

Johnson & Johnson Pharmaceutical Research & Development

Millennium Pharmaceuticals, Inc.

Statistical Analysis Plan

A Randomised, Open-Label, Multicentre Phase 3 Study of the Combination of Rituximab, Cyclophosphamide, Doxorubicin, VELCADE, and Prednisone (VcR-CAP) or Rituximab, Cyclophosphamide, Doxorubicin, Vincristine, and Prednisone (R-CHOP) in Patients With Newly Diagnosed Mantle Cell Lymphoma who are not Eligible for a Bone Marrow Transplant

Protocol 26866138-LYM-3002; Phase 3

JNJ-26866138 (VELCADE* [Bortezomib] for Injection)

* VELCADE is the exclusive trademark of Millennium Pharmaceuticals, Inc., registered in the United States and internationally.

Issue/Report Date: 8 January 2014
Prepared by: Janssen Research & Development, LLC
Document No.: EDMS-ERI-28105966:2.0

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

Confidentiality Statement

The information in this document contains trade secrets and commercial information that are privileged or confidential and may not be disclosed unless such disclosure is required by applicable law or regulations. In any event, persons to whom the information is disclosed must be informed that the information is privileged or confidential and may not be further disclosed by them. These restrictions on disclosure will apply equally to all future information supplied to you that is indicated as privileged or confidential.

TABLE OF CONTENTS

1.	INTRODUCTION	4
1.1.	Trial Objectives	4
1.2.	Study Design	4
1.3.	Statistical Hypotheses for Trial Objectives	6
1.4.	Sample Size Justification.....	7
1.5.	Randomization and Blinding.....	7
2.	GENERAL ANALYSIS DEFINITIONS	8
2.1.	Imputation of Missing Data	8
2.1.1.	Missing/Partial Dates in Screening Visit.....	8
2.1.2.	Missing/Partial Dates in Adverse Events/Concomitant Therapies/Subsequent Therapies	8
2.1.3.	Missing/Partial Dates in Death	9
2.2.	Analysis Sets	10
2.2.1.	Efficacy Analysis Set(s).....	10
2.2.1.1.	Primary Efficacy Analysis Set	10
2.2.1.2.	Secondary Efficacy Analysis Set	10
2.2.1.2.1.	Per-Protocol Analysis Set	10
2.2.1.2.2.	Response Evaluable Population	10
2.2.1.2.3.	Biomarker Population.....	10
2.2.2.	Safety Analysis Set	10
2.3.	Definition of Subgroups	10
2.4.	Other General Definitions	11
2.4.1.	Treatment Groups	11
2.4.2.	Individual Study Start (Day 1), Study Day and Month.....	12
2.4.3.	End of Follow-up and Length of Follow-up	12
2.4.4.	Treatment Start, Treatment End, and Time on Study Treatment	12
2.4.5.	Number of Treatment Cycles	13
2.4.6.	Properties of Treatment Cycles.....	13
2.4.7.	Dose Intensity and Relative Dose Intensity	14
2.4.8.	Age and Age Categories	15
2.4.9.	Baseline Values.....	15
2.4.10.	Categories of Extranodal Sites of Disease	15
2.4.11.	Category of International Prognosis Index (IPI).....	15
2.4.12.	Stage of MCL at Diagnosis	16
2.4.13.	Creatinine Clearance and Its Category	16
2.4.14.	Time since Initial Diagnosis.....	16
2.4.15.	Relationship of Adverse Events to Study Medication	16
2.4.16.	Study Medication-Related Death	17
2.4.17.	Dose Discontinuation of A Specific Study Drug Due to Adverse Events ...	17
2.4.18.	Adverse Events Leading to Treatment Discontinuation, and Study Medication-Related Treatment Discontinuation	17
2.4.19.	Linking of Adverse Events.....	17
2.4.20.	Concomitant Medication During Treatment	18
2.4.21.	Categories of Body Surface Area.....	18
2.4.22.	Censoring Reasons for PFS	18
2.4.23.	Censoring Reasons for Overall Survival	19
3.	INTERIM ANALYSIS AND DATA MONITORING COMMITTEE REVIEW	19
4.	SUBJECT INFORMATION	20
4.1.	Demographics and Baseline Characteristics.....	20
4.2.	Disposition Information	21
4.3.	Extent of Exposure	22
4.4.	Protocol Deviations.....	22

4.5.	Concomitant Medications	22
4.6.	Subsequent Therapies for MCL.....	23
5.	EFFICACY.....	23
5.1.	Analysis Specifications	23
5.1.1.	Level of Significance	23
5.1.2.	Data Handling Rules	23
5.1.3.	Definitions for Some Efficacy Variables	24
5.1.3.1.	Tumor Assessment by Independent Radiology Review Committee (IRC).....	24
5.1.3.2.	Date of Progression	24
5.1.3.3.	Complete Response Assessment.....	25
5.1.3.4.	Tumor Assessment by Investigator.....	25
5.1.3.5.	Best Overall Response	25
5.2.	Primary Efficacy Endpoint(s)	25
5.2.1.	Definition	25
5.2.2.	Analysis Methods	26
5.3.	Major Secondary Endpoints	27
5.3.1.	Definition	27
5.3.2.	Analysis Methods	29
5.4.	Other Efficacy Variables	30
5.4.1.	Definition	30
5.4.2.	Analysis Methods	30
6.	SAFETY.....	31
6.1.	Adverse Events	31
6.1.1.	All Adverse Events	31
6.1.2.	Adverse Events of Clinical Interest	33
6.1.2.1.	Peripheral Neuropathies	33
6.1.2.2.	Cardiac Rhythm and Conduction Abnormalities	34
6.1.2.3.	Heart Failure	34
6.1.2.4.	Pericardial Disease	35
6.1.2.5.	Acute Diffuse Infiltrative Pulmonary Disease.....	35
6.1.2.6.	Reactivation of Hepatitis B.....	35
6.1.2.7.	Acute Hypersensitivity.....	36
6.1.2.8.	Central Nervous System Disorders.....	36
6.1.2.9.	Herpes Zoster Infection.....	36
6.1.2.10.	Thrombocytopenic Bleeding Events	37
6.1.2.11.	Neutropenic Sepsis and Infections Events	38
6.1.3.	Death.....	38
6.2.	Clinical Laboratory Tests	39
6.3.	Vital Signs and Physical Examination Findings.....	39
6.4.	Electrocardiogram and ECHO/MUGA Scans	40
7.	PATIENT-REPORTED OUTCOMES (PRO) ENDPOINTS.....	40
7.1.	Definition.....	40
7.2.	Analysis Methods	40
8.	BIOMARKER EVALUATIONS.....	42
8.1.	Definition.....	42
8.2.	Analysis Methods	42
	REFERENCES	43
	ATTACHMENT.....	44

1. INTRODUCTION

This statistical analysis plan (SAP) contains definitions of analysis sets, derived variables and statistical methods for the analysis of efficacy and safety.

1.1. Trial Objectives

Primary Objective

To determine which regimen of chemotherapy rituximab, cyclophosphamide, doxorubicin, VELCADE, and prednisone (VcR-CAP) or rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone (R-CHOP) provides greater benefit in newly diagnosed mantle cell lymphoma (MCL) subjects with Stage II, III, or IV disease, as assessed by significant prolongation of progression-free survival (PFS).

Secondary Objectives

The secondary objectives are:

- To determine time to progression (TTP)
- To determine the duration of response (DoR), time to next treatment (TNT), and treatment-free interval
- To determine overall response (complete response [CR]+CR unconfirmed [CRu]+partial response [PR]) and CR rates (CR+CRu)
- To determine overall survival (OS)
- To determine the 18-month survival rate
- To evaluate the safety of VcR-CAP compared to R-CHOP

Exploratory Objectives

- To evaluate patient-reported outcomes (PROs) utilizing the European Organisation for Research and Treatment of Cancer quality of life questionnaire (EORTC-QLQ-C30), EQ-5D, and Brief Fatigue Inventory (BFI) instruments
- To evaluate medical resource utilization (MRU) information which may be used in future economic evaluation models
- To identify subject populations that are more or less likely to respond to VcR-CAP or R-CHOP through the evaluation of biomarker analyses

1.2. Study Design

This is a randomized, open-label, multicenter, prospective study to compare the efficacy and safety of the combination of VcR-CAP to that of R-CHOP in subjects who have newly diagnosed MCL and who are ineligible for bone marrow transplantation.

At least 486 subjects will be randomized into 1 of 2 arms (Treatment Arm A or Treatment Arm B) in a 1:1 ratio taking into account the following stratification factors: International Prognostic Index (IPI) and stage of disease at diagnosis.

Treatment Arm A (VcR-CAP): rituximab 375 mg/m² intravenous (IV) on Day 1, cyclophosphamide 750 mg/m² IV on Day 1, doxorubicin 50 mg/m² IV on Day 1, VELCADE 1.3 mg/m² IV on Days 1, 4, 8 and 11, and prednisone 100 mg/m² per os (PO) on Day 1 to Day 5 of a 21-day (3-week) cycle for 6 cycles (or 8 cycles if a response is first documented at the Cycle 6 assessment).

Treatment Arm B (R-CHOP): rituximab 375 mg/m² IV on Day 1, cyclophosphamide 750 mg/m² IV on Day 1, doxorubicin 50 mg/m² IV on Day 1, vincristine 1.4 mg/m² (maximum total of 2 mg) IV on Day 1, and prednisone 100 mg/m² PO on Day 1 to Day 5 of a 21-day (3-week) cycle for 6 cycles (or 8 cycles if a response is first documented at the Cycle 6 assessment).

The study will include a Screening Phase, a Treatment Phase, a Short-term Follow-up Phase, and a Long-term Follow-up Phase. The Screening Phase will be up to 28 days (56 days for bone marrow evaluation) prior to randomization. The Treatment Phase will extend from randomization until 6 cycles of treatment have been given (or 2 cycles beyond a response documented in Cycle 6). The Short-term Follow-up Phase will extend from the End-of-Treatment Phase to progressive disease (PD; or relapse if subject achieves a CR or CRu), initiation of alternate antineoplastic therapy, decision by the subject to completely withdraw from the study and refuse any further study related procedure, or death. Subjects who are withdrawn from the study treatment due to adverse events, or reasons other than above and are willing to continue study follow-up procedures can be followed as per protocol for PD or death.

The Long-term Follow-up Phase will be used to assess survival and will document when the subject has died.

Three interim analyses are planned for this study. The first interim analysis will occur after the first 100 subjects have been randomized into the study and will assess safety and the concordance rate of the diagnosis of MCL when central review is compared with the investigator assessment of the diagnosis. The concordance rate of the diagnosis of MCL will also be reviewed in a blinded manner at 50% accrual. This may be used to recalculate the sample size to ensure an adequate number of PFS events (approximately 280) in those subjects with a centrally confirmed MCL diagnosis at the time of the final analysis (295 PFS events in total). A review of these data will occur by the IDMC.

A second interim analysis for safety will occur after 100 subjects in each arm (200 subjects in total) have either completed the study treatment or discontinued the study treatment, which allows sufficient exposure for review of cumulative toxicity.

The third interim analysis is planned after at least 148 events have occurred. If, at the time of the third interim analysis, the pre-specified superiority boundaries for PFS are met, then the study will be terminated and superiority of the experimental arm (VcR-CAP) will be declared over the comparator arm (R-CHOP). If the pre-specified futility boundaries for PFS are met, then the study will be terminated due to futility.

If pre-specified boundaries are not met at the time of the third interim analysis, the final analysis of the study will occur after 295 events have been observed in the ITT population.

An Independent Data Monitoring Committee (IDMC) has been commissioned for this study.

1.3. Statistical Hypotheses for Trial Objectives

The primary statistical hypothesis is that there is no difference between the two treatment groups (Treatment Group A: VcR-CAP, and Treatment Group B: R-CHOP) with regard to PFS.

Secondary statistical hypotheses are as follows:

- 1) There is no difference between the two treatment groups with respect to TTP
- 2) There is no difference between the two treatment groups with respect to TNT;
- 3) There is no difference between the two treatment groups with respect to the complete response (CR+CRu) rate;
- 4) There is no difference between the two treatment groups with respect to OS;
- 5) There is no difference between the two treatment groups with respect to overall response (CR+CRu+PR) rate (ORR);
- 6) There is no difference between the two treatment groups with respect to 18-month survival rate.
- 7) There is no difference between the two treatment groups with respect to treatment free interval

Exploratory hypotheses include:

There is no difference between the two treatment groups with respect to the PRO endpoints as defined in Section 7.

There is no difference between the two treatment groups within each biomarker subgroup as defined in Section 8.

Analysis of DoR will be descriptive instead of inferential. This is due to the fact that the analyses can only be performed on subsets of subjects determined by a post-randomization outcome.

1.4. Sample Size Justification

The sample size calculation for the study population is based on the following assumptions. The median PFS of Treatment Arm B (R-CHOP) is 18 months. Assuming that Treatment Arm A (VcR-CAP) can improve the median PFS by 40%, i.e. from 18 months to 25 months, a total number of 295 events (PD or death) provides 80% power (alpha = 0.05, two-sided) to detect such improvement. Assuming a 24-month accrual and 18-month follow-up, 486 subjects in total are needed for the study (243/arm).

1.5. Randomization and Blinding

Subjects will be assigned in a 1:1 ratio to 1 of the 2 treatment groups based on a computer-generated randomization schedule prepared by the sponsor before the study. The sponsor and the investigational sites will be blinded to all data reviewed by the IDMC.

Permuted blocks central randomization will be implemented in conducting this study. The randomization will be stratified by two stratification factors: IPI (0-1, 2, 3, and 4-5) and stage of disease at diagnosis (II, III, and IV) per the American Joint Committee on Cancer NHL staging system. The stratified randomization minimizes the imbalance in the distribution of treatment numbers within the levels of each individual stratification factors.

The IPI is the total score of the following risk factors (score 1 for each positive factor): age >60 years, MCL Stage III or IV at diagnosis, Eastern Cooperative Oncology Group (ECOG) performance status score >1, >1 extranodal site involvement, and lactate dehydrogenase (LDH) above normal limits. For stratification, the scores will then be categorized (low [0-1 factor], low-intermediate [2 factors], high-intermediate [3 factors] and high [4-5 factors]).

The stage of disease at diagnosis will also be used for stratification and will be assessed using the American Joint Committee on Cancer NHL staging system (see Attachment 11 of Protocol 26866138LYM3002).

2. GENERAL ANALYSIS DEFINITIONS

2.1. Imputation of Missing Data

Unless otherwise specified, no data imputation will be applied for missing values and all available data for all subjects will be presented. For PRO data, missing elements may be substituted with the average of non-missing items per the published methods of analysis using well-established scaling programs, and missing scores will be handled using statistical techniques as outlined in the section for PRO data.

2.1.1. Missing/Partial Dates in Screening Visit

The following rules apply to dates recorded in the screening visits except for prior therapies (excluding prior systemic therapies for MCL).

- If only the day-component is missing, the first day of the month will be used if the year and the month are the same as those for the first treatment. Otherwise, the 15th will be used unless it is later than the first treatment.
- If only a year is present, and it is the same as the year of the first treatment, 15 January will be used unless it is later than the first treatment, in which case the date of 1 January will be used.
- If only a year is present, and it is not the same as the year of the first treatment, 15 June will be used, unless other data indicate that the date is earlier. For instance, the diagnosis date of MCL should be used to impute missing month/day of the start dates for prestudy systemic therapy for MCL if the diagnosis date is later than 15 June.

2.1.2. Missing/Partial Dates in Adverse Events/Concomitant Therapies/Subsequent Therapies

Every effort will be made to avoid missing/partial dates in on-study data. In case that the resolution date of a resolved adverse event, or the stop date of a concomitant therapy is missing, the following rules are to be used unless conflicting data exists: if month and year are present and the day of the month is missing, the last day of the month is imputed. If only a year is present, 31 December will be used. After imputation, the imputed dates will be compared against the date of death, if available. If the date is later than the date of death, the date of death will be used as the imputed date instead.

In case that the onset date of an adverse event is missing completely or partially, the following imputation rules will be used.

1) When month and year are present and the day of the month is missing,

- If the onset month and year are the same as the month and year of first treatment with study drug, the day of first treatment or the day-component of the resolution date are imputed, whichever is earliest.

- If the onset month and year are not the same as the month and year of first treatment with study drug, the first day of the month is imputed.

2) When only a year is present, or no components of the onset date are present,

- If the imputed resolution date is available, the later of the imputed resolution date (possibly imputed) and the date of first treatment will be used.
- If the imputed resolution date is missing, and the onset-year is the same as the year of first treatment with study drug, then the date of first treatment with study drug is used.
- Otherwise if only a year is present, 1 January of that year is imputed.

3) If none of the previous rules can be applied, then the date of first treatment with study drug is imputed as the onset date.

The imputation rules for missing/partial start dates of concomitant therapies will be the same as the above with the exception as follows.

For prior therapy data recorded in the concomitant medication page, no imputation will be done for start dates.

The imputation rules for missing/partial start dates of subsequent therapies will be the same as the above with the exception as follows.

1) When month and year are present and the day of the month is missing,

- If the month and year of the start date are the same as the month and year of treatment termination, the day of treatment termination visit or the day-component of the stop date is imputed, whichever is earliest.
- If the start month and year are not the same as the month and year of treatment termination visit, the first day of the month is imputed.

2) When only a year is present, or no components of the start date are present, the date will not be imputed.

2.1.3. Missing/Partial Dates in Death

If month and year are present and the day of the month is missing, the last day of the month is imputed.

2.2. Analysis Sets

2.2.1. Efficacy Analysis Set(s)

2.2.1.1. Primary Efficacy Analysis Set

The primary efficacy analysis set is the intent-to-treat (ITT) population, which includes all randomized subjects.

2.2.1.2. Secondary Efficacy Analysis Set

2.2.1.2.1. Per-Protocol Analysis Set

The per-protocol (PP) population is defined as all subjects who are randomized to treatment who meet all inclusion and exclusion criteria, who received at least 1 dose of study drug, and undergo at least one post-baseline disease assessment (post-baseline tumor assessment by independent radiology reviewers [IRC]). Subjects in this population will be analyzed according to the treatment to which they were randomized. The analyses using the PP population will be considered as sensitivity analyses.

2.2.1.2.2. Response Evaluable Population

The response-evaluable population is defined as all subjects in the ITT population who receive at least 1 dose of study drug, have at least 1 measurable tumor mass (greater than 1.5 cm in the longest dimension and greater than 1.0 cm in the short axis) at baseline, and have at least 1 post-baseline tumor assessment by IRC, before any subsequent anti-lymphoma treatment.

2.2.1.2.3. Biomarker Population

The biomarker population is defined as subjects who are randomized to treatment (ITT), whose biomaterial is available, who have consented to participate in the study's biomarker and pharmacogenomic evaluations or future research, and have biomarker data generated for the biomarker evaluation.

2.2.2. Safety Analysis Set

The safety population, which includes all randomized subjects who receive at least 1 dose of study drug, will be used in the safety analyses.

2.3. Definition of Subgroups

In general, the following subgroup analyses are to be performed for efficacy and safety. Additional subgroup analyses will also be performed for selected efficacy and/or safety endpoints, which are described in Sections 5 and 6.

Subgroup	Definition of Group	Analysis Type
IPI risk	low [0-1 factor], low-intermediate [2 factors], high-intermediate [3 factors], high [4-5 factors]	E, S
Sex	Male, Female	E, S
Race	White, Non-White	E, S
Region	European Union, North America Region, Rest of World	E, S
Age	≤65, >65	E, S
Stage of Disease at Diagnosis	II, III, IV	E, S
LDH	Normal, Elevated	E
White Cell Count (WBC)	<6.7, 6.7-<10, 10-<15, ≥15	E
ECOG	0: asymptomatic, 1: symptomatic, Fully Ambulatory, 2: symptomatic, in Bed<50% of the Day	E
Renal Function (mL/min)	Baseline creatinine clearance: Renal impairment (≤60), Normal renal function (>60)	S
Liver Function	Baseline ALT and AST<1.5*ULN, ALT/AST≥1.5*ULN	S

E= efficacy (primary, secondary endpoints); S= Safety (adverse events)

In order to demonstrate applicability of study results to the US population, a US/Canada/Western Europe subset will be created where Western Europe includes Austria, Belgium, France, Germany, Italy, Portugal, and Spain. Summary data for demographics, baseline disease characteristics, key efficacy endpoints (progression free survival by IRC and Investigator, progression free survival for US/Canada subset, time to tumor progression, overall survival, best overall response), and treatment-emergent adverse events will be generated for the US/Canada/Western Europe subset.

2.4. Other General Definitions

Unless specified otherwise, the dates used in the derivations are those obtained after applying the rules specified in Section 2.2.

2.4.1. Treatment Groups

There are two variables for treatment group identification (Treatment Group A: VcR-CAP or Treatment Group B: R-CHOP). The first variable identifies the treatment group to which a subject is randomized. All randomized subjects will have a nonmissing value for this variable.

The second variable denotes the actual treatment received. This variable is derived from the dosing data of the study medication. A subject is in the VcR-CAP group if the subject receives at least one dose of VELCADE; otherwise, the actual treatment group is R-CHOP. Those subjects without any dose of the study medications (VELCADE, rituximab, cyclophosphamide, doxorubicin, vincristine, or prednisone) will have this variable set as missing.

2.4.2. Individual Study Start (Day 1), Study Day and Month

Individual Study Start (Day 1) is the day of randomization.

Study day is equal to the date of procedure minus the date of randomization plus one.

A month is defined as $365.25/12$ days.

2.4.3. End of Follow-up and Length of Follow-up

The end of follow-up is either the date of death or the end of survival follow-up, defined as the maximum date of the following study evaluations: randomization, vital signs, ECOG performance status, CT scans (neck, chest, abdomen, and pelvis), dates of assessable lesions or new lesions, physical examination, bone marrow evaluations, labs (hematology, chemistry), study drug administration, subsequent therapy, investigator-assessed PD, relapse from CR/CRu, investigator determined response, and survival follow-up.

Length of follow-up (in months) equals to the end of follow-up minus study start, then divided by $365.25/12$.

2.4.4. Treatment Start, Treatment End, and Time on Study Treatment

Treatment start is the date of first dose of study medication (dose > 0 for any of the study drugs: VELCADE, rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone). Treatment end is the date of the last dose of the study medication (dose > 0 for any of the six drugs: VELCADE, rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone). Time (weeks) on study treatment equals treatment end minus treatment start plus 1, then divided by 7.

For subjects randomized to VcR-CAP, the start date of VELCADE is the date of first nonzero dose of VELCADE, and the stop date of VELCADE is the date of last nonzero dose of VELCADE. Time (in weeks) on VELCADE equals the stop date minus the start date plus 1, then divided by 7. Note that time on VELCADE can only be determined for subjects who have received at least one dose of VELCADE.

For subjects randomized to R-CHOP, the start date of vincristine is the date of first nonzero dose of vincristine, and the stop date of vincristine is the date of last nonzero dose of vincristine. Time (in weeks) on vincristine equals the stop date minus the start date plus 1, then divided by 7. Note that time on vincristine can only be determined for subjects who have received at least one dose of vincristine.

The start date of rituximab is the date of first nonzero dose of rituximab, and the stop date of rituximab is the date of last nonzero dose of rituximab. Time (in weeks) on rituximab equals the stop date minus the start date plus 1, then divided by 7.

The start date of cyclophosphamide is the date of first nonzero dose of cyclophosphamide, and the stop date of cyclophosphamide is the date of last nonzero dose of cyclophosphamide. Time (in weeks) on cyclophosphamide equals the stop date minus the start date plus 1, then divided by 7.

The start date of doxorubicin is the date of first nonzero dose of doxorubicin, and the stop date of doxorubicin is the date of last nonzero dose of doxorubicin. Time (in weeks) on doxorubicin equals the stop date minus the start date plus 1, then divided by 7.

The start date of prednisone is the date of first nonzero dose of prednisone, and the stop date of prednisone is the date of last nonzero dose of prednisone. Time (in weeks) on prednisone equals the stop date minus the start date plus 1, then divided by 7.

2.4.5. Number of Treatment Cycles

The first treatment cycle is the first cycle in which a subject receives any nonzero dose of the study drugs (VELCADE, rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone). The last treatment cycle is the last cycle in which a subject receives any nonzero dose of the study drugs. Number of treatment cycles equals the last cycle number minus the first cycle number plus 1.

For each of the study drugs (VELCADE, rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone), the first (or last) treatment cycle on this drug is the first (or last) cycle in which a subject receives any nonzero dose of the study drug, and the number of cycles on this study drug equals the last cycle minus the first cycle plus 1.

Note that number of cycles on rituximab, cyclophosphamide, doxorubicin, or prednisone can be determined for all treated subjects, while the number of cycles on VELCADE can only be determined for subjects randomized to receive VcR-CAP, and the number of cycles on vincristine can only be determined for subjects randomized to receive R-CHOP.

2.4.6. Properties of Treatment Cycles

For each treatment cycle, the following variables are to be derived.

- Cycle delay (≥ 5 days) — to flag if this cycle is delayed for at least 5 days (date of Day 1 in the current cycle – date of Day 1 in the previous cycle is ≥ 26 days) and checked on the case report form (CRF) by investigator as cycle delayed.
- Reasons for cycle delay (≥ 5 days) — to record reasons for cycle delay of at least 5 days as captured in CRF (AE or Other).
- Cycle start — the date of the first dose of any of the study drugs (including zero doses).
- Cycle stop — For cycles other than the last one, cycle end is defined as the day before the start of the next cycle; for the last cycle, it is defined as the visit date of the end of

treatment visit (or the cycle start plus 20 days in case that the end of treatment visit is not performed yet).

- Dose reduction — to flag if the dose of the study medication is reduced (each drug has its own flag[s]). For VELCADE, a dose is considered reduced if a nonzero dose is marked as reduced on the CRF and the prescribed dose is lower than the previous non-zero dose level. For rituximab, a nonzero dose is considered as reduced if it is >25% lower than the prescribed dose. For other drugs, a dose is considered as reduced if a non-zero dose is marked on the CRF as reduced. Dose adjustment due to BSA change is not counted as dose reduction as long as the prescribed dose is not lower than the protocol specified level. For Prednisone, due to the tapering regimen, dose after DAY 5 in each cycle is not used in dose reduction consideration.
- Dose withholding — to flag if the dose of the study medication is withheld at least once (one flag for each drug), excluding those due to schedule changes for VELCADE. A dose not given due to schedule change will not be counted as withholding for VELCADE. The number of withheld doses will be recorded for VELCADE and prednisone.
- Schedule change — to flag if a schedule change of VELCADE is initiated within the cycle, i.e., if the VELCADE dose schedule is changed from twice weekly to once weekly. Note: this is applicable only for those subjects randomized to receive VcR-CAP.
- Dose modification — to flag if the dose of the study medication is modified (delayed, reduced or withheld, schedule change [for VELCADE only]) at least once (one flag for each drug).
- Number of doses — to record the number of nonzero actual doses given (one variable for each drug).
- Total dose (mg) — to record the sum of nonzero doses (one variable for each drug).
- Total doses (mg/m²) — to record the sum of nonzero doses per m² (one variable for each drug), which equals the total dose (mg) divided by body surface area in that cycle. When the BSA is not available for the cycle of interest, the most recent value is to be used.

2.4.7. Dose Intensity and Relative Dose Intensity

Dose intensity (mg/m²/cycle) will be calculated separately for the study drugs (VELCADE, rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone) by cycle and overall.

For each of the study drugs, the dose intensity equals the sum of total dose (mg/m²) received in all cycles divided by the number of cycles on that drug.

The dose intensity per cycle is equal to the sum of the total dose (mg/m²) received in that cycle.

The overall planned dose intensity is equal to the sum of the planned total dose (mg/m^2) in all cycles divided by the number of cycles received.

The planned dose intensity per cycle is equal to the sum of the planned total dose (mg/m^2) in the cycle. When a subject receives all scheduled doses in a cycle, the planned dose intensity per cycle for VELCADE is $1.3 \times 4 = 5.2$ ($\text{mg}/\text{m}^2/\text{cycle}$); for rituximab, is $375 \times 1 = 375$ ($\text{mg}/\text{m}^2/\text{cycle}$); for doxorubicin, is $50 \times 1 = 50$ ($\text{mg}/\text{m}^2/\text{cycle}$); for cyclophosphamide, is $750 \times 1 = 750$ ($\text{mg}/\text{m}^2/\text{cycle}$); for vincristine, is $1.4 \times 1 = 1.4$ ($\text{mg}/\text{m}^2/\text{cycle}$); for prednisone, is $100 \times 5 = 500$ ($\text{mg}/\text{m}^2/\text{cycle}$).

The relative dose intensity is the ratio of the dose intensity and the planned dose intensity.

2.4.8. Age and Age Categories

Age will be calculated at the date of randomization. Integer values are used. The following categories of age will be adopted: ≤ 65 years, > 65 years.

2.4.9. Baseline Values

The baseline value is defined as the value collected at the time closest to, but prior to, the date of first dose (inclusive). If a subject is not dosed, the date of randomization will be used.

In case that laboratory values from both the central lab and local lab are available before randomization, the values from central lab will be used as baseline, even if the sample date of central lab result is prior to that of local lab value.

2.4.10. Categories of Extranodal Sites of Disease

Two categories will be used for the number of extranodal sites: ≤ 1 or > 1 .

2.4.11. Category of International Prognosis Index (IPI)

The IPI will be assessed according to the following risk factors: age, stage of disease, performance status, number of extranodal sites, and LDH level. A subject's IPI ranges from 0-5 and equals the number of the following applicable conditions:

- Age > 60 years
- MCL stage III or IV at diagnosis
- ECOG performance status score > 1 at baseline
- More than one extranodal site involvement at baseline
- Baseline LDH $>$ normal limits

IPI at baseline will be classified into the following categories: low (0-1 factor), low-intermediate (2 factors), high-intermediate (3 factors) and high (4-5 factors).

Categories of IPI are also used to determine the appropriate stratum in which a subject is to be randomized. Since evaluations at screening may be used to determine the stratification, a separate variable is used to capture the value used in randomization.

2.4.12. Stage of MCL at Diagnosis

The staging of MCL at diagnosis is to be assessed using the American Joint Committee on Cancer NHL staging system. Those subjects with Stages II, III, or IV disease are eligible for this study.

Stage of MCL disease is also used to determine the appropriate stratum in which a subject is to be randomized. A separate variable is used to capture the value used in randomization.

2.4.13. Creatinine Clearance and Its Category

Creatinine clearance is to be calculated using the following formulae.

For male subjects,

$$\text{Creatinine clearance} = (140 - \text{Age [yrs]}) * \text{Weight [kg]} / 72 / (\text{Serum creatinine}/88.4)$$

For female subjects,

$$\text{Creatinine clearance} = 0.85 * (140 - \text{Age [yrs]}) * \text{Weight [kg]} / 72 / (\text{Serum creatinine} / 88.4).$$

Integer values will be used. When available, creatinine clearance values are divided into 3 categories: ≤ 30 mL/min, 31-60 mL/min, and >60 mL/min.

2.4.14. Time since Initial Diagnosis

Time since initial diagnosis is defined for all subjects and is calculated from the date of diagnosis to the date of randomization (month).

2.4.15. Relationship of Adverse Events to Study Medication

For each adverse event, its relationship to each of the six study medications is determined by the investigator and recorded on the CRF.

An adverse event that occurs in the VcR-CAP arm is considered as related to study medication if it is possibly, probably or very likely related to any one or combinations of the five study drugs (VELCADE, rituximab, cyclophosphamide, doxorubicin, and prednisone). The category of the relationship to study medication is to be the maximum of the relationship to the five individual study drugs (“Very Likely”>“Probable”>“Possible”> “Doubtful”> “Not Related”).

An adverse event that occurs in the R-CHOP arm is considered as related to study medication if it is possibly, probably or very likely related to any one or combinations of the five study drugs (rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone). The category of the relationship to study medication is to be the maximum of the relationship to the five individual study drugs.

An adverse event that occurs in the VcR-CAP-treated subjects is considered as related to VELCADE if it is possibly, probably or very likely related to VELCADE.

2.4.16. Study Medication-Related Death

A death is considered as related to study medication if the cause is recorded as adverse event, at least one of which is related to study medication.

2.4.17. Dose Discontinuation of A Specific Study Drug Due to Adverse Events

The dose of a specific drug (VELCADE, rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone) may be discontinued due to adverse events while at least one of the other study drugs may be continuing. Such dose discontinuations will have Action Taken Regarding Trial Med being “permanently stopped” for the study drug of interest. The cycle in which the dose is discontinued is to be the cycle in which the last nonzero dose of this study drug is given.

2.4.18. Adverse Events Leading to Treatment Discontinuation, and Study Medication-Related Treatment Discontinuation

The CRF collects adverse events that lead to discontinuation of any of the six study drugs. For an adverse event to be considered as the reason for treatment discontinuation, the following two conditions must be met:

- 1) The action taken regarding trial medication due to this adverse event should be “permanently stopped” for at least one of the study drugs; and
- 2) This subject should be indicated by the investigator to have discontinued study treatment due to adverse events.

A subject is considered as discontinued due to study medication-related adverse events if (1) the reason for treatment discontinuation is adverse event, and (2) at least one of the adverse events leading to treatment discontinuation, as defined above, should be related to study medication.

2.4.19. Linking of Adverse Events

To analyze new onsets of adverse events, overlapping/adjacent adverse event records of the same preferred term may need to be linked. This is due to the AE collection procedure

as follows. For an adverse event recorded on the CRF, whenever a feature of the adverse event changes (e.g., changes in grade/seriousness/action taken), a new adverse event record will be entered to reflect the change. For example, if the toxicity of an AE changes, the altered toxicity must be recorded on a new line. Resolved date and Onset date must not overlap from 1 line to another for more than 1 day. For the first degree of toxicity, the Resolved date must be filled. A new line should list the same AE, a new Onset date, and the new degree of toxicity. All the adverse event records of the same preferred term from the same subject are to be linked by the onset date and the resolution date. If a resolved event is followed by another event of the same preferred term with the onset date within 1 day of the resolution date of the previous record, these 2 records should be linked together and considered as 1 event.

2.4.20. Concomitant Medication During Treatment

Concomitant medications administered during study treatment include those concomitant medication records that are indicated in the CRF as not being prior therapy, and have the onset date, possibly imputed, no earlier than treatment start and no later than the treatment end.

2.4.21. Categories of Body Surface Area

Three categories of body surface area are of interest: $<1.5 \text{ m}^2$, $1.5\text{-}2 \text{ m}^2$, and $>2 \text{ m}^2$.

2.4.22. Censoring Reasons for PFS

Those subjects who are censored for PFS will be classified into the following categories: subsequent therapy, lost to follow-up, missed tumor assessments before death, and study cut-off.

The Censoring reasons for PFS are defined as below:

- 1) Subsequent therapy: if subsequent therapy is started before PFS events (i.e., PD or death).
- 2) Lost to follow-up: if the reason for end of data collection is lost to follow-up, subject choice, or other.
- 3) Missed tumor assessments before death: if a subject died more than 6-month after last adequate tumor assessment.
- 4) Date of clinical cut-off/study closed by sponsor: for subjects without PD at the time of clinical cut-off/ study closed by sponsor for analysis.

Refer to the Table 1 in Section 5.2.1 for more details of censoring method.

2.4.23. Censoring Reasons for Overall Survival

Those subjects who are censored for OS will be classified into the following categories:

- 1) Lost to follow-up: the reason for end of data collection is lost to follow-up, subject choice, or other.
- 2) Study cut-off: subjects who are still alive as of the date of clinical cut-off.

3. INTERIM ANALYSIS AND DATA MONITORING COMMITTEE REVIEW

An Independent Data Monitoring Committee (IDMC) has been commissioned for this study. There will be three interim analyses planned in the study.

The first interim analysis will occur after the first 100 subjects have been randomized into the study and will assess the safety and concordance rate of the diagnosis of MCL when central review is compared to the investigator assessment of the diagnosis.

The concordance rate of the diagnosis of MCL will also be reviewed at 50% accrual in a blinded manner. If the observed concordance rate is $\geq 95\%$ (95% CI: [91%, 97%]), the sample size for the study will not be changed. If the observed difference is less than 95%, the sample size may be adjusted to provide adequate events (approximately 280) in those subjects with a centrally confirmed MCL diagnosis for the final analysis (295 PFS events in total). There is no alpha adjustment for the 1st interim analysis since no efficacy analyses will be performed.

The first interim analysis will also review the safety data collected to date. Safety data will be reviewed after at least one cycle for the first 100 treated subjects, irrespective of whether they complete treatment within that first cycle.

The second interim analysis will review the safety data and will be performed after 100 subjects per arm have either completed or discontinued study treatment, which allows sufficient exposure for review of cumulative toxicity.

The third interim analysis has been planned for this study after at least 148 events have occurred. If, at the interim analysis, pre-specified boundaries for PFS are met then the study may be terminated and superiority of the experimental arm (VcR-CAP) will be declared over the comparator arm (R-CHOP). If the observed hazard ratio (R-CHOP vs. VcR-CAP) for PFS in the third interim analysis is equal to or less than 1.03 (a value of >1 favoring VcR-CAP), then the study may be terminated due to futility. The conditional power would be $<30\%$ with the observed hazard ratio (HR) ≤ 1.03 at the third interim analysis and the future HR of 1.4, which suggests low probability of success at the end of the study. There will also be a review of the safety data at the third interim analysis.

The stopping boundary for the interim is to be calculated using the observed number of PFS events at the data cutoff using the O'Brien-Fleming spending function. The alpha allocated for the interim is 0.003 (2-sided) with 148 events, and is 0.049 (2-sided) for the final analysis.

If pre-specified boundaries are not met at the interim stage, the final analysis of the study will occur after 295 events.

4. SUBJECT INFORMATION

4.1. Demographics and Baseline Characteristics

All demographic and baseline characteristic variables will be summarized for both the ITT and the PP populations.

Subject enrollment will be summarized by region, country, and investigator. Also to be summarized is subject eligibility by meeting inclusion/exclusion criteria.

Stratification factors used in randomization (IPI category and stage of disease at baseline) will be tabulated. The stratified logrank test and stratified cox model use the stratification factors in randomization.

Demographic and baseline characteristic data will be summarized, including sex, age, race, ECOG performance status score, weight, height, and body surface area (m^2).

The following categories of disease characteristics at baseline will be summarized:

- 1) Diagnosis information of MCL: including confirmed diagnosis by central review (Yes, No, or indeterminate), stage of disease at diagnosis, type of histology, cellularity, over-expression of cyclin D1 in association with CD20 and CD5 (Yes or No), and chromosome translocation t(11; 14) (Yes or No).

- 2) Disease characteristics at baseline: including IPI at baseline, bone marrow aspirate results (Positive, Negative, Indeterminate, or Not Evaluable), bone marrow biopsy results (Positive, Negative, Indeterminate, or Not Evaluable), baseline LDH (absolute value and category [$>ULN$, $\leq ULN$], and reason for transplantation ineligibility.

Baseline IPI score and baseline stage of disease will be derived or obtained from CRF. These baseline derived stratification factors will be tabulated and they will be used in the subgroup analyses.

Two panels of baseline clinical laboratory tests will be summarized: hematology and serum chemistry. Descriptive summaries for creatinine clearance and categories of creatinine clearance will be provided. Frequencies of CTC grade are to be provided for the following laboratory tests at baseline:

Hematology: hemoglobin, WBC, absolute neutrophil count (ANC), absolute lymphocyte count (ALC), and platelets.

Chemistry: aspartate transaminase (AST), alanine transaminase (ALT), bilirubin, creatinine, alkaline phosphatase, glucose, albumin, bicarbonate, magnesium, LDH, urea, blood urea nitrogen, uric acid, electrolytes (sodium, potassium, calcium, and phosphate), and creatinine clearance.

Medical history will be summarized. The findings (normal or abnormal [clinically significant or clinically insignificant]) will also be summarized for baseline 12-lead ECGs, and ECHO/MUGA scans.

4.2. Disposition Information

A CONSORT-type flow chart of subject disposition will be provided for the intent-to-treat population.

The number of subjects discontinuing and completing the study treatment will be reported for the intent-to-treat population. The reasons for discontinuation as indicated by the investigators will be summarized by the number of subjects reporting. Those subjects who discontinue study treatment due to adverse events will be further tabulated by relationship to study medication. The number and percentage of subjects in the safety population and the per-protocol populations will also be reported.

Descriptive statistics (mean, standard deviation, median, and range) will be provided for study duration (in weeks) as well as length of follow-up (in months) for the intent-to-treat and the per-protocol populations.

4.3. Extent of Exposure

All extent-of-exposure summaries are to be presented for the safety population.

Frequency of subjects who receive at least a certain number of treatment cycles will be tabulated by the number of cycles. Number of treatment cycles will be summarized along with time on study treatment, and dose intensity for the study drugs.

Summaries will be provided for number of doses and total doses of study drug (mg/m²) by cycle. Also to be presented by cycle are the number and percentage of subjects on treatment, and the number and percentage of subjects with cycle delays (>5 days). For each of the study drugs, the following summaries are to be reported: number of subjects dosed, number of subjects with dose reduced, dose withheld, schedule change (only applicable to subjects treated with VELCADE), and any change in dosing (reduced, withheld, schedule change [VELCADE only]).

Reasons for dose modifications will be summarized by cycle as well as across the study for each of the following changes: cycle delay, dose reduction, dose withholding, and schedule change (VELCADE only).

For each type of the dose changes mentioned above, adverse events leading to these changes in doses of study treatment will be summarized for each of the study drugs.

Adverse events leading to the discontinuation of VELCADE will be summarized in a separate summary. In this summary, the incidences of the following AEs are to be reported: AEs leading to the discontinuation of VELCADE alone, AEs leading to the discontinuation of VELCADE and at least one other study drug, and AEs leading to the discontinuation of study treatment.

4.4. Protocol Deviations

Major protocol deviations will be summarized by types of deviations.

4.5. Concomitant Medications

Concomitant medications will be coded using the latest version of the WHO Drug Dictionary. The concomitant therapies will be provided by therapeutic class, pharmacological class, and preferred term. A summary of all concomitant medications will be summarized, including the ones that (1) started prior to study treatment and continued during study treatment (2) started during study treatment. A separate summary will be provided for concomitant therapies started during study treatment.

4.6. Subsequent Therapies for MCL

Subsequent therapies for MCL, including antineoplastic agents and other systemic therapies such as steroids, will be summarized by pharmacological class and preferred term. In addition, the number and percentage of subjects who receive VELCADE as subsequent therapy will be presented. These summaries are to be provided for both the ITT and the PP populations.

Best responses to subsequent anti-lymphoma therapies will be tabulated.

5. EFFICACY

5.1. Analysis Specifications

5.1.1. Level of Significance

All tests will be 2-sided. The primary population is the ITT population, with sensitivity analyses to be performed on the centrally confirmed MCL population and PP populations.

The primary hypothesis is to be tested at the 0.05 significance level (overall). The significance level at an interim analysis is to be determined by the observed number of events at the time of the interim analysis per the O'Brien-Fleming spending function.

The secondary hypotheses are to be tested sequentially at the nominal 0.05 significance level. The order of these hypotheses is as follows: 1) TTP, 2) TNT, 3) CR rate and 4) OS. A secondary hypothesis will be tested if and only if the primary hypothesis is rejected along with all the secondary hypotheses that precede it.

OS will be analyzed regardless of what happened to the other endpoints in the hierarchical test.

For exploratory modeling, main effects other than the treatment group in a proposed model will be tested at the 0.05 significance level, while interactions will be examined at the 0.10 significance level. Non-significant interaction terms will be removed from the model.

The clinical cutoff of the third interim analysis was 5 July 2011. At the time of clinical cut-off, there were 173 PFS events. Based on the original design with an O'Brien-Fleming spending function (the Lan-DeMets approach) and a total of 295 events, an alpha allocated for the 3rd interim analysis is 0.006 (2-sided), and for the final analysis is 0.048 (2-sided).

5.1.2. Data Handling Rules

Unless specified otherwise, missing values for efficacy criteria will be handled using the last-observation-carried forward method. Baseline values of efficacy criteria will not be carried forward.

5.1.3. Definitions for Some Efficacy Variables

5.1.3.1. Tumor Assessment by Independent Radiology Review Committee (IRC)

Tumor assessment by radiological images (computed tomography [CT] scans) is the primary efficacy evaluation. All radiological images will be collected and reviewed centrally by Perceptive Informatics, an Independent Radiology Review committee (IRC). The radiological images will be collected at baseline, Cycles 2, 4, 6 (and 8, if applicable), End-of-Treatment, and thereafter every 6 weeks for 18 weeks and every 8 weeks until PD/initiation of alternate therapy/withdrawal. With the approval of Amendment INT-3, when the subject is recorded to have an event of PD, a repeat CT scan to confirm PD must be undertaken at least 30 days after the scan that was used to determine PD. In the event a subject starts subsequent anti-lymphoma treatment, it is strongly recommended that this repeat CT scan be performed before the subject starts treatment. The repeat CT scan must be done using i.v. and oral contrast and must be of the neck, chest, abdomen, and pelvis. If the subject is intolerant of i.v. contrast agents, the scan may be performed with only oral contrast.

The IRC will consist of a minimum of 3 board certified radiologists assigned to Study LYM-3002. Two primary radiologists will independently read each case (double read) with a third radiologist serving as adjudicator, when necessary.

The primary radiologists will independently interpret imaging studies for subjects enrolled in the study, using modified International Workshop to Standardize Response Criteria for non-Hodgkin's Lymphoma (IWRC) [Cheson 2007]). Detailed criteria for response categories can be found in the attachment to this SAP. The radiologists will be blinded to treatment assignment.

The IRC primary radiologists will assess study imaging to determine overall tumor response at each timepoint (CR, CRu, PR, SD, or PD). The adjudicator will evaluate cases where disagreement exists between the primary radiologists and will choose one of the primary reviewer's assessments as the final assessment. The IRC primary radiologists will review all timepoints available for each subject in a sequential fashion.

5.1.3.2. Date of Progression

For analyses based on the IRC tumor assessment, the date of progression will be the earliest date of PD from imaging studies.

For analyses based on the investigator's tumor assessment, the date of progression will be the earliest date of PD from imaging study (collected on Evaluation of Response page on

CRF) or date of PD other than imaging study (collected on PD page of CRF), whichever is first reported.

5.1.3.3. Complete Response Assessment

Overall complete response refers to CR and CRu, defined by meeting the requirements for CR/CRu per the modified IWRC criteria (Attachment) .

A CR must meet the CR criteria based on radiological imaging study, and subsequently verified by the laboratory tests with normal LDH and negative bone marrow at screening or at the time of the radiological CR (from central pathology laboratory Phenopath/Covance). A CRu is CRu based on radiological imaging study, and subsequently verified by laboratory tests with normal LDH and negative bone marrow at screening or normal LDH and negative/indeterminate bone marrow at the time of the radiological CRu.

Radiological overall complete response refers to radiological CR/CRu verified or not verified by LDH and bone marrow. Radiological CR/CRu is defined by CR/CRu meeting all the criteria for CR/CRu based on radiological imaging studies, but subsequent verification by LDH and bone marrow findings is not required. The date of radiological CR/CRu is the first date that CR/CRu is documented (tumor assessment per IRC) based on radiological imaging study.

5.1.3.4. Tumor Assessment by Investigator

Tumor assessment will be performed by investigators locally based on radiological images and using the same modified IWRC criteria. Overall tumor assessment by the investigators at each timepoint is collected in the CRF. In addition, tumor assessment of PD can also be done based on clinical evidence other than radiological images, which is collected in the CRF. Analyses for investigator tumor assessment are secondary analyses.

5.1.3.5. Best Overall Response

The best overall response is the best response recorded from the start of the treatment until disease progression or the start of subsequent anti-neoplastic therapy.

5.2. Primary Efficacy Endpoint(s)

5.2.1. Definition

The primary endpoint is PFS, which is defined as the interval between the date of randomization and the date of PD or death, whichever is first reported, in the ITT population. Subjects who withdraw from the study (i.e., withdrawal of consent, lost to follow-up) or start anti-neoplastic therapy without documented PD will be censored at the time of the last adequate (CT scan available) disease assessment. Subjects who complete the study, have not

progressed, and are still alive at the cut-off date of the final analysis will be censored at the last adequate disease assessment. For PD documented between scheduled evaluations, the actual date of PD will be used, not the date of the next scheduled evaluation. For death no more than 6-month after last disease assessment, the date of the event will be the date of death. For death more than 6-month after last disease assessment, the PFS will be censored at the last adequate disease assessment. The censoring method is summarized in Table 1.

Table 1: Censoring Method

Situation	Date of Progression or Censoring	Outcome
No baseline tumor assessments	Randomization	Censored
Progression documented between scheduled visits	Earliest date of radiological assessment documenting progression	PFS event
No progression	Date of last visit with adequate assessment	Censored
New anticancer treatment started	Date of last visit with adequate assessment (prior to subsequent anticancer treatment)	Censored
Death before first PD assessment	Date of death	PFS event
Death between adequate assessment visits	Date of death	PFS event
Death within 6-month after last disease assessment	Date of death	PFS event
Death more than 6-month after last disease assessment	Date of last visit with adequate assessment	Censored

The primary analysis for PFS will be based on IRC assessment of PD. Progression-free survival derived from investigator data will be evaluated and used as supportive evidence. Sensitivity analysis on PFS using IRC alternative assessments of transient fluid collection or transient lesion as basis for PD will be performed.

The primary analysis for PFS will also be conducted among the subjects with centrally confirmed MCL.

Sensitivity analysis for PFS will be conducted by not censoring a subject who started subsequent antineoplastic therapy without documented PD. Sensitivity analysis for PFS will also be conducted by censoring a subject whose PD is after more than one missing adequate disease assessment and censoring date is the last adequate disease assessment before PD.

5.2.2. Analysis Methods

The Kaplan-Meier method will be used to estimate the distribution of overall PFS for each treatment group. The primary treatment comparison is based on a stratified log-rank test. Hazard ratio and its 95% confidence interval are to be estimated based on a stratified Cox's

model with treatment as the explanatory variable. Stratification factors include International Prognostic Index (IPI: 0-1, 2, 3, and 4-5) and stage of disease at diagnosis (II, III, and IV). The sensitivity analyses for PFS among centrally confirmed MCL subjects, for PFS without censoring subsequent therapy, and for PFS censoring after more than one missing adequate disease assessment will be performed similarly.

The unstratified log-rank test will also be performed as a sensitivity analysis.

The subgroup analyses for PFS will be done as specified in Section 2.3.

A Cox's regression model will be used in sensitivity analyses. It is intended to estimate the effect size of VcR-CAP vs. R-CHOP after accounting for prognostic factors. The following prognostic factors will be included in the analysis: age (>65 , ≤ 65), sex, race (white, non-white), baseline ECOG performance status (≤ 1 , >1), baseline extranodal site involvement (≤ 1 , >1), baseline LDH (normal, above normal limits), and WBC (normal, above normal limits), together with stage of disease at diagnosis (II and III vs. IV).

Concordance rate between investigator PD and IRC PD will be provided. Agreement on PD status includes PD by both IRC and investigator or no PD by both of them. In addition, PD by both IRC and investigator will be categorized into sub categories of complete agreement, agreement with later date, or agreement with early date. Disagreement on PD status includes PD by investigator but not IRC and PD by IRC but not investigator.

The total number of PD events adjudicated will be given. For subjects without PD, the main reason for censoring will also be tabulated.

5.3. Major Secondary Endpoints

5.3.1. Definition

Time to Progression

Time to Progression is defined as the duration from the date of randomization until the date of first documented evidence of PD or relapse for subjects who experience CR or CRu on this study. Death due to PD will be considered as an event if the date of death is within 6-month after last disease assessment, otherwise, death will be censored at the date of last disease assessment. Subjects who withdraw from the study (i.e., withdrawal of consent, lost to follow-up) or start subsequent therapy or die of causes other than PD will be censored at the time of the last disease assessment.

Overall Response Rate (ORR)

Overall Response Rate is defined as the proportion of subjects who achieve CR, CRu, or PR relative to the response evaluable population. Disease response and progression will be evaluated according to the modified IWRC.

Complete Response (CR/CRu) Rate

Complete response rate is defined as the proportion of subjects who achieve CR or CRu relative to the response evaluable population.

Radiological complete response rate is defined as the proportion of subjects who achieve radiological CR or radiological CRu relative to the response evaluable population.

Time to Next Antilymphoma Treatment

Time (days) to next anti-lymphoma treatment is measured from the date of randomization to the start date of new anti-lymphoma treatment. Death due to disease progression prior to subsequent therapy is considered as an event. Otherwise, time to next anti-lymphoma treatment is censored at the date of death or the last date known to be alive.

Overall Survival (OS)

Overall survival is measured from the date of randomization to the date of the subject's death. If the subject is alive or the vital status is unknown, OS will be censored at the date that the subject is last known to be alive.

Duration of Response and Duration of Complete Response

Duration of radiological response (CR, CRu, or PR) will be calculated from the date of initial documentation of a response to the date of first documented evidence of PD or death due to PD. Responders who withdraw from the study (i.e., withdrawal of consent, lost to follow-up) or who start subsequent anti-neoplastic therapy without documented PD will be censored at the time of the last adequate (CT scan available) disease assessment. Responders who complete the study, have not progressed, and who are still alive at the cut-off date of the final analysis will be censored at the last adequate disease assessment. For death due to PD more than 6-month after last disease assessment, the duration will be censored at the last adequate disease assessment before PD or death.

Duration of complete response (CR, CRu) will be calculated from the date of initial documentation of a complete response to the date of first documented evidence of PD or death due to PD. Subjects who have a complete response without disease progression or death due to PD will be censored in the same way as in duration of response.

5.3.2. Analysis Methods

Time to Progression

The analysis method will be similar to that used for the primary PFS. The analysis of TTP is considered sensitivity analyses for PFS.

Response-Related Endpoints

The number and percentage of subjects in the following response categories will be presented by treatment group: complete response (CR), unconfirmed complete response (CRu), overall complete response (CR+CRu), partial response (PR), overall response (CR+CRu+PR), stable disease (SD), and progressive disease (PD).

Overall Response Rate (ORR)

Overall response rate is to be presented by treatment group. Stratified Cochran-Mantel-Haenszel (CMH) test will be used to test treatment difference. The Mantel-Haenszel estimate of odds ratio and its 95% confidence interval are to be presented as well.

Both ORR based on IRC assessment (primary) and on investigator assessment (supportive) will be evaluated and analyzed for the response evaluable population.

Complete Response (CR/CRu) Rate

Both complete response rate and radiological complete response rate are to be presented by treatment group. Stratified CMH test will be used to test treatment difference. The Mantel-Haenszel estimate of odds ratio and its 95% confidence interval are to be presented as well.

Analyses will be based on the response evaluable population.

Time to Next Antilymphoma Treatment

The Kaplan-Meier method will be used to estimate the distribution of time to next anti-lymphoma treatment for each treatment group. The primary treatment comparison is based on a stratified log-rank test. The hazard ratio and its 95% confidence interval are to be estimated based on the Cox's regression model.

Overall Survival (OS)

The Kaplan-Meier method will be used to estimate the distribution of OS for each treatment group. The primary treatment comparison is based on a stratified log-rank test. Hazard ratio and its 95% confidence interval are to be estimated based on a stratified Cox's model with treatment as the explanatory variable. Stratification factors include International Prognostic Index (IPI: 0-1, 2, 3, and 4-5) and stage of disease at diagnosis (II, III, and IV).

In addition, the probability of survival at 18 months and 3-year after randomization, based on Kaplan-Meier estimates, will be presented for the 2 treatment groups, along with their 95% confidence interval.

Duration of Response and Duration of Complete Response

The Kaplan-Meier method will be used to descriptively summarize duration of response and duration of complete response.

5.4. Other Efficacy Variables

5.4.1. Definition

Time to Response

Time (days) to response is measured from the date of randomization to the date of initial response (CR, CRu, or PR). Those subjects without response (CR/CRu/PR) will be censored either at PD or at the last efficacy assessment before the start of subsequent therapy.

Rate of Durable Response

A durable response is defined as a response (CR, CRu, PR) with a duration ≥ 6 months. The rate of durable response is the proportion of subjects who achieve durable response relative to the response evaluable population.

Duration of Treatment Free Interval

Duration (days) of treatment free interval is measured from the date of last dose plus 1 to the start date of the new treatment.

Death due to disease progression prior to subsequent therapy is considered as an event. Otherwise, treatment-free interval is censored at the date of death or the last date known to be alive.

5.4.2. Analysis Methods

Time to Response

The Kaplan-Meier method will be used to estimate the cumulative distribution over time for time to response in the response evaluable population. The stratified log-rank test will be used to compare the 2 treatment groups. In addition, descriptive summaries are to be provided for time to response for responding subjects (CR/CRu/PR).

Rate of Durable Response

For rate of durable response, the analysis method will be similar to the one used for ORR. For subjects who don't have durable response, numbers of subjects whose best response is worse than PR, numbers of subjects whose best response are better or equal PR with

censored value and duration <6 months, and numbers of subjects whose best response are better or equal PR with event and duration <6 months will be given separately.

Duration of Treatment Free Interval

The Kaplan-Meier method will be used to descriptively summarize the duration of the treatment-free interval.

Subsequent Anti-Lymphoma Therapy

Frequency tabulation will be given, indicating how many subjects in each group started subsequent therapy.

6. SAFETY

Safety of the treatment arms will be compared using the incidence and severity of adverse events, laboratory test results, ECG/echo or MUGA results, vital sign measurements, physical examination results, mortality, and ECOG performance status.

Safety analyses will be performed by the treatment actually received. That is, those subjects who are randomized to VcR-CAP but never receive VELCADE will be analyzed in the R-CHOP group (Treatment Arm B), and those who are randomized to R-CHOP but mistakenly receive at least one dose of VELCADE will be analyzed in the VcR-CAP group (Treatment Arm A).

MRU data will be analyzed in a separate report.

6.1. Adverse Events

Treatment-emergent adverse events are defined as any adverse event that occurs on or after administration of the first dose of study medication (VELCADE, rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone) and through the End of Treatment visit to 30 days after the last dose of study drug, any event that is considered drug-related (very likely, probably or possibly related) regardless of the start date of the event, or any event that is present at baseline but worsens in intensity or is subsequently considered drug-related by the investigator.

The most recent MedDRA dictionary will be used to code the investigator's adverse event terms. Adverse event tables will summarize subject incidence by treatment group actually received. The severity of adverse events will be assessed using NCI CTC v3.0.

6.1.1. All Adverse Events

The incidence of treatment-emergent adverse events will be summarized. The following tabulations will also be presented:

- Overview of treatment-emergent adverse events
- Incidence of treatment-emergent adverse events, especially the most commonly reported events (i.e., those events reported by at least 10% of subjects) by MedDRA SOC and preferred term.
- Incidence of treatment-emergent adverse events by SOC, high level term, and preferred term.
- Incidence of treatment-emergent adverse events considered by the investigator to be related to study medication, by SOC and preferred term.
- Incidence of treatment-emergent adverse events by relationship to study medication.
- Incidence of treatment-emergent adverse events by CTC toxicity grade.
- Incidence of treatment-emergent adverse events by CTC toxicity grade and relationship to study medication.
- Incidence of AEs leading to withdrawal from treatment by SOC and preferred term
- Incidence of AEs leading to dose reduction, dose withholding, and cycle delay by SOC and preferred term
- Incidence of treatment-emergent adverse events leading to death by system organ class and preferred term.
- Incidence of AEs leading to the discontinuation of VELCADE alone, AEs leading to the discontinuation of VELCADE and at least one other study drug, and AEs leading to the discontinuation of study treatment.

Incidence of treatment-emergent adverse events (overview, all events, Grade ≥ 3 , and serious) will also be summarized by Cycles 1, 2-3, 4-6, vs. >6 . In case a subject experiences multiple adverse events of the same preferred term during the study, all new onsets will be summarized. For treatment-emergent adverse events with CTC toxicity grade of 3 or higher, the following summaries are to be produced.

- Incidence of Grade ≥ 3 treatment-emergent adverse events, by SOC and preferred term.
- Incidence of Grade ≥ 3 treatment-emergent adverse events considered by the investigator to be related to study medication, by SOC and preferred term.
- Incidence of Grade ≥ 3 treatment-emergent adverse events by relationship to study medication.

For treatment-emergent serious adverse events, the following summaries are to be produced.

- Incidence of treatment-emergent serious adverse events, by SOC and preferred term.
- Incidence of treatment-emergent serious adverse events considered by the investigator to be related to study medication, by SOC and preferred term.

- Incidence of treatment-emergent serious adverse events by relationship to study medication.

Analyses of adverse events leading to treatment discontinuation are specified in Section 2.4.18.

6.1.2. Adverse Events of Clinical Interest

Unless specified otherwise, the adverse events in this section are treatment-emergent ones.

6.1.2.1. Peripheral Neuropathies

Definition

Peripheral neuropathies (PN) refer to the adverse events in the high level term of peripheral neuropathies NEC.

All the PN event records of the same preferred term from the same subject are to be linked by the onset date and the resolution date. If a resolved event is followed by another event of the same preferred term with the onset date within 1 day of the resolution date of the previous record, these 2 records should be linked together and considered as 1 event.

A PN event is considered as resolved if its final outcome is resolved with no subsequent PN event of the same preferred term occurring on the resolution date or the day before and after.

A PN event is considered as improved if the event improves from the maximum grade. That is, all the grades recorded after the maximum grade is less than the maximum grade.

Time to first onset of peripheral neuropathy events (any grade, grade ≥ 2 , and grade ≥ 3) is measured from the first dosing date to the initial onset date of event.

Time to improvement and time to resolution are to be defined for each PN event.

Time to resolution is defined as the number of days from the initial onset date (inclusive) to the resolution date for resolved events. Those events that are not resolved will be censored at the date of last study drug dose plus 30 days for Grade 1 events (or the data cut-off if earlier) and the date of last study drug dose plus 180 days for Grade ≥ 2 events (or the data cut-off if earlier). This is in line with the protocol specifications on adverse event reporting, indicating that unresolved neuropathic adverse events Grade 2 or higher should be followed for a maximum of 6 months, while general adverse events are reported until at least 30 days following last dose of study medication (Protocol 26866138LYM3002, Section 9.6)

For improved events, time to improvement is defined as the number of days from the initial onset date (inclusive) of the maximum grade to the first onset date that the toxicity grade is below the maximum grade with no higher grade thereafter, or the resolution date, whichever occurs first. For those events that stay at the maximum grade until the end of follow-up, time to improvement will be censored at the date of last study drug dose plus 180 days if the maximum grade is at least 2, and the date of last study drug dose plus 30 days for Grade 1 events.

Method

The incidence of treatment-emergent peripheral neuropathies will be summarized by high level term and preferred term.

Kaplan-Meier estimates of probability of developing peripheral neuropathy events (any grade, grade ≥ 2 , and grade ≥ 3) will be presented for time to the first onset in the two treatment arms, and for cumulative VELCADE dose up to the first onset in the VcR-CAP arm. Subjects without any peripheral neuropathy events will be censored at 30 days after last dose, or at total VELCADE dose received in these analyses.

Furthermore, time to improvement and time to resolution of peripheral neuropathy events will be summarized by outcome (improvement or resolution) using the Kaplan-Meier method. This analysis is event based, thus 1 subject could contribute multiple observations if the subject has more than 1 PN event.

6.1.2.2. Cardiac Rhythm and Conduction Abnormalities

Definition

Cardiac rhythm and conduction abnormalities include adverse events in the following high level terms (HLTs): cardiac conduction disorders, cardiac signs and symptoms NEC (palpitations only), rate and rhythm disorders NEC, supraventricular arrhythmias, ventricular arrhythmias, and cardiac arrest.

Method

The incidence of treatment-emergent adverse events associated with cardiac rhythm and conduction abnormalities will be tabulated.

6.1.2.3. Heart Failure

Definition

This endpoint refers to those adverse events coded to the following preferred terms: acute left ventricular failure, acute pulmonary oedema, acute right ventricular failure, cardiac asthma, cardiac failure, cardiac failure acute, cardiac failure chronic, cardiac failure congestive, cardiac failure high output, cardiogenic shock, cardiopulmonary failure,

cardiorenal syndrome, chronic left ventricular failure, chronic right ventricular failure, cor pulmonale, cor pulmonale acute, cor pulmonale chronic, ejection fraction decreased, hepatic congestion, hepatojugular reflux, left ventricular failure, low cardiac output syndrome, neonatal cardiac failure, pulmonary oedema, pulmonary oedema neonatal, ejection fraction decreased, and diastolic dysfunction.

Method

Incidence of treatment-emergent heart failure events will be summarized.

6.1.2.4. Pericardial Disease

Definition

Pericardial disease includes adverse events in the high level group term (HLGT) of “Pericardial disorders” and high level term of noninfectious pericarditis.

Method

The incidence will be presented.

6.1.2.5. Acute Diffuse Infiltrative Pulmonary Disease

Definition

This endpoint refers to those adverse events coded to the following preferred terms: acute interstitial pneumonitis, acute respiratory distress syndrome, allergic granulomatous angiitis, alveolar proteinosis, alveolitis, alveolitis allergic, alveolitis necrotising, bronchiolitis, organising pneumonia, diffuse alveolar damage, diffuse panbronchiolitis, eosinophilic pneumonia, eosinophilic pneumonia acute, eosinophilic pneumonia chronic, idiopathic pneumonia syndrome, idiopathic pulmonary fibrosis, interstitial lung disease, loeffler’s syndrome, lower respiratory tract inflammation, lung induration, lung infiltration, acute lung injury, necrotising bronchiolitis, obliterative bronchiolitis, pneumonitis, progressive massive fibrosis, pulmonary eosinophilia, pulmonary fibrosis, pulmonary necrosis, pulmonary toxicity, pulmonary alveolar haemorrhage, pulmonary haemorrhage, and non-cardiogenic pulmonary oedema.

Method

The incidence will be presented.

6.1.2.6. Reactivation of Hepatitis B

Definition

Reactivation of Hepatitis B includes adverse events in the following preferred terms: Hepatitis B, Acute hepatitis B, Hepatitis acute, Asymptomatic viral hepatitis, Chronic hepatitis B, Hepatitis post transfusion, Hepatitis Viral.

Prophylactic use of anti-virals of hepatitis B includes lamivudine, entecavir, entecavir monohydrate, thymalfasin, with prophylaxis checked in the CRF.

Method

The incidence will be presented. The incidence of hepatitis B by prophylactic use of anti-virals of hepatitis B will be tabulated.

6.1.2.7. Acute Hypersensitivity

Definition

This endpoint refers to those adverse events coded to the following preferred terms: allergic oedema, anaphylactic reaction, anaphylactic shock, anaphylactoid reaction, anaphylactoid shock, angioedema, circumoral oedema, drug hypersensitivity, eyelid oedema, face oedema, hypersensitivity, laryngeal oedema, laryngotracheal oedema, lip oedema, lip swelling, oedema mouth, periorbital oedema, small bowel angioedema, tongue oedema, type I hypersensitivity, type IV hypersensitivity reaction, swelling face, infusion site urticaria, and urticaria.

Method

The incidence will be presented. Incidence of study drug related treatment-emergent acute hypersensitivity adverse events will be tabulated by study drugs. Listing of treatment-emergent acute hypersensitivity adverse events will be provided.

6.1.2.8. Central Nervous System Disorders

Definition

This endpoint refers to those adverse events coded to the following preferred terms: Leukoencephalopathy, Progressive multifocal leukoencephalopathy, Posterior reversible encephalopathy syndrome, Guillain-Barre Syndrome, Encephalitis and Encephalopathy.

Method

The incidence will be presented.

6.1.2.9. Herpes Zoster Infection

Definition

This endpoint refers to those adverse events coded to the following preferred terms: Encephalitis herpes, Herpes zoster multi-dermatomal, Herpes ophthalmic, Herpes zoster ophthalmic, Herpes sepsis, Herpes zoster oticus, Herpes zoster, Keratitis herpetic, Herpes zoster disseminated, Meningitis herpes, Herpes zoster infection neurological, Meningoencephalitis herpetic.

Prophylactic use of anti-virals of herpes zoster includes aciclovir, valaciclovir, famciclovir, valaciclovir hydrochloride, aciclovir sodium, with prophylaxis checked in the CRF.

Method

The incidence will be presented. The incidence by prophylactic use of anti-virals of herpes zoster will be tabulated.

6.1.2.10. Thrombocytopenic Bleeding Events

Definition

This endpoint refers to the adverse events with the following preferred terms:

Abdominal wall haematoma, Abdominal wall haemorrhage, Acute haemorrhagic conjunctivitis, Adrenal haematoma, Adrenal haemorrhage, Anal haemorrhage, Anal ulcer haemorrhage, Application site haematoma, Application site haemorrhage, Basal ganglia haemorrhage, Bloody discharge, Brain stem haematoma, Brain stem haemorrhage, Brain stem microhaemorrhage, Breast haematoma, Breast haemorrhage, Bronchial haemorrhage, Catheter site haematoma, Catheter site haemorrhage, Central nervous system haemorrhage, Cerebellar haematoma, Cerebellar haemorrhage, Cerebellar microhaemorrhage, Cerebral haematoma, Cerebral haemorrhage, Cerebral microhaemorrhage, Cerebrovascular accident, Cervix haematoma uterine, Cervix haemorrhage uterine, Choroidal haematoma, Choroidal haemorrhage, Ciliary body haemorrhage, Colonic haematoma, Conjunctival haemorrhage, Corneal bleeding, Cystitis haemorrhagic, Duodenitis haemorrhagic, Ear haemorrhage, Enterocolitis haemorrhagic, Epistaxis, Exsanguination, Extradural haematoma, Extravasation blood, Eye haemorrhage, Eyelid bleeding, Gastric haemorrhage, Gastroduodenal haemorrhage, Gastroduodenitis haemorrhagic, Gastrointestinal haemorrhage, Gastrointestinal ulcer haemorrhage, Genital haemorrhage, Gingival bleeding, Haemarthrosis, Haematemesis, Haematochezia, Haematocoele, Haematoma, Haematosalpinx, Haematotympanum, Haemoptysis, Haemorrhage, Haemorrhage intracranial, Haemorrhage urinary tract, Haemorrhagic anaemia, Haemorrhagic cerebral infarction, Haemorrhagic disorder, Haemorrhagic stroke, Intestinal haematoma, Intestinal haemorrhage, Intracranial haematoma, Intracranial tumour haemorrhage, Intraocular haematoma, Intraventricular haemorrhage, Iris haemorrhage, Lacrimal haemorrhage, Large intestinal haemorrhage, Large intestinal ulcer haemorrhage, Lower gastrointestinal haemorrhage, Melaena, Meningorrhagia, Mesenteric haematoma, Mesenteric haemorrhage, Mouth haemorrhage, Ocular retrobulbar haemorrhage, Oesophageal haemorrhage, Oesophagitis haemorrhagic, Optic disc haemorrhage, Optic nerve sheath haemorrhage, Ovarian haematoma, Ovarian haemorrhage, Pelvic haematoma, Pelvic haematoma obstetric, Pelvic haemorrhage, Penile haematoma, Penile haemorrhage, Perineal haematoma, Periorbital haematoma, Periorbital haemorrhage, Pituitary

haemorrhage, Prostatic haemorrhage, Putamen haemorrhage, Renal haematoma, Renal haemorrhage, Retinal haemorrhage, Retinopathy haemorrhagic, Scleral haemorrhage, Scrotal haematocoele, Scrotal haematoma, Spermatic cord haemorrhage, Spinal cord haemorrhage, Spinal epidural haemorrhage, Spinal haematoma, Subarachnoid haemorrhage, Subdural haematoma, Subdural haemorrhage, Subretinal haematoma, Testicular haemorrhage, Thalamus haemorrhage, Traumatic intracranial haemorrhage, Ureteric haemorrhage, Urethral haemorrhage, Uterine haematoma, Uterine haemorrhage, Vaginal haematoma, Vaginal haemorrhage, Vitreous haematoma, Vitreous haemorrhage, Vulval haematoma, Vulval haemorrhage, rectal haemorrhage, intra-abdominal haematoma, post procedural haemorrhage, pulmonary haemorrhage, pulmonary alveolar haemorrhage, conjunctival haemorrhage, upper gastrointestinal haemorrhage, haemorrhoidal haemorrhage.

Method

The incidence will be presented. To review thrombocytopenic bleeding AEs, all grade 3 or more bleeding events where bleeding was associated with grade 3 or above thrombocytopenia will be reviewed. The thrombocyte counts prior to the bleeding event will be taken into consideration. The platelet transfusion during the cycle where thrombocytopenic bleeding AEs took place will also be reviewed. These events will be reviewed by 2 different physicians (project physician and GMS-P) independently.

6.1.2.11. Neutropenic Sepsis and Infections Events

Definition

This endpoint refers to the adverse events in the SOC of infections and infestations.

Method

The incidence will be presented. To review neutropenic infection/sepsis AEs, Infection SOC will be taken as a starting point and all grade 3 and above infections will be taken into account if patients who have experienced grade 3 or above infection/sepsis has also grade 3 or above neutropenia at the time of the infection. The G-CSF use during the cycle where the neutropenic infection/sepsis took place will also be reviewed. These events will be reviewed by 2 different physicians (project physician and GMS-P) independently.

6.1.3. Death

A tabulation will be produced for all-cause mortality within 30 days and 60 days after the first dose, within 30 days of the last dose, as well as all deaths for all subjects treated. Cause of death will be summarized as well in this table. In particular, frequencies of deaths that are due to study treatment-related adverse events will also be reported.

6.2. Clinical Laboratory Tests

Laboratory data of hematology and clinical chemistry up to 30 days after last dose or the end of treatment visit date, whichever is later, will be reported in SI units. Normal ranges from local laboratories used in this study will be listed.

Summary statistics (mean, standard deviation, median and range) will be calculated for the raw data and for their changes from baseline at each time point of assessment as well as for the changes from baseline to the last value.

Graphical displays of over-time summaries will be presented for the following hematology parameters: hemoglobin, neutrophils, lymphocytes, platelets.

Shift tables for each cycle will be produced for neutrophils and platelets. These tables will summarize by cycle the number of subjects with each baseline CTC grade and changes to the maximum CTC grade in the cycle.

Changes from baseline to worst value on study (from treatment start to 30 days after last dose or the end of treatment visit date, whichever is later) will also be provided. The worst toxicity grade during the study will be tabulated.

Due to the asymmetric assessment schedule between the two treatment groups for hematology laboratory tests, the shift tables and the summary of worst toxicity grade will be repeated excluding the samples taken on Days 4 and 8 of each cycle for those subjects treated with VcR-CAP.

6.3. Vital Signs and Physical Examination Findings

Over-time summary statistics (mean, standard deviation, median and range) of vital signs will be provided. A separate summary will be produced for vital signs at baseline, maximum, change to maximum, last value, and change to last value.

In addition, any significant vital sign changes will be tabulated. Significant findings are defined as follows:

- Systolic blood pressure: changes of +/- 20% or greater from baseline (separately)
- Diastolic blood pressure: changes of +/- 20% or greater from baseline (separately)

In order to be included in the table, a subject must have both a baseline value and a value for the given post-baseline time point.

Abnormal physical examination findings will be tabulated by body system.

6.4. Electrocardiogram and ECHO/MUGA Scans

A tabulation of ECG assessments will be made for the clinically significant changes in ECG evaluation during the study versus the baseline ECG assessment. Similar analysis will be done for ECHO/MUGA scans.

7. PATIENT-REPORTED OUTCOMES (PRO) ENDPOINTS

7.1. Definition

Validated scoring algorithms will be used to derive the scores for the 3 PRO multi-item scales: 1) EORTC QLQ-C30 scales: one global scale (global health status/quality of life), 5 functional scales (physical functioning, role functioning, emotional functioning, cognitive functioning, and social functioning), plus 3 symptom scales (fatigue, nausea and vomiting, and pain) and six single items (dyspnea, insomnia, appetite loss, constipation, diarrhea and financial difficulties) 2) BFI (severity and fatigue); and 3) EQ-5D.

7.2. Analysis Methods

The number of available/missing assessments will be reported by scheduled visit and assessment/missing type.

Absolute values of the EORTC QLQ-C30 and BFI scale scores (worst fatigue and interference scales) and changes from baseline will be summarized (N, mean, SD, minimum, maximum) at each scheduled post-baseline assessment visit, along with graphs.

For each of the scales in the EORTC QLQ-C30 and worst fatigue scale of the BFI, a repeated-measures model to estimate changes from baseline to post-baseline assessments will be conducted. Treatment, time, and the interaction between time and treatment will be included as discrete parameters. Serial correlation will be assumed using an unstructured covariance matrix to account for the correlations between all time points. Change and corresponding standard error in each outcome measure on each time point up to Week 18 will be estimated for each treatment group separately and treatment and control will be compared.

Graphics on EORTC QLQ-C30 and worst fatigue scale of the BFI will be displayed with visits as the X-axis, and means (SEs) of EORTC QLQ-C30 and worst fatigue scale of the BFI at each visit as the Y-axis.

Cumulative distribution curves will be constructed for global health status in EORTC QLQ-C30 and worst fatigue scale of the BFI scales and displayed graphically. The curves will show a continuous plot of the absolute change of the last PRO measurement from baseline on the X-axis and the percent of subjects experiencing that change on the Y-axis for each of the treatment groups.

Additional PRO Endpoints

The EQ-5D assesses 5 dimensions (mobility, self care, usual activity, pain or discomfort, and anxiety or depression) of health related to PRO. Each dimension is measured on a 3-point ordinal scale ranging between 1 (no limitation) and 3 (greatest limitation) where a higher score corresponds to a worse health state.

Responses in EQ-5D enable subjects to be assigned to 1 of 243 possible health states for which a global value (i.e., a "utility score") is assigned. The utility score has been previously determined based on the preferences of a sample of 3,395 members of the United Kingdom general population. A utility score of 1 corresponds to the highest degree of HRQOL, and score of 0 is compatible with a level of HRQOL equivalent to death.

EQ-5D utility score

Definition

The EQ-5D utility scores will be computed using the UK value set obtained using Time Trade-Off (TTO) methodology from the table below.

Value set for Calculation of the EQ-5D Scoring.	
UK TTO Value set	
	Weight
Full Health (11111)	1
At least one 2 or 3 (constant)	-0.081
At least one 3	-0.269
Mobility = 2	-0.069
Mobility = 3	-0.314
Self care = 2	-0.104
Self care = 3	-0.214
Usual activities = 2	-0.036
Usual activities = 3	-0.094
Pain/Discomfort = 2	-0.123
Pain/Discomfort = 3	-0.386
Anxiety/Depression = 2	-0.071
Anxiety/Depression = 3	-0.236

For example, the EQ-5D utility score for the health state 21232 is:

$$\text{Score}_{21232} = 1 - 0.081 - 0.269 - 0.069 - 0.036 - 0.386 - 0.071 = 0.088$$

Analysis Methods

Absolute values of the utility score for each of the 5 dimensions of EQ-5D, VAS, and their changes from baseline will be summarized at each scheduled post-baseline assessment visit (N, mean, SD, minimum, maximum). Exploratory analysis will be performed when necessary.

8. BIOMARKER EVALUATIONS

Biomarker analysis will be restricted to the following two assessments based on results from previous studies [Coiffier 2013; Goy 2010] (1) Pharmacogenomics and (2) Immunohistochemistry. The goal of the biomarker analysis is to provide some insight as to how subjects with certain prognostic markers (e.g. Ki67) respond to treatment or validate/confirm observations from previous studies that identified markers (e.g. PSMB) associated with better response to treatment.

8.1. Definition

Pharmacogenomic Analysis

Genetic variability in drug target genes may influence drug binding and subsequently, response to treatment. Suspected drug target genes for VELCADE, including *PSMB1(P11A)* and *PSMB5(R24C)*, which are proteasome subunits will be prospectively analyzed using whole blood DNA samples collected from consented subjects to determine any associations with response and PFS.

Immunohistochemistry Analysis

Paraffin-embedded, formalin-fixed tumor or fresh frozen tissue will be used for immunohistochemical analysis to quantify the tissue protein expression of the previously reported prognostic marker, Ki67.

8.2. Analysis Methods

Pharmacogenomic Analysis

The pharmacogenomic exploratory statistical analysis will be performed at the completion of the trial in order to inform subsequent lymphoma trials. Results from PSMB gene analysis will include investigation of association of these genes with clinical endpoints including PFS and complete response. Analyses will be performed within each treatment group and stratified by demographic and or baseline features such as race, tumor burden, etc. Results of this prospective analysis will be presented in a separate report with key summary points included in the CSR.

Immunohistochemistry Analysis

The protein levels of Ki67 will be measured as scores on an ordinal scale with a cut off $\leq 10\%$ as negative and $>10\%$ as positive. The impact of protein expression level on response endpoints will be tested using standard categorical tests or survival analysis methods, as appropriate. Results of Ki67 analyses will be presented in a separate report with key summary points included in the CSR.

REFERENCES

1. Cheson BD, Pfistner B, Juweid ME, et al. Revised response criteria for malignant lymphoma. *J Clin Oncol* 2007; 25 (5):579-586.
2. Coiffier B, Henitz ED, Karkera J, et al. Pre-specified candidate biomarkers identify subjects with relapsed or refractory follicular lymphoma who achieved longer progression-free survival after treatment with bortezomib plus rituximab compared to rituximab alone. Under review 2013.
3. Goy A, Bernstein SH, McDonald A, et al: Potential biomarkers of bortezomib activity in mantle cell lymphoma from the phase 2 PINNACLE trial. *Leuk Lymphoma* 2010; 51(7):1269-1277.

ATTACHMENT

Tumor Assessment Criteria (Modified IWRC)

All tumor assessments will be done according to modified International Workshop to Standardize Response Criteria for non-Hodgkin's Lymphoma (IWRC) [Cheson 2007]). Relapsed disease is classified the same as PD by IRC. The criteria that must be met for each disease response category, CR, CRu, PR, SD, disease relapse and progression, are provided below.

Complete Response (CR) requires ALL of the following:

- Complete disappearance of all detectable clinical and radiological evidence of disease and disease-related symptoms and normalization of biochemical abnormalities definitely assignable to lymphoma (e.g., LDH) if present before therapy.
- All measurable lymph nodes and nodal masses must have regressed on CT to normal size (≤ 1.5 cm in their greatest transverse diameter for nodes > 1.5 cm before therapy).
- Non-measurable and assessable nodes that were 1.1 to 1.5 cm in their greatest transverse diameter before treatment must have decreased to ≤ 1 cm in their greatest transverse diameter after treatment, or by more than 75% in the sum of the products of the greatest diameters (SPD), as visually estimated.
- The spleen or liver, if considered enlarged due to involvement with lymphoma before therapy on the basis of a physical examination or CT scan, should not be palpable on physical examination and should be considered normal size by imaging studies. Similarly, other organs considered to be enlarged before therapy due to involvement by lymphoma such as kidneys, must have decreased in size.
- If the bone marrow was involved by lymphoma, indeterminate or not adequately assessed during screening, an adequate aspirate and biopsy of the same site must be clear of lymphoma.
- All extranodal sites of disease must have completely disappeared.

Unconfirmed Complete Response (CRu) requires:

That the first and fourth criteria for CR be satisfied, however:

- Any residual lymph node mass > 1.5 cm in longest transverse dimension or extranodal site of disease (irrespective of size) must have regressed by more than 75% of the product of the longest perpendicular dimensions compared to the pretreatment baseline.
- The bone marrow aspirate may be indeterminate (contain increased number or size of lymphoid aggregates without cytologic or architectural atypia).
- If there are residual masses in a subject who would otherwise be considered to have achieved a CR or CRu, the subject should be classified as a partial responder.

Partial Response (PR) requires ALL of the following:

- At least a 50% decrease in sum of the product of the diameters (SPD) of the measurable sites of disease.

- No increase should be observed in any site of disease that meet the criteria for relapsed or progressive disease.
- Non-measurable nodes and nodules must regress by $\geq 50\%$ in their SPD or, for single non-measurable lesions, in the greatest transverse diameter, as visually estimated.
- Bone marrow assessment is irrelevant for determination of a PR if the sample was positive before treatment. However, subjects who achieve a CR by the above criteria, but who have persistent morphologic bone marrow involvement will be considered partial responders. When the bone marrow was involved before therapy and a clinical CR was achieved, but with no bone marrow assessment after treatment, subjects should be considered partial responders.
- No new sites of disease should be observed that meet the criteria for relapsed or progressive disease.

Stable disease (SD) is defined as the following:

- A subject is considered to have SD when he or she fails to attain the criteria needed for a CR or PR, but does not fulfill those for progressive disease (see below).

Progressive Disease (after PR/SD) or Relapsed Disease (after CR/CRu)

Progressive or relapsed disease requires any one of the following:

1. A) $\geq 50\%$ increase from nadir in the SPD of all measurable sites of disease at the time that progressive or relapsed disease is identified and the absolute change in at least 1 dimension is ≥ 0.5 cm for at least 1 lesion; or B) $\geq 50\%$ increase in the long axis of any measurable site of disease at the time that progressive or relapsed disease is identified and the absolute change in the long axis is ≥ 0.5 cm.
2. A) $\geq 50\%$ increase from nadir in the SPD of all non-measurable sites of disease (excluding truly assessable disease), as visually estimated, and the absolute change in at least 1 dimension is ≥ 0.5 cm for at least 1 non-measured lesion as estimated visually; or B) $\geq 50\%$ increase in the long axis of any non-measurable site of disease (excluding truly assessable disease), and the absolute change in the long axis is ≥ 0.5 cm, as estimated visually.
3. $\geq 50\%$ increase from nadir in any truly assessable site of disease, as visually estimated.
4. Appearance of any new lymph node site of disease that measures >1.5 cm in long axis and >1.0 cm in short axis, any new unequivocal extranodal site of disease (irrespective of size), or unequivocal evidence of a new site of assessable disease (for example effusions, ascites, masses with indistinct borders, new involvement of the bone marrow).
5. Appearance of a new organ enlargement or unequivocal increase of an organ enlargement that was present since baseline.

Measurable extranodal disease should be assessed in a manner similar to that for nodal disease. For these recommendations, the spleen is considered nodal disease. Disease that is only assessable (e.g., pleural effusions, bone lesions) will be recorded as no change, increased, decreased, or new, unless, while an abnormality is still noted by imaging studies or physical examination, it is found to be histologically negative.