DISCLOSURE

REDACTED STATISTICAL ANALYSIS PLAN

CC-5013-MCL-002

A PHASE 2, MULTICENTER, RANDOMIZED OPEN-LABEL STUDY TO DETERMINE THE EFFICACY OF LENALIDOMIDE (REVLIMID®) VERSUS INVESTIGATOR'S CHOICE IN PATIENTS WITH RELAPSED OR REFRACTORY MANTLE CELL LYMPHOMA

The information contained in the attached report is the property of Celgene and should not be shared or used for any purpose other than that for which it was provided.

Celgene is committed to providing information about its clinical trials to researchers and patients with the goal of furthering science and enhancing healthcare worldwide. Laws and regulations require, however, that Celgene protects patient privacy. The company may further have legal or contractual obligations not to disclose commercial or technical information provided by or related to certain partner companies or vendors.

The attached report is presented in its original format, but certain information has been redacted in order to comply with the aforementioned obligations or to protect Celgene's confidential commercial information. The redactions are based on the following principles:

- Redacted information has been replaced by grey space, maintaining original spacing and pagination.
- Any information that might allow the identification of individuals has been redacted for anonymization.
- Attachments to this report that contain confidential information are not made available. Such attachments include those that contain identifiable patient information, such as subject listings, narratives, and profiles. They also may contain confidential commercial information such as methodologies, and hypothesis generating and exploratory analyses.
- Cross-references to these attachments (such as links to subject listings in Section 16.2) are not redacted from the body of the report. However, the hyperlinks in the electronic document are no longer functional.
- Information about Celgene vendors and their services are redacted because many contracts prohibit disclosure of that information. Further, laws and regulations prevent us from disclosing certain information about our vendors or their services because it is protected by copyright.

Information about Celgene's redaction policies and the availability of additional data from this report may be found at http://www.celgene.com/research-development/clinical-trials/clinical-trials-data-sharing/.

STATISTICAL ANALYSIS PLAN

A PHASE 2, MULTICENTER, RANDOMIZED OPEN-LABEL STUDY TO DETERMINE THE EFFICACY OF LENALIDOMIDE (REVLIMID®) VERSUS INVESTIGATOR'S CHOICE IN SUBJECTS WITH RELAPSED OR REFRACTORY MANTLE CELL LYMPHOMA

STUDY DRUG:	Lenalidomide/CC-5013
PROTOCOL NUMBER:	CC-5013-MCL-002
DATE FINAL:	26JUL2010
Date Amendment 1	18JUL2011
Date Amendment 2	19DEC2012

Prepared by:
Celgene International
Route de Perreux, 1
2017 Boudry, Switzerland

TABLE OF CONTENTS

LIST OF TABLES

SIGNATURE PAGE

Study Statistician Signature _ Printed Name _
Therapeutic Head of Statistics Signature Printed Na
Clinical Research Physician Signature Printed Nam

1. LIST OF ABBREVIATIONS

Table 1: Abbreviations and Specialist Terms

Abbreviation or Specialist Term	Explanation
AE	Adverse event
AML	Acute Myelogenous Leukemia
BSA	Body surface area
CI	Confidence Interval
CR	Complete response
CRu	Complete response unconfirmed
CRF	Case report form
CTCAE	Common Terminology Criteria for Adverse Events
DMC	Data monitoring committee
DOB	Date of birth
ECG	Electrocardiogram
ECHO	Echocardiogram
ECOG	Eastern Cooperative Oncology Group
FAS	Full Analysis Set
FCBP	Female of childbearing potential
GCP	Good Clinical Practice
HR	Hazard ratio
IC	Informed consent
ICH	International Conference on Harmonization
IEC	Independent Ethics Committee
InvC	Investigator's choice
IRRC	Independent radiology review committee
ITT	Intent-to-treat
IV	Intravenous
IVRS	Interactive voice response system
KM	Kaplan Meier
Len	Lenalidomide

Abbreviation or Specialist Term	Explanation
MCL	Mantle Cell Lymphoma
MDS	Myelodysplastic syndrome
MedDRA	Medical Dictionary for Drug Regulatory Activities
NCI-CTC	National Cancer Institute Common Toxicity Criteria
ORR	Overall response rate
OS	Overall survival
PD	Progressive disease
PFS	Progression-free survival
PI	Principal Investigator The investigator who leads the study conduct at an individual study center. Every study center has a principal investigator.
PPS	Per Protocol Set
PR	Partial response
QoL	Quality of Life
SAE	Serious adverse event
SAP	Statistical analysis plan
SCT	Stem Cell Transplant
SD	Stable disease
SDev	Standard deviation
SPM	Second Primary Malignancy
TEAE	Treatment emergent adverse event
TTF	Time to treatment failure
TTP	Time to tumor progression
TTR	Time to response
WHO	World Health Organization

2. SUMMARY OF CHANGES

2.1. From version 1.2

Table 2: Summary of Changes

Section N°	Change
5.4 Sample Size	Increase of sample size was reported Rationale: Implemented following recommendation from the third study DMC to allow a more reliable estimation of potential PFS differences between the study arms after the DMC observed that the outcome in the control arm of the study was different from the initial assumptions used to calculate the sample size.
5.5 Changes to the Planned Protocol Analysis	Initially it was planned not to present all the populations if they differ by less than 10% in size. All the populations will be presented regardless of the difference in size.
7 Subject disposition	PPS population was added into the list of populations presented in the disposition table.
8.1 Demographics	Initially it was planned not to present all the populations if they differ by less than 10% in size for the demographic and baseline characteristic. All the populations will be presented regardless of the difference in size for the demographic and sensitivity and secondary analyses.
8.2 Baseline Clinical Characteristics	Update of the section with more details on the variables described.
9.3 Dose intensity	Change the denominator to total of days dosed instead of treatment duration
9.4 Relative Dose Intensity	Add 10 mg as reference dose for the moderate renal impairment.
9.5 Dose Modification	Update of the section with more details on the variables described.

Table 2: Summary of changes (Continued)

10 Concomitant medications	Update of the section with more details on the listings presented.
11 Efficacy Analysis	Clarify which population will be used for the primary analysis and for the supportive analyses. Update of the section by adding additional subgroups based on the protocol amendment #4.
12 Safety Analysis	Clarify which population will be used for the safety analyses. Update of the section by using targeted events instead of events of interest and adding paragraph on analyses of second primary malignancies
14.1 General Information	Update of the section by adding the additional DMC after 200 subjects have been treated for at least 2 cycles.
14.2 Statistical Approaches for Control of Alpha	Update of the section by adding method to control alpha level after DMC made an unplanned sample size reassessment.
16.5 Tables and listings shells	This appendix has been separated to the core SAP.

3. INTRODUCTION

This statistical analysis plan (SAP) describes the analyses and data presentations for Celgene's protocol CC-5013-MCL-002 "A Phase 2, Multicenter, Randomized Open-Label Study To Determine The Efficacy Of Lenalidomide (REVLIMID®) Versus Investigator's Choice In Subjects With Relapsed Or Refractory Mantle Cell Lymphoma." It contains definitions of analysis populations, derived variables and statistical methods for the analysis of efficacy and safety.

These analyses include one interim analysis for futility and one final analysis. Throughout this SAP, the treatment arms will be referred to as lenalidomide (Len) and Investigator's choice (InvC). The purpose of the SAP is to ensure the credibility of the study findings by specifying the statistical approaches to the analysis of study data prior to database lock for the final analysis. This SAP will be finalized and signed prior to the clinical database lock for the futility analysis. All statistical analyses detailed in this SAP will be conducted using *SAS® Version 9.1*.

4. OBJECTIVES

The objective of the statistical analyses will be to investigate the efficacy and safety of lenalidomide monotherapy in subjects with relapsed or refractory mantle cell lymphoma.

Primary objective

To compare the progression-free survival (PFS) of lenalidomide monotherapy versus investigator's choice single agent in subjects with mantle cell lymphoma (MCL) who are refractory to their regimen or have relapsed once, twice or three times..

Secondary objective

To determine the overall response rate (ORR) of lenalidomide monotherapy or investigator's choice single agent in subjects with relapsed or refractory MCL.

To evaluate the safety of lenalidomide monotherapy or investigator's choice single agent in subjects with relapsed or refractory MCL.

To determine the time to tumor progression (TTP), and overall survival (OS) of subjects with relapsed or refractory MCL who have received treatment with lenalidomide or investigator's choice single agent treatment.

To investigate the health-related quality of life (QoL) of subjects treated with lenalidomide or investigator's choice single agent treatment.

5. INVESTIGATIONAL PLAN

5.1. Overall Study Design and Plan

This is a multicenter, randomized, open-label, comparative, active controlled phase II study to determine the efficacy and safety of single agent lenalidomide over a concurrent control group treated with an investigator's choice monotherapy in subjects with MCL who are refractory to their regimen or have relapsed once, twice or three times.

This multicenter study design aims to determine the PFS hazard ratio (HR) of lenalidomide over a single agent of investigator's choice.

The investigator's choice in the control arm comprises the monotherapy treatment with one of the following: chlorambucil, cytarabine, rituximab, fludarabine, or gemcitabine. The investigator shall choose the single agent of choice in the control arm for each subject prior to randomization on to the study. Subjects in the investigator's choice arm will have the option to switch to lenalidomide at the time of progressive disease (cross over phase).

Subjects are stratified according to:

- Time since diagnosis (< 3 years or ≥ 3 years)
- Time since the end of last treatment (< 6 months [refractory] or ≥ 6 months)
- Prior stem cell transplant (yes or no).

This study will be conducted in four phases:

- A Screening Phase
- An open label randomized Treatment Phase
- A cross over phase (subjects in the investigator's choice Arm B only)
- Follow-up phase.

5.1.1. Discussion of Study Design

- Duration of Treatment

Subjects will be treated until progression in this study.

- Study Population

Subjects with MCL who have relapsed once, twice or three times are being assessed to evaluate the efficacy of lenalidomide. There is no standard of care for such subjects.

- Blinding

The open-label option was chosen because the five alternatives therapies on the InvC arm are given either intravenous (IV) or orally.

- Cross-over

Due to ethical concerns and based on the fact there is no alternative for subjects on the InvC arm who progress the possibility to cross-over to Len is offered. It is clearly a source of bias in the analysis endpoints such as OS. Cross-over will not affect the primary endpoint (PFS) as it is allowed only after progression occurs.

5.2. Study Endpoints

Following the implementation of Amendment #2 (dated 14 Dec 2009), PFS became the primary endpoint comparing PFS in subjects who receive lenalidomide monotherapy versus investigator's choice. Despite the change in the primary endpoint from ORR to PFS, there was no impact on the sample size calculation.

Primary

PFS is defined as the time from randomization to the first observation of disease progression or death due to any cause. If the subject has not progressed or died, PFS will be censored at the time of last completed assessment when the subject was known not to have progressed. Subjects who receive a new treatment without documented progression will be censored at the last assessment date that the subject is known to be progression-free.

Secondary

- ORR (complete response [CR], complete response unconfirmed [CRu], and partial response [PR]) will be assessed by a modification of the International Workshop Lymphoma Response Criteria (IWRC), which will hereafter be described as modified IWRC (Cheson,1999).
- Duration of response
- Duration of stable disease
- Tumor control rate (Rates for CR, CRu, PR, and stable disease [SD])
- Time to tumor progression (TTP)
- Time to treatment failure (TTF)
- Time to tumor response (TTR)
- Overall survival (OS)
- Safety
- Quality of Life (EORTC QLQ-C30 questionnaire)

5.3. Stratification , Randomization and Blinding

An Interactive Voice Response System (IVRS) will be utilized to ensure a 2:1 central randomization based on a permuted-block randomization method. Subjects will be randomized according to the following stratification factors: Time since diagnosis (< 3 years or ≥ 3 years), Time since last treatment (< 6 months [refractory] or ≥ 6 months), Prior stem cell transplant (yes or no).

5.4. Sample Size

Following the approval of Amendment #4 (dated 27SEP2011), the sample size increase was implemented after recommendation from the third DMC(held on July 22, 2011) to allow a more reliable estimation of potential PFS differences between the study arms after the DMC observed that the outcome in the control arm of the study was different from the initial assumptions used to calculate the sample size.

The DMC therefore recommended increasing the sample size from 174 (number of subjects randomized at the time of the DMC was held) to 250 subjects and conducting the underpowered primary analysis one year after the last subject being randomized. This would allow a more reliable estimation of the treatment effect aiming at demonstrating a clinically meaningful improvement in PFS between the experimental arm and the control arm. In addition the proposed timing for the analysis will allow a reasonable follow up for overall survival so full information on the clinical benefit of lenalidomide can be provided at the time of PFS analysis.

The main objective of the study changed to demonstrate the efficacy of lenalidomide over a single agent of investigator's choice based on PFS after protocol amendment #2. This change was already reflected in the SAP amendment 1.

The primary analysis for the study is to compare PFS between lenalidomide and investigator's choice monotherapy.

Protocol amendment # 4:

After protocol amendment #4 the PFS assumptions used to determine sample size and the timing of the primary analysis could not be used anymore. The primary analysis will be conducted one year after the last subject is randomized.

Protocol amendment # 2:

For the primary efficacy variable PFS, a hazard ratio (HR) of 1.7 leading to an improvement in median PFS from 2.5 months for the control arm to at least 4.25 months for lenalidomide is considered to be clinically relevant.

These assumptions are supported by recent data published in relapsed refractory MCL (Witzig *et al.* 2007 and Hess *et al.*, 2009).

With a hazard ratio of 1.7, full information necessary for a one-sided log rank test with an overall alpha of 0.025, to have 80% power, will be achieved when approximately 128 subjects have progressed or died (PFS).

Initially, the sample size was calculated to estimate the ORR; however, it remains adequate to estimate PFS.

Original protocol:

The sample size calculation is based on the width of the 95% confidence interval around a certain point estimate for ORR that is considered significant clinical activity.

Based on preliminary data, a response rate in the range of 30% to 40% can reasonably be expected.

A sample size equal to 100 allows the construction of a two sided 95% confidence interval with a width of 9% (one direction) for an expected proportion of 30%.

The lower observed confidence interval limit would be about 21%, which is still considered to be clinically meaningful.

This sample size allows a width (one direction) of 9.6% for an expected proportion of 40% and 9.3% for an expected proportion of 35% (Table 3).

Table 3: Confidence Intervals

	95% Two-Sided Confidence Intervals (N = 100)		
Response rate	30%	35%	40%
CI[a]	0.214 – 0.400	0.257 – 0.452	0.303 – 0.503

a Exact confidence intervals based on the Clopper-Pearson method

Using a one group chi-square ($\chi 2$) test with a 0.050 two-sided significance level a sample size of 100 subjects will have 81% power to detect the difference between the Null hypothesis proportion of 20% and the alternative proportion of 32%.

No formal sample size calculation will be done for the control arm. With a 2:1 ratio 50 subjects are needed in the control arm.

Assuming that 10% of subjects will be lost to follow up, 167 subjects will need to be randomized.

5.5. Changes to the Planned Protocol Analysis

Initially it was planned not to present all the populations if they differ by less than 10% in size for the demographic and baseline characteristic. All the populations will be presented regardless of the difference in size for the demographic and any efficacy analyses.

6. GENERAL STATISTICAL CONSIDERATIONS

6.1. Reporting Conventions

Summary tables, listings, and any supportive SAS output will include a "footer" of explanatory notes that will indicate, at a minimum, the following:

- Program source (e.g., SAS program name, including the path, that generates the output)
- Data extraction date (e.g., the database lock date, run date).
- Source listing(s) for the summary tables

The purpose of the data extraction date is to link the output to a final database, either active or archived, that is write-protected for replication and future reference. An output date will also appear on each output page and will indicate the date the output was generated by the analysis program. Individual source listings will display all the relative values supporting the corresponding tables or figures.

6.1.1. Dates and Partial Dates Imputation

Dates will be stored as numeric variables in the SAS analysis files and reported in DDMMMYYYY format (ie, the Date9. datetime format in SAS) if complete dates are expected, for information (e.g: medical history, prior medications) dates can be reported in the YYYY-MM-DD format. Dates in the clinical database are classified into the categories of procedure dates, log dates, milestone dates, outcome dates, and special dates.

- **Procedure Dates** are the dates on which given protocol-specified procedures are performed. They include the dates of laboratory testing, physical examinations, tumor scans, etc. They should be present whenever data for a protocol-specified procedure is present and should only be missing when a procedure is marked as NOT DONE in the database. Procedure dates will not be imputed.
- **Log Dates** are dates recorded in case report form (CRF) data logs. Specifically, they are the start and end dates for adverse events and concomitant medications/procedures. They should not be missing unless an event or medication is marked as *ongoing* in the database. Otherwise, incomplete log dates will be imputed according to the rules in Appendix 16.1 Date Imputation Guideline (eg, for duration or cycle assignment etc). However, in listings, log dates will be shown as recorded without imputation.
- **Milestone Dates** are dates of protocol milestones such as randomization, study drug start date, study termination, etc. They should not be missing if the milestone occurs for a subject. They will not be imputed.
- **Outcome Dates** are dates corresponding to study endpoints such as survival, progression, etc. In most cases they are derived either from a milestone (eg, the

survival date is derived from the death date), or a procedure date (eg, the progression date is derived from the date of the tumor scan that was used to determine progression or from the date of clinical evaluation). They may be subject to endpoint-specific censoring rules if the outcome did not occur, but are not otherwise subject to imputation.

Dates recorded in comment fields will not be imputed or reported in any specific format.

6.1.2. Calculations Using Dates

Calculations using dates (e.g., subject's age or relative day after the first dose of study medication) will adhere to the following conventions:

Study days after the start day of study drug will be calculated as the difference between the date of interest and the first date of dosing of study medication (eg, lenalidomide) plus 1 day. The generalized calculation algorithm for relative day is as follows: STUDY DAY = [(TARGET DATE – DSTART) + 1] where DSTART = the start day of study drug. Note that Study Day 1 is the first day of treatment of study drug. Negative and zero study days are reflective of observations obtained during the baseline/screening period. Note: in general, a partial date for the first date of study drug intake will not be imputed. All efforts should be made to avoid an incomplete study drug start date.

- Age (expressed in days) is calculated: AGE = CONSENT – DATE of BIRTH + 1. In practice, age will be transformed to years by dividing the difference by 365.25 days, then truncating.

The calculated age from the clinical database will be used preferentially. When not available, it will be permissible to use the calculated age from the CRF or IVRS

Partial birth date: impute missing day as 15th of the month; impute missing month as July; set missing age for missing year

- Intervals that are presented in weeks will be transformed from days to weeks by using (without truncation) the following conversion formula:

WEEKS = DAYS / 7.

- Intervals that are presented in months will be transformed from days to months by using (without truncation) the following conversion formula:

MONTHS = DAYS / 30.4167.

6.1.3. Calculation of Cycles

The start date of each treatment cycle will be calculated based on study drug exposure records for each subject. The start date of the first cycle will be the date when the subject receives any study drug.

The cycle will be derived following the algorithm given in Appendix 16.2. Rituximab drug has a cycle length of 56 days, whereas the other treatments have a 28 days schedule. Any programming should be adapted accordingly.

Once the cycles are derived, they can be used to attribute adverse events to a specific cycle based upon the following rules: the start date of the adverse event should be included in a given cycle or if the start date does not fit in any defined cycle, the adverse event will be attributed to the previous cycle if the AE has started within 30 days after the end of the cycle.

6.2. Analysis Populations

The primary efficacy analyses will be performed on the Intent-to-Treat (ITT) population for the primary and secondary endpoints. Sensitivity analyses will be conducted for the primary and secondary endpoints based on the Full Analysis Set (FAS) population and the additional populations defined in this section.

The safety analyses will be conducted only on the Safety population.

6.2.1. Intent-to-Treat Population

The ITT population is defined as all subjects who are randomized, independent of whether they received study treatment or not. The ITT population will be used for the primary efficacy analysis.

Subjects will be analyzed according to the initial treatment to which they are assigned.

6.2.2. Full Analysis Set Population

The FAS population includes all randomized subjects that have received at least one single treatment dose with centrally confirmed histology of MCL as well as documented progression at entry.

6.2.3. Per Protocol Set Population

The PPS population includes all randomized subjects that have received at least one single treatment dose with centrally confirmed histology of MCL as well as documented progression at entry without protocol violations.

Protocol violations will be presented in data listing and summarized by categories in a table. Protocol violations will be determined by careful medical review of the data prior to database lock and conduct of statistical analyses. Physician and clinical research scientist will be responsible to produce the final protocol violation file (formatted as a Microsoft Excel file) in collaboration with the clinical operation and the data management group. This file will include a description of the protocol violations and deviations.

6.2.4. As treated Population

The As treated population analysis set is defined as all randomized subjects who receive at least two cycles of treatment regardless the treatment arm received. The subjects will be analyzed in the treatment group they actually received.

6.2.5. Safety Population

The Safety population or "All Treated Subjects" analysis set is defined as all randomized subjects who receive at least one dose of the study treatment (either lenalidomide or InvC).

Drug exposure and all safety analyses (including AEs, labs and deaths) will be based on the Safety population. Subjects will be analyzed according to the initial treatment actually received.

7. SUBJECT DISPOSITION

All subjects enrolled/randomized will be included in the analyses listed below. Subject disposition will be summarized for the following analysis populations by initial dosing regimen:

- ITT Population
- FAS Population
- PPS Population
- As Treated Population
- Safety Population.

A separate listing will be provided for subjects not randomized.

The number of subjects who will cross-over will be also displayed.

The primary reasons for **ending study treatment** will be collected on the CRF and will be summarized for all treatment discontinued subjects using the following categories:

- Adverse events
- Disease progression
- Withdrew consent
- Lost to follow-up
- Death
- Protocol violation
- Other.

The primary reasons for study discontinuation in all randomized subjects will also be collected on the CRF and summarized using the categories outlined above.

The number of subjects who will enter the follow-up phase with or without progression will be displayed.

A summary tabulation will be provided for subjects enrolled by study center and country. Duration of study participation, defined as from date of first dose to date of discontinuation (or the date of last visit), will be summarized.

8. DEMOGRAPHICS AND BASELINE CHARACTERISTICS

Demographics and baseline characteristics will be summarized for the all the defined populations in section 6.2 Analysis Populations .

Baseline clinical characteristics are defined as the latest data collected on or before day 1 (or randomization day in case day 1 is missing). When there are retested values, the retest values will be used for the analysis. Individual subject listings will be provided to support the tables.

8.1. Demographics

Summary statistics (mean, standard deviation, median, minimum and maximum) will be provided for variables measured on a continuous scale, e.g. age, height, and weight. The frequency distribution (n, %) will be provided for those variables measured on a nominal scale.

Subject demographics (age, age group (<65, ≥65), sex, race/ethnicity, renal function status at baseline: severe (<30 ml/min) or moderate (30 to <60 ml/min) or normal (≥60 ml/min) will be tabulated with summary statistics: N, mean, SD, median, minimum and maximum for continuous variables, and N, percent for categorical variables.

Baseline ECOG and baseline ECG (normal/abnormal) will be summarized with frequency counts.

8.2. Baseline Clinical Characteristics

Time from the latest disease progression to the first study treatment will be calculated in months and summarized descriptively.

Other baseline characteristics, including number of prior treatment regimens (< 3 or ≥ 3), prior high-intensity therapy (defined as transplant or hyper-CVAD or R-hyper-CVAD [yes or no]), prior transplant (yes or no), time from last prior therapy (< 6 months, ≥ 6 months), MCL staging at randomization (stage I+II or stage III+IV), MIPI score at randomization (high or intermediate or low), baseline ECOG score, baseline LDH, baseline WBC count, time from initial diagnosis to first dose (< 3 years vs > 3 years), prior bone marrow involvement, baseline tumor burden (high vs. low; high tumor burden defined as at least one lesion that is ≥ 5 cm in diameters or 3 lesions that are 3 cms or greater in diameter), bulky disease (Yes vs. No; bulky disease defined as at least one lesion that is ≥ 7cm in the longest diameter by central radiology review) will also be tabulated.

Histological diagnosis (investigator's assessment and confirmed histology) will be summarized.

8.3. Medical History

A summary of medical history will be presented by MedDRA system organ class (SOC). A similar summary will be generated for the concomitant disease (currently active medical history events).

8.4. Prior Anti-Lymphoma Therapies

Frequency tabulations of the number of subjects with at least one prior therapy and the different types of previous therapies (chemotherapies, immunotherapies, Stem cell transplant (SCT) separating autologus from allogenic if possible, radiation, and so on) will be given.

8.5. Prior Medications

Prior medications are defined as medications that were started before the start of the study treatment (whether or not ended before the start of the study treatment). Prior medications that continue into study treatment period will be also reported as concurrent therapy. Frequency tabulations of the number of subjects with at least one prior medication and the different types of prior medications will be provided.

9. STUDY TREATMENTS AND EXTENT OF EXPOSURE

9.1. Treatment Duration

Duration of treatment is the total number of days on treatment (date of last dose – date of first dose + 1). For subjects ongoing in the treatment phase of the study on the cut-off date for the final analysis, the cut-off date will be used as the date of last dose.

The treatment duration (weeks) is defined as:

[(The study treatment end date) – (the first study drug start date) + 1] / 7

Summary statistics will be provided for treatment duration by treatment group.

An additional table will be created for all treated subjects, displaying number of subjects by cycle and the treatment arm.

Days Dosed is the total number of days study drug was taken, adjusted for any treatment interruptions. Days dosed will be calculated by cycle and overall.

Cycles Dosed is the total number of treatment cycles during which a subject took study drug. Any cycle which has started will be counted as a complete one.

Average Number of Days dosed per cycle is calculated as days dosed (overall) divided by cycles dosed.

9.2. Cumulative Dose

Overall Cumulative dose is the total dose received during the treatment period.

Cumulative dose will be calculated separately for lenalidomide or InvC. The cumulative dose during the treatment is defined as the sum of all doses taken across the treatment period (in milligrams).

Cumulative dose at each cycle is calculated as the total dose received up to and including the current cycle.

9.3. Dose Intensity

Dose intensity during the treatment is defined as the cumulative dose divided by the total number of days dosed. Dose intensity will be calculated separately for lenalidomide and for InvC single agents separately.

The dose intensity per cycle is calculated as the overall cumulative dose divided by the number of cycles dosed.

9.4. Relative Dose Intensity

Relative dose intensity is the dose intensity divided by the planned dose intensity:

- For lenalidomide the planned dose intensity is 25 mg/per 28 days (10 mg/per 28 days if moderate renal impairment).
- For InvC single agents:

Table 4: InvC Drug Recommended Dose

Investigator's choice	mg/m^2	days	q days	max # cycles
Chlorambucil PO	40 (total monthly dose)	Split over 3-10 days	28	until PD or toxicity
Rituximab IV	375	1, 8, 15, 22	56[#]	until PD or toxicity
Cytarabine IV	1000 -2000 once or twice per day	1, 2	28	6
Gemcitabine IV	1000	1, 8, 15	28	6
Fludarabine IV	25	1-5	28	6
Fludarabine PO	40	1-5	28	6

IV = intravenous; PD = progressive disease; PO = oral; q = days, repeated; y = years.
[#] Rituximab (single agent) is to be repeated every 56 days after Day 56 (given only on day 1 of every 56 days cycle). For the prevention of cytokine release syndrome associated with the treatment of Rituximab ≤ 125 mg of methylprednisolone or equivalent are accepted on C1D1.

Descriptive statistics (n, mean, SD, median, max, min) of treatment duration, cumulative dose, dose intensity and relative dose intensity will be presented by treatment arm.

9.5. Dose Modification

Dose reduction overall is the number and percent of subjects with at least one dose reduction during the treatment period. Dose reduction by cycle is the number and percent of subjects with only one (subjects with one dose reduction) or more than one (number of subjects with two or more dose reductions) during a given cycle.

In order for a subject to have two or more dose reductions, study drug must have been reduced for one or more days, followed by another period of one or more days during which study drug dosing was reduced.

Number of subjects who have one dose reduction, two dose reductions, three dose reductions, or four or more dose reductions due to AEs will be summarized.

Only dose interruption due to AE will be considered. It is defined as the number and percent of subjects who interrupted treatment with study drug for one or more days at least once during the treatment period due to AE. Dose interruption due to AE by cycle is the number and percent of subjects with only one (subjects with one dose interruption) or more than one (number of subjects with two or more dose interruptions) during a given cycle. In order for a subject to have two or more dose interruptions, study drug must have been stopped for one or more days, followed by dosing for one or more days, followed by another period of one or more days during which study drug dosing was stopped due to AE.

Time to first dose reduction/interruption for subjects with at least one dose reduction/interruption, is the time in days elapsed from the date of first dose of treatment to the date that first dose reduction/interruption occurred.

Interval between dose reduction/interruptions is the time from the start of the first dose reduction/interruption to the start of the second dose reduction/interruption.

Dose reduction/interruption will be summarized by Initial dosing regimen group. Summaries include subjects who have at least one dose reduction/interruption, time to first dose reduction/interruption. Additional descriptive statistics will include first and second dose reduction/interruption due to AE, duration of first and second dose reduction/interruption due to AE and interval between first and second dose reduction/interruption due to AE.

10. CONCOMITANT MEDICATIONS

Concomitant mediations are defined as non-study medications that are started after the start (start included) but before the end of the study treatment, or started before the start of the study treatment and ended or remain ongoing during the study treatment.

All concomitant treatments documented during the study period will be summarized in frequency tabulations. The Anatomical Therapeutical Chemical (ATC) coding scheme of the World Health Organization (WHO) will be used to group medications into relevant categories for these tabulations.

By-subject listings will be provided for all prior and concomitant medications and therapies taken during the trial.

Other tables such as therapies received after study discontinuation, baseline diagnostic tests and baseline physical exam are provided.

11. EFFICACY ANALYSIS

All efficacy analyses will be performed for the ITT, FAS, PPS and As treated populations. Efficacy analyses on the FAS, PPS and As treated populations are supportive to ensure robustness of efficacy findings.

The primary efficacy analyses will be based on data from the central review by the IRC with the ITT population. Data from investigator assessments will be used as supportive analysis for the primary and key secondary efficacy endpoints.

Statistical comparisons will be made between lenalidomide and InvC.

11.1. General Statistical Methods

11.1.1. Time to Event Endpoints

Kaplan-Meier (KM) survival analysis will be performed (unadjusted for the stratification variables). The resulting survival estimates will be presented graphically for selected endpoints. The median, 25th and 75th percentile time-to-event data will be presented with 95% CIs (if they exist) unadjusted by strata and within strata for the primary endpoint only.

The numerical difference (and CI of the difference) in the median, 25th and 75th percentiles between lenalidomide and InvC will be presented for the unstratified analysis.

In addition, the event rates at specific time-points will be computed, along with the standard errors (Greenwood's formula; Klein, 1997), based on which the event rates at specific timepoints can be compared if necessary.

The groups will be compared using the stratified log-rank test in order to assess superiority and the unstratified log-rank test as supportive analysis. New statistic test will also be built to take into account the unplanned sample size reassessment recommended by the DMC and it will be used as primary analysis (see section 14.2 Statistical Approaches for Control of Alpha for more details).

The stratified Cox proportional hazard regression models will be used to estimate the hazard ratios (HR) and associated 95% CIs for the HRs.

If any imbalances appear in the demographic or baseline characteristics variables considered as important prognostic factors to clinical outcomes, the variables will be analyzed as the stratification factors. They will be included as covariates in the Cox model to provide estimate of lenalidomide treatment effect adjusted for these factors and in the stratified log-rank test.

The MIPI score at randomization will be analyzed as a stratification factor.

More detailed methods (for example exploring additional covariates such as age) will be investigated as a robustness check and may also be presented.

Underlying model assumptions will be investigated for the analysis of the PFS endpoint using diagnostic statistics and graphical methods and, if necessary, an alternative analysis technique may be used.

Any changes in methodology will be documented in the clinical study report (CSR), including the rationale for use. The proportional hazards assumption will be assessed by graphical display of the log-log survival distribution function versus the log event time for each treatment group. The censoring distributions of the treatment groups will be assessed for possible differences using KM methods.

KM curves will be displayed graphically by treatment group. Additionally, reasons for censoring will be summarized (n, percent) for the TTP, PFS, and OS endpoints.

Individual survival time information and mortality data will be provided in subject data listings.

In order to evaluate the effect of cross-over on overall survival a Mantel-Byar approach will be used. The estimation of the control arm without cross-over and the mixed arm (lenalidomide arm and cross-over subjects) will be estimated. A weighted log rank test or Wilcoxon test will be used to take into account the effect of cross-over.

Multivariate models will be used to identify those baseline and prognostic factors (eg absolute leukocyte count, time from diagnosis, number of prior therapies and any other variables suggested in the literature prior to the analysis) most predictive of response (Colett, 1991).

11.1.2. Categorical Endpoints

Categorical analyses will be based on the Fisher exact test. Sensitivity analyses will be performed using the Cochran-Mantel-Haenszel test with the stratification factors as strata. The p-values will be presented. The null hypothesis is that incidence is equal for both treatment arms, the alternative hypothesis being that incidence is not equal.

The probability of response rates will be estimated using the proportion of subjects with responses with exact two-sided 95% confidence intervals.

The response rate in the cross-over portion of subjects in Arm B will be analyzed in a descriptive manner.

11.1.3. Quality of Life Endpoints

The EORTC QLQ-C30 will be analyzed using change from baseline and percentage of change from baseline according to the functional scores and the recommendations in the EORTC scoring manual. Statistical tests could be performed in an exploratory manner.

The EORTC Reference Data Manual will also be used to descriptively check comparability between subjects in our trial and with other comparable populations.

Mann-Whitney tests for simple comparison and longitudinal data modeling techniques (ie Proc mixed in SAS) will be used to analyze QoL scores.

Subjects who drop out of the study without being evaluated for QoL will be counted as non-responders. Responses from subjects after they receive other anti-cancer treatments will be treated as non responders.

An update of the time to events endpoints and overall survival will be done at the end of the follow up. No adjustment for multiplicity will be applied.

11.2. Expert Review of Disease Response

The independent review committee (IRC) is composed of two external independent radiologists (with an additional radiologist adjudicator in the event of a tie) and a hematologist/oncologist. The IRC will perform a blinded, independent assessment of radiological response (including assessment of Stable Disease [SD] and Progressive Disease [PD]), as well as reviewing the tumor response data and the dates of disease progression for each subject.

For further details please refer to Central Radiology Manual.

No central review will be performed for the subjects who crossed over after progression.

11.3. Analysis of Primary Efficacy Endpoint

PFS is defined as the time from randomization to the first observation of disease progression or death due to any cause. If the subject has not progressed or died, PFS will be censored at the time of last completed assessment when the subject was known not to have progressed. Subjects who will receive a new treatment without documented progression will be censored at the last assessment date that the subject is known to be progression-free. These rules are based on FDA guidance for cancer trial endpoints (FDA guidance, 2007).

Table 5: Censoring Rules for Time to Event (Progression and/or Death) Endpoints

Situation	Date of Progression or Censoring	Situation Outcome
No baseline assessments	Randomization	Censored
Progression documented	First adequate assessment determined by central review	Progressed
No progression	Date of last adequate assessment with evidence of no progression	Censored
Study discontinuation for reasons other than disease progression or death	Date of last adequate assessment with evidence of no progression	Censored
New anti-lymphoma / non-protocol treatment started prior to progression	Date of last adequate assessment with evidence of no progression prior to the start of new anti-lymphoma treatment	Censored
Death before first PD assessment while on study	Date of death	Progressed
Death between adequate assessment visits	Date of death	Progressed
Progression after an extended lost-to-follow-up time (two or more missed assessments)	Date of last adequate assessment with evidence of no progression	Censored

PD = progressive disease.
See details in Appendix 16.3 .

As additional robustness checks of the primary endpoint and censoring definition, the time to event analysis will be repeated for the ITT population using the following modified definitions of PFS:

- Using the investigator assessment instead the IRC central review
- Considering progression or death under a new anti-lymphoma treatment as an event (Table 5)
- Considering death or progression after an extended lost-to-follow-up time (two or more missed assessments) as an event (Table 6)
- One using the earliest progression date either in investigator set or IRC set.

Table 6: Method b: Censoring Rules for Time to Event (Progression and/or Death) Endpoints Sensitivity Analysis

Situation	Date of Progression or Censoring	Situation Outcome
No baseline assessments	Randomization	Censored
Progression documented	First adequate assessment determined by central review	Progressed
No progression	Date of last adequate assessment with evidence of no progression	Censored
Study discontinuation for reasons other than disease progression or death	Date of last adequate assessment with evidence of no progression	Censored
New anti-lymphoma / non-protocol treatment started prior to progression	First adequate assessment determined by central review	Progressed
Death before first PD assessment while on study	Date of death	Progressed
Death between adequate assessment visits	Date of death	Progressed
Progression after an extended lost-to-follow-up time (two or more missed assessments)	Date of last adequate assessment with evidence of progression	Progressed

PD = progressive disease.

Table 7: Method c: Censoring Rules for Time to Event (Progression and/or Death) Endpoints Sensitivity Analysis

Situation	Date of Progression or Censoring	Outcome
No baseline assessment	Randomization	Censored
Progression documented between scheduled visits	As defined for the main analysis	Progressed
No progression	Date of last visit with adequate assessment	Censored
Investigator claim of clinical progression	Scheduled visit (or next scheduled visit if between visits)	Progressed
Treatment discontinuation for toxicity or other reason	Date of last visit with adequate assessment	Censored
New anticancer treatment started with no claim of progression	Date of last visit with adequate assessment	Censored
Death before first PD assessment	Date of death	Progressed
Death between adequate assessment visits or after subject misses one assessment visit	Date of death	Progressed
Progression after an extended lost-to-follow-up time (two or more missed assessments)	Last visit with adequate assessment	Censored

PD = progressive disease.

11.4. Analyses of Secondary Efficacy Endpoints

The secondary endpoints will be analyzed in an exploratory manner (see details in Appendix 16.3).

11.4.1. Overall Response Rate

The Overall response rate will include best response of CR, CRu, or PR. The modified IWRC will be used and response assessed by CT scan/MRI every 56 days (± 7 days) for the first 6 months and then every 90 days ± 15 days thereafter.

Subjects who discontinue before any post-randomization efficacy assessments will be considered to be non-responders.

11.4.2. Duration of Response

Duration of response will be measured from the time of initial response (at least PR) until documented tumor progression or death. Subjects who do not progress at the time of analysis will be censored at the last assessment date that the subject is known to be progression-free. Subjects who will receive a new treatment without documented progression will be censored at the last assessment date that the subject is known to be progression-free.

This analysis will be restricted to the subgroup of responding subjects.

11.4.3. Tumor Control Rate

Rates for CR, CRu, PR, and SD.

CR/CRu/PR/SD will be determined using the modified IWRC (Cheson, 1999).

11.4.4. Duration of Stable Disease

SD is defined as less than PR but is not progressive disease or relapsed disease.

Duration of SD will be calculated as the time from the first evidence of SD to documented disease progression or documented response or death. Subjects who do not progress or respond at the time of analysis will be censored at the last assessment date that the subject is known to be progression-free. Subjects who will receive a new treatment without documented progression will be censored at the last assessment date that the subject is known to be progression-free.

This analysis will be restricted to the subgroup of SD subjects.

11.4.5. Time to Progression (TTP)

TTP will be defined as the time from randomization until objective tumor progression. TTP will not include deaths. Subjects without progression at the time of analysis will be censored at the last assessment date that the subject is known to be progression-free. Subjects who will receive a new treatment without documented progression will be censored at the last assessment date that the subject is known to be progression-free.

11.4.6. Time to Treatment Failure (TTF)

TTF is defined as the time from the first dose of study drug to discontinuation of treatment for any reason, including disease progression, treatment toxicity or death. Subjects who will be on drug or complete the treatment according to the protocol will be censored at the last date of drug intake.

11.4.7. Time to Tumor Response (TTR)

This analysis will be restricted to the subgroup of responding subjects. Time to tumor response (TTR) will be defined as the time from randomization until initial response (CR, CRu, PR) if it has been confirmed.

11.4.8. Overall Survival (OS)

OS will be defined as the time from randomization until death from any cause. Subjects alive or lost to follow up at the time of analysis will be censored at the last date the subject was known to be alive.

11.5. Subgroup Analysis

Subgroup analyses will be conducted in an exploratory manner for the following endpoints PFS, ORR and OS in the following subgroups:

- Time since diagnosis (< 3 years or ≥ 3 years)
- Time since the end of last treatment (<6 months [refractory] or ≥ 6 months)
- Prior stem cell transplant (yes or no)
- Age(<65;≥65)
- Gender
- MCL International Prognostic Index (MIPI) score at initial time of diagnosis
- MIPI score at randomization
- Number of prior treatment lines (with description of treatment regimens, including first line treatment and stem cell transplant) (1,> 1), (<2;≥2), (<3; ;≥3) and by single category (1, 2, 3, 4, and >4)
- Ki-67 index (as defined by Determann, 2008) labeling in the original pathology specimen at diagnosis, if available at time or at time of relapse
- Absolute lymphocyte count at baseline (< 800/mm^3 or ≥ 800/mm^3)
- Time since last rituximab to cycle 1 day 1 (< 230 days or ≥ 230 days)
- Type of prior regimens received (rituximab containing regimens; Ara-C containing regimens; and fludarabine containing regimens).
- Number of relapses (1 vs > 1), (<2 vs ;≥2 relapses), (<3; ;≥3 relapses)
- Any demographic or baseline characteristic considered as important prognostic factors with imbalance in two treatment arms.

In addition any other possible prognostic factors may be used as exploratory analyses.

The methods described in previous sections will be used for each subgroup separately.

12. SAFETY ANALYSIS

The safety analyses will be performed using the safety population.

Adverse event and medical history coding will be performed using the Medical Dictionary for Drug Regulatory Activities (MedDRA) Version 15.1.

12.1. Adverse Events

Adverse events (AEs) will be coded according to the Medical Dictionary for Regulatory Activities (MedDRA) version 15.1. The intensity of AEs will be graded according to the National Cancer Institute (NCI) Common Terminology Criteria for Adverse Events (CTCAE) version 3.0 (same as protocol).

Treatment-emergent adverse events (TEAEs) are defined as any AE occurring or worsening on or after the first treatment of the study medication and within 30 days after the last dose. TEAEs, TEAEs leading to study medication discontinuation, TEAEs leading to dose reduction/interruption, TEAEs related to study medication and treatment-emergent serious adverse events (SAEs) will be summarized by system organ class and preferred term for each treatment group. A summary of TEAEs with NCI CTCAE grade 3 or higher and most frequent preferred terms will also be provided.

If a subject experiences the same preferred term multiple times then the subject will be counted only once and by greatest severity.

Listings for the corresponding summary tables will be presented separately.

For the cross-over subjects, any adverse events starting after the start of the drug will be assigned to lenalidomide even if the event has started within 30 days after the InvC end date. Additional summary tables will be provided for these TEAEs (cross-over lenalidomide).

Analysis of targeted TEAEs, time to targeted TEAEs will be presented. Targeted TEAEs will be defined using standardized MedDRA Queries (SMQs). The SMQ will be selected based on the internal expertise on the drug and the disease.

Second Primary Malignancies (SPMs) will be summarized by frequency for the following categories:

- Invasive SPMs will include hematologic malignancies and solid tumor malignancies. Hematological malignancies will be subcategorized into B-cell malignancies, acute myelogenous leukemia (AML), myelodysplastic syndrome (MDS), MDS to AML and other hematological malignancies.
- Non-invasive SPMs (non-melanoma skin cancers)

Incidence per 100 person-years will be calculated. Confidence intervals will be provided using the Mid-P exact test. The following formulae will be applied:

The notation for the formulae is:

a = the observed numerator

PT = is the observed denominator in person-time units

rate = a/PT

$Z_{1-\alpha/2}$= the two-sided Z value (eg. Z=1.96 for a 95% confidence interval).

Exact Tests (Mid-P and Fisher)

The limits for 'a' with 100(1-α) percent confidence are the iterative solutions $\underline{a}$ *and* $\overline{a}$.

Computing iterative solutions $\underline{a}$ *and* $\overline{a}$ is below……

Mid-P exact test (see Rothman and Boice):

Lower bound:

$$\left(\frac{1}{2}\right)\frac{e^{-\underline{a}}\underline{a}^{a}}{a!}+\sum_{k=0}^{a-1}\frac{e^{-\underline{a}}\underline{a}^{k}}{k!}=1-\alpha/2$$

Upper bound:

$$\left(\frac{1}{2}\right)\frac{e^{-\overline{a}}\overline{a}^{a}}{a!}+\sum_{k=0}^{a-1}\frac{e^{-\overline{a}}\overline{a}^{k}}{k!}=\alpha/2$$

If appropriate, competing risk analysis considering death as competing event will be conducted.

If the number of events observed is sufficient, multivariate analysis to identify demographic and clinical factors predictive of SPMs will be conducted. Univariate Cox regression analysis will be used to select the subset of relevant factors for inclusion in a multivariate model, with variable selection set at the $p < 0.20$ level. Backward selection procedures will be used for final model determination, with $p < 0.10$ selected as the p value for retention in the multivariate model. These analyses will be performed for all invasive SPMs as well as for hematologic malignancies and solid tumors separately.

The table output will display the number of cycles received, the overall total dose, the treatment duration and the time to onset of the SPM. A supportive listing will include, at minimum, the verbatim and the SPM category, the treatment duration, start and stop date, action taken on treatment, gender and age.

Hospitalizations will be summarized by the primary cause of hospitalization.

The FCBP status along with pregnancy test results will be summarized for female subjects.

12.2. Clinical Laboratory Evaluations

Clinical laboratory values will be graded according to NCI CTCAE version 3.0 for applicable tests (see Appendix 16.4 Laboratory Values for details on standardization of the laboratory values). Frequency distributions for shift from baseline in severity grade to most extreme post-baseline value will be displayed in cross-tabulations for the whole study period for each treatment. Normal ranges will be used to determine the categories if High, Low, and Normal for lab tests that have no severity grade. Listings of clinical laboratory data with abnormal flags will be provided by subjects and tests.

12.3. Vital Sign Measurements

For vital signs, shift from baseline to worst during the treatment in below, within and above the normal ranges will be displayed in cross-tabulations for each treatment. Summary statistics (N, Mean, SDev, Median, Minimum, and Maximum) of observed and change from baseline values will be presented.

12.4. Physical Examination

Shift tables regarding the evolution of the spleen, liver, lymphadenopathy will be displayed by treatment groups.

Presence or absence of lymphoma symptoms will be summarized by visit and treatment groups.

12.5. Electrocardiogram

The overall ECG interpretation will be summarized by presenting the number and percentage of subjects with 'Normal', 'Abnormal, not clinically significant', and 'Abnormal, clinically significant' by treatment. Shift from baseline to worst during the treatment in the overall ECG interpretation will be displayed in cross-tabulations for each treatment.

13. QUALITY OF LIFE ANALYSIS

13.1. EORTC QLQ-C30

QoL instrument: EORTC QLQ-C30 (see attachment 21.9)

The version 3.0 of the EORTC QLQ-C30 is a 30-item scale. The EORTC QLQ-C30 is composed of both multi-item scales and single-item measures. These include five functional scales, three symptom scales, a global health status / QoL scale, and six single items (Table 8).

Table 8: EORTC QLQ-C30 Scores

FUNCTIONAL SCALES (15 QUESTIONS)	SYMPTOM SCALES (7 QUESTIONS)	SINGLE ITEMS (6 QUESTIONS)	GLOBAL QUALITY OF LIFE (2 QUESTIONS)
PHYSICAL (ITEMS 1 TO 5)	**FATIGUE (ITEMS 10, 12, 18)**	**CONSTIPATION (ITEM 16)**	**GLOBAL QOL (ITEMS 29, 30)**
ROLE (ITEMS 6, 7)	**PAIN (ITEMS 9, 19)**	**DIARRHEA (ITEM 17)**	
COGNITIVE (ITEMS 20,25)	**NAUSEA / VOMITING (ITEMS 14, 15)**	**SLEEP (ITEM 11)**	
EMOTIONAL (ITEMS 21 TO 24)		**DYSPNEA (ITEM 8)**	
SOCIAL (ITEMS 26, 27)		**APPETITE (ITEM 13)**	
		FINANCIAL (ITEM 28)	

All of the scales and single-item measures range in score from 0 to 100. A higher scale score represents a higher level of well being and better ability of daily functioning. A 10-point change in the scoring is considered to be meaningful change in QoL (Osoba, 1998).

Thus a high score for a functional scale represents a high/healthy level of functioning; a high score for the global health status/QoL represents a high QoL, but a high score for a symptom scale/item represents a high level of symptomatic problems.

Data collection and timing

Questionnaires must be filled out by the subject at 6 pre-specified time points (see below) when the subject comes for a scheduled visit. They will be handed out by a nurse or the treating physician/investigator, and ideally are collected again soon after the subject has filled it out.

Subjects will be asked to complete the questionnaire as completely and accurate as possible. The average time to complete the questionnaire is around 10to 15 minutes. The reasons for not completing the questionnaire will be recorded.

QoL will be assessed at 6 time points:

- at screening/baseline (within 7 days prior to randomization)
- after cycle 2 (C3D1)
- after cycle 4 (C5D1)
- after cycle 6 (C7D1)
- after cycle 8 (C9D1)
- and at time of discontinuation from treatment.

A ± 7 days time window will be allowed.

Compliance & Missing data

During the study compliance will be investigated at each time point. The compliance rate will be descriptively compared between the two arms.

In the case that a subject still on study has not filled in the questionnaire, no more than ± 7 days delay is accepted.

The rate of compliance will be assessed after 40, 80, and 120 and 200 subjects enrolled.

Missing data will be reported as recommended in the EORTC scoring manual (Fayers, 2002).

Scales scores: if at least 50% of the items have been answered, the scale scores will be calculated according to the standard equations given on the manual (any items with missing values will be ignored). In other cases, scores will be set to missing.

Single-item measures: the score will be set to missing.

Statistical considerations

No calculation in terms of sample size will be performed based on changes in QoL.

In the absence of more specific hypothesis, the global score will be used as the primary QoL outcome and physical functional score and fatigue item will be used as secondary outcomes.

Expectations are that the global score and the physical functional score are improved by lenalidomide. The fatigue item may be worse initially under lenalidomide, but improve in mid and long term in responding subjects due to tumor control compared to control arm with shorter response duration

A difference of 10 points between the two arms is considered as clinically significant.

Derivation

For all scales, the Raw Score, RS, is the mean of the component items:

$$RawScore = RS = (I_1 + I_2 + ... + I_n)/n$$

For **functional scales**:

$$Score = \left\{1 - \frac{(RS - 1)}{range}\right\} \times 100$$

For **symptom scales / items and Global health status / QoL**:

$$Score = \{(RS - 1)/range\} \times 100$$

Range is defined as the difference between the maximum possible value of RS and the minimum possible value, eg, most items are quoted from 1 to 4, and the range is equal to 3.

13.2. Analysis of Quality of Life Scales

The EORTC QLQ-C30 will be analyzed using change from baseline and percentage of change from baseline according to the functional scores and the recommendations in the EORTC scoring manual. Statistical tests could be performed in an exploratory manner.

The EORTC Reference Data Manual will also be used to descriptively check comparability between subjects in our trial and with other comparable populations.

Mann-Whitney tests for simple comparison and longitudinal data modeling techniques (ie Proc mixed in SAS) will be conducted to analyze QOL scores.

14. INTERIM ANALYSIS

14.1. General Information

An independent external Data Monitoring Committee (DMC) will review on an ongoing basis safety data throughout the study and efficacy for futility at a pre-defined time point. Specifics are outlined in the DMC charter. Summaries of safety information will be prepared for Data Monitoring Committee (DMC) review. The first safety analysis will occur after the first 40 subjects have received at least 2 cycles of treatment or have discontinued prior to completing 2 cycles. The second safety analysis, as well as an efficacy analysis for futility, will occur after 80 subjects complete 2 cycles or withdraw before completing 2 cycles. A third safety analysis will occur after 120 subjects complete 2 cycles or withdraw before completing 2 cycles.

After amendment #4, an additional DMC will be added when the first 200 subjects complete 2 cycles or withdraw before completing 2 cycles.

The futility analysis will be conducted when approximately 80 subjects complete 2 cycles or withdraw before completing 2 cycles (around 54 subjects in the lenalidomide arm) on investigator and central assessment.

The DMC will conduct an analysis for futility on PFS and ORR. No specific stopping rules will be given to the DMC for the PFS; the following rules might be used for ORR.

Based on a 95%CI approach, if the upper bound is below than 20% in the lenalidomide arm, the DMC should recommend stopping the trial.

Table 9: Response Percentage and 95% Exact Confidence Interval

FREQUENCY OF RESPONSE	BINOMIAL PROPORTION P	EXACT LOWER CL, BINOMIAL PROPORTION[A]	EXACT UPPER CL, BINOMIAL PROPORTION[A]
1	**1.8519**	**0.04687**	**9.8915**
2	**3.7037**	**0.45173**	**12.7472**
3	**5.5556**	**1.16068**	**15.3885**
4	**7.4074**	**2.05510**	**17.8933**
5	**9.2593**	**3.07528**	**20.3002**
6	**11.1111**	**4.18838**	**22.6313**
7	**12.9630**	**5.37430**	**24.9012**
8	**14.8148**	**6.61976**	**27.1198**
9	**16.6667**	**7.91544**	**29.2941**
10	**18.5185**	**9.25455**	**31.4297**

CI = confidence interval.
a Exact confidence intervals based on the Clopper-Pearson method.

14.2. Statistical Approaches for Control of Alpha

14.2.1. Primary Analysis

The sample size increase recommended by the DMC was not supported by any planned interim analyses; therefore adjustment for controlling the alpha level was not planned in the original protocol and has to be described in this section.

The final level will be determined using an α-spending function of the O'Brien-Fleming type and a new group sequential test procedure will also be used to preserve the type I error following method developed by Cui et al. and adapted by Wassmer for time to event endpoints.

Final alpha level:

At the time of the DMC#3, 67 events have been reported on the 128 needed at this time for the final analysis. If one interim analysis had been planned at 52% of the information the upper boundary for superiority would have been based on an α-spending function of the O'Brien-Fleming type with overall $\alpha = 0.025$, one-tailed.

Table 10: P-Value for Rejecting Null Hypothesis (Superiority)

	P-Value for Rejecting Null Hypothesis (Superiority)
Interim 1 (52%)	0.002
Final	0.024
Note: « program East Version 4 » (Cytel Statistical Software Company).	

New test procedure :

$$Z_2^* = \frac{\sqrt{d_2}\,LR_2^* - \sqrt{d_1}\,LR_1^*}{\sqrt{d_2 - d_1}}$$

where

d is the observed number of events

LR is the LogRank test

Estimating the Hazard ratio and the confidence interval :

$$\exp\left(\frac{\sum_{k=1}^{2} w_k z_k}{h} \pm \frac{u_k \left(\sum_{k=1}^{2} w_k^2\right)^{1/2}}{h}\right)$$

$$with\ h = \sum_{k=1}^{2} \frac{w_k}{\sqrt{d_k - d_{k-1}}}\left(d_k \frac{\sqrt{r_k}}{1+r_k} - d_{k-1}\frac{\sqrt{r_{k-1}}}{1+r_{k-1}}\right)$$

$$w_1 = \xi_1 \ and \quad w_2 = \sqrt{\xi_2 - \xi_1}$$

where ξ is the exp*ected number of events*

Despite the change in the test statistic the usual KM median will be provided.

14.2.2. Interim Analysis

Formal adjustment is needed as an unplanned sample size reassessment has been made in addition to the planned interim analysis for futility.

15. REFERENCES

Cheson B, Horning S, Coiffier B, Shipp M, Fisher R, Connors J, et al. Report of an international workshop to standardize response criteria for non-Hodgkin's lymphoma. J Clin Oncol.1999; 17: 1244-53.

Collett, D. *Modelling Binary Data.* 1991; London: Chapman & Hall.

Collett, D. *Modelling Survival Data in Medical Research Second Edition.* 2003; London: Chapman & Hall.

Cui L, Hung HMJ, Wang SJ ; Modification of sample size in group sequential clinical trials. Biometrics 1999; 55:853-857.

Determann O et al.; European Mantle Cell Lymphoma Network and the German Low Grade Lymphoma Study Group.Ki-67 predicts outcome in advanced-stage mantle cell lymphoma subjects treated with anti-CD20 immunochemotherapy: results from randomized trials of the European MCL Network and the German Low Grade Lymphoma Study Group. Blood. 2008 Feb 15; 111(4):2385-7.

Fayers PM, Weeden S, Curran D on behalf of the EORTC Quality of Life Study Group. EORTC QLQ-C30 Reference Values. In: EORTC, 1998. ISBN: 2-930064-11-0.

Guidance for Industry Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics. FDA , May2007

Hess G, et al., Phase III Study to Evaluate Temsirolimus Compared With Investigator's Choice Therapy for the Treatment of Relapsed or Refractory Mantle Cell Lymphoma. J Clin Oncol 27:3822-3829.

Klein J.P, Moeschberger ML. Survival Analysis: Techniques for Censored and Truncated Data, 1997; New York: Springer-Verlag.

Mosteller RD: Simplified Calculation of Body Surface Area. N Engl J Med 1987 Oct 22;317(17):1098 (letter)

Osoba D, Rodrigues G, Myles J, Zee B, Pater J. Interpreting the significance of changes in health-related quality-of-life scores. J Clin Oncol 1998;16(1):139-44.

Stead ML, Brown JM, Velikova G, Kaasa S, Wisloff F, Child JA, et al. Development of an EORTC questionnaire module to be used in health-related quality-of-life assessment for subjects with multiple lymphoma. European Organization for Research and Treatment of Cancer Study Group on Quality of Life. Br J Haematol 1999 Mar;104(3):605-11.

Wassmer G, Planning and Analyzing Adaptive Group Sequential Survival Trials. Biometrical Journal 48 (2006) 4, 714–729.

Witzig TE, Reeder CB, Polikoff J, et al. Initial Results from an International Study in Relapsed/Refractory Aggressive Non-Hodgkin s Lymphoma to Confirm the Activity, Safety and Criteria for Predicting Response to Lenalidomide Monotherapy. Blood (ASH Annual Meeting Abstracts), Nov 2007; 110: 2572.

16. APPENDICES

16.1. Date Imputation Guideline

16.1.1. Impute Missing AE/ Prior or Concomitant Medications Start Dates

Missing day and month

- If the year is **same** as the year of first day on study medication, then the day and month of the start date of study medication will be assigned to the missing fields
- If the year is **prior to** the year of first day on study medication, then December 31 will be assigned to the missing fields.
- If the year is **after** the year of first day on study medication, then January 1 will be assigned to the missing fields.

Missing day only

- If the month and year are **same** as the year and month of first day on study medication, then the start date of study medication will be assigned to the missing day.
- If the month and year are **before** the year and month of first day on study medication, then the last day of the month will be assigned to the missing day.
- If the month and year are **after** the year and month of first day on study medication, then the first day of the month will be assigned to the missing day.
- If the stop date is non-missing and the imputed start date is after the stop date, the start date will be imputed by the stop date.

Missing day, month, and year
Included as TEAE

Incomplete Stop Date: If the imputed stop date is before the start date then the imputed stop date will be equal to the start date.

Missing day and month

- If the year of the incomplete stop date is the **same** as the year of the last dose date of study medication, then the day and month of the last dose date will be assigned to the missing fields.
- If the year of the incomplete stop date is **prior to** the year of the last dose date of double-blind study medication, then December 31 will be assigned to the missing fields.
- If the year of the incomplete stop date is **after** the year of the last dose date of study medication, then January 1 will be assigned to the missing fields.

Missing day only

- If the month and year of the incomplete stop date are the **same** as the month and year of the last dose date of study medication, then the day of the last dose date will be assigned to the missing day.
- If the month and year of the incomplete stop date are **before** the month and year of the last dose date of the study medication, then the last day of the month will be assigned to the missing day.

- If the month and year of the incomplete stop date are **after** the month and year of the last dose date of study medication, then the first day of the month will be assigned to the missing day.

16.2. Cycle Derivation Guideline

Subjects on Rituximab treatment have cycles of 56 days in duration past cycle 1. Cycle 2 day 1 will in fact correspond to day 56. Programming need to be adapted accordingly.

eCRF pages completion	C1D1	C2D1	C3D1	C4D1	C5D1	C6D1	C7D1
days on study	1	28	56	84	112	140	168
Tumor assessment			X		X		X
Rituximab cycles	C1D1		C2D1		C3D1		C4D1

A cycle is defined as follows:

A new cycle is defined when the delay between the previous drug start date and the current start date is lower or equal to -26.

del <=-26 and del=lag1(SDRSTDT)-SDRSTDT;

If the data do not fulfil the previous condition, a cumulative duration (dur = Σ SDRENDT – SDRSTDT + 1) is calculated (including the previous record) to take into account the fact that several drug modifications could have appeared during one cycle. If the cumulative duration does not exceed 28 days, there is no new cycle.

16.3. Endpoints Derivation

Efficacy evaluation window:

All efficacy assessments will be taken into account if they are occurred after the randomization date

to the end the study, for subjects who do not stop study medication and not start other non study anti-cancer therapy;

Best response

Best response is defined as best efficacy benefit response during the "window" for efficacy evaluation.

PFS:

PFS = PFS Ending Date – PFS Starting Date + 1

Starting date: the randomization date

Ending date: the earliest date among the following 4 events during the "window" for efficacy evaluation:

a. Death;

b. The earlier assessment date for the cycle with overall response is PD or Relapse;

c. Study discontinue due to Disease progression specified;

d. Treatment discontinue due to Disease progression specified;

For subjects who have more than 1 consecutive missing regular assessments right before the earliest of the four events date and this non-assessed period last more than 118 days (this length is study-dependent), then, the date of last adequate assessment indicating no progression before the date will be used, and the PFS will be censored.

If subjects who do not have any of the 4 events listed above, then ending date is defined as the latest CT/MRI date for target or non-target lesion that indicating no progression, if subjects do not have such response, they will be censored at the day after the randomization date.

Duration of response (DoR)

DoR = Ending Date – Starting Date + 1

Starting date: The date of the first occurrence of CR, CRu or PR during the Window for efficacy evaluation

Ending date: the earliest date among the following 4 events during the "window" for efficacy evaluation:

a. Death;

b. The earlier assessment date for the cycle with overall response is PD or Relapse;

c. Study discontinue due to Disease progression specified;

d. Treatment discontinue due to Disease progression specified;

For subjects who have more than 1 consecutive missing regular assessments right before the earliest of the four events date and this non-assessed period last more than 118 days, then, the last adequate at least PR or better response assessment date before the date will be used, and the DoR will be censored.

If subjects who do not have any of the 4 events listed above, then ending date is defined as the latest CT/MRI date for target or non-target lesion with at least PR or better response after the DoR start date, if subjects do not have such response, they will be censored at the day after the 1st response date

16.4. Laboratory Values

The following calculations need to be performed for all subjects in treatment arms A and B. The lab values will be standardized using the normal range values collected in the eCRF.

Step 1: Assign an upper (ULN) and lower (LLN) normal range value to each analyte at each visit collected by merging the ranges collected in the datasets labrng. Use the labseq variable to assign the appropriate normal ranges to a value.

If the ranges could not be assigned then use the default SI ranges after the conversion of the value..

Convert any raw lab values to the unit that corresponds to the normal range values as necessary. These converted (if necessary) or raw (if conversion not needed) lab values and their associated ULN and LLN values will be referred to as the "base" values in the calculations given in Step 2.

Step 2: Convert all base values (lab, ULN, and LLN values) to the standardized units given (please refer to the Data Transfer Guidelines in the data management plan). The standardized units are the units associated with the normal ranges to be used for mapping the lab value to a standardized value. Calculations in Step 3 are based on these converted lab and normal range values (referred to as converted values).

Step 3: Determine, for each (converted) lab value, if it is within the (converted) normal range and assign a flag indicating where the value falls relative to the upper and lower limits of normal (eg, low (below lower limit normal), normal, or high (above upper limit normal).

Step 4: Determine p, the value to be used for mapping the converted lab value to a standardized value as follows:

a) If the converted lab value is within converted normal range then

p = (lab value – LLN value) / (ULN value – LLN value)

b) If the converted lab value is less than converted LLN then

p = lab value / LLN value

c) If the converted lab value is greater than converted ULN then

p = lab value / ULN value

Step 5: Use the value of p calculated in Step 4 and the standardized ULN and LLN values

a) If the converted lab value is within normal range then

SLV = p(sULN value – sLLN value) + sLLN

b) If the converted lab value is less than LLN then

SLV = p(sLLN value)

c) If the converted lab value is greater than ULN then

SLV = p(sULN value)

Step 6: Use the SLV and the sLLN and sULN to assign the NCI-CTC grade and for summary statistics in all tables.

Example calculations for standardizing lab values to the standard normal ranges

Subject X at CxD1 visit:

SGPT = 22 UI/L

Step 1: normal range

LLN = 7 UI/L

ULN = 40 UI/L

Step 2:

Since the conversion factor for converting UI/L to U/L is 1.0, the converted values associated with unit=U/L remain the same as the original values.

converted SGPT = 22 U/L

converted LLN = 7 U/L

converted ULN = 40 U/L

Step 3:

Value of 22 U/L is within the normal range.

Normal range flag = N

Step 4: Calculate p for 22 U/L based on formula a):

p = (lab value – LLN value) / (ULN value – LLN value)

= (22 – 7) / (40 – 7)

= 15 / 33

= 0.45

Step 5: Calculate SLV for p = 0.45 and sLLN = 6 U/L and sULN = 35 U/L.

SLV = p(sULN value – sLLN value) + sLLN

= 0.45(35 – 6) + 6

= 19.2 U/L

Step 6: NCI-CTC grade = 0, since the value is within normal limits.