory manual. Nominal time may be used for the interim PK analysis and actual collection times will be used in the final analysis and reporting. For intensive PK profile, PK parameters will be derived using standard non-compartmental methods. In addition, PK data from this study, along with data from other trials will be included in the population PK analysis and will be reported separately.

For intensive PK profile, plasma zanubrutinib concentration-time data will be summarized and displayed in both tabular and graphical form. Individual and mean plasma concentration versus time data will be tabulated and plotted by dose level. The PK parameter will be estimated based on noncompartmental methods and will be computed in WinNonlin® Enterprise v.5.2 or higher. The PK parameter estimates for a single dose profile and after multiple doses will be summarized, including $AUC_{0\text{-}inf}$, AUC_{last}, C_{max}, t_{max}, CL/F, V_z/F, accumulation ratio and $t_{1/2}$ [as appropriate for data collected]). Additional PK parameters may be calculated if deemed appropriate. Estimates for these parameters will be tabulated and summarized by the dose level, schedule and collection day (i.e., n, Mean, StD, CV%, Median, Min, and Max, GeoMean and GeoCV%).

Dose proportionality of AUC and C_{max} for zanubrutinib will be assessed using the power model as described below and evaluated visually in graphical form.

A linear regression model with the logarithm of the PK parameter as the dependent variable and the logarithm of the dose as the independent variable ($\log[PK]=\alpha+\beta*\log[Dose]$) will be fitted. The model parameters (slope [β] and intercept [α]) will be estimated using least square regression. A minimum of 3 values per dose must be available for a given PK parameter to estimate dose proportionality with the power model. Point estimates and corresponding 2-sided 95% confidence intervals for the slope parameter and the intercept parameter will be provided.

6.7 OTHER ANALYSES

The BTK inhibition activity of zanubrutinib in PBMCs and lymph nodes via BTK occupancy assay in selected cohorts (mandatory for Cohort 2a and optional for Cohorts 2b to 2j) will be summarized using summary statistics and median plots unless significant protocol deviations affect the data analysis or if key dosing, dosing interruption, or sampling information is missing.

The methods for the summary and analysis of the following exploratory endpoints will be provided in a separate document.

[redacted]

[redacted]

7 INTERIM ANALYSIS

No formal interim analyses are planned for this study. Data from the individual dose escalation cohorts in Part 1 of the study were reviewed by the study monitoring committee at the end of the DLT observation period for each cohort. Summaries and analyses of subsets of the study data will be performed as needed for submission to regulatory agencies, to professional meetings and for internal decision-making.

8 CHANGES IN THE PLANNED ANALYSIS

The efficacy endpoints for combined Part 1 and Part 2 cohorts which were not defined in the protocol were added to the SAP. The SAP clarifies that efficacy endpoints for patients in selected disease subtypes will be based on disease assessment by IRC.

9 REFERENCES

Brookmeyer, R. and Crowley, J. (1982) *A confidence interval for the median survival time*. Biometrics, 38, 29-41.

Cheson BD, Byrd JC, Rai KR, et al. (2012) Novel Targeted Agents and the Need to Refine Clinical Endpoints in Chronic Lymphocytic Leukemia. J Clin Oncol. 30 (23): 2820-2822.

Cheson, BD et al. (2014) Recommendations for Initial Evaluation, Staging, and Response Assessment of Hodgkin and Non-Hodgkin Lymphoma: The Lugano Classification. J Clin Oncol. 32:3059–3068.

Fleiss, J.L. (1981) *Statistical Methods for Rates and Proportions, Second Edition,* New York: John Wiley & Sons, Inc.

Greenwood, M. (1926) *The natural duration of cancer*. Public health and medical patients, 33: 1-26.

Hallek M, Cheson BD, Catosvsky D, 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. 2008;111(12):5446-5456

Schemper, M, and Smith, T.L. (1996) *A note on quantifying follow-up in studies of failure time*. Controlled Clinical Trials, 17:343-346

Appendix A: WM Response Assessment Methods

	Overall Combined Assessment (Modified Owens 2013)	Overall IgM Assessment
Complete Response (CR)	• Normal serum IgM values • Disappearance of monoclonal protein by immunofixation • No histological evidence of bone marrow involvement • Complete resolution of lymphadenopathy/splenomegaly if present at baseline	Meeting the CR criteria for IgM/Serum M protein/paraprotein assessment only (after taking into account immunofixation, plasmapheresis*, cryoglobulinemia and dose hold* information).
Very Good Partial Response (VGPR)	• Monoclonal IgM protein is detectable • ≥90% reduction in serum IgM level from baseline (or normal serum IgM level) • Improvement of extramedullary disease, lymphadenopathy/splenomegaly if present at baseline • No new signs or symptoms of active disease	Meeting the VGPR criteria for IgM/Serum M protein/paraprotein assessment only (after taking into account immunofixation, plasmapheresis*, cryoglobulinemia and dose hold* information).
Partial Response (PR)	• ≥50% reduction but <90% reduction in of serum IgM from baseline • Reduction in lymphadenopathy/splenomegaly if present at baseline • No new signs or symptoms of active disease	Meeting the PR criteria for IgM/Serum M protein/paraprotein assessment only (after taking into account immunofixation, plasmapheresis*, cryoglobulinemia and dose hold* information).

Minor Response (MR)	• 25% but < 50% reduction from baseline • No new signs or symptoms of active disease	Meeting the MR criteria for IgM/Serum M protein/paraprotein assessment only (after taking into account immunofixation, plasmapheresis*, cryoglobulinemia and dose hold* information).
Stable Disease (SD)	Not meeting the criteria for Overall Combined Assessment of CR, VGPR, PR, MR or PD.	Not meeting the criteria for Overall IgM/Serum M protein/paraprotein Assessment of CR, VGPR, PR, MR or PD.
Progressive Disease (PD): At least one	≥ 25% increase in serum IgM level from nadir (requires confirmation on ≥ 2 consecutive measurements at least 4 weeks apart if the only sign of progression) and total increase of ≥500 mg/dl from nadir on treatment. Progression of extramedullary disease New splenomegaly or abnormal lymph node New symptomatic disease New or recurrent involvement in bone marrow associated with WM Plasmapheresis that occurs after cycle 2	Meeting the PD criteria for IgM/Serum M protein/paraprotein assessment only (after taking into account immunofixation, plasmapheresis*, cryoglobulinemia and dose hold* information).

Dose hold: Study drug hold with at least 7 consecutive days will be taken into account to assess disease flare when there is a rapid rise in serum IgM level or an increase in known extramedullary disease.

Plasmapheresis: If plasmapheresis has been performed for a subject, pre-plasmapheresis IgM or M protein/paraprotein level obtained during the screening will serve as the baseline value for response assessment. IgM/M-protein values within 4 weeks post plasmapheresis can't be considered in nadir determination. Plasmapheresis that occurs after cycle 2 will be adjudged to have progressive disease

Appendix B: WM International Prognostic Scoring System (IPSS)

Score:

- Age > 65 yr
- Hemoglobin ≤ 11.5 g/dL
- Platelet count ≤ 100 x 10^9/L
- β2-microglobulin > 3 mg/L
- Monoclonal IgM concentration > 7.0 g/dL

IPSS:

- Low 0 or 1 (except age) score
- Intermediate age or 2 scores
- High ≥ 3 scores

Source: International prognostic scoring system for Waldenstrom macroglobulinemia, ***Blood*. 2009;113:4163-4170**

BeiGene

STATISTICAL ANALYSIS PLAN APPENDIX

Study Protocol Number:	**BGB-3111-AU-003**
Study Protocol Title:	**A Phase I, Open-Label, Multiple-Dose, Dose Escalation and Expansion Study to Investigate the Safety and Pharmacokinetics of the BTK Inhibitor BGB-3111 in Subjects with B-Cell Lymphoid Malignancies**
Date:	**Mar 23, 2021**
Version:	**1.0**

TABLE OF CONTENTS

LIST OF TABLES

REVISION HISTORY

Revision	Effective Date	Summary of Revisions
1.0	23-Mar-2021	Effective version

INTRODUCTION

This Statistical Analysis Plan (SAP) appendix provides additional details on the efficacy and safety analyses other than those specified in the SAP Version 1.0 (dated on 19 October 2018) for the BGB-3111-AU-003 Study (referred to as AU003 in the following sections).

This document provides the most updated information for safety-related sections and summarizes the efficacy-related information based on below 3 disease-specific appendixes:

- SAP appendix for MCL Version 1.0 dated on 01Mar2019.
- SAP appendix for MZL Version 1.0 dated on 09Nov2020.
- SAP appendix for WM Version 2.0 dated on 27Nov2019.

1. EFFICACY ANALYSIS SET

Per AU003 SAP, the Safety Set, which includes all patients received any dose of zanubrutinib, will be the primary analysis set of the final efficacy analysis except for WM patients.

For clarity of the presentation, Efficacy Evaluable Set defined same as above safety set will be referenced and used for the final efficacy analyses except for WM patients. For WM patients, as defined in SAP, the final efficacy analyses will be based on the Efficacy Evaluable Set which includes all WM patients in the safety set with baseline IgM (or M-protein) >= 5 g/L and with no prior exposure to a BTK inhibitor.

Unless specified otherwise, efficacy analyses will be presented by specific indications, including CLL/SLL, WM, MCL, MZL, FL, DLBCL, RT, and HCL. For CLL/SLL, WM, MCL and DLBCL, the efficacy will be presented by the following subgroups:

- CLL/SLL patients will be presented by their status of Del17p status, i.e., Del17p+, Del17p- (wild type) and total (including both Del17p+, Del17p- and patients with missing Del17p status).
- WM patients will be presented by treatment naïve (TN) vs. relapsed/refractor (R/R) status as presented in the WM specific CSR.
- MCL patients will be presented by TN vs. R/R status, and within R/R patients, the efficacy will be further presented by those treated with 320 mg total daily dose and total regardless of the doses received. These are based on those presented in the MCL filing specific analyses and CSR.
- DLBCL patients will be presented by non-GCB DLBCL and GCB DLBCL patients.

2. DETERMINATION OF INVESTIGATOR RESPONSE DATE

If investigator assessed timepoint response date can't be fully cleaned at the final database lock and the study team decides to derive the timepoint response date, below principals will be followed:

- Assessment dates of all measurements that could impact the investigator's determination of response will be compared with investigator assessed response date in EDC. (Note: the measurements vary by indications and need be confirmed with medical monitor)
- The last date of the measurements that is on or before the investigator assessed timepoint response date will be the derived timepoint response date. If the response is PD, then the derived timepoint response date will be the PD date from "Disease Progression" CRF page for that corresponding timepoint.

To determine the timepoint response date, the assessment dates of all measurements that could impact the investigator's determination of response along with the date of the investigator response assessment will be placed into a tall skinny file and sorted by Subject and Date.

Table 1 is an example of how the tall skinny file would look like:

Table 1 Example

Subject	Assessment	Date
Sxxxx-xx-xx	Target Lesion	02 June 2017
	Non-Target	02 June 2017
	New Lesion	NA
	PET Scan	03 June 2017
	Liver and Spleen Measurements	02 Jun 2017
	Bone Marrow	NA
	Study Biopsy	NA
	Investigator Response	03 Jun 2017

The last assessment date of all considered measurements (Table 2) on or before the Investigator Response date will be the derived date associated with the Investigator's response, except that PD date from "Disease Progression" CRF page will be used when the corresponding timepoint response is PD.

The measurements to be considered are listed below for different indications in Table 2.

Table 2 Measurements to be considered to derive the timepoint response date

Measurements	CLL	WM	MCL/ MZL	RT & other NHL (i.e., FL, DLBCL, SLL)	HCL
Scan date of (Target Lesions, Non-Target Lesions, New Lesion)	X	X	X	X	X
PET-CT Scan	X		X	X	X
Liver and Spleen measurements (CT, MRI, PET/CT only in lvr_spl)	X	X (Spleen only)	X (Spleen only)	X (Spleen only)	X (Spleen only)
Physical Examination-Targeted (Spleen and Lymph nodes from PETAR)	X (Liver also)	X		X	X
Bone Marrow (% tumor cell involvement: 2 fields (BMLYMPHO, BMGENM) & other morphological findings (BMFIND) & BMTUMYN, excluding blank/NAs)	X	X	X	X	
Bone Marrow (% Hairy Cell Involvement (BMHCL), non-missing ones)					X
Study biopsy	X	X	X	X	X
Serum IgM (non-missing ones, excluding Not Done/Not reported/ Unknown/Not available/blank)		X			
Serum SPEP (only if serum paraprotein was used to determine overall response (SIMPROT=1), non-missing ones)		X			
Serum Immunofixation (only if serum paraprotein was used to determine overall response, Positive/Negative ones, excluding NAs/blanks)		X			
Hematology (Hemoglobin, Platelet Count, Neutrophils, non-missing ones)	X (Lymphocytes also)				X (Hairy Cell Count also)

3. PFS/DOR EVENT AND CENSORED DATES

Step 1: First check whether there is a PD or death.

- If yes, go to step 2.
- If no, go to step 4

Step 2: Define PFS event date as earlier date of following two dates and PFS censoring flag=0

- PD date = Date for the first documented PD from PD form (or IRC raw.RS domain for IRC PFS)
- Death date

Step 3: Compare PFS event date from step 2 to: 1) new anti-cancer therapy start date and 2) the skipped assessment date.

- If PFS event date > new anti-cancer therapy start date OR the PFS event date > the skipped assessment date,
    - PFS censored date = Last disease assessment date prior to the earlier of PFS event date and new anticancer therapy start date.
    - PFS censoring flag = 1

Under protocol version 6.0 or earlier, no regular response assessments (unless PD) were required after Week 52 for patients on treatment. To avoid under count of PDs/death that due to lack of protocol-defined regular assessments, the patients were assumed to be followed till the end of treatment. The potential skipped assessments will be calculated from the maximum of the last assessment and the end of treatment dates. Under protocol version 7.0, dated 02 October 2017, imaging for response assessment after Week 52 (beginning Week 76) was required to be conducted regularly every 24 weeks or when a significant change in response was suspected (progressive disease [PD] or upgrade of response), the skipped assessment will be counted from the very last assessment date:

a) For patients with PFS event date within Week 52 (i.e., ≤365 days) from first dose, the skipped assessment date is defined as 6 months from the last disease assessment date prior to PFS event.
b) For patients with PFS event date after Week 52 (i.e., >365 days) from first dose,
    - If the last disease assessment date prior to PFS event date was earlier than protocol version 7.0 effective date (02 October 2017), the skipped assessment date is defined as 12 months from the maximum of (last disease assessment date prior to PFS event, and EOT date), assuming the response remains unchanged as long as the treatment continues.
    - If the last disease assessment date prior to PFS event date was later than 02 October 2017, the skipped assessment date is defined as 12 months from the last disease assessment date prior to PFS event.

6 months interval ~ 24 weeks + 2 weeks window =182 days

12 months interval ~48 weeks + 2 weeks window=350 days

Step 4: Define PFS censored date as below and PFS censoring flag = 1

- If no disease assessment at baseline or post-baseline, then PFS censored date = first dose date, otherwise PFS censored date = Last disease assessment date

Definition of Last disease assessment date for censored patients and for determining the skipped assessment date:

- Last disease assessment date by Investigator= Last response date (See Section 2) where response is evaluable and not PD prior to new anti-cancer therapy start date.
- Last disease assessment date by IRC= Last non-PD date by IRC if it is prior to new anti-cancer therapy start, otherwise last timepoint assessment date prior to new anti-cancer therapy start date.
- If all timepoint responses are "Not evaluable" (by Investigator or IRC), then Last disease assessment date = Last timepoint assessment date prior to new anti-cancer therapy start date.

3.1. PFS/DOR censoring reasons

Check against the SAP if ">1 missed assessment' apply.

1) If subject is CNSR=1 and has no baseline or post baseline response assessment, set to 'No baseline/post-baseline assessment'.
2) Else if subject is CNSR=1 and disease progression or death right after more than one missed scheduled disease assessment at the time of data cutoff, set to 'Progressive disease/death after >1 missed assessment'.
3) Else if subject is CNSR=1 and disease progression or death after new anti-cancer therapy, set to 'Progressive disease/death after non-protocol anti-cancer therapy'.
4) Else if subject is CNSR=1 and started new anti-cancer therapy and last tumor assessment is on or before new anti-cancer therapy at the time of data cutoff, set to "No documented progressive disease/death".
5) Else if subject is CNSR=1 and started new anti-cancer therapy and last tumor assessment is after new anti-cancer therapy at the time of data cutoff, set to "No documented progressive disease /death: Non-protocol anti-cancer therapy".
6) Else if subject is CNSR=1 and disc study due to patient withdrew consent or lost to follow-up, then set to ' No documented progressive disease/death: Withdrew consent/lost to follow-up'.
    a. Comment: Only report study discontinuation due to withdrew consent or lost to follow up. For other reasons, report as "No documented progressive disease /death" if patient does not have PD or death.
7) Else if subject is CNSR=1 and alive without disease progression at the time of data cutoff, set to "No documented progressive disease/death".

4. DETERMINATION OF BASELINE EXTRAMEDULLARY DISEASE FOR WM

Baseline extramedullary disease (EMD) is defined as "Yes" if any of the below criteria is "Yes":

Lymphadenopathy:

- Lymph Node in raw.TLB [Target Lesions – Baseline]: if any "type of lesion" = "Lymph node" with LDi ("longest perpendicular diameter measurement") >1.5 (cm) then "Yes".
- Lymph Node in raw.PETAR [Physical Examination – Targeted] or raw.PEB [Physical Examination]: if baseline lymph node assessment is "Abnormal" then "Yes". Baseline lymph node is defined as "body system" = "Lymph nodes" with "folder" = "WK1D1" from raw.PETAR, if a patient has missing record or missing result or non-"Abnormal" result then use "body system"="Hematologic/Lymphatic" with "folder"="Screening" from raw.PEB.

Splenomegaly:

- Spleen in raw.LVR_SPL [Liver and Spleen Measurements]: if spleen is enlarged then "Yes". Baseline is defined per SAP using date of assessment.
- Spleen in raw.PETAR [Physical Examination – Targeted] or raw.PEB [Physical Examination]: if baseline spleen assessment is enlarged then "Yes". Baseline is defined as "folder" = "WK1D1" from raw.PETAR, if a patient has missing record or missing result or not enlarged result then use "folder"="Screening" from raw.PEB.

Other:

- Extranodal in raw.TLB [Target Lesions – Baseline]: if any "type of lesion" = "Extranodal" with LDi ("longest perpendicular diameter measurement") >1.0 (cm) then "Yes".

Otherwise, baseline EMD is "No".

5. DETERMINATION OF BASELINE LYMPHADENOPATHY, SPLENOMEGALY, AND HEPATOMEGALY FOR CLL/SLL

For CLL/SLL patients,

- Baseline Lymphadenopathy and Splenomegaly were defined the same as those defined above for WM.
- Baseline Hepatomegaly is defined as "Yes" if any of the below criteria is "Yes":
    a) Liver in raw.LVR_SPL [Liver and Spleen Measurements]: if liver is enlarged then "Yes". Baseline is defined per SAP using date of assessment.
    b) Liver in raw.PETAR [Physical Examination – Targeted] or raw.PEB [Physical Examination]: if baseline liver assessment is enlarged then "Yes". Baseline is defined as "folder" = "WK1D1" from raw.PETAR, if a patient has missing

record or missing result or not enlarged result then use “folder”=”Screening” from raw.PEB.

6. BONE MARROW INVOLVEMENT AT BASELINE

Use last bone marrow record on or before the first dose date (from the bone marrow eCRFs)

- Was there tumor involvement in the bone marrow? (Yes, No, Indeterminate, Not Available).
- If no bone marrow data on or before the first dose date, patient will be counted in Not Available.
- Only patients with this question checked as ‘Yes’ will be considered as with bone marrow involvement at baseline.

7. EXTRA-NODAL LESION AT BASELINE

- Yes, if any lesion in “Target Lesions - Baseline” or “Non-Target Lesions - Baseline” with lesion type = “Extranodal”, or bone marrow involvement = “Yes” at baseline per bone marrow biopsy/aspiration (see Section 6).
- No, otherwise.

8. REFRACTORY DISEASE

Refractory disease is derived from the last regimen of prior anticancer therapy,

- Yes, if best overall response =”SD” or “PD”.
- No, if best overall response = “CR” or “PR” (for CLL patients, also consider best overall response of “CRi”, “nPR” and “PR-L”).
- Unknown if best overall response is missing or not evaluable.

9. BEST RESPONSE OF LAST PRIOR REGIMEN

To determine the best response of last prior regimen,

- First, pick the maximum regimen number from clinical review spreadsheet of ‘prior anticancer therapies’.
- Then, select the best response among the last regimens. (CR> CRi> VGPR> nPR> PR> PR-L> MR> SD> PD>Not evaluable >Not done > Missing)

10. DETERMINATION OF ACTUAL DOSE

The actual dose group is determined as follows.

- Remove all records from raw.sda with non-missing enddate and with enddate-startdate<=2.
- Remove all records with dose missed/held, i.e., exreas1 ne “No” and exreas1 ne “”. Records with this info missing will be considered as no dose missed/held.
- Remove all records with missing intended or actual dose, i.e., exdose2 eq “” or sdaad eq “”.
- Remove all records with no actual dose, i.e., sdaad eq “0 mg”.

For each subject select the record with the earliest start date and use the corresponding value for the dose variable as the actual dose received.

11. DEFINITION OF RELATIVE DOSE INTENSITY

Below steps should be followed to calculate Relative Dose Intensity (RDI):

1. Determine the first dose date from raw.SDA as the earliest start date among records with total dose per day>0.
2. Determine the treatment end date as the end date of last dosing record from raw.SDA. If the end date of last dosing record is after the cutoff date, use the cutoff date. Partial dates are imputed as per Section 15.A.7. If this date is completely missing, use the minimum of
   - The end of treatment date from the treatment discontinuation page
   - The start date for the last dosing record with a missing end date + 30 days
   - The data cutoff date

   For patients who have discontinued study treatment, if the last dosing record has a total dose per day of zero, then determine the treatment end date as the end date of last non-zero dosing record from raw.SDA before cutoff date.
3. Actual total dose: To each day that is between first dose date and treatment end date, calculate the total daily dose. Sum over all days to obtain the total dose received. For cases where there are dose records that overlap in time assign the smallest non-missing total daily dose for the overlapping days. Note that there are some instances where the am and pm dose for the same day are recorded as separate entries. These entries need to be added together to obtain the total daily dose.
4. Planned total dose: To each day that is between first dose date and treatment end date, assign the total planned daily dose. This will be the same number for every day during this period. For Part 1 patients, it is the actual initial total daily dose determined in Section 10; for Part 2 patients, it is 320 mg (per day). Calculate the duration of treatment as (treatment end date -first dose date +1) and multiply by the total planned daily dose to obtain the total expected dose.

5. Divide the total dose received by the total expected dose and multiply by 100 to determine the relative dose intensity.

12. DOSE REDUCTION

Below steps should be followed to calculate number of patients with dose reductions:

1) Select records in raw.SDA where dose change reason ne “” (SDARRPD ne “”).
2) Among records in step 1, select:
   - For Part 2 patients: total daily dose < original assigned total daily dose (e.g., 160mg BID = 320mg).
   - For Part 1 patients: total daily dose < original assigned total daily dose (e.g., 160mg BID=320mg) or total daily dose < maximum total daily dose before this record.
3) Exclude records where dose missed/held reason ne “”.
4) Exclude records where total daily dose = 0mg or missing.
5) If dose change reason = “Other” (SDARRPD= “Other”), then only keep those with dose reduction duration >= 3 days (EXENDTC1-EXSTDTC1+1>=3). Note if dose change reason is adverse event or PI decision, then checking if dose reduction duration is >=3 days is not needed.

13. DOSE INTERRUPTION DUE TO AE

To calculate number of patients with dose interruptions due to AE, select records in raw.SDA where

1) sdarmcd = “Adverse Event” AND
2) total daily dose=0 or 0mg (raw.SDA. sdacalcdose) AND
3) duration > 1 day (stop date-start date+1 > 1 day; if start/stop date is partial then count it as duration > 1 day)

Note: Some patients had dose held due to same AE as multiple dosing records, those need be added up to decide whether it is greater than 1 day. i.e. Among records with raw.sda.sdarmcd=”Adverse Event”, look for records where Raw.sda.sdaaedsl2 is not equal to “ “ AND raw.sda.sdaaedsl2 are same AND total daily dose=0 or 0mg AND start date of the next dosing record is 1 day after the stop date of the previous dosing record (i.e. dates are consecutive).

14. UNKNOWN DATES

Sites used 1900/01/01 to denote an unknown date.

Sites used xxxx/01/01 to denote and unknown month for some records.

Sites used xxxx/xx/01/to denote an unknown day for some records.

If an end date is before a start date and year is missing for an end date, then impute the end date as follows:

- Start date and end date have the same year and the end date is of the form xxxx/01/01, then write the end date as xxxx/unk/un
- Start date and end date have the same year and month and the end date is of the form xxxx/xx/01 then write the end date as xxxx/xx/un
- End date has the form 1900/01/01 then write the end date as unkn/unk/un

15. MISSING DATE IMPUTATION RULE

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

A.1 Prior/Concomitant Medications/Procedures

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

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

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

If end date of a medication is 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 is completely missing, do not impute.

A.2 Adverse Events

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

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

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

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

- If both month and day are missing, then set to December 31
- If only day is missing, then set to last day of the month

If end date is completely missing, do not impute.

A.3 Deaths

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

- If both month and day are missing, then the imputed month and day will be January 01 or the last date of the last date of subject 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.

A.4 New Anti-cancer 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 (refer to #2 derivation of investigator response date, use the derived last tumor assessment where the response is not NA, NE or not done) if they have the same year.
- If only day is missing, then the imputed day will be the first day of the month.

A.5 Diagnosis

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

A.6 Prior Therapy/Response to Prior Therapy

If a prior therapy or response to prior therapy date is partially missing, impute as follows:

- If only day is missing, then set to the 15th of the month

No imputation will be performed for all other types of missing dates.

A.7 Dose Stop Date

Sort all dose records by start and stop date and impute the partial missing dose stop date as follows:

- If dose stop date is not the last dosing record, then impute the stop date as start date of next dosing record -1. (used when calculating the total dose received)
- If dose stop date is the last dosing record and partially missing
    - If only day is missing, impute the stop date as the earliest of last day of the month, end of treatment date (if end of treatment date is complete), end of study date, death date and data cutoff date.

16. NORMAL RANGE FOR NEUTROPHIL COUNT AND LYMPHOCYTE COUNT DERIVED FROM %

When deriving the absolute neutrophil count from a measure reported as %, the percent value is multiplied by the total WBC count. The standard normal range from Mosby's Manual of Diagnostic and Laboratory Tests will be used to evaluate the toxicity grade of such neutrophil values which is 2.5-8.0 x 10^9/L. The same method will be used for the absolute lymphocyte count whose standard normal range is 1.0-4.0 x 10^9/L.