BeiGene

STATISTICAL ANALYSIS PLAN

Study Protocol Number:	BGB-3111-210
Study Protocol Title:	A Phase 2, Single-Arm, Open-Label, Multicenter Study of Bruton's Tyrosine Kinase (BTK) Inhibitor BGB-3111 in Chinese Subjects with Relapsed/ Refractory Waldenström's Macroglobulinemia (WM)
Date:	15-May-2019
Version:	2.0

SIGNATURE PAGE

Author:

[redacted]
[redacted]
BeiGene (Beijing) Co., Ltd.

[redacted]

Approval

[redacted]
[redacted]
BeiGene USA, Inc.

[redacted]

[redacted]
[redacted]
BeiGene USA, Inc.

[redacted]

TABLE OF CONTENTS

LIST OF ABBREVIATIONS AND DEFINITIONS OF TERMS

Abbreviation	Term
AEs	adverse events
AESI	AE of special interest
ALT	alanine aminotransferase
ALP	alkaline phosphatase
aPTT	Activated Partial Thromboplastin Time
AST	aspartate aminotransferase
ATC	Anatomic Therapeutic Chemical
BID	twice a day
BOR	best overall response
BTK	Bruton's tyrosine kinase
CI	confidence interval
CR	complete response
CT	computed tomography
CTCAE	Common Terminology Criteria for Adverse Events
DOMR	duration of major response
DOR	duration of response
DOV	date of visit
EAIR	exposure-adjusted incidence rate
ECG	electrocardiogram
ECOG	Eastern Cooperative Oncology Group
eCRF	electronic case report form
EOT	end of treatment
FDA	Food and Drug Administration
HBcAb	hepatitis B core antibody
HBsAb	hepatitis B surface antibody
HCV	hepatitis C virus
INR	international normalized ratio
IPSS	International Prognostic Scoring System
IRC	independent review committee
IWWM	International Workshop for Waldenström's Macroglobulinemia
MedDRA	Medical Dictionary for Regulatory Activities

MRR	major response rate
NDA	new drug application
ORR	overall response rate
OS	overall survival
PD	progressive disease
PFS	progression-free survival
PP	per-protocol
PR	partial response
PT	preferred term
R/R	relapsed or refractory
SAEs	serious adverse events
SAP	statistical analysis plan
SMC	safety monitoring committee
SOC	system organ class
SPD	sum of products of diameters
TBL	total bilirubin
TEAEs	treatment emergent adverse events
TTR	time to response
ULN	upper limit of normal
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 data analysis in evaluation of safety and efficacy of zanubrutinib (BGB-3111) for BGB-3111-210. This document is based on the protocol version 3.0 dated 30-Mar-2018.

The analysis details for biomarker analyses are not described within this SAP. Any changes made to the planned analyses that are in the protocol will be identified and documented in Clinical Study Report.

The SAP version 2.0 is an amendment to the initial version of the SAP (version1.0 dated Sep-19-2018).

2 STUDY OVERVIEW

This is a single-arm, open-label, multi-center Phase 2 study in patients with Waldenström's macroglobulinemia (WM) requiring therapy using the consensus panel criteria updated at the 6th International Workshop for Waldenström's Macroglobulinemia (IWWM) (Owen et al 2013; NCCN Guidance Insights, 2012). The study is composed of an initial screening phase (up to 28 days), a single-arm treatment phase, and a follow-up phase.

Approximately 40 patients will be enrolled. The primary efficacy analysis will be conducted at up to 12 months after the last patient receives the first dose of study drug. Tumor response will be assessed by independent review committee (IRC) according to the response criteria updated at the 6th IWWM (Owen et al 2013; NCCN Guidance Insights, 2012). IgM level will be measured at screening, on Day 1 of every cycle for the first 52 weeks, and every 3 cycles thereafter. Assessment by computed tomography (CT) scan will occur every 12 weeks for the first 48 weeks, and then every 24 weeks until disease progression. Bone marrow will be assessed by aspirate and biopsy at screening, every 24 weeks, at time of suspected complete response (CR) and as clinically indicated.

All patients will be followed for adverse events（AEs) for 30 ± 7 additional days after the last dose of study drug. All treatment-related AEs and serious adverse events (SAEs) will be followed until resolution or stabilization.

Screening phase: Screening evaluations will be performed within 28 days prior to the first dose of study drug. Patients will sign the informed consent form prior to any screening evaluations. Please refer to Table 3 in the protocol for details on screening procedures. Screening evaluations can be repeated within the screening period.

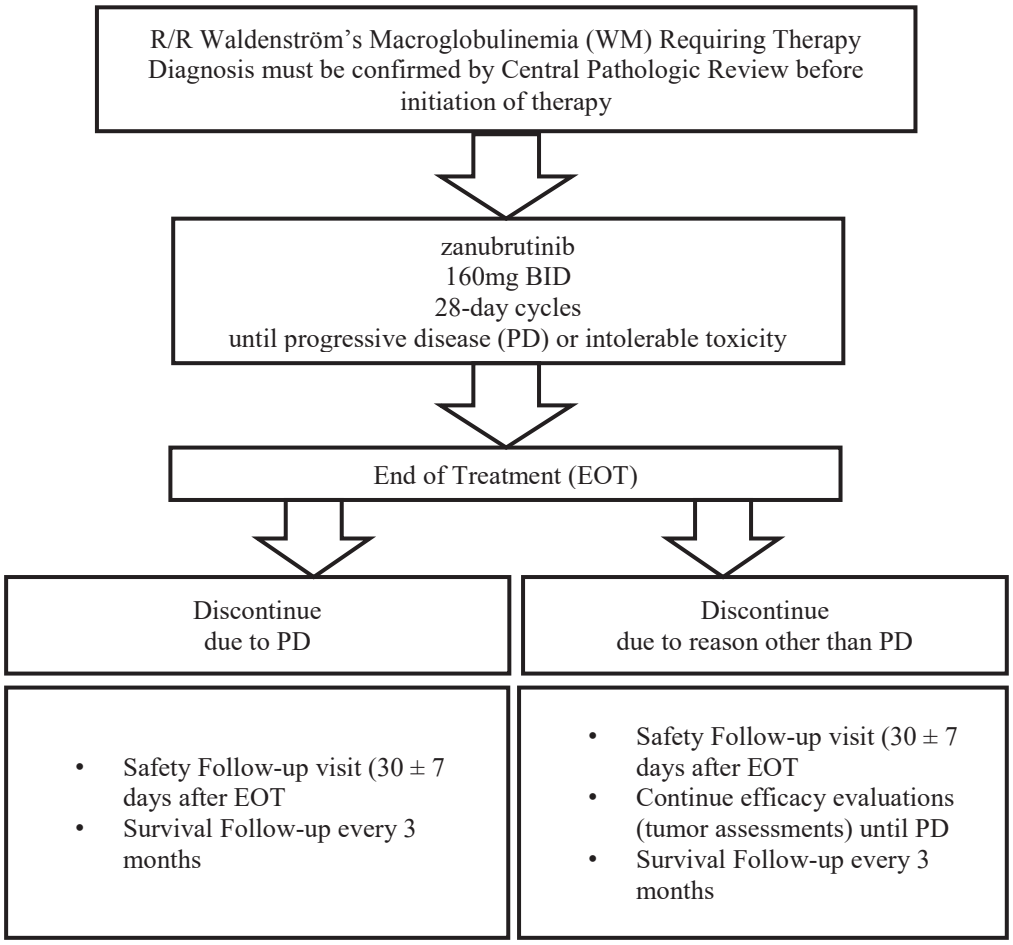
Treatment phase: Patients will receive the first dose of zanubrutinib at Cycle 1 Day 1. All patients will be treated with 160 mg, administered orally, twice a day (BID), about 12± 2 hours apart, and will continue to be treated until disease progression, unacceptable toxicity, death,

withdrawal of consent, or the study is terminated by the sponsor for final analysis. A treatment cycle is 28 days.

Follow-up phase: Patients will return 30 ± 7 days after the last dose of study drug for safety follow-up visit(s). Assessments to be performed are presented in Table 3 in the protocol. Efficacy evaluations will continue until documented disease progression. If a patient discontinues study drug due to reasons other than disease progression, efficacy evaluations will continue until patient exhibits first progression, withdrawal of consent, death, lost to follow-up, or study termination by sponsor, whichever occurs first.

Survival Follow-up phase: After attending the last visit, the patient/guardian will be contacted by telephone every 3 months for follow up, and this will continue up until the patient withdraws informed consent, lost to follow up, death, or final data cut off.

Figure 1 Schema for Study BGB-3111-210



3 STUDY OBJECTIVES

3.1 PRIMARY OBJECTIVES

The primary objective of the study is to determine the efficacy of zanubrutinib in Chinese patients with relapsed or refractory (R/R) WM as measured by the major response rate (MRR) defined as the proportion of patients who achieves complete response (CR) + very good partial response (VGPR) + partial response (PR), to be assessed by an IRC according to an adaptation of the response criteria updated at the 6th IWWM (Owen et al 2013; NCCN Guidance Insights, 2012).

3.2 SECONDARY OBJECTIVES

- To determine the efficacy of zanubrutinib in patients with R/R WM as measured by progression-free survival (PFS)

- To determine the efficacy of zanubrutinibin patients with R/R WM as measured by overall response rate (ORR)
- To determine the efficacy of zanubrutinib in patients with R/R WM as measured by duration of major response（DOMR)
- To characterize the clinical benefit of zanubrutinib in patients with WM as determined by the frequency of resolution of treatment-precipitating symptoms
- To characterize the magnitude of improvement in bone marrow involvement with lymphoplasmacytoid lymphocytes in patients with R/R WM treated with zanubrutinib, as measured by the maximum decrease in percentage of lymphoplasmacytoid lymphocytes by bone marrow biopsy
- To determine the safety and tolerability of zanubrutinib in patients with R/R WM as determined by the frequency and severity of AEs according to Common Terminology Criteria for Adverse Events (CTCAE) v4.03, and the rate of discontinuation of treatment, dose reduction and dose interruption of study drug due to AEs.

3.3 EXPLORATORY OBJECTIVES

- To evaluate overall survival (OS) in patients with WM
- To determine the CR plus VGPR rates in WM patients with different genotype (MYD88 L265P mutation, MYD88 L265P mutation/CXCR4 WHIM, MYD88 L265P mutation/CXCR4 wild-type, MYD88 wild-type).
- To evaluate drug resistance mechanisms through comparison of bone marrow aspiration at screening, and on relapse

4 STUDY ENDPOINTS

Response of WM patients will have 2 types of assessment by IRC and 1 type of assessment by investigator, respectively:

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

All efficacy endpoints will be analyzed in the revised safety analysis set as defined in Section 6.1.

The differences between these 2 types of assessment by IRC are summarized in Appendix B.

4.1 PRIMARY ENDPOINT

The primary endpoint of the study is the MRR, defined as the proportion of patients who achieve CR + VGPR + PR, to be assessed by an IRC according to an adaptation of the response criteria updated at the 6th IWWM (Owen et al 2013; NCCN Guidance Insights, 2012) (hereafter referred as overall combined assessment by IRC).

For regulatory submission, response assessment based on IgM alone (i.e. overall IgM assessment by IRC) may be used in the primary analysis, depending on the discussion with regulatory agency in pre-NDA meeting.

4.2 SECONDARY ENDPOINTS

Efficacy:

- PFS: defined as time from first dose of zanubrutinib until first documentation of progression or death, whichever comes first.
  - Per overall combined assessment by IRC
  - Per overall IgM assessment by IRC
  - Per overall combined assessment by investigator
- ORR: defined as the proportion of patients who achieve minor, partial, very good partial, and complete response.
  - Per overall combined assessment by IRC
  - Per overall IgM assessment by IRC
  - Per overall combined assessment by investigator
- MRR: defined as the proportion of patients who achieve PR, VGPR, or CR
  - Per overall IgM assessment by IRC
  - Per overall combined assessment by investigator
- Rate of VGPR or CR: defined as the proportion of patients who achieve VGPR or CR
  - Per overall combined assessment by IRC
  - Per overall IgM assessment by IRC
- Duration of response (DOR): defined as the time from the date of the earliest qualifying response to the date of PD or death (whichever occurs earlier).
  - Duration of overall response
    - Per overall IgM assessment by IRC
  - Duration of major response
    - Per overall combined assessment by IRC

  - Per overall IgM assessment by IRC
  - Per overall combined assessment by investigator
- Duration of VGPR or CR
  - Per overall IgM assessment by IRC

- Time to response (TTR): defined as time from the first dose of the study drug to the date of the earliest qualifying response.
  - Time to overall response
    - Per overall IgM assessment by IRC
  - Time to major response
    - Per overall combined assessment by IRC
    - Per overall IgM assessment by IRC
    - Per overall combined assessment by investigator
  - Time to VGPR or CR
    - Per overall IgM assessment by IRC
- Resolution of treatment precipitating symptoms, defined as absence of symptoms (see Table 2)
- Anti-lymphoma effect is defined as any reduction in bone marrow involvement by lymphoplasmacytoid lymphocytes and/or size of lymphadenopathy and/or splenomegaly by CT scan by IRC. Lymphadenopathy is defined as any node with longest diameter (LDi) > 1.5 cm and splenomegaly is defined as vertical spleen length > 13 cm.

Safety:

To evaluate the safety and tolerability of zanubrutinib as defined by:

- The incidence and severity of treatment emergent adverse events (TEAEs), SAEs and treatment-related AEs according to CTCAE v4.03
- The incidence, severity, and causation of AEs leading to study drug discontinuation, dose reduction and dose interruption

4.3 EXPLORATORY ENDPOINTS

- OS defined as the time from the date of the first dose of zanubrutinib until date of death from any cause
- CR plus VGPR rates in 4 subgroups: patients with MYD88 L265P mutation, MYD88 L265P mutation/CXCR4 WHIM, MYD88 L265P mutation/CXCR4 wild-type, MYD88 wild-type.

- Identification of potential drug resistance biomarkers and mechanisms: paired bone marrow aspiration (at screening and at relapse) will be used to identify potential biomarkers and mechanisms.
- Rate of VGPR or CR
    - overall combined assessment by investigator
- Duration of overall response and time to overall response
    - overall combined assessment by IRC
    - overall combined assessment by investigator
- Duration of VGPR or CR and time to VGPR or CR
    - per overall combined assessment by IRC
    - per overall combined assessment by investigator
- Change of IgM level from baseline
- Change of lymph node size from baseline
- Change of spleen size from baseline
- Change of hemoglobin from baseline

5 SAMPLE SIZE CONSIDERATIONS

The sample size calculation of 40 patients was based on the precision of an MRR estimate and the power of the comparison to the historical rate, under assumed MRR of 60% in the study as compared to 30% in the historical control (95% CI is 43.3%, 75.1%). Using a binomial exact test, the power is >0.969 with 40 patients to demonstrate statistical significance at a 1-sided alpha of 0.025 under above assumption.

If the observed MRR is 60%, its 95% exact CI from 40 patients will be (43.3%, 75.1%).

6 STATISTICAL METHODS

6.1 ANALYSIS SETS

The safety analysis set (SA) includes all patients who received at least one dose of zanubrutinib. It will be the primary analysis set for the safety analyses.

The revised safety analysis set includes patients with pathologically confirmed WM and with baseline IgM (or M-protein) ≥ 5 g/L among those in the safety analysis set. The analysis set is the primary efficacy evaluable analysis set.

The Per-protocol (PP) Analysis Set includes patients in the revised safety set without important protocol deviations. Criteria for exclusion from the PP analysis set will be determined and documented before the database lock for the primary analysis. This analysis set may be used as a sensitivity analysis of the efficacy endpoints as necessary.

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

6.2 DATA ANALYSIS GENERAL CONSIDERATIONS

6.2.1 Definitions and Computations

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

Study day: Study day will be calculated in reference to the date of the first dose of study drug. For assessments conducted on or after the date of the first dose of study drug, study day will be calculated as (assessment date – date of 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 first dose of study drug). 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 Appendix A.

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

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

All calculations and analyses will be conducted using SAS version 9.4 or higher.

6.2.2 Conventions

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

- 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'.
- Missing efficacy or safety data will not be imputed unless otherwise specified.
- For by-visit observed data analyses, percentages will be calculated based on the number of patients with non-missing data as the denominator, unless otherwise specified.
- For continuous endpoints, summary statistics will include n, mean, standard deviation, median, Q1, Q3 and range (minimum and maximum). The minimum and maximum will be reported to the same number of decimal places as the raw data recorded in the database. The mean, median, Q1 and Q3 will be reported to one more decimal place than the raw data recorded in the database. The standard deviation will be reported to two more decimal places than the raw data recorded in the database. In general, the maximum number of decimal places reported shall be four for any summary statistic.

- For discrete endpoints, summary statistics will include frequencies and percentages. Percentages will be presented to one decimal place.
- Data collected at an unscheduled visit will not be included in by visit analysis. For by visit analysis, if multiple assessments happened in the same visit window in table 3 of the protocol, the closest assessment to the planned date/time (if collected) will be used in analysis. If multiple records have the same distance to the planned date/time, the last non-missing assessment will be used.

6.2.3 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 prior/concomitant medications/procedures, subsequent anti-cancer therapies, adverse events and deaths as provided in Appendix A 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.

By-visit summary of variables with missing data will use only non-missing data, not imputed one, unless otherwise specified. Unscheduled visits will not be included in by-visit summaries.

6.2.4 Adjustment for Covariates

Not applicable.

6.2.5 Multiplicity Adjustment

Not applicable.

6.2.6 Data Integrity

Before any 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. 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 for all enrolled patients:

- Number of patients enrolled (screened and inclusion/exclusion criteria met)
- Number of patients enrolled but not treated
- Number of treated patients
- Number (%) of treated patients who discontinued treatment
- Reason(s) for treatment discontinuation
- Number (%) of treated patients who discontinued study
- Reason(s) for study discontinuation

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 (whichever occurs earlier) for patients discontinued from study, or the database cutoff date for ongoing patients. Study follow-up time will be estimated by median and range.

6.3.2 Protocol Deviations

The CRO (Parexel) will assess potential protocol deviations as potential minor or potential major protocol deviations according to the protocol deviation specification in the study and review them with the sponsor (BeiGene) medical monitor and statistician to determine the major protocol deviations. The final determination of important protocol deviations from major protocol deviations will be made by BeiGene team. Important protocol deviations will be summarized by deviation category for the safety analysis set. Both major protocol deviations and important protocol deviation will be listed.

6.3.3 Demographic and Other Baseline Characteristics

Demographic and other baseline characteristics will be summarized in the safety analysis set using descriptive statistics. Continuous variables include age, weight, height, body mass index, categorical variables include gender, race, age group (< 65 years vs. ≥ 65 years), ECOG-performance status, viral serology (including HBcAb, HBsAb and HCV antibody). A listing of demographic and other baseline characteristics will be provided.

6.3.4 Disease History and Characteristics

Disease history and baseline disease characteristic, as recorded on the eCRF, will be summarized by descriptive statistics for the safety analysis set. Disease characteristics include time since initial diagnosis of WM to first dose, prognostic group at initial diagnosis and study entry, WM International Prognostic Scoring System (IPSS) score at study entry (low, intermediate, high),

Genotype status for MYD88 (L265P mutation, wild type), CXCR4 (WHIM mutation, wild-type), extramedullary disease (yes vs no), time from last disease progression to first dose, indications for treatment (see Table 2), cytopenia, baseline bone marrow lymphoplasmacytic cell involvement (lower limit ≥ 50% vs. higher limit < 50%), serum β2 microglobulin (≤ 3 mg/L vs. > 3 mg/L), hemoglobin, absolute neutrophil count, serum immunoglobulin (IgM, IgA and IgG), SPEP, serum immunofixation. A listing of disease history will be provided.

6.3.5 Prior Anti-Cancer Therapies

The number of prior systemic regimens, prior systemic therapies, best response to last systemic regimen, duration of last regimen (i.e. most recent therapy), time (months) from the end of last systemic regimen to first dose of study drug, number (%) of patients with prior radiotherapy will be summarized and listed for the safety analysis set.

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

6.3.6 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 are 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 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 Appendix A will be used.

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

6.3.7 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 eCRF, will be summarized by MedDRA system organ class (SOC) and preferred term (PT) based on the safety analysis set. A listing of medical history will be provided.

6.4 EFFICACY ANALYSIS

All efficacy analyses will be based on the revised safety analysis set.

6.4.1 Primary Efficacy Endpoint

Major response rate per overall combined assessment by IRC

The primary endpoint of the study is MRR, defined as the proportion of patients who achieved Best Overall Response (BOR) of CR or VGPR or PR assessed by an IRC according to an adaptation of the response criteria updated at the 6th IWWM (Owen et al 2013; NCCN Guidance Insights 2012).

A point estimate and a 2-sided Clopper-Pearson 95% confident interval (CI) of MRR will be provided. A binomial exact test will be performed to test against the null hypothesis H_0: MRR=0.30 using the significant level of 0.025 (1-sided).

A patient's best overall response (BOR) is the best response recorded throughout the study (prior to data cutoff). 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 (due to any reason) will be considered as non-responders. The number and proportion of patients who achieved each BOR category will be calculated.

The primary efficacy analysis will be conducted at up to 12 months after the last patient receives the first dose of study drug. Subsequent analyses will be performed when mature secondary efficacy endpoints are available.

For regulatory submission, response assessment based on IgM alone (i.e. overall IgM assessment by IRC) may be used as the primary analysis, depending on the discussion with regulatory agency in the pre-NDA meeting.

6.4.2 Secondary Efficacy Endpoints

6.4.2.1 Response rate

The ORR is defined as the proportion of patients whose BOR meet the overall response.

The VGPR or CR rate is defined as the proportion of patients whose BOR meet the VGPR or CR.

A point estimate and a 2-sided Clopper-Pearson 95% CI will be provided for ORR and VGPR or CR, respectively.

6.4.2.2 Progression Free Survival (PFS)

PFS is defined as the time (in months) from the first dose date of study drug to the date of first documentation of disease progression or death of any cause, whichever occurs first:

PFS = (The earlier of disease progression or death date – the date of first study dose +1)/ 30.4375

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 1 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) (FDA 2007).

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

Table 1 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 visits	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 visit	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; Klein and Moeschberger 1997), whereas the 95% confidence interval for PFS rate at landmark times will be generated by using Greenwood formula (Greenwood 1926; Kalbfleisch and Prentice1980). 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 censor, and reason.

6.4.2.3 Duration of response

Duration of response (DOR) for responders (those who satisfy the criteria for overall response, major response, VGPR or CR, 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 the same as those for PFS. Only responders are included in the analysis.

6.4.2.4 Time to response

TTR for responders (those who satisfy the criteria for overall response, major response, VGPR or CR, respectively) 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 mean, median and range for responders only.

6.4.2.5 Resolution of treatment precipitating symptoms

The number (%) of patients who achieve resolution of baseline signs or symptoms that precipitated the need for treatment (see Table 2) will be summarized.

Table 2 Indications for initiation of therapy in patients with WM[a]

Clinical indications for initiation of therapy
Recurrent fever, night sweats, weight loss
Fatigue
Hyperviscosity
Lymphadenopathy which is either symptomatic or bulky (≥5 cm in maximum diameter)
Symptomatic hepatomegaly and/or splenomegaly
Symptomatic organomegaly and/or organ or tissue infiltration
Peripheral neuropathy due to WM
Laboratory indications for initiation of therapy
Symptomatic cryoglobulinemia
Cold agglutinin anemia
Immunehemolytic anemia and/or thrombocytopenia
Nephropathy related to WM
Amyloidosis related to WM
Hemoglobin ⩽10 g/dL
Platelet count ≤100 × 10^9/L

[a] Seventh IWWM (Dimopoulos et al, 2014)

6.4.2.6 Anti-lymphoma effect

Any reduction in bone marrow involvement by lymphoplasmacytoid lymphocytes and/or any reduction in size of lymphadenopathy and/or any reduction in size of splenomegaly by CT scan will be summarized descriptively with incidence and percentage. Lymphadenopathy is defined as any node with longest diameter (LDi) > 1.5 cm and splenomegaly is defined as vertical spleen length > 13 cm.

6.4.3 Exploratory Efficacy Endpoints

6.4.3.1 Overall Survival

Overall survival (OS) is defined as the time (in months) from the date of 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.

6.4.3.2 VGPR or CR rate

VGPR or CR rate will be determined in patients with MYD88 L265P mutation, MYD88 L265P mutation/CXCR4 WHIM, MYD88 L265P mutation/CXCR4 wild-type, MYD88 wild-type as the proportion of patients who achieve CR or VGPR. The analysis methods for VGPR or CR rate will be same as those for the rate in Section 6.4.2.1.

6.4.3.3 Changes of lymph node size from baseline

Change of lymph node size from baseline will be summarized for patients with baseline lymphadenopathy by maximum reduction in sum of products of diameters (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.

6.4.3.4 Change of spleen size from baseline

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.

6.4.3.5 Change of IgM level from baseline

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 also be 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.

6.4.3.6 Change of hemoglobin from Baseline

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 revised safety analysis 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.

6.4.3.7 Other exploratory endpoints

Rate of VGPR or CR per overall combined assessment by investigator as mentioned in Section 4.3 will be analyzed using same method in Section 6.4.2.1.

Duration of overall response and duration of VGPR or CR as mentioned in Section 4.3 will be analyzed using same method in Section 6.4.2.3.

Time to overall response and time to VGPR or CR as mentioned in Section 4.3 will be analyzed using same method in Section 6.4.2.4.

6.4.4 Sensitivity Analyses

Sensitivity analyses may be performed by repeating analysis of selected efficacy endpoints in the PP analysis set as necessary.

6.4.5 Subgroup Analyses

Primary and selected secondary efficacy endpoints will be summarized in the subgroups, including below, as appropriate (ie. when there is sufficient number of patients in the subgroup, otherwise relevant subgroups may be combined): gender, age group (< 65 years vs. ≥ 65 years), ECOG-performance status (0 vs. ≥ 1), MYD88/CXCR4 mutation status (MYD88 L265P mutation, MYD88 L265P mutation/CXCR4 WHIM, MYD88 L265P mutation/CXCR4 wild-type, MYD88 wild-type/CXCR4 wild-type, MYD88 wild-type), baseline bone marrow involvement (lower limit ≥ 50% vs. higher limit < 50% vs. other), WM IPSS score (low, intermediate, high), baseline serum β2 microglobulin level (≤ 3 mg/L vs. > 3 mg/L), number of prior systemic regimens (1-2 vs >=2), baseline IgM level (<40 g/L vs >=40 g/L), baseline platelet count (≤ 100 x 10^9/L vs > 100 x 10^9/L), baseline hemoglobin (≤ 110 g/L vs >110 g/L), and baseline extramedullary disease (yes vs no) identified by either CT scan or physical exam per investigator. Subgroup results will be presented in forest plots also. The subgroup variables and their respective categories are subject to change if warranted to better represent the data.

6.4.6 Biomarker/Pharmacodynamic Endpoints

Numbers (%) of patients with MYD88 and/or CXCR4 mutation will be presented.

Correlation analysis between the mutation status and response status might be performed. An independent biomarker SAP will be generated to describe the detailed analyses for these endpoints.

6.5 SAFETY ANALYSES

All safety analyses will be based on the safety analysis set. The study will set up a Safety Monitoring Committee (SMC). The SMC will monitor safety data according to the SMC charter (Appendix C) throughout the study.

6.5.1 Extent of Exposure

Extent of exposure to the study drug will be summarized descriptively with respect to the following:

- Number of treatment cycles received: defined as duration of exposure (days)/28.
- Duration of exposure (days): defined as the duration (in days) from the date of the first dose to the last dose of the study drug for patients who discontinued treatment, or the database cutoff date for ongoing patients.

  Categorical distribution of duration of exposure will also be provided.
- Total dose received (g): defined as the cumulative dose of the study drug during the treatment period of the study.

  Note: If the actual dose received during a period is unknown (e.g. diary not returned, and no other information can be used), it will be considered as the study drug was administrated as prescribed conservatively.
- Actual dose intensity (mg/day): defined as the total dose (in mg) received by a patient divided by the duration of treatment (in day).

  Note: For patients who are ongoing at the data cutoff, the actual dose intensity is calculated based on study drug administration data prior to data cutoff as the total dose (in mg) received by a patient prior to the cutoff divided by the duration from first dose date to the end date of the last study drug administration record.
- Relative dose intensity: defined as the ratio of the actual dose intensity (mg/day) and the planned dose intensity in percentage. Planned dose intensity is 320 mg/day.

The number (%) of patients requiring dose reductions, dose interruption, dose missed, and drug discontinuation due to AEs will be summarized. The cycle in which the first dose reduction/interruption occurred will be summarized using descriptive statistics. Frequency of dose reductions and interruptions as well as reason for dose reductions and interruptions will be summarized descriptively.

The cycles used in the dose exposure analysis will be derived from the date of visit (DOV) if the cycle information is not available. The start date of each cycle will be set as the DOV for that cycle, and the end date will be one day before the DOV for next cycle. When DOV for certain cycle is not available (patient missed certain cycle visit), cycle duration before and after the

missing visit will be assigned evenly. For example, if for one patient, cycle 2 visit is missing, and the duration between cycle 1 and cycle 3 (C3D1-C1D1+1) is 58 days, the duration for cycle 1 and cycle 2 will be assigned both as 29 days, with imputed cycle 2 DOV as C1D1+29-1.

Dose reduction is defined as any planned dose reduced from the original planned dose (160mg BID), regardless of reasons. Dose interruption is defined as any dose temporary discontinuation due to AE or held for procedure. Dose missing is defined as any dose temporary discontinuation due to reasons other than AE or held for procedure. Since dose interruption also need to be analyzed by cycle, for patients with dose interruption but with missing start/stop date, the end date will be imputed with the corresponding AE stop date, the dose interruption start date will be back calculated based on the dosage missing reported. If the dosage is also missing, the start and stop date of the dose interruption will be imputed with the AE start/stop date.

6.5.2 Adverse Events

AEs will be graded by the investigators using CTCAE v4.03. 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 adverse event (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. Only those AEs that were treatment emergent will be included in summary tables. All AEs, treatment emergent or otherwise, will be presented in patient data listings.

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

The incidence of TEAEs will be summarized based the number (%) of patients with TEAEs by SOC, PT and grade. A patient will be counted only once by the highest severity grade according to CTCAE v.4.03 and strongest causal relationship to zanubrutinib 1) within a given SOC and 2) within a given SOC and PT, even if the patient experienced more than one TEAE within a specific SOC and PT.

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 TEAEs that led to death
- With at least one treatment-related TEAE that led to death

- With at least one serious TEAE
- With at least one treatment-related serious TEAE
- With at least one TEAE that led to treatment discontinuation
- With at least one treatment-related TEAE that led to treatment discontinuation
- With at least one TEAE that led to dose modification
- With at least one treatment-related TEAE that led to dose modification
- With at least one TEAE that led to dose reduction
- With at least one treatment-related TEAE that led to dose reduction
- With at least one TEAE that led to treatment interruption
- With at least one treatment-related TEAE that led to treatment interruption
- With at least one TEAE of special interest
- With at least one treatment-related TEAE of special interest

The number (%) of patients with TEAE, serious TEAE, TEAEs with grade 3 or above, TEAEs that led to death, and TEAEs that led to treatment discontinuation, dose modification (reduction or interruption), treatment-related TEAEs, treatment-related TEAE with grade 3 or above and treatment-related serious TEAE, treatment-related TEAEs that led to treatment discontinuation, dose modification (reduction or interruption) and death will be summarized by SOC, PT and grade.

Summaries of all TEAEs, SAE, grade 3 or higher TEAEs will also be summarized by PT in descending order.

AE of special interest (AESI) will be defined and summarized by AESI category name and PT for all AESI, treatment related AESI, AESI of grade 3 or higher, serious AESI, AESI leading to treatment discontinuation, dose reduction/treatment interruption, and death. Besides, AESI will be summarized by category name, PT and grade. The AESI categories and detailed search criteria will be provided separately.
Same summaries will be provided for other adverse events by organ system or syndrome if necessary.

Exposure-Adjusted Incidence Rates (EAIR)

Exposure-adjusted incidence rate (EAIR) by category will also be calculated for AESIs. In calculating EAIR, 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 independent 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 first event has occurred, or the treatment has ended, whichever is earlier (non-censored data), or 2) by the total duration of exposure from first dose date to last dose date or the data cutoff date (whichever is earlier) in case the patient does not experience the event (censored data). Depending on whether a patient has an event or not, the duration of exposure enters the denominator in its non-censored

or censored form, respectively. The EAIR per event considers the first event per patient only, and the corresponding exposure time in the denominator:

$$EAIR_{event} = \frac{\sum_{i=1}^{n} AESI_{event,i}}{\sum_{i=1}^{n} t_{event,i}}$$

Whereby $AESI_{event,i}$ represents if patient i experienced the event (1) or not (0), and $t_{event,i}$ as time when the first AESI occurs or the treatment ends, whichever is earlier (non-censored data) or total duration of treatment if no event occurs (censored data).

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. Other exploratory analysis will be performed if needed.

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.

Patient data listings of deaths, all AEs, SAEs, treatment-related AEs, grade 3 or above AEs, AEs that led to death, AEs that led to dose modification (reduction/interruption) and AEs that led to treatment discontinuation will be provided.

6.5.3 Laboratory Values

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

Laboratory parameters that are graded in CTCAE (v.4.03) will be summarized by shifts table assessing the CTCAE toxicity grades at baseline versus worst toxicity recorded post-baseline. 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.

Patient data listings of following selected (not limited to) laboratory parameters will be provided. Similar listing with only grade 3 and 4 lab assessments will also be provided.

- Hematology laboratory assessments: Hemoglobin, Platelets, Leukocytes, Neutrophils, Lymphocytes.
- Coagulation profile: Activated Partial Thromboplastin Time (aPTT), Intl. Normalized Ratio (INR).
- Chemistry laboratory assessments: Alkaline Phosphatase (ALP), Alanine Aminotransferase (ALT), Aspartate Aminotransferase (AST), Albumin, Total Bilirubin (TBL), Calcium, Creatinine, Glucose, Potassium, Sodium, Magnesium, Phosphate, Urate.

- Urine laboratory assessments: Urine pH, Urine Glucose, Urine Protein, Urine Ketones, Urine Occult Blood, Urine Specific Gravity.

IgA, IgG and their changes from baseline will be summarized by visit.

For hypocalcemia and hypercalcemia, serum calcium will be corrected using the following 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.

A summary of the number (%) of patients with grade 3 or higher toxicity will be provided for selected laboratory parameter of interest. Box and whiskers plots will be generated for parameters of interest.

Hy's Law criteria is defined with ALT or AST > 3xULN and TBL > 2xULN and ALP < 2xULN; TBL and ALP were both within 28 days after ALT or AST elevation. Incidence of patients who met one or more of the Hy's law criteria will be summarized. A listing of patients that met one or more of the Hy's law criteria will be generated.

6.5.4 Vital Signs

Descriptive statistics for vital sign parameters (systolic and diastolic blood pressure, heart rate, respiratory rate, temperature, weight) and changes from baseline will be presented at each scheduled visit. Vital signs will be listed by patients and visits. A shift table assessing the toxicity grade at baseline versus worst toxicity recorded post-baseline will also be presented for systolic and diastolic blood pressure. Box and whiskers plots will be generated for actual value and change from baseline for systolic and diastolic blood pressure.

6.5.5 Electrocardiograms (ECG)

ECG assessments will be performed at Screening and end of treatment (EOT) visit. Patient listing for ECG data will be provided.

6.5.6 ECOG

A shift table assessing the Eastern Cooperative Oncology Group (ECOG) performance status at baseline versus worst performance status post-baseline will be presented. ECOG performance status will be summarized and listed by visit.

6.6 OTHER ANALYSES

Additional 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.

7 INTERIM ANALYSIS

There will be one interim safety analysis based on an earlier data cutoff at 02-DEC-2018, which is about 7 months after last patient first dosed, for the purpose of regulatory filing. The trial will be continued regardless of the interim analysis result and therefore no multiplicity adjustment will be needed.

8 CHANGES IN THE PLANNED ANALYSIS

The following changes to the planned analysis were made:

- The revised safety analysis set (ie, the analysis set used for efficacy) is updated to exclude the patients with baseline IgM (or M-protein) <5 g/L.
- The per-protocol analysis set is updated to include patients in the revised safety analysis set without important protocol deviations.
- For anti-lymphoma effect, "hepato-" is removed from "hepatosplenomegaly" as the adaptation of the response criteria per the Sixth IWWM consensus guideline (Owen et al 2013) does not include evaluation for hepatomegaly.
- Some efficacy endpoints which were not defined in the protocol are added to the SAP.

9 REFERENCES

1. Brookmeyer, R. and Crowley, J. (1982), A confidence interval for the median survival time, Biometrics, 38: 29-41.
2. Common Toxicity Criteria Version 4.03. (2010), Cancer Therapy Evaluation Program. 14 June 2010.
3. Dimopoulos MA, Kastritis E, Owen RG, et al. (2014), Treatment recommendations for patients with Waldenstrom macroglobulinemia (WM) and related disorders: IWWM-7 consensus, Blood, 124:1404-11.
4. Food and Drug Administration Center for Drug Evaluation Research CDER and Center for Biologics Evaluation and Research (2007), FDA Guidance for Industry Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics.
5. Greenwood, M. (1926) The natural duration of cancer. Public health and medical patients, 33: 126.
6. Kalbfleisch, J. D. and Prentice, R. L. (1980), The Statistical Analysis of Failure Time Data, New York: John Wiley & Sons.
7. Klein, J. P. and Moeschberger, M. L. (1997), Survival Analysis: Techniques for Censored and Truncated Data, New York: Springer-Verlag.
8. Schemper, M. and Smith T.L., (1996), A note on quantifying follow-up in studies of failure time, Controlled Clinical Trials, 17: 343-346.
9. Owen RG, Pratt G, Auer RL, et al. Response assessment in Waldenstrom macroglobulinaemia: update from the VIth International Workshop. Br J of Haematol, 2013;160:171-6.
10. NCCN Guidance Insights, JNCCN 2012;10:1211-1218

10 APPENDIX

Appendix A: Missing Data Imputation Rule

In general, missing or partial dates will not be imputed at 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, 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.

If the imputed end date > death date or end of study date, then set to the death date or end of study date, whichever occurs first.

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 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, then the imputed day and month will be January 01 or the first dosing date if they have the same year, whichever is later.
- If only day is missing, then the imputed day will be the first day of the month or the first dosing date if they have the same month and year, whichever is later
- If 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.

If the imputed AE end date > death date or end of study date, then set to the death date or end of study date, whichever occurs first.

A.3 Deaths

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

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

A.4 Subsequent Anti-cancer Therapies

If the start date 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.

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 Bone Marrow Involvement

Bone marrow will be assessed by biopsy at screening, every 24 weeks, at time of CR and clinical indicated. Values for neoplastic B-lymphocytes and plasma cells from biopsy will be added to determine the baseline bone marrow involvement for subgroup analysis. Given that the test values in the original reports are shown in various formats, following rules will be applied to better capture the magnitude of the bone marrow involvement at baseline. If any of the values are not captured in following cases, appropriate rules will be applied and stated in the study report.

1) If the value is shown as one single numeric value, i.e. X, assign lower limit and upper limit of bone marrow involvement for the patient both as X.
2) If the value is shown as a range, i.e. X-Y, assign lower limit as X and upper limit as Y for the patient.
3) If the value is shown with an upper limit only, i.e. <X or <=X, assign lower limit as 0 and upper limit as X for the patient.
4) If the value is shown as "*easy to see"*, assign lower limit as 0 and upper limit as 20 for the patient.
5) If the value is shown with "*ND"*, assign both lower limit and upper limit as 0 for the patient.

Total bone marrow involvement will be obtained by adding the lower and upper limit values of neoplastic B-lymphocytes and plasma cells together, respectively. For each patient, if lower limit of the combined value is ≥ 50%, then he/she will be categorized to the subgroup of baseline bone marrow involvement lower limit ≥ 50%; If upper limit of the combined value is < 50%, then he/she will be categorized to the subgroup of baseline bone marrow involvement higher limit < 50%.

To determine the anti-lymphoma effect, it's required to check if there is any reduction in bone marrow involvement at any post-baseline visits onwards. If the post-baseline bone marrow involvement and baseline bone marrow involvement are both reported as a range (eg. baseline reported with lower limit a and upper limit b, postbaseline reported with a lower limit c and upper limit d), the following rules will be used to determine if there is reduction:

- If a<b<c<d, then no reduction
- If a<c<b<d, then no reduction
- If a<c<d<b, indeterminate
- If c<a<b<d. indeterminate
- If c<a<d<b, slight reduction
- If c<d<a<b, then reduction

A.8 WM International Prognostic Scoring System (IPSS)

Score:

- Age > 65 year
- 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

Appendix B: Response Assessment

B1. WM Response Assessment Methods

Owens (IWWM 6th) Criteria	Overall Combined Assessment	Overall IgM Assessment
Monoclonal IgM by immunofixation	X	X
IgM level (Include M-protein)*	X	X
Bone marrow	X	
Extramodular disease per CT or PE	X	
New Signs and symptoms of active disease, clinical features of PD	X	
Drug holds	X	X

* Patients may undergo plasmapheresis, when clinically indicated, during the first two cycles of study treatment. A pre-plasmapheresis serum total IgM and M-protein must be obtained during the screening period, and will serve as the baseline value for response assessment throughout the study. Response determination will commence 4 weeks following the last plasmapheresis procedure and will be based upon the baseline pre-plasmapheresis serum total IgM or M-protein value (according to the guidelines above). Patients requiring plasmapheresis after cycle 2 will be adjudged to have progressive disease.

B2. Categorical Response Definitions (modified Owen 2013)

Response category	Definition
Complete response (CR)	Absence of serum monoclonal IgM protein by immunofixation Normal serum IgM level Complete resolution of extramedullary disease, ie, lymphadenopathy and splenomegaly if present at baselines No histological evidence of bone marrow involvement
Very good partial response (VGPR)	Monoclonal IgM protein is detectable ≥ 90% reduction in serum IgM level from baseline [a] Improvement in extramedullary disease, ie, lymphadenopathy/splenomegaly if present at baseline No new signs or symptoms of active disease
Partial response (PR)	Monoclonal IgM protein is detectable ≥ 50% but < 90% reduction in serum IgM level from baseline [a] Improvement in extramedullary disease,

	ie, lymphadenopathy/splenomegaly if present at baseline No new signs or symptoms of active disease
Minor response (MR)	Monoclonal IgM protein is detectable ≥ 25% but< 50% reduction in serum IgM level from baseline [a] No new signs or symptoms of active disease
Stable disease (SD)	Not meeting criteria for CR, VGPR, PR, MR, or progressive disease
Progressive disease (PD)	At least one of the following: • Confirmed ≥ 25% increase in serum IgM and total increase of ≥ 500 mg/dL from nadir (on treatment) [a,b] • New lymph nodes >1.5 cm, or ≥ 50% increase from nadir in SPD of >1 node, or ≥ 50% increase in longest diameter of a previously identified node >1 cm in short axis • New splenomegaly or ≥50% increase from nadir in enlargement • New extranodal disease • New or recurrent involvement in bone marrow • New symptomatic disease

[a] Sequential changes (separated by at least 4 weeks) in IgM levels should be determined by the IgM value from the quantitative serum immunoglobulin assay, unless for assay limitations this is not possible, in which case the M protein level by densitometry (SPEP) will be used

[b] Isolated increase in serum IgM levels during periods of study drug withholding will not be considered as progressive disease unless confirmed by a repeat serum IgM level at least 10 weeks after restarting study drug administration and accompanied by a total increase of at least 500mg/Dl from lowest nadir. Please see guidelines for specific clinical or laboratory circumstances below.

c For response assessments that occur during cycles where a CT scan is not required then results from prior scans (up to 12 weeks during the first 48 weeks and up to 24 weeks thereafter) can be carried forward in those patients with extramedullary disease at baseline.

Refer to the IRC charter (appendix E) for detailed instructions.

Appendix C: Safety Monitoring Committee Charter

The safety monitoring committee charter is provided in a separate document.

Appendix E: Independent Review Charter

The independent review charter is provided in a separate document.