**CLINICAL STUDY PROTOCOL**

NCT Number: NCT01659658

Study Title: A Phase 3, Randomized, Controlled, Open-label, Multicenter, Safety and Efficacy Study of Dexamethasone Plus MLN9708 or Physician's Choice of Treatment Administered to Patients With Relapsed or Refractory Systemic Light Chain (AL) Amyloidosis

Study Number: C16011

Statistical Analysis Plan

Amendment Version 2.0: 15-July-2015

# STATISTICAL ANALYSIS PLAN

***A Phase 3 Randomized, Controlled, Open-label, Multicenter, Safety and Efficacy Study of Dexamethasone plus MLN9708 or Physician's Choice of Treatment Administered to Patients With Relapsed or Refractory Systemic Light Chain (AL) Amyloidosis***

Protocol #: C16011

| SAP Version: | Date of Statistical Analysis Plan: |
|---|---|
| Final | 7 December 2011 |
| Amendment 1 | 05 June 2012 |
| Amendment 2 | 14 July 2015 |

## Approval Signatures

| | |
|---|---|
| Date | Prepared by: [redacted], Ph D<br>[redacted], Global Biostatistics |
| Date | Reviewed by: [redacted], PhD<br>[redacted], Global Biostatistics |
| Date | Approved by: [redacted], MD, PhD<br>[redacted], Global Biostatistics |
| Date | Approved by: [redacted], MD, PhD<br>[redacted], Oncology Clinical Research |
| Date | Approved by: [redacted]<br>[redacted], Oncology Clinical Research |

### Rationale for Amendment 2

The main purpose of this amendment is to incorporate the statistically-relevant changes in Protocol C16011, amendment 4, dated 09 July 2015.

Purposes for amendment 2 are to:

- Enrich the study population for proteasome inhibitor-naïve patients to better reflect the global population of patients with relapsed or refractory AL amyloidosis
- Delay the first interim analysis (IA), which is also the final analysis (FA) for hematologic response, to allow time for proteasome inhibitor-naïve patients to enroll
- The fall-back alpha spending approach is used to strongly control type I error between the testing of hypothesis in the first family (hematologic response) and those in the second family which consists of 2-year organ deterioration and mortality rate, overall survival and complete hematologic response. The closed sequential approach is used to control type I error of the testing of hypothesis within the second family.
- Remove the requirement to terminate the study at the second IA if the test for the primary endpoint is not statistically significant.

# TABLE OF CONTENTS

LIST OF ABBREVIATIONS AND GLOSSARY OF TERMS ..... 5
1. INTRODUCTION ..... 8
1.1 Study Design ..... 8
1.2 Study Objectives ..... 9
2. POPULATIONS FOR ANALYSIS ..... 11
2.1 Intent-to-Treat Population ..... 11
2.2 Hematologic Response-Evaluable Population ..... 11
2.3 Safety Population ..... 11
3. HYPOTHESES AND DECISION RULES ..... 11
3.1 Statistical Hypotheses ..... 11
3.2 Statistical Decision Rules ..... 12
4. INTERIM ANALYSIS ..... 15
4.1 Interim Analysis ..... 15
4.2 Independent Data Monitoring Committee (IDMC) ..... 16
4.3 Adjudication Committee (AC) ..... 16
5. STATISTICAL METHODOLOGY ..... 16
5.1 Sample Size Justification ..... 16
5.2 Randomization and Stratification ..... 17
5.3 Unblinding ..... 18
5.4 Data Handling ..... 18
5.4.1 Methods for Handling Missing Data ..... 18
5.4.2 Definition of Baseline Values ..... 21
5.4.3 Windowing of Visits ..... 21
5.4.4 Justification of Pooling ..... 21
5.4.5 Withdrawals, Dropouts, Loss to Follow-up ..... 21
5.5 Patient Disposition ..... 21
5.6 Demographics and Baseline Disease Characteristics ..... 22
5.6.1 Demographics ..... 22
5.6.2 Medical History ..... 22
5.6.3 Baseline Disease Status ..... 22
5.7 Treatments and Medications ..... 23
5.7.1 Concomitant Medications ..... 23
5.7.2 Study Treatments ..... 24
5.8 Efficacy Analyses ..... 26
5.8.1 Primary Efficacy Endpoint ..... 26
5.8.2 Key Secondary Efficacy Endpoints ..... 28
5.8.3 Other Secondary Efficacy Endpoints and Analyses ..... 30
5.9 Pharmacokinetic, Pharmacodynamic, and Biomarker Analysis ..... 33
5.9.1 Pharmacokinetic Analyses ..... 33
5.9.2 Pharmacodynamic Analyses ..... 33
5.9.3 Biomarker Analysis ..... 33
5.10 Resource Utilization and Patient Reported Outcome Analysis ..... 34
5.10.1 Patient Reported Outcomes (PROs) ..... 34

5.10.2 Health Economics (Health Care Resource Use) ..... 35
5.11 Safety Analyses ..... 36
5.11.1 Adverse Events ..... 36
5.11.2 Laboratory Data ..... 37
5.11.3 Electrocardiograms ..... 38
5.11.4 Vital Signs ..... 39
5.11.5 Eastern Cooperative Oncology Group (ECOG) Performance Status ..... 39
5.11.6 Other Safety Assessments ..... 39
6. CHANGES TO PLANNED ANALYSES FROM PROTOCOL ..... 39
7. PROGRAMMING CONSIDERATIONS ..... 39
7.1 Statistical Software ..... 39
7.2 Rules and Definitions ..... 39
8. APPENDIX ..... 40
8.1 Appendix 1: Amyloid-Related Hematologic and Organ Criteria ..... 40
8.2 Appendix 2: Proof of Strong Control of Type I Error Rate for Key Secondary Endpoints ..... 42
8.3 Appendix 3: Amendment 2 Detailed Summary of Changes ..... 45
8.4 Rationale for Amendment 1 ..... 54

# LIST OF ABBREVIATIONS AND GLOSSARY OF TERMS

| Abbreviation | Term |
|---|---|
| AC | Adjudication Committee |
| AE | adverse event |
| ACE | Angiotensin converting enzyme (ACE) |
| AL amyloidosis | primary systemic light chain amyloidosis |
| ALT | alanine aminotransferase |
| ANC | absolute neutrophil count |
| AST | aspartate aminotransferase |
| AUC | area under the plasma concentration versus time curve |
| $AUC_{0\text{-}24hr}$ | area under the plasma concentration versus time curve zero to 24 hours |
| $AUC_{0\text{-}tau}$ | area under the plasma concentration versus time curve zero to next dose |
| AV | atrioventricular |
| BIW | biweekly (ie, twice weekly) |
| BP | bodily pain |
| BSA | body surface area |
| BUN | blood urea nitrogen |
| CBC | complete blood count |
| CDF | cumulative distribution function |
| CHF | congestive heart failure |
| CI | confidence interval |
| CL | clearance |
| $CL_b$ | blood clearance |
| $CL_P$ | plasma clearance |
| $C_{max}$ | maximum plasma concentration |
| $CO_2$ | carbon dioxide |
| CR | complete response |
| CT | computed tomography |
| CYP | cytochrome $P_{450}$ |
| DDI | drug-drug interaction |
| dFLC | serum differential free light chain concentration; difference between amyloid forming and non amyloid forming FLC |
| DNA | deoxyribonucleic acid |
| DOR | duration of response |
| DVT | deep vein thrombosis |
| ECG | electrocardiogram |
| ECOG | Eastern Cooperative Oncology Group |
| eCRF | electronic case report form |
| EDC | electronic data capture |
| eGFR | estimated glomerular filtration rate; eGFR according to the recently recommended CKD-EPI equation |
| $E_{max}$ | maximum effect |
| EMG | electromyography |
| EDC | electronic data capture |

| Abbreviation | Term |
|---|---|
| eGRF | estimated glomerular filtration rate |
| EOT | End of Treatment (visit) |
| ESRD | end-stage renal disease |
| EQ-5D | EuroQol 5-Dimensional (questionnaire) |
| F | bioavailability |
| FA | final analysis |
| FISH | fluorescent in situ hybridization |
| FLC | free light chain |
| GADD34 | Growth Arrest DNA Damage 34 |
| GCP | good clinical practice |
| G-CSF | granulocyte colony stimulating factor |
| GH | general health |
| GI | gastrointestinal |
| GM-CSF | granulocyte macrophage-colony stimulating factor |
| hERG | human ether-à-go-go related gene |
| HDT-SCT | high-dose therapy- stem cell transplantation |
| HIV | human immunodeficiency virus |
| HU | health utilization |
| HemR | hematologic response |
| MLN9708 IB | investigator's brochure |
| IA | interim analysis |
| ICF | informed consent form |
| ICH | International Conference on Harmonisation |
| IDMC | independent data monitoring committee |
| IEC | independent ethics committee |
| IEM | immunogold electron microscopy |
| IHC | immunohistochemistry |
| IMiDs | immunomodulatory drugs |
| IRB | institutional review board |
| ISA | international society of amyloidosis |
| ITT | intent-to-treat (population) |
| K-M | Kaplan-Meier |
| LVEF | left ventricular ejection fraction |
| MCS | mental component summary |
| MedDRA | Medical Dictionary for Regulatory Activities |
| Millennium | Millennium Pharmaceuticals, Inc., and its affiliates |
| MM | multiple myeloma |
| MPD | maximum planned dose |
| MR | minimal response |
| MRI | magnetic resonance imaging |
| msec | Millisecond |
| NCCN | National Comprehensive Cancer Network |
| NCI | National Cancer Institute |

| Abbreviation | Term |
|---|---|
| NCI CTCAE | National Cancer Institute Common Terminology Criteria for Adverse Events |
| NDMM | newly-diagnosed multiple myeloma |
| NSAID | nonsteroidal anti-inflammatory drug |
| NT-proBNP | N-terminal proBNP |
| NYHA | New York Heart Association |
| ODM | organ deterioration and mortality |
| OS | overall survival |
| PCS | physical component summary |
| PD | progressive disease (disease progression) |
| PF | physical function |
| PFS | progression free survival |
| P-gp | P-glycoprotein |
| PK | pharmacokinetic(s) |
| PN | peripheral neuropathy |
| PO | *per os*; by mouth (orally) |
| PR | partial response |
| PRO | patient-reported outcome |
| RRMM | relapsed and refractory multiple myeloma |
| QALYs | quality adjusted life years |
| QOL | quality of life |
| QTc | rate-corrected QT interval (millisecond) of electrocardiograph |
| Q3C | every 3 cycles |
| RBC | red blood cell |
| RE | role emotional |
| RP | role physical |
| RRAL | relapsed or refractory AL amyloidosis |
| SAE | serious adverse event |
| SAP | statistical analysis plan |
| SCT | stem cell transplant |
| SF | social function |
| SMA | safety management attachment |
| SPEP | serum protein electrophoresis |
| TEAE | treatment-emergent adverse event ; may or may not be treatment related |
| TTF | time to treatment failure |
| ULN | upper limit of the normal range |
| UPEP | urine protein electrophoresis |
| US | United States |
| VT | vitality |
| VGPR | very good partial response |
| $V_{ss}$ | volume of distribution at steady state |
| WBC | white blood cell |
| WHO | World Health Organization |

# 1. INTRODUCTION

In general, the purpose of the Statistical Analysis Plan (SAP) is to provide a framework that addresses the protocol objectives in a statistically rigorous fashion, with minimized bias or analytical deficiencies. Specifically, this plan has the following purpose:

To prospectively (a priori) outline the types of analyses and data presentations that will addresses the study objectives outlined in the protocol, and to explain in detail how the data will be handled and analyzed, adhering to commonly accepted standards and practices of biostatistical analysis in the pharmaceutical industry.

## 1.1 Study Design

This is a phase 3, randomized, controlled, open-label, multicenter study of the oral formulation of dexamethasone plus MLN9708 compared with treatment chosen by the investigator from a prespecified list of regimens available in clinical practice. Treatment options will include: dexamethasone alone, dexamethasone plus an alkylating agent (melphalan or cyclophosphamide), or dexamethasone plus an immunomodulatory drug ([IMiD] thalidomide or lenalidomide) in patients with relapsed or refractory AL amyloidosis. Crossover to the investigational treatment arm is not permitted during participation in this study.

Eligible patients must have: 1) biopsy-proven AL-amyloidosis with relapsed or refractory disease despite 1 or 2 prior therapies; 2) disease requiring further treatment; 3) measureable disease as defined by serum differential free light chain concentration (dFLC); and 4) objective and measurable vital organ involvement (ie, cardiac or renal) as defined by the standard International Society of Amyloidosis (ISA) criteria. Patients enrolled based on protocol amendment 4 must not have been previously treated with proteasome inhibitors. (The Sponsor reserves the right to open the trial to proteasome inhibitor-exposed patients in the future, at some time point after the first interim analysis [IA].) The definition of relapsed is documented hematologic progressive disease (PD) after a response to prior therapy [PD more than 60 days of last dose]. The definition of refractory is documented absence of hematologic response or hematologic progression on or within 60 days of last dose of prior therapy.

Physicians will choose a treatment regimen from a list of options provided by the sponsor. Before randomization, physicians will declare which treatment regimen they plan to select for each screened patient; the selection will be recorded in the database. To maintain a balanced representation of the disease characteristics, patients enrolled in this study will be stratified by: 1) Cardiac Risk Stage: 1 versus 2 versus subgroup Cardiac Risk Stage 3(ie, both NT-proBNP and troponin T over threshold , but NT-proBNP < 8000 pg/mL); 2) relapsed versus refractory to last prior therapy; and 3) proteasome inhibitor naïve versus exposed.

Eligible patients will be randomized in a 1:1 ratio into 1 of the 2 study arms:

Arm A: dexamethasone plus MLN9708

Arm B: Physician's choice

In both treatment arms, each patient will continue to receive sequential cycles of therapy until disease progression/death, unacceptable toxicity, or until the study is terminated, whichever occurs first.

Response to therapy will be evaluated by an adjudication committee (AC) which will include the assessment of hematologic response and organ response according to the criteria outlined in the Revised Consensus Response Criteria of the ISA (Appendix 1), and vital organ (that is, heart and kidney) deterioration. An independent data monitoring committee (IDMC) will review safety and efficacy data at the interim analyses.

Safety will be assessed through adverse events (AEs), clinical laboratory tests, and vital sign measurements. In addition, quality of life (QOL) and health utilization (HU) will be assessed using questionnaires.

After disease progression, patients will be followed for survival, vital organ deterioration, and subsequent therapy at least every 12 weeks.

## 1.2 Study Objectives

The 2 primary objectives are:

- To determine whether dexamethasone plus MLN9708 improves hematologic response (PR + VGPR + CR) versus a Physician's choice of a chemotherapy regimen as selected from the list of offered treatment options in patients diagnosed with relapsed or refractory AL amyloidosis

- To determine whether dexamethasone plus MLN9708 improves 2-year vital organ (that is, heart or kidney) deterioration and mortality rate versus a Physician's choice of a chemotherapy regimen as selected from the list of offered treatment options in patients diagnosed with relapsed or refractory AL amyloidosis. Cardiac deterioration is defined as the need for hospitalization for heart failure. Kidney deterioration is defined as progression to end-stage renal disease (ESRD) with the need for maintenance dialysis or renal transplantation.

The key secondary objectives are:

- To determine overall survival (OS)
- To determine the complete hematologic response rate (CR)

Other secondary objectives are:

- To determine progression-free survival (PFS)
- To measure hematologic disease PFS
- To measure time to vital organ deterioration or death
- To measure vital organ response
- To measure vital organ PFS
- To measure the duration of hematologic response (DOR)
- To evaluate safety
- To measure time to treatment failure (TTF)
- To measure time to subsequent anticancer therapy
- To describe the impact of treatment on quality of life (QOL) using the Medical Outcomes Study 36- item Short Form General Health Survey (SF-36 v2), FACT-neurotoxicity subscale (FACT/GOG-Ntx) and a symptom scale questionnaire
- To evaluate health utilization (HU) and collect EuroQol 5-Dimensional (EQ-5D) data
- To collect PK data to contribute to population PK analyses
- To assess the association between clinical outcomes and gene polymorphisms that may relate to MLN9708 activity, such as NF-kB pathway-related genes NFKB2 and TRAF-3 in blood samples.

# 2. POPULATIONS FOR ANALYSIS

## 2.1 Intent-to-Treat Population

The Intent-to-Treat (ITT) population is defined as all patients who are randomized. Patients will be analyzed according to the treatment they were randomized to receive, regardless of any errors of dosing.

The ITT population will be used for the primary, secondary efficacy analyses, and resource utilization and patient reported outcome analysis.

## 2.2 Hematologic Response-Evaluable Population

The hematologic response-evaluable population is defined as patients who have measurable disease at baseline, who receive at least 1 dose of any treatment drug (study drug[Arm A] or standard drug [Arm B]), and have at least 1 postbaseline hematologic response assessment by an AC

Hematologic response-evaluable population will be used for sensitivity analysis of hematologic response rate.

## 2.3 Safety Population

The safety population is defined as all patients who receive at least 1 dose of any treatment drug.

Safety population will be used for all safety related analyses such as AE, concomitant medication, laboratory tests, and vital signs.

# 3. HYPOTHESES AND DECISION RULES

## 3.1 Statistical Hypotheses

There are two primary endpoints in this study.

The null and alternative hypothesis for hematologic response is:

$H_0$: Hematologic Response rate in Arm A = Hematologic Response Rate in Arm B

$H_a$: Hematologic Response rate in Arm A > Hematologic Response Rate in Arm B

The null and alternative hypothesis for 2-year vital organ deterioration and mortality rate is:

> $H_0$: 2-year vital organ deterioration and mortality rate in Arm A = 2-year vital organ deterioration and mortality rate in Arm B
>
> $H_a$: 2-year vital organ deterioration and mortality rate in Arm A < 2-year vital organ deterioration and mortality rate in Arm B

There are two key secondary efficacy endpoints in this study.

The null and alternative hypothesis for OS is:

> $H_0$: OS in Arm A = OS in Arm B
>
> $H_a$: OS in Arm A > OS in Arm B

The null and alternative hypothesis for complete hematologic response rate (CR) is:

> $H_0$: CR in Arm A = CR in Arm B
>
> $H_a$: CR in Arm A > CR in Arm B

### 3.2 Statistical Decision Rules

Hematological response will be tested using unstratified Cochran-Mantel-Haenszel (CMH) test at 0.04 significance level at the first IA; if p value is ≤ 0.04, then at the second IA, 2-year vital organ deterioration and mortality rate will be tested at 0.05 level; otherwise it will be tested at 0.01 level.

Hypothesis for 2-year vital organ deterioration and mortality rate will be tested at the second IA. If the test is not significant, the study may still continue, however no formal testing will be conducted. If the test is significant, the null hypothesis will be rejected, and the key secondary endpoint OS will be tested at the second IA. Using O'Brien-Fleming boundaries, if the alpha allocated to OS is 0.05, the trial may be stopped for overwhelming efficacy if the observed p value is less than 0.027 at the second IA assuming there are exactly 120 death events. The final analysis (FA) will be tested at 2-sided alpha level of 0.0423 (corresponding to nominal alpha of 0.023). If the alpha allocated to OS is 0.01, OS will be tested using a similar approach as above.

Hypothesis for the second key secondary endpoint, complete hematologic response rate, will be tested sequentially when OS hypothesis is rejected either at second IA or at FA. The testing level of significance will be the same as those used for OS. For instance if OS is

significant at 0.023 level at the FA, complete hematologic response rate will also be tested at 2-sided alpha level of 0.023 at the FA.

The type I error for all primary endpoints and key secondary endpoints using the testing procedure presented in Figure 1 is strongly controlled. Those endpoints are grouped into two ordered families 1) hematologic response, and 2) organ deterioration and mortality, OS and CR. The closed sequential testing approach will be used for strong type I error rate control in the second family. That is, OS will be only tested using the same alpha level assigned to 2-year organ deterioration and mortality rate if the test on 2-year organ deterioration and mortality rate is statistically significant, and then the alpha assigned to OS is allocated using O'Brien-Fleming boundaries for OS tests at second IA and FA; and CR will be only tested at the same alpha level of OS at second IA or FA if the test on OS is significant at second IA or FA. For testing the two families, the fall-back approach (Wiens B.L., 2003) is used to control type I error rate.

This approach for the key secondary endpoints is similar to strategy 2 from Hung, Wang, and O'Neill (2007) for family-wise Type I error control; that is, testing the second endpoint conservatively using the same rejection boundary as the first endpoint to be tested. In Hung's paper, they assume that the correlation (information fraction) for the second endpoint between the IA and FA is the same as that for the first endpoint. However, this might not be the case in our study since complete hematologic response rate is gated by OS. By the time of the second IA, all patients will have been enrolled and contributed to the complete hematologic response assessment, which leads to the fact that the correlation between test statistics for the second endpoint is much higher than that of the first endpoint OS. Therefore, instead of using the same rejection boundary as the first endpoint, it is proposed to use the same alpha level for the second endpoint. The proof of strong control of family-wise error rate for both first and second endpoints is provided in Appendix 2 of this document.

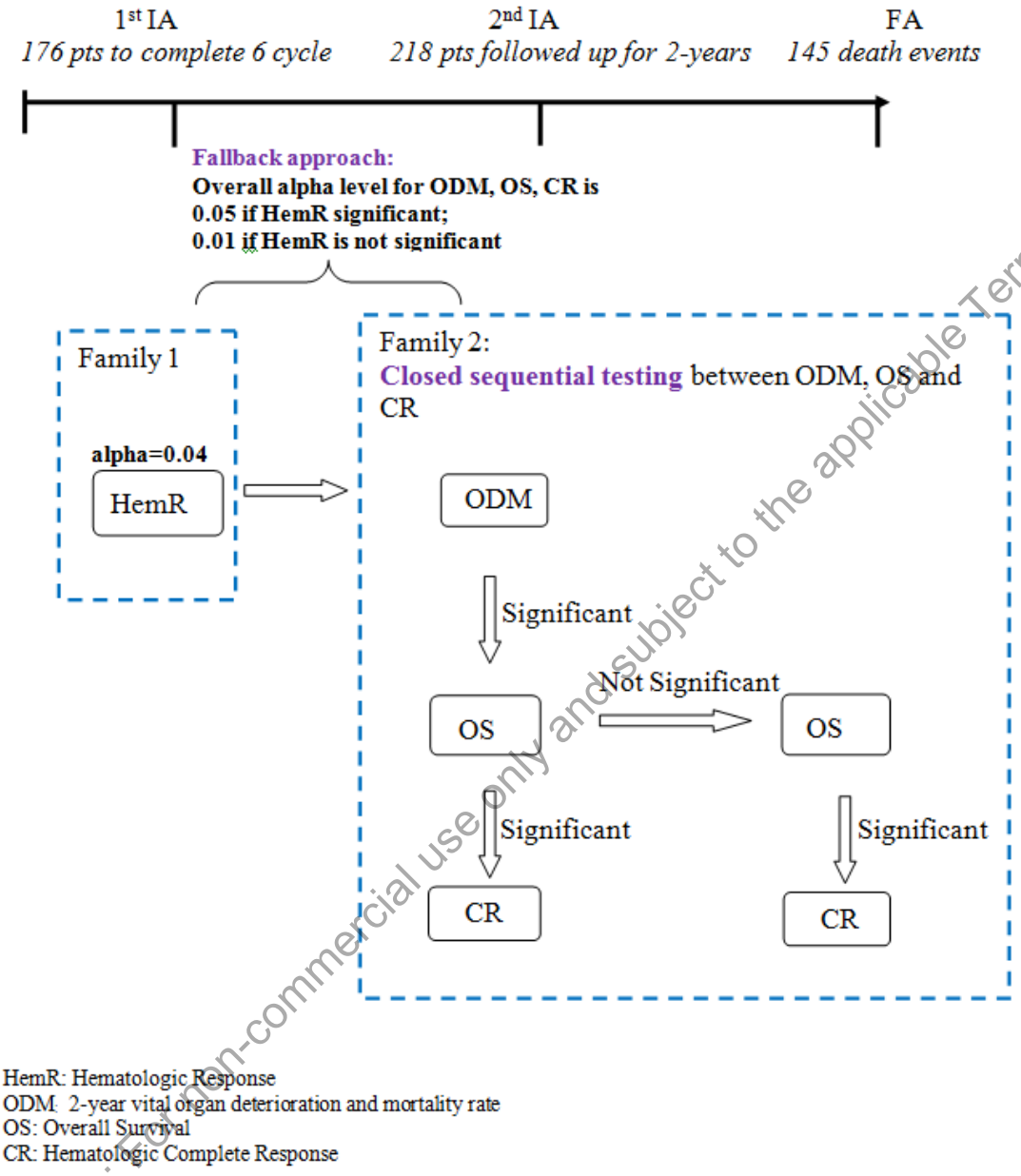


Figure 1: Statistical Testing Procedure and Alpha Spending Schema

The rejection boundary for the secondary endpoints at the FA will be calculated based on the actual correlation between the test statistics at the second IA and FA. Examples of critical values at the FA based on different correlations between the test statistics at IA and FA using a 2- sided alpha of 0.023 are presented below. The formula to calculate the critical values is also shown in the proof of strong control of error rate in Appendix 2.

| Information fraction( $n_1/n_2$) | Correlation ($\sqrt{n_1/n_2}$) | Critical value |
|---|---|---|
| 2/3 | 0.82 | 2.09 |
| 5/6 | 0.91 | 2.03 |
| 1 | 1 | 1.96 |

where $n_1$ and $n_2$ are the size of patient population at the second IA and FA, respectively

Since in the family 2, the testing of key secondary endpoints is gated by the significance of the second primary endpoint, and between family 1 and family 2, the type I error rate is controlled at 2-sided alpha level by using the fall-back approach, the overall type I error rate for all endpoints, i.e. both primary and all key secondary endpoints, is also controlled strongly at 2-sided 0.05 level.



# 4. INTERIM ANALYSIS

## 4.1 Interim Analysis

There are two planned formal IAs. The first IA will be performed when approximately 176 patients have been enrolled and have had the opportunity to complete 6 cycles of treatment or have discontinued study treatment before receiving 6 cycles of treatment. At that time, relevant safety and efficacy data from these patients will be queried and cleaned; all analyses at this IA will be based on these patients whose data have been cleaned. This will be the FA for hematologic response. At the first IA, hematologic response will be tested using unstratified CMH at a 2-sided alpha level of 0.04. If the test is statistically significant, the study will continue to test organ deterioration and mortality rate at a 2-sided alpha of 0.05 at the second IA; otherwise organ deterioration and mortality rate will be tested at the second IA with a 2-sided alpha of 0.01. This IA is expected to occur approximately 54 months after the first patient is enrolled. During this IA, enrollment to the study will continue.

The second IA will be performed when approximately 218 patients enrolled have had the opportunity to complete 2 years of treatment or followed up for at least 2 years if discontinuing treatment before receiving 2 years of treatment. This IA is expected to occur approximately 98 months from the first patient is enrolled. At that time, relevant safety and efficacy data from those patients will be queried and cleaned; all analyses at this IA will be based on these patients whose data have been cleaned.

If the test for 2-year vital organ deterioration and mortality rate is significant at the second IA, the analyses on OS will be performed and the test significance level on OS for the second IA and FA will be determined using O'Brien-Fleming boundaries. Assuming there are exactly 120 death events at second IA, the trial will be stopped for overwhelming

efficacy if the observed p value is less than 0.027. The FA will be tested at 2-sided alpha level of 0.0423 (corresponding to nominal alpha of 0.023). If the alpha allocated to OS is 0.01, OS will be tested using a similar approach as above using O'Brien-Fleming boundaries.

### 4.2 Independent Data Monitoring Committee (IDMC)

An IDMC will periodically review safety and efficacy data at regularly scheduled meetings pre-specified in the IDMC charter. A detailed charter outlining all activities of the IDMC (eg, type of data reviewed, frequency of meetings, location of meetings, etc.) will be finalized during its initial meeting. At the time of the interim analyses, the IDMC will be responsible to provide a recommendation regarding study continuation based on the safety and efficacy parameters.

### 4.3 Adjudication Committee (AC)

Hematologic response/progression and organ response/progression will be assessed by an Adjudication Committee (AC) who are blinded to the treatment assignment. Assessment of hematologic response and organ response will be based on central laboratory results and will follow the criteria outlined in the Revised Consensus Response Criteria of the international Society of Amyloidosis (ISA). The detailed criteria are included in Appendix 1. Vital organ deterioration will also be assessed by the AC.

## 5. STATISTICAL METHODOLOGY

In general, summary tabulations will be presented that display the number of observations, mean, standard deviation, median, minimum, and maximum for continuous variables, and the number and percent (of nonmissing) per category for categorical data, unless specified otherwise.

### 5.1 Sample Size Justification

Two primary endpoints, hematologic response and 2-year vital organ deterioration and mortality rate, along with key secondary endpoints of OS and complete hematologic response rate, will be sequentially tested in this study.

The total sample size was calculated to provide 80% power for the assessment of OS with the allocated alpha 0.05. The study is also adequately powered to test both primary endpoints: hematologic response rate and 2-year vital organ deterioration and mortality rate. There are two planned IAs and one FA.

The parameters used for sample size calculation on OS are a 2-sided test at the significance level of $\alpha = 0.05$, power of 80%, a control arm median OS of 26 months, and testing arm median OS of 41.6 months (assuming exponential distribution, hazard ratio 0.625). With 1 IA (second) and 1 FA occurring at information time of 0.89 and 1, an average enrollment rate of 3 patient /month for the first 5 months, and 3-5 patient/month thereafter, an additional 28 months follow up for all patients after last patient is enrolled, and approximately 10% drop out rate, a total of 248 patients will need to be randomized in a 1:1 ratio into 2 treatment arms in order to achieve 145 death events. The O'Brien-Fleming stopping boundary (Lan-DeMets method) will be used to assign alpha level to the second IA and FA.

The parameters used for hematologic response endpoint are a 2-sided test at the significance level of $\alpha = 0.04$, power of 90%, a null hypothesis hematologic response rate 40%, and an alternative hypothesis hematologic response rate 65%. Approximately 176 patients are needed for two-sample test of the difference of proportions. Testing for hematologic response hypothesis will be performed when approximately 176 patients enrolled have had opportunity to either complete 6 cycles of treatment or discontinue treatment before receiving 6 cycles of treatment. This is the first IA for the study, also the FA for hematologic response for statistical testing purpose, with the opportunity to claim hematologic response rate benefit. If the test is statistically significant in favor of the dexamethasone plus MLN9708 arm, the second IA for organ deterioration and mortality will be tested at 2-sided $\alpha = 0.05$ based on fallback approach, otherwise it'll be tested at 2-sided $\alpha = 0.01$.

The second IA will be performed when approximately 218 patients enrolled have had the opportunity to complete 2 years of treatment or discontinue treatment before receiving 2 years of treatment. This will be the FA for 2-year vital organ deterioration and mortality rate. With 218 patients, the endpoint of 2-year vital organ deterioration and mortality rate is powered at 90% at a 2-sided alpha level of 0.05 with the assumption of 80% deterioration and mortality rate in the control arm and 60% rate in the MLN9708 arm. If first IA result is not significant, power of 2-year vital organ deterioration and mortality rate will be 74% at a 2-sided alpha level of 0.01. If the test for 2-year vital organ deterioration and mortality rate is not significant, there will not be any formal testing; otherwise, OS will be tested. If the test for OS is statistically significant, the complete hematologic response rate will be tested at the same alpha level as that for OS and the study may be stopped for evidence of efficacy; if the test for OS is not statistically significant, the study will continue to the FA.

## 5.2 Randomization and Stratification

Randomization scheme will be generated by an independent biostatistician at Millennium who is not on the study team. Before dosing, a randomization number will be assigned to each patient. The randomization assignment will be implemented by an Interactive Voice Response System (IVRS).

Eligible patients will be randomized in a 1:1 ratio into 1 of the 2 treatment arms, stratified by: 1) Cardiac Risk Stage: 1 versus 2 versus subgroup Cardiac Risk Stage 3(ie, both NT-proBNP and troponin T over threshold, but NT-proBNP < 8000 pg/mL); 2) relapsed versus refractory to last prior therapy; and 3) proteasome inhibitor naïve versus exposed.

## 5.3 Unblinding

This is an open-label study; investigators and patients will know the individual treatment assignments. However, the Millennium and CRO study team, investigators, and patients will be blinded to the aggregate efficacy results. Efficacy data will be masked with dummy patient identification number during the study conduct for data review/cleaning purposes. Only limited Millennium and CRO personnel will have access to un-blinded individual patient level data in the electronic data capture system. The periodic safety analyses will be generated for the IDMC by CRO's statistical group. Two formal interim efficacy analyses will be conducted by an independent statistical center (ISC) for the IDMC.

Refer to section 4 for the roles and responsibilities of IDMC and AC.

## 5.4 Data Handling

### 5.4.1 Methods for Handling Missing Data

All available efficacy and safety data will be included in data listings and tabulations. Data that are potentially spurious or erroneous will be examined under the auspices of standard data management operating procedures.

In general, missing data will be treated as missing and no data imputation will be applied, unless otherwise specified.

For patient reported outcomes data, primarily missing data imputation will be based on published instrument specific methods. Other missing data imputation method such as Last Observation Carry Forward (LOCF), random slope model, and pattern mixture model may be explored as sensitivity analyses for patient reported outcomes data.

For the primary analysis on the first primary endpoint (hematologic response rate), missing value is defined as no post-baseline hematologic response assessment either due to lost to follow-up or withdrawal by patients. If the hematologic response assessment in either arm is missing on comparing hematologic response rates, it will be counted as a failure (non-responder) instead of a missing value.

For the primary analysis on the second primary endpoint (2-year organ deterioration and mortality rate), missing value is defined as no documented vital organ deterioration or death event within the first 2 years and no further information available regarding vital organ

deterioration or death event on or after 2 years from the date of first dosing of the treatment drug (study drug [Arm A] or standard drug [Arm B]). All event (vital organ deterioration or death) occurred within the first 2 years will be counted as one event. If the 2-year vital organ deterioration assessment or death information is missing in either arm, it will be counted as a failure instead of a missing value; that is, patient who has a missing event will be counted as an event towards vital organ deterioration or death.

#### 5.4.1.1 Missing/Partial Dates in Screening Visit

The following rules apply to dates recorded in the screening visits.

- 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 dose of treatment drug. Otherwise, the 15th will be used.

- If only a year is present, and it is the same as the year of the first dose of treatment drug, the 15th of January will be used unless it is later than the first dose, in which case the date of the first of January will be used, unless other data indicates that the date is earlier.

- If only a year is present, and it is not the same as the year of the first dose of treatment drug, the 15th of June will be used, unless other data indicates that the date is earlier.

#### 5.4.1.2 Missing/Partial Dates in Adverse Events/Concomitant Therapies/Subsequent Therapies

Every effort will be made to avoid missing/partial dates in on-study data.

Adverse events with stop dates that are completely or partially missing will be imputed as follows:

- If the stop date has month and year but day is missing, the last day of the month will be imputed
- If the stop date has year, but day and month are missing, the 31$^{th}$ of December will be imputed

After the 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.

Adverse events with start dates that are completely or partially missing will be imputed as follows:

- If the start date has month and year but day is missing, the first day of the month will be imputed
    - If this date is earlier than the first dose date, then the first dose date will be used instead
    - If this date is later than the stop date (possibly imputed), then the stop date will be used instead
- If the start date has year, but day and month are missing, the 15th of June will be imputed
    - If this date is earlier than the first dose date, then the first dose date will be used instead
    - If this date is later than the stop date (possibly imputed), then the stop date will be used instead

If the start date of an event is completely missing, then it is imputed with the first dose date.

Concomitant therapies with start dates that are completely or partially missing will be analyzed as follows:

- If the start date has month and year but day is missing, the therapy will be included in the summary table if the month and year of the start date of the event are:
    - On or after the month and year of the date of the first dose of treatment drug

    and

    - On or before the month and year of the date of the last dose of treatment drug plus 30 days.
- If the start date has year, but day and month are missing, the therapy will be included in the summary table if the year of the start date of the event is:
    - On or after the year of the date of the first dose of treatment drug

    and

    - On or before the year of the date of the last dose of treatment drug plus 30 days.

If the start date of an event is completely missing, then the therapy will be included in the summary table.

Subsequent therapies with start dates that are completely or partially missing will be analyzed as follows:

- 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 last dose with treatment drug, the day of last dose + 1 will be imputed.
    - If the onset month and year are not the same as the month and year of last dose with treatment drug, the first day of the month is imputed.
- When only a year is present,
    - If the onset year is the same as the year of last dose with treatment drug, the date of last dose + 1 will be imputed.
    - If the onset year is not the same as the year of last dose with treatment drug, the first day of the year is imputed.
- If no components of the onset date are present the date of last dose + 1 will be imputed.

### 5.4.2 Definition of Baseline Values

Unless otherwise specified, the baseline value is defined as the value collected at the time closest to, but prior to, the start of treatment drug administration.

### 5.4.3 Windowing of Visits

All data will be categorized based on the scheduled visit at which it was collected. These visit designators are predefined values that appear as part of the visit tab in the eCRF.

### 5.4.4 Justification of Pooling

All data from all sites will be pooled. Study center or treatment-by-center interaction will not be included in any statistical analysis.

### 5.4.5 Withdrawals, Dropouts, Loss to Follow-up

Time to event parameters will be censored if patients withdraw, drop out, or are lost to follow-up before documentation of the events (progressive disease / death). Rules for censoring are detailed in section 5.8.

## 5.5 Patient Disposition

A disposition of patients includes the number and percentage of patients for the following categories: patients in each of the study population, patients discontinued from the

treatment, primary reason to discontinue from the treatment, patients discontinued from the study, and primary reason to discontinue from the study. All percentages will be based on the number of patients in the ITT population.

A listing will present data concerning patient disposition.

## 5.6 Demographics and Baseline Disease Characteristics

### 5.6.1 Demographics

Demographics will be summarized by treatment arms in a descriptive fashion in the ITT population. Baseline demographic data to be evaluated will include age, sex, race, height, weight, and other parameters as appropriate. Patient enrollment by region and country will also be summarized by treatment arms.

### 5.6.2 Medical History

Medical history will be presented in a by-patient listing. Specific neurologic (11 questions with grade) and cardiac (7 questions with grade) medical history information will be summarized by treatment group in the ITT population. Smoking history will also be summarized similarly in the ITT population.

### 5.6.3 Baseline Disease Status

Disease status at initial diagnosis will be summarized by the treatment arms in the ITT population, including cardiac biomarker stage, NYHA classification, and sites of amyloid involvement,

Baseline disease status will be summarized by the treatment arms in the ITT population, including time since initial diagnosis (months), type and number of organ involvement, type and quantity of involved free light chain, dFLC, serum κ/λ ratio, serum and urine m-protein quantity, serum and urine immunofixation results, skeletal survey results, cardiac biomarker stage, Eastern Cooperative Oncology Group (ECOG) performance status, ß2-microglobulin, serum creatinine, serum creatinine clearance, total urine creatinine, urine creatinine clearance, serum albumin, alkaline phosphatase, ALT, AST, serum cardiac markers (NT-proBNP, BNP, troponin T), NYHA classification, and other parameters as appropriate. Separate by-patient listings will also be presented.

Creatinine clearance is to be calculated using the Cockcroft-Gault formulas as follows:

For male patients:

$$\text{creatinine clearance} = \frac{(140 - \text{Age[yrs]}) \times \text{weight[kg]}}{72 \times (\text{serum creatinine[mg/dL]})}$$

For female patients:

$$\text{creatinine clearance} = 0.85 \times \frac{(140 - \text{Age[yrs]}) \times \text{weight[kg]}}{72 \times (\text{serum creatinine[mg/dL]})}$$

Integer values will be used.

A separate table will summarize the numbers and percentages of patients who had prior systemic therapy, prior transplant, prior surgery, and prior radiation therapy. This table will also include regimens and the number of lines of prior systemic therapies, best organ and/or hematologic responses to prior systemic therapy, relapse/refractory status to prior systemic therapy, and months since progression from last prior systemic therapy and months since diagnosis for each treatment group in the ITT population. By-patient listing will also be presented for prior systemic therapy.

Months since diagnosis for each treatment is calculated by

$$\frac{\text{randomization date - date of diagnosis}}{365.25/12}$$

Distribution of stratification factors will also be summarized.

## 5.7 Treatments and Medications

### 5.7.1 Concomitant Medications

Concomitant medications will be coded by preferred term using the World Health Organization (WHO) Drug Dictionary. The number and percentage of patients taking concomitant medications from screening through the end of the on-treatment period will be tabulated by Anatomical Therapeutic Chemical (ATC) classification pharmacological subgroup and WHO drug generic term for each treatment group in the safety population. By-patient listing will also be presented for concomitant medications.

Concomitant procedures will not be coded, but will be presented in a data listing in the safety population.

### 5.7.2 Study Treatments

Prior to randomization, the physician will choose a treatment regimen from the list of options provided by the sponsor for each screened patient. The physician's choice will be collected and recorded in the database. Eligible patients will be randomized to 1 of 2 study arms in a 1:1 ratio. Once randomized, patients must start therapy within 5 business days.

Arm A: Patients will receive MLN9708 (4.0 mg) PO on Days 1, 8, and 15 plus dexamethasone 20 mg/day PO weekly on Days 1, 8, 15, 22 of each 28-day cycle; dexamethasone may be increased up to 40 mg/day after 4 weeks, if tolerated. Patients may continue to receive treatment until PD or unacceptable toxicity, whichever comes first.

Arm B: Patients will receive **one** of the following treatment options as selected by the physician:

- **Dexamethasone:** Dexamethasone 20 mg/day PO on Days 1-4, 9-12, 17-20 of each 28-day cycle; dexamethasone may be increased up to 40 mg/day after 4 weeks, if the lower dose is tolerated without any > Grade 2 dexamethasone-related toxicities. Dose adjustments are possible based on toxicities experienced.

- **Dexamethasone plus melphalan:** Dexamethasone 20 mg/day PO on Days 1-4 of each 28-day cycle; dexamethasone may be increased up to 40 mg/day after 4 weeks if the lower dose is tolerated without any > Grade 2 dexamethasone-related toxicities, plus melphalan 0.22 mg/kg PO on Days 1-4 every 28 days. Melphalan has to be dose adjusted in patients with renal function impairment**.** Dose adjustments are possible based on toxicities experienced.

- **Dexamethasone plus cyclophosphamide:** Dexamethasone 20 mg/day PO weekly on Days 1, 8, 15, and 22 of each 28-day cycle; dexamethasone may be increased up to 40 mg/day after 4 weeks, if the lower dose is tolerated without any > Grade 2 dexamethasone-related toxicities, plus cyclophosphamide 500 mg PO Days 1, 8, and 15, every 28 days. Dose adjustments are possible based on toxicities experienced.

- **Dexamethasone plus thalidomide**: Dexamethasone 20 mg/day PO weekly Days 1, 8, 15, and 22 of each 28-day cycle; dexamethasone may be increased up to 40 mg/day after 4 weeks, if the lower dose is tolerated without any > Grade 2 dexamethasone-related toxicities, plus thalidomide total dose 200 mg/day PO (starting dose 50 mg/day, increased as tolerated to a maximum of 200 mg/day). Dose adjustments are possible based on toxicities experienced.

- **Dexamethasone plus lenalidomide**: Dexamethasone 20 mg/day PO weekly on Days 1, 8, 15, and 22 of each 28-day cycle; dexamethasone may be increased up to

40 mg/day after 4 weeks, if the lower dose is tolerated without any > Grade 2 dexamethasone-related toxicities, plus lenalidomide 15 mg/day for 21 days every 28 days. Dose adjustments are possible based on toxicities experienced.

Patients may continue to receive treatment until PD or unacceptable toxicity.

#### 5.7.2.1 Extent of Exposure

An overall summary of drug exposure will be presented including number of treated cycles, numbers and percentages of patients who had ≥1, ≥2, …, and ≥12 treated cycles, for each treatment group in the safety population. Aggregate summary of numbers and percentages of patients who had 1-6, 7-12, ≥13 treated cycles will also be presented in the same table. For patients in arm B, summarization will be based on the actual treatment option received.

Additionally exposure to dexamethasone will be characterized by total amount of dose taken in mg, total number of dose taken, number of treated cycles, numbers and percentages of patients who had ≥1, ≥2, …, and ≥12 treated cycles, and relative dose intensity (%) for each treatment group in the safety population. Aggregate summary of numbers and percentages of patients who had 1-6, 7-12, ≥13 treated cycles will also be presented in the same table. For patients in arm B, summarization will be based on the actual treatment option received.

MLN9708, melphalan, cyclophosphamide, thalidomide, and lenalidomide exposure will be summarized similarly as dexamethasone for the applicable treatment group/option.

A treated cycle is defined as a cycle in which the patient received any amount of any treatment drug.

A treated cycle for a specific drug is defined as a cycle in which the patient received any amount of the specific drug.

Prescribed dose for MLN9708, melphalan, cyclophosphamide, and lenalidomide is determined by the dose level to which a patient is enrolled at the onset of the study. Prescribed dose for dexamethasone will be increased up to 40 mg/day after 4 weeks if a dose increase is applicable for a patient. Prescribed dose for thalidomide will be increased up to 200 mg/day according to the actual dose increase schedule for a patient.

Relative dose intensity (%) is defined as 100 x (total dose received in mg) / (sum of prescribed dose over all treated cycles).

Dosing data will also be presented in a by-patient listing.

#### 5.7.2.2 Treatment Modifications

Action on each treatment drug will be summarized by each of the Cycle 1 through 12, sum of the remainder Cycles, Cycles 1-6, Cycles 7-12, and total for each treatment group in the safety population. For patients in arm B, summarization will be based on the actual treatment option received.

## 5.8 Efficacy Analyses

All efficacy evaluations will be conducted using the ITT population. At the IAs, relevant safety and efficacy data from these patients for the primary analyses will be queried and cleaned; all analyses at this IA will be based on these patients whose data have been cleaned. Hematologic response-evaluable population will be used for sensitivity analyses for hematologic response as specified below.

### 5.8.1 Primary Efficacy Endpoint

There are 2 primary endpoints: hematologic response and 2-year vital organ (that is, heart or kidney) deterioration and mortality rate.

Hematologic Response

Hematologic response rate is defined as the proportion of patients who achieved PR or better in the ITT population. Hematologic response is assessed by AC using central laboratory results and ISA criteria.

2-year vital organ (that is, heart or kidney) deterioration and mortality rate

Cardiac deterioration is defined as the need for hospitalization for heart failure. Kidney deterioration is defined as progression to end-stage renal disease (ESRD) with the need for maintenance dialysis or renal transplantation. Vital organ deterioration is assessed by AC as well.

#### 5.8.1.1 Primary Efficacy Analysis

Two primary efficacy endpoints will be tested sequentially in the order of 1) hematologic response; 2) 2-year vital organ (that is, heart or kidney) deterioration and mortality rate.

Hematologic Response

If the hematologic response assessment in either arm is missing on comparing hematologic response rates, it will be counted as a failure (non-responder) instead of a missing value.

The unstratified CMH test will be used to compare hematologic response rate between the 2 treatment arms when 176 ITT patients have had the opportunity to receive 6 cycles of treatment, or discontinued treatment prior to 6 cycles. If the hematologic response rate is higher in the dexamethasone plus MLN9708 arm (arm A), and the 2-sided CMH test p value is $\leq 0.04$, null hypothesis will be rejected. Therefore statistical significance will be claimed for dexamethasone plus MLN9708.

A logistic regression model will be used to estimate the treatment effect in terms of odds ratio. The odds ratio and its associated 95% confidence intervals (CIs) will be presented.

2-year vital organ deterioration and mortality rate

If the 2-year vital organ deterioration assessment or death information is missing in either arm, it will be counted as a failure instead of a missing value; that is, patient who has a missing event will be counted as an event towards vital organ deterioration or death.

The unstratified CMH test will be used to make comparisons between the 2 treatment arms at the second IA when approximately 218 patients enrolled have had the opportunity to complete 2 years of treatment or have discontinued treatment before receiving 2 years of treatment. If the 2-year organ deterioration and mortality rate is lower in the dexamethasone plus MLN9708 arm (arm A), and the 2-sided CMH test p value is $\leq 0.05$ when hematologic response endpoint is significant or $\leq 0.01$ when hematologic response endpoint is not significant, null hypothesis will be rejected. Therefore statistical significance will be claimed for dexamethasone plus MLN9708.

A logistic regression model will be used to estimate the treatment effect in terms of odds ratio. The odds ratio and its associated 95% confidence intervals (CIs) will be presented.

#### 5.8.1.2 Sensitivity Analyses for Primary Endpoints

It is expected that the missing values will be limited and therefore, the primary analysis on both primary endpoints will not be driven by the handling of the missing data. To further assess the impact of missing values and potential other factors, a range of sensitivity analyses are proposed as below:

1) Hematologic response and vital organ deterioration assessed by investigator in the ITT population

2) Stratified CMH method

3) Binomial test with standard error estimate for the difference in rates for both primary endpoints
Let $n_A$ and $n_B$ be the total number of subjects for the respective arms and $n_A^*$ and $n_B^*$ are the total number of known outcomes (missing values will be excluded) for the respective arms; and $\hat{p}_A$ and $\hat{p}_B$ are the proportion of successes (known as no vital organ deterioration or death within 2-years) used in the analysis based on all subjects in the respective arms. Then the test statistics will be:

$$(\hat{p}_A - \hat{p}_B) \Big/ \sqrt{\frac{n_A \hat{p}_A (1-\hat{p}_A)}{(n_A^*)^2} + \frac{n_B \hat{p}_B (1-\hat{p}_B)}{(n_B^*)^2}}$$

4) Subgroup analyses for hematologic response and 2-year vital organ deterioration and mortality rate will include the stratification factors, and other baseline prognostic factors to be determined before database lock.

5) Hematologic response assessed by AC in the hematologic response-evaluable population

### 5.8.2 Key Secondary Efficacy Endpoints

There are 2 key secondary endpoints: OS and complete hematologic response rate.

Overall Survival

OS is defined as time from the date of randomization to the date of death. Patients without documentation of death at the time of analysis will be censored at the date last known to be alive.

Complete Hematologic Response Rate

Complete hematologic response rate is defined as the proportion of patients who achieved CR in the ITT population. Complete hematologic response is assessed by AC using central laboratory results and ISA criteria.

#### 5.8.2.1 Key Secondary Efficacy Analysis

Two key secondary efficacy endpoints will be tested sequentially in the order of 1) OS; 2) complete hematologic response rate when the test of 2-year vital organ deterioration and mortality rate is statistically significant at the second IA. The alpha level for complete hematologic response rate will be the same alpha level as that for the OS analysis either at the second IA or at the FA.

Overall Survival

OS will only be tested after statistical significance is achieved for dexamethasone plus MLN9708 in 2-year vital organ deterioration and mortality rate analysis. A 2-sided, stratified log-rank test will be used to compare the treatment arms with respect to OS. OS will be tested at the second IA and FA if necessary. O'Brien-Fleming boundaries will be calculated using the Lan-DeMets method. The information fraction at the second IA is calculated as (number of death events/145). In addition, an unadjusted stratified Cox regression model will be used to estimate the hazard ratio and its 95% CIs for the treatment effect using the stratification factors. The Kaplan-Meier (K-M) survival curves and K-M medians (if estimable), along with their 2-sided 95% CIs, will also be provided for each treatment group. If the log-rank test statistics cross the efficacy boundary at the second IA in favor of the dexamethasone plus MLN9708 arm (arm A), statistical significance will be claimed and the study will be stopped at the second IA. Otherwise, the study will continue until 145 death events are observed. At the FA if the log-rank test statistics crosses the efficacy boundary in favor of the dexamethasone plus MLN9708 arm (arm A), statistical significance will be claimed.

Complete Hematologic Response Rate

The unstratified CMH test will be used to compare complete hematologic response rates between the 2 treatment arms in the ITT population. A logistic regression model will be used to estimate the treatment effect in terms of odds ratio. The odds ratio and its associated 95% CIs will be presented. If the complete hematologic response rate is higher in the dexamethasone plus MLN9708 arm (arm A), and the 2-sided CMH test statistics cross the efficacy boundary, null hypothesis will be rejected. Statistical significance will be claimed for dexamethasone plus MLN9708.

#### 5.8.2.2 Sensitivity Analyses for Key Secondary Endpoints

Complete hematologic response assessed by investigator will be summarized in the ITT population. Complete hematologic response rate will also be summarized in the hematologic response-evaluable population. In addition, the binomial test described in section 5.8.1.2 as the sensitivity analysis method for primary endpoints will be performed if necessary.

Sensitivity analyses for OS include: 1) adjustment for prognostic factors in Cox regression model; 2) unstratified log-rank test and unstratified Cox regression model; 3) Subgroup analysis based on prognostic factors.

### 5.8.3 Other Secondary Efficacy Endpoints and Analyses

Other secondary efficacy parameters include PFS, duration of hematologic response, hematologic disease PFS, time to vital organ deterioration or death, vital organ response rate, vital organ PFS, TTF, and time to subsequent anticancer therapy. In addition, organ response rate and organ PFS will be presented for any involved organ at the baseline.

PFS

PFS is defined as the time from the date of randomization to the date of first documentation of hematologic disease progression, or organ (cardiac and renal) progression, or death due to any cause, whichever occurs first.

Patients without documentation of hematologic PD and vital organ progression will be censored at the date of last hematologic response assessment that is SD or better, or the date of last vital organ assessment SD or better, whichever occurs last.

PFS will be analyzed using the similar method as OS.

Duration of Hematologic Response

Duration of hematologic response is defined as the time from the date of first documentation of a hematologic PR or better to the date of first documentation of hematologic PD for hematologic responders.

Hematologic responders without documentation of hematologic PD will be censored at the date of last hematologic response assessment that is SD or better.

Duration of hematologic response will be summarized descriptively using the Kaplan-Meier method.

Hematologic Disease PFS

Hematologic disease PFS is defined as the time from the date of randomization to the date of first documentation of hematologic PD, or death due to any cause, whichever occurs first.

Patients without documentation of hematologic PD will be censored at the date of last hematologic response assessment that is SD or better.

Hematologic disease PFS will be analyzed using the similar method as OS.

Time to Vital Organ Deterioration or Death

Time to vital organ deterioration or death defined as the time from randomization to vital organ (heart or kidney) deterioration or death, whichever occurs first. Cardiac deterioration is defined as the need for hospitalization for heart failure. Kidney deterioration is defined as progression to ESRD with the need for maintenance dialysis or renal transplantation.

Patients without documentation of organ deterioration or death will be censored at the date of last assessment.

Time to vital organ deterioration or death will be analyzed using the similar method as OS.

Vital Organ Response Rate

Vital organ (heart and kidney) response rate is defined as the proportion of patients who achieved vital organ response in the ITT population. Vital organ response will be assessed by the AC using central laboratory results and ISA criteria. Changes from baseline in the vital organs will be documented as "response", "no change", or "progression". An overall determination of vital organ response will then be documented for the given time point.

- Progression of one of the two vital organs will equate to organ progression
- Response of one or two of the involved vital organs with no change from baseline in the rest of other involved vital organs will equate to vital organ response
- No change from baseline in all involved vital organs will equate to stable organ disease.

The vital organ response rate will be analyzed using the similar method as hematologic response rate.

Vital Organ PFS

Vital organ PFS is defined as the time from the date of randomization to the date of first documentation of vital organ progression, or death due to any cause, whichever occurs first.

Patients without documentation of vital organ progression will be censored at the date of last vital organ assessment SD or better.

Vital organ PFS will be analyzed using the similar method as OS.

Time to Treatment Failure

TTF is defined as the time from randomization to the date of first documented treatment failure. The following events are considered as treatment failure:

- Death due to any cause
- Hematologic progression by AC
- Vital organ (cardiac and renal) progression by AC
- Hematologic response with stable but clinically morbid organ disease requiring additional therapy
- Withdrawn from study for any reason

Patients without documentation of treatment failure will be censored at the date of last response assessment.

TTF will be analyzed using the similar method as OS.

Time to Subsequent Anticancer Therapy

Time to subsequent anticancer therapy is defined as the time from randomization to the first date of subsequent anticancer therapy.

Patients without documented subsequent anticancer therapy will be censored at the date of death or last known to be alive.

Time to subsequent anticancer therapy will be analyzed using the similar method as OS.

Organ Response Rate

Organ response rate is defined as the proportion of patients who achieved organ response in the ITT population. Organ response will be assessed by the AC using central laboratory results and ISA criteria. Changes from baseline in any involved organs will be documented as "response", "no change", or "progression". An overall determination of organ response will then be documented for the given time point.

- Progression of any involved organs will equate to organ progression
- Response of any involved organs with no change from baseline in the rest of involved organs will equate to vital organ response

- No change from baseline in all involved organs will equate to stable organ disease.

The organ response rate will be analyzed using the similar method as hematologic response rate.

Organ PFS

Organ PFS is defined as the time from the date of randomization to the date of first documentation of organ progression, or death due to any cause, whichever occurs first.

Patients without documentation of organ progression will be censored at the date of last organ assessment SD or better.

Organ PFS will be analyzed using the similar method as OS.

## 5.9 Pharmacokinetic, Pharmacodynamic, and Biomarker Analysis

### 5.9.1 Pharmacokinetic Analyses

Plasma concentration-time data will be presented in listings. PK data will be used to perform population PK analysis using a nonlinear mixed effects modeling approach and to assess the effect of various covariates on PK after including data from other studies, if possible. The analysis plan for the population PK analysis will be separately defined and the results of these analyses will be reported separately.

### 5.9.2 Pharmacodynamic Analyses

Not applicable.

### 5.9.3 Biomarker Analysis

Germline DNA Polymorphisms Analysis

Exploratory analysis may examine the association between polymorphisms in the major NFκB pathway-related genes and hematologic response and OS. Characterization of the potential effects of genetic variation in these genes on the activity of MLN9708 in amyloidosis may require analysis of data from this study in combination with the data from other clinical studies of MLN9708.

## 5.10 Resource Utilization and Patient Reported Outcome Analysis

### 5.10.1 Patient Reported Outcomes (PROs)

Patient-reported outcome assessments using FACT/GOG-NTX, the symptom questionnaire, and SF-36, will be analyzed to determine if response to therapy and side effects of therapy are accompanied by measurable changes in the PROs.

The FACT/GOG-NTX (Functional Assessment of Cancer Therapy/Gynecologic Oncology Group-Neurotoxicity [FACT/GOG-Ntx]) comprises 11 individual items evaluating symptoms of neurotoxicity on a scale of 0 (not at all) to 4 (very much).

The symptom scale questionnaire contains 3 items, each rated on an 11-point numerical rating scale of symptom severity. The questionnaire yields individual symptom scores such as swelling, shortness of breath, dizziness and lighted headedness, as well as a total symptom score.

SF-36 v2 is a multi-purpose, short-form health survey with only 36 questions. It yields an 8-scale profile of functional health and well-being scores as well as psychometrically-based physical and mental health summary measures. Physical component summary (PCS) is mostly contributed by physical function (PF), role physical (RP), bodily pain (BP), and general health (GH). Mental component summary (MCS) is mostly contributed by mental health (MH), role emotional (RE), social function (SF), and vitality (VT).

ITT population will be used for patient reported outcomes related analyses.

Missing data pattern will be examined by the proportion of missing responses for each individual scale, as well as summary scores, over time.

Item-level missing data will be handled based on the instrument developer's guideline. SF36 v2 scores will be calculated by software provided by the instrument developer.

For FACT/GOG-NTX, if there are missing items, subscale scores can be prorated. This is done by multiplying the sum of the subscale by the number of items in the subscale, then dividing by the number of items actually answered. When there are missing data, prorating by subscale in this way is acceptable as long as at least 50% of the items were answered. If more than 50% of the items in the subscale are missing, the subscale score will be missing.

Summary statistics of observed values will be presented over time for FACT/GOG-NTX (including 11 individual subscales, sensory scale (sum of NTX1-4), and the summary score), the symptom questionnaire (including 3 individual symptom scores, and the total symptom score) and SF-36 (including PF, RP, BP, GH, MH, RE, SF, VT, PCS, and MCS). Graphical

display will include plot of mean and standard deviation over time for each score by treatment arms.

To compare the 2 treatment arms with respect to maximum improvement from baseline in SF-36 (including PF, RP, BP, GH, MH, RE, SF, VT, PCS, and MCS), as well as symptom questionnaire (including 3 individual symptom scores, and the total symptom score), an analysis of covariance (ANCOVA) model will be fitted. Maximum improvement is defined as the difference between the best value on study and the baseline value. Summaries from ANCOVA with baseline and other relevant clinical or demographic variables as covariates will include F-test p values, mean change scores, and 95% CI for the differences between the treatment arms.

As a sensitivity analysis, a longitudinal model that accounts for repeated measures over time, such as mixed model with repeated measures, will be fitted to compare the SF-36 scores change, as well as the symptom scale questionnaire score change from baseline between the 2 treatment arms. Different imputation methods for missing data including Last Observation Carry Forward (LOCF), random slope model, and pattern mixture model may be evaluated if appropriate after examining missing data patterns.

For PCS, MCS, FACT/GOG-NTX, and individual and total symptom score, a graphical display with maximum improvement from baseline on the x-axis, and the percent of patients experiencing that change on the y-axis will be presented by treatment arms. These plots will be informative to examine the cumulative distribution function (CDF) of responses between treatment arms to characterize the treatment effects. Similar figures will also be presented for change from baseline at each planned data collection time point if there are sufficient data available.

In addition, patient reported outcome assessments may be analyzed to determine if response to therapy and side effects of therapy (such as AE, change from baseline of vital signs and key laboratory parameters) are accompanied by measurable changes in the PROs.

### 5.10.2 Health Economics (Health Care Resource Use)

EQ-5D scores will be summarized in descriptive statistics for treatment arms.

HU data will be summarized in descriptive statistics of medical encounters (length of stay, inpatient, outpatient, and reason), number of missing days from work or other activities by patient and care-giver for treatment arms.

Further modeling will be performed separately at post hoc analyses

## 5.11 Safety Analyses

Safety evaluations will be based on the incidence, intensity, type of AEs, clinically significant changes in the patient's physical examination findings, ECGs, vital sign measurements, and clinical laboratory results.

These analyses will be performed using the safety population. All analyses will be performed for each treatment arm.

### 5.11.1 Adverse Events

#### 5.11.1.1 Adverse Events

Adverse events will be coded using MedDRA version 14.0 or higher. All AEs will be presented in a by-patient listing. Treatment-emergent AEs are AEs that occur after administration of the first dose of any treatment drug and through 30 days after the last dose of any treatment drug.

AEs will be tabulated according to the MedDRA by system organ class, high level terms and preferred terms and will include the following categories:

- Treatment-emergent AEs
- Drug-related treatment-emergent AEs
- Grade 3 or higher treatment-emergent AEs
- Grade 3 or higher drug-related treatment-emergent AEs
- The most commonly reported treatment-emergent AEs (ie, those events reported by ≥ 10% of patients in either treatment group)

Patients with the same AE more than once will have that event counted only once within each body system, once within each high level term, and once within each preferred term.

Drug-related treatment-emergent AEs will also be summarized by the National Cancer Institute Common Toxicity Criteria (NCI CTCAE) version 4.03. Patients with the same AE more than once will have the maximum intensity of that event counted within each body system, once within each high level term, and once within each preferred term.

The most commonly reported treatment-emergent AEs (ie, those events reported by ≥10% of any treatment arm) will be tabulated by preferred term. Patients with the same AE more than once will have that event counted only once within each preferred term.

An overall summary AE table will include numbers and percentages of patients who had any AE, drug-related AE, grade 3 or higher AE, grade 3 or higher drug-related AE, serious AE (SAE), drug-related SAE, AE resulting in discontinuation, and on-study deaths. On-study death is defined as the death that occurs between the first dose of any treatment drug and 30 days of the last dose of any treatment drug.

#### 5.11.1.2 Serious Adverse Events

The number and percentage of patients experiencing at least one treatment-emergent SAE will be summarized by MedDRA primary system organ class, high level term, and preferred term. Drug-related SAE will be summarized similarly.

In addition, a by-patient listing of the SAEs will be presented (the patient listing will contain all SAEs regardless of treatment-emergent AE status).

#### 5.11.1.3 Deaths

A by-patient listing of the deaths will be presented. All deaths occurring on-study and during follow-up will be displayed (regardless of treatment-emergent AE status).

#### 5.11.1.4 Adverse Events Resulting in Discontinuation of Study Drug

A by-patient listing of treatment-emergent AEs resulting in discontinuation of study drug will be presented.

### 5.11.2 Laboratory Data

For the purposes of summarization in both the tables and listings, all laboratory values will be converted to standardized units. If a lab value is reported using a non-numeric qualifier (eg, less than (<) a certain value, or greater than (>) a certain value), the given numeric value will be used in the summary statistics, ignoring the non-numeric qualifier.

Laboratory test results from the central laboratory will be used when they are available. Laboratory test results from local laboratory will only be used when no central laboratory test results exist at the same scheduled sample collection time point.

If a patient has repeated laboratory values for a given time point, the value from the last evaluation will be used.

Laboratory test results will be summarized according to the scheduled sample collection time point. Change from baseline will also be presented. Unscheduled laboratory test results will be listed and included in laboratory shift tables. The parameters to be analyzed are as follows:

- Hematology: hemoglobin, hematocrit, ANC, ALC, monocytes, eosinophils, basophils, platelets, and white blood cell (WBC) count

- Serum chemistry: blood urea nitrogen, creatinine, total bilirubin, uric acid, LDH, albumin, alkaline phosphatase, AST, ALT, glucose, calcium, sodium, potassium, chloride, $CO_2$, magnesium, phosphate, and PT.

Shift tables will be constructed for laboratory parameters to tabulate changes in NCI CTCAE for toxicity (version 4.03) from baseline to post baseline worst CTC grade. Parameters to be tabulated will include:

- Hematology: ALC, ANC, hemoglobin, platelets, WBC
- Serum chemistry: ALT, AST, alkaline phosphatase, creatinine, total bilirubin, calcium, $CO_2$, magnesium, potassium, sodium, and phosphate.

Summary statistics will also be presented for shift from baseline urinalysis values.

Box plots over time through Cycle 12 for key lab parameters will be produced, including but not limited to ANC, platelets, and liver function tests (ALT/SGPT, AST/SGOT, alkaline phosphatase, and total bilirubin).

By-patient listings to be presented include hematology, serum chemistry, urinalysis, urine total protein, and urine creatinine.

### 5.11.3 Electrocardiograms

Descriptive statistics for the actual values and changes from baseline in ECGs will be tabulated by time point including any unscheduled measurements.

QTc interval will be calculated using Bazett's correction and Fridericia's correction. The formulas are:

$$QTc\ (Bazett) = QT / (RR^{0.5})$$

$$QTc\ (Fridericia) = QT / (RR^{0.33})$$

where RR = 60 / heart rate (bpm)

In addition, a categorical analysis of QTc intervals will be performed for each time point. The number and percentage of patients in each QTc interval (< 450 msec, 450-480 msec, 481-500 msec, and > 500msec) will be summarized at baseline and each of the subsequent time points. Categories of changes from baseline (≥ 30 msec and ≥ 60 msec) will be summarized as well.

Maximum QTc intervals and maximum changes from baseline will also be summarized similarly in a separate display.

ECG abnormalities will be presented in a data listing.

### 5.11.4 Vital Signs

The actual values of vital sign parameters including oral temperature, pulse rate, systolic and diastolic blood pressure, and weight, will be summarized over time for each treatment arm. Change from baseline will also be presented.

A by-patient listing will also be presented.

### 5.11.5 Eastern Cooperative Oncology Group (ECOG) Performance Status

Eastern Cooperative Oncology Group performance status and change from baseline will be summarized. Shifts from baseline to the worst postbaseline score will be tabulated by treatment arm.

### 5.11.6 Other Safety Assessments

Pregnancy testing results will be presented in a by-patient listing.

Additional safety analyses may be performed to most clearly enumerate rates of toxicities and to further define the safety profile of MLN9708.

## 6. CHANGES TO PLANNED ANALYSES FROM PROTOCOL

Reference materials for this statistical plan include Clinical Study Protocol C16011 (Protocol Amendment 4 dated 09 July 2015).

## 7. PROGRAMMING CONSIDERATIONS

### 7.1 Statistical Software

SAS version 9.1 (or higher) will be used for all analyses.

### 7.2 Rules and Definitions

Patient populations are defined in Section 2.

Baseline values are defined in Section 5.4.2.

# 8. APPENDIX

## 8.1 Appendix 1: Amyloid-Related Hematologic and Organ Criteria

Amyloid-Related Hematologic and Organ Criteria for Involvement, Stabilization, Response, and Progression

| Organ System | Involvement | Stable[a] | Response | Progression |
|---|---|---|---|---|
| **Hematologic** | Measurable disease defined as serum differential free light chain concentration (dFLC) > 50mg/dL<br><br>dFLC is the difference between the amyloid forming (involved) and nonamyloid forming (uninvolved) free light chain.<br><br>A nephelometric assay measures both kappa and lambda light chains and identifies an excess production of FLC isotype produced from the clonal plasma cell. This FLC isotype is involved in the pathogenic process of misfolding into the amyloid fibrils. | No CR, VGPR, PR, no progression | **Complete Response (CR):**<br>Negative serum & urine immunofixation, normal kappa:lambda FLC ratio,<br><br>**Very Good Partial Response (VGPR)**:<br>dFLC <40 mg/dL<br><br>**Partial Response (PR)**:<br>dFCL decrease ≥ 50% | Progression from CR: any detectable monoclonal protein or abnormal FLC ratio; involved free light chain must double.<br><br>Progression from VGPR, PR or stable disease:<br>Involved free light chain increase of 50% to > 10 mg/dL (100 mg/L) from its lowest measured level<br><br>50% increase in serum M-protein to > 0.5 g/dL from nadir/baseline, or 50% increase in urine M-protein to > 200 mg/24 hours with a visible peak present; |

| | | | | |
|---|---|---|---|---|
| **Involvement of Heart or Kidney (at least one) required for study entry** | | | | |
| **Cardiac** | Echo: mean interventricular septal wall thickness >12mm, no other cardiac cause, or<br>NT-proBNP > 332 ng/L in the absence of renal failure<br><br>Note: a baseline NT-proBNP ≥650 ng/L was required for NT-proBNP response to be evaluable | Stable disease is defined when none of the criteria for response or for worsening disease are met. | NT-proBNP >30% and >300 ng/L decrease if baseline NT-proBNP ≥650 ng/L)<br><br>Echo: mean interventricular septal wall thickness decrease by 2mm, or<br>20% improvement in LVEF,<br>Improvement by ≥ 2 NYHA classes without an increase in diuretic use and no increase wall thickness, or | Increase in NT-proBNP that is both >30% and >300 ng/L, or<br><br>Echo: interventricular septal thickness increased by 2mm over baseline, or<br>An increase in NYHA class by 1 grade with decreasing LVEF of ≥10%,<br><br>Note: Worsening wall thickness and LVEF while harbinger of poor outcome, not useful clinical endpoint |
| **Renal** | 24- hour protein > 0.5 g/day, predominantly albumin | Stable disease is defined when none of the criteria for response or for worsening disease are met. | At least a 50% reduction in 24-hour urine protein (must be at least 0.5 g/day) without worsening of creatinine or creatinine clearance by 25% over baseline | 50% increase in urinary protein loss (at least 1 g/24 hours) or 25% worsening of creatinine or creatinine clearance |
| **Involvement of any of the below allowed in addition to the requirement for involvement of one or more of the above organs** | | | | |
| **Liver** | Alkaline phosphatase value > 1.5 × ULN, or liver span > 15 cm radiographically in the absence of heart failure,. | Stable disease is defined when none of the criteria for response or for worsening disease are met. | ≥50% decrease in alkalaline phosphatase from baseline<br>Decrease in liver size by at least 2 cm (radioagraphic determination | ≥50% increase of alkaline phosphatase above lowest level. |
| **Gastrointestinal** | Direct biopsy verification with symptoms such as diarrhea, frank bleeding, early satiety, malabsorption, GI motility disturbances, and weight loss (BJH 2004) | Stable disease is defined when none of the criteria for response or for worsening disease are met.<br><br>Clinical changes based on NCI CTC Version 4.02 criteria may be useful | Reliable, quantitative methods for defining response do not to exist.<br><br>Improvement in clinical findings such as diarrhea, motility disturbances and weight loss may be useful[b] | Progression of signs and symptoms not attributable to therapy under study |
| **Nerve, Peripheral** | Evidence of amyloid involvement alternative site plus clinical symptoms such as symmetric lower extremity sensory peripheral neuropathy (often with pain) on neurologic examination, parasthesiae, numbness or muscle weakness. Motor neuropathy is rare. (BJH 2004)<br><br>Note: EMG and nerve conduction velocity are relatively insensitive in detecting involvement. | Stable disease is defined when none of the criteria for response or for worsening disease are met.<br><br>Clinical changes based on NCI CTC Version 4.02 criteria may be useful. | Reliable, quantitative methods for defining response do not to exist..<br><br>Improvement in clinical findings with decrease in the Ntx score, signs and/or symptoms of peripheral neuropathy, and decrease in neuropathic pain may be useful[b] | Progressive neuropathy not attributable to therapy under study.<br><br>Progressive neuropathy by EMG or nerve conduction velocity may show progression. |

| | | | | |
|---|---|---|---|---|
| **Nerve, Autonomic** | Based on clinical history of autonomic dysfunction and symptoms such orthostasis, gastric emptying disorder, early satiety, impotence/ erectile dysfunction, bowel or bladder dysfunction not related to direct organ infiltrate, anhidrosis or gustatory sweating (BJH 2004) | Stable disease is defined when none of the criteria for response or for worsening disease are met.<br><br>Clinical changes based on NCI CTC Version 4.02 criteria may be useful | Reliable, quantitative methods for defining response do not to exist.<br>Improvement in clinical examination findings including such symptoms as improvement in orthostasis and other signs and/or symptoms related to autonomic dysfunction[b] | Progression neuropathy not attributable to therapy under study |
| **Soft tissue and lymphatic** | Based on classic physical findings such as:<br>• Macroglossia (enlarged tongue),<br>• Arthropathy,<br>• Claudication, presumed vascular amyloid<br>• Shoulder pad sign,<br>• Periorbital purpura (Raccoon eyes),<br>• Carpal tunnel syndrome,<br>• Synovial enlargement,<br>• Lymphadenopathy, biopsy verification.<br>• Skin thickening<br>• Myopathy by biopsy or pseudohypertrophy | Stable disease is defined when none of the criteria for response or for worsening disease are met.<br><br>Clinical changes based on NCI CTC Version 4.02 criteria may be useful | Reliable, quantitative methods for defining response do not to exist.<br><br>Clinical changes based on NCI CTC Version 4.02 criteria may be useful | Progression of signs and symptoms not attributable to therapy under study |
| **Lung** | Direct biopsy verification with symptoms. Diffuse interstitial radiographic pattern usually by CT scan | Stable disease is defined when none of the criteria for response or for worsening disease are met.<br><br>Clinical changes based on NCI CTC Version 4.02 criteria may be useful | Reliable, quantitative methods for defining response do not to exist.<br><br>Radiographic improvement | Progression of radiographic findings not attributable to therapy under study |

a Unless clinically indicated to check sooner, stabilization of organ function must be confirmed visit in the absence of worsening of any other organs unless worsening is considered a treatment-emergent adverse event.

b Based on NCI CTC Version 4.02 criteria.

## 8.2 Appendix 2: Proof of Strong Control of Type I Error Rate for Key Secondary Endpoints

*Proof of strong control of Type I error rate for both key secondary endpoints (OS and complete hematologic response rate):*

With the proposed testing procedure of the two key secondary endpoints (OS and complete hematologic response rate), this is to prove the strong control of overall Type I error rate for both endpoints at one-sided 0.025 level.

To facilitate the probability presentation, we introduce the following notations. Let P, S indicate the first and second key secondary, OS and complete hematologic response rate (CR). Denote the family of null hypotheses of interest as: $H_0^P : \theta^P = 0$ (OS is not efficacious); $H_0^S : \theta^S = 0$ (CR is not efficacious). Let $T_1^P$, $T_2^P$ be the log-rank test statistic for OS at the second interim and FA, and $T_1^S, T_2^S$ be the square root of the CMH chi-square test statistics for CR at the second interim and FA. Denote $z_\alpha$ as the upper $\alpha$ quantile from the standard normal distribution.

We use the fundamental multiple testing technique, partitioning principle to prove the overall Type I error rate control for both endpoints.

Table 1: Partition hypotheses for testing both key secondary endpoints

| Index | Partition Hypothesis |
|---|---|
| 1 | $\theta^P = 0, \theta^S = 0$ |
| 2 | $\theta^P = 0, \theta^S > 0$ |
| 3 | $\theta^P > 0, \theta^S = 0$ |

By partition principle, as long as each of the partition hypothesis is tested at level 0.025, the overall Type I error rate is also controlled at the same level.

1. Under partition hypothesis $\theta^P = 0, \theta^S = 0$, a false rejection is {reject P or reject S}. Due to the closed sequential testing between P and S at both second IA and FA, rejecting S implies rejecting P, therefore probability of false rejection is simply the probability of rejecting P, i.e.

$$P(T_1^P > z_{\alpha_1} \text{ or } T_2^P > z_{\alpha_2})$$

where $\alpha_1$ and $\alpha_2$ are calculated based on the actual information fraction at second IA from O'Brien-Fleming boundary (Lan-DeMets method). Base on this boundary, the probability is no greater than 0.025.

2. Under partition hypothesis $\theta^P = 0, \theta^S > 0$, the probability of false rejection: Pr{reject P} is again controlled at 0.025 by O'Brien Flemming boundary.
3. Under partition hypothesis $\theta^P > 0, \theta^S = 0$, a false rejection is {reject S}={(reject P at IA and reject S at IA) or (failed to reject P at IA and reject P at FA and reject S at

FA)} which is a subset of {reject S at IA or reject S at FA}. Therefore, the probability of false rejection is bounded by:

$$P(T_1^S > z_{\alpha_1} \text{ or } T_2^S > c) = P(T_1^S > z_{\alpha_1}) + \mathrm{P}(T_1^S \le z_{\alpha_1} \text{ and } T_2^S > c) \qquad (1)$$

The critical value c will be calculated as follows

$$\mathrm{P}(T_1^S \le z_{\alpha_1} \text{ and } T_2^S > c) = 0.025 - \alpha_1$$

to maintain the probability equation (1) at 0.025 level using following asymptotic multivariate normal distribution:

$$\begin{pmatrix} T_1^S \\ T_2^S \end{pmatrix} \sim MVN\left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \sqrt{\frac{n_1}{n_2}} \\ \sqrt{\frac{n_1}{n_2}} & 1 \end{pmatrix} \right)$$

where $n_1$ and $n_2$ are the size of the patient population at the second IA and FA, respectively.

## 8.3 Appendix 3: Amendment 2 Detailed Summary of Changes

THE PRIMARY SECTION(S) OF THE SAP AFFECTED BY THE CHANGES IN AMENDMENT 2 ARE INDICATED. THE CORRESPONDING TEXT HAS BEEN REVISED THROUGHOUT THE SAP.

**Purpose:** Enroll only proteasome inhibitor naïve patients.

The primary change occurs in Section 1.1, Study Design:

Formerly read: Eligible patients must have: 1) biopsy-proven AL-amyloidosis with relapsed or refractory disease despite 1 or 2 prior therapies; 2) disease requiring further treatment; 3) measureable disease as defined by serum differential free light chain concentration (dFLC); and 4) objective and measurable vital organ involvement (ie, cardiac or renal) as defined by the standard International Society of Amyloidosis (ISA) criteria. ~~Patients may be proteasome inhibitor-exposed or naïve, but cannot be refractory to proteasome inhibitor therapy.~~ The definition of relapsed is documented hematologic progressive disease (PD) after a response to prior therapy [PD more than 60 days of last dose]. The definition of refractory is documented absence of hematologic response or hematologic progression on or within 60 days of last dose of prior therapy.

Now reads: Eligible patients must have: 1) biopsy-proven AL-amyloidosis with relapsed or refractory disease despite 1 or 2 prior therapies; 2) disease requiring further treatment; 3) measureable disease as defined by serum differential free light chain concentration (dFLC); and 4) objective and measurable vital organ involvement (ie, cardiac or renal) as defined by the standard International Society of Amyloidosis (ISA) criteria. **Patients enrolled based on protocol amendment 4 must not have been previously treated with proteasome inhibitors. (The Sponsor reserves the right to open the trial to proteasome inhibitor-exposed patients in the future, at some time point after the first interim analysis [IA].)** The definition of relapsed is documented hematologic progressive disease (PD) after a response to prior therapy [PD more than 60 days of last dose]. The definition of refractory is documented absence of hematologic response or hematologic progression on or within 60 days of last dose of prior therapy.

**Purpose:** Revise statistical decision rules to split alpha between the first and second families and remove stopping rule.

The primary change occurs in Section 3.2, Statistical Decision Rules:

Formerly read: ~~Closed sequential testing procedure will be used to strongly control type I error rate at 2-sided 0.05 for the primary endpoints and key secondary endpoints. The hypothesis for hematologic response will be tested at the first~~

~~interim analysis (IA). If the unstratified Cochran-Mantel-Haenszel (CMH) test p value is < 0.05, the null hypothesis will be rejected, and study will continue to full enrollment. Otherwise the study will be deemed unsuccessful and terminated.~~

~~Hypothesis for 2-year vital organ deterioration and mortality rate will be tested at the second IA. If the unstratified CMH test p value is >0.05, no further analysis will be conducted and the study will be terminated. If the unstratified CMH test p value is <0.05, the null hypothesis will be rejected, and the key secondary endpoint OS will be tested at the second IA with approximately 120 death events and again at the final analysis (FA) with approximately 145 events if necessary. O'Brien-Fleming boundary will be calculated using the Lan-DeMets method. The information fraction for second IA is equal to the number of death events observed divided by 145. If there are exactly 120 death events at second IA, null hypothesis for OS will be rejected if the observed p value of the stratified log-rank test is less than 0.027. Otherwise OS hypothesis will be tested again at FA. If the observed p value of the stratified log-rank test is less than 0.0423 at FA (corresponding to nominal alpha of 0.023); the null hypothesis for OS will be rejected.~~

Hypothesis for the second key secondary endpoint, complete hematologic response rate, will be tested sequentially when OS hypothesis is rejected either at second IA or at FA. The testing level of significance will be the same as those used for OS. For instance if OS is significant at 0.023 level at the FA, complete hematologic response rate will also be tested at 2-sided alpha level of 0.023 at the FA.

---

Now reads:

**Hematological response will be tested using unstratified Cochran-Mantel-Haenszel (CMH) test at 0.04 significance level at the first IA; if p value is ≤ 0.04, then at the second IA, 2-year vital organ deterioration and mortality rate will be tested at 0.05 level; otherwise it will be tested at 0.01 level.**

**Hypothesis for 2-year vital organ deterioration and mortality rate will be tested at the second IA. If the test is not significant, the study may still continue, however no formal testing will be conducted. If the test is significant, the null hypothesis will be rejected, and the key secondary endpoint OS will be tested at the second IA. . Using O'Brien-Fleming boundaries, if the alpha allocated to OS is 0.05, the trial may be stopped for overwhelming efficacy if the observed p value is less than 0.027 at the second IA assuming there are exactly 120 death events. The final**

**analysis (FA) will be tested at 2-sided alpha level of 0.0423 (corresponding to nominal alpha of 0.023). If the alpha allocated to OS is 0.01, OS will be tested using a similar approach as above.**

Hypothesis for the second key secondary endpoint, complete hematologic response rate, will be tested sequentially when OS hypothesis is rejected either at second IA or at FA. The testing level of significance will be the same as those used for OS. For instance if OS is significant at 0.023 level at the FA, complete hematologic response rate will also be tested at 2-sided alpha level of 0.023 at the FA.

**The type I error for all primary endpoints and key secondary endpoints using the testing procedure presented in Figure 1 is strongly controlled. Those endpoints are grouped into two ordered families 1) hematologic response, and 2) organ deterioration and mortality, OS and CR. The closed sequential testing approach will be used for strong type I error rate control in the second family. That is, OS will be only tested using the same alpha level assigned to 2-year organ deterioration and mortality rate if the test on 2-year organ deterioration and mortality rate is statistically significant, and then the alpha assigned to OS is allocated using O'Brien-Fleming boundaries for OS tests at second IA and FA; and CR will be only tested at the same alpha level of OS at second IA or FA, if the test on OS is significant at second IA or FA. For testing the two families, the fall-back approach (Wiens B.L., 2003) is used to control type I error rate.**

Sections that also contain this change are:

- Section 4.1, Interim Analysis
- Section 5.1, Sample Size Justification
- Section 5.8.1.1, Primary Efficacy Analysis

**Purpose:** Add schema to describe the alpha controlled testing procedure.

The primary change occurs in Section 3.2, Statistical Decision Rules:

| | |
|---|---|
| Added Figure: | Figure 1. Statistical Testing Procedure and Alpha Spending Schema |

**Purpose:** Further emphasize the type I error control between family 1 and family 2.

The primary change occurs in Section 3.2, Statistical Decision Rules:

Formerly read:

Since the testing of key secondary endpoints is gated by the significance of the primary endpoint, and the type I error rate for ~~all key secondary endpoints~~ is controlled at 2-sided alpha level, the overall type I error rate for all endpoints, i.e. both primary and all key secondary endpoints, is also controlled strongly at 2-sided 0.05 level.

Now reads:

Since **in** the **family 2, the** testing of key secondary endpoints is gated by the significance of the **second** primary endpoint, and **between family 1 and family 2**, the type I error rate is controlled at 2-sided alpha level **by using the fall-back approach**, the overall type I error rate for all endpoints, i.e. both primary and all key secondary endpoints, is also controlled strongly at 2-sided 0.05 level.

Sections that also contain this change are:

- Section 4.1, Interim Analysis
- Section 5.1, Sample Size Justification
- Section 5.8.1.1, Primary Efficacy Analysis

**Purpose:** Update the sample size, alpha spending, and projected time at interim analysis.

The primary change occurs in Section 4.1, Interim Analysis:

Formerly read:

There are two planned formal IAs. The first IA will be performed when approximately ~~94~~ patients have been enrolled and have had the opportunity to complete 6 cycles of treatment or have discontinued study treatment before receiving 6 cycles of treatment. At that time, relevant safety and efficacy data from these patients will be queried and cleaned; all analyses at this IA will be based on these patients whose data have been cleaned. ~~The study will continue only if the test for hematologic response rate is statistically significant at 2-sided alpha level of 0.05. Otherwise, the study will be terminated.~~ This IA is expected to occur approximately ~~27~~ months after the first patient is enrolled. During this IA, enrollment to the study will continue.

~~If the study continues,~~ the second IA will be performed ~~on 2-year vital organ deterioration and mortality rate~~ when approximately 218 patients enrolled have had the opportunity to complete 2 years of treatment or ~~have discontinued~~ treatment before receiving 2 years of treatment. At that time, relevant safety and efficacy data from these patients will be queried and cleaned; all analyses at this IA will be based on these patients whose data

have been cleaned. ~~This IA is expected to occur approximately 70 months after the first patient is enrolled~~.

If the test for 2-year vital organ deterioration and mortality rate is significant at the second IA, the analyses on OS will be performed and the test significance level on OS for the second IA and FA will be determined using O'Brien-Fleming boundaries. Assuming there are exactly 120 death events at second IA, the trial will be stopped for overwhelming efficacy if the observed p value is less than 0.027. ~~Also at the second IA, the hematologic response rate data will be considered as matured and a non-inferential analysis for hematologic response rate will be performed.~~

---

Now reads:

There are two planned formal IAs. The first IA will be performed when approximately **176** patients have been enrolled and have had the opportunity to complete 6 cycles of treatment or have discontinued study treatment before receiving 6 cycles of treatment. At that time, relevant safety and efficacy data from these patients will be queried and cleaned; all analyses at this IA will be based on these patients whose data have been cleaned. **This will be the FA for hematologic response. At the first IA, hematologic response will be tested using unstratified CMH at a 2-sided alpha level of 0.04. If the test is statistically significant, the study will continue to test organ deterioration and mortality rate** at a 2-sided alpha of 0.05 **at the second IA; otherwise organ deterioration and mortality rate will be tested at the second IA with a 2-sided alpha of 0.01.** This IA is expected to occur approximately **54** months after the first patient is enrolled. During this IA, enrollment to the study will continue.

The second IA will be performed when approximately 218 patients enrolled have had the opportunity to complete 2 years of treatment or **followed up for at least 2 years if discontinuing** treatment before receiving 2 years of treatment. **This IA is expected to occur approximately 98 months from the first patient is enrolled.** At that time, relevant safety and efficacy data from **those** patients will be queried and cleaned; all analyses at this IA will be based on these patients whose data have been cleaned.

If the test for 2-year vital organ deterioration and mortality rate is significant at the second IA, the analyses on OS will be performed and the test significance level on OS for the second IA and FA will be determined using O'Brien-Fleming boundaries. Assuming there are exactly 120 death events at second IA, the trial will be stopped for overwhelming efficacy if the observed p value is less than 0.027. **The FA will be tested at 2-sided alpha level of 0.0423 (corresponding to nominal alpha of 0.023). If the alpha allocated**

**to OS is 0.01, OS will be tested using a similar approach as above using O'Brien-Fleming boundaries.**

Sections that also contain this change are:

- Section 3.2 Statistical Decision Rules
- Section 5.1 Sample Size Justification
- Section 5.8.1.1, Primary Efficacy Analysis

**Purpose:** Update the sample size calculation based on statistical design

The primary change occurs in Section 5.1, Sample Size Justification:

Formerly read:

The total sample size was calculated to provide 80% power for the assessment of OS. The study is also adequately powered to test both primary endpoints: hematologic response rate and 2-year vital organ deterioration and mortality rate. There are two planned IAs and one FA.

The parameters used for sample size calculation on OS are a 2-sided test at the significance level of $\alpha = 0.05$, power of 80%, a control arm median OS of 26 months, and testing arm median OS of 41.6 months (assuming exponential distribution, hazard ratio 0.625). With 1 IA (second) and 1 FA occurring at information time of 0.89 and 1, an average enrollment rate of 3 patient /month for the first 5 months, and 5 patient/month thereafter, an additional 28 months follow up for all patients after last patient is enrolled, and approximately 10% drop out rate, a total of 248 patients will need to be randomized in a 1:1 ratio into 2 treatment arms in order to achieve 145 death events. The O'Brien-Fleming stopping boundary (Lan-DeMets method) will be used to assign alpha level to the second IA and FA.

The parameters used for hematologic response endpoint are a 2-sided test at the significance level of $\alpha =$ ~~0.05~~, power of ~~85%~~, a null hypothesis hematologic response rate ~~45%~~, and an alternative hypothesis hematologic response rate ~~75%~~. Approximately ~~94~~ patients are needed for two-sample test of the difference of proportions. Testing for hematologic response hypothesis will be performed when approximately ~~94~~ patients enrolled have had opportunity to either complete 6 cycles of treatment or discontinue treatment before receiving 6 cycles of treatment. This is the first IA for the study, also the FA for hematologic response for statistical testing purpose, with the opportunity to claim hematologic response rate benefit. If the test is statistically significant in favor of the dexamethasone plus MLN9708 arm, ~~the study will continue; otherwise, the study will be terminated~~.

The second IA will be performed when approximately 218 patients enrolled have had the opportunity to complete 2 years of treatment or discontinue treatment before receiving 2 years of treatment. This will be the FA for 2-year vital organ deterioration and mortality rate. With 218 patients, the endpoint of 2-year vital organ deterioration and mortality rate is powered at 90% at a 2-sided alpha level of 0.05 with the assumption of 80% deterioration and mortality rate in the control arm and 60% rate in the MLN9708 arm. If the test for 2-year vital organ deterioration and mortality rate is not significant, ~~the study will be terminated~~; otherwise, OS will be tested. If the test for OS is statistically significant, the complete hematologic response rate will be tested at the same alpha level as that for OS and the study may be stopped for evidence of efficacy; if the test for OS is not statistically significant, the study will continue to the FA.

---

Now reads:

The total sample size was calculated to provide 80% power for the assessment of OS **with the allocated alpha 0.05**. The study is also adequately powered to test both primary endpoints: hematologic response rate and 2-year vital organ deterioration and mortality rate. There are two planned IAs and one FA.

The parameters used for sample size calculation on OS are a 2-sided test at the significance level of $\alpha = 0.05$, power of 80%, a control arm median OS of 26 months, and testing arm median OS of 41.6 months (assuming exponential distribution, hazard ratio 0.625). With 1 IA (second) and 1 FA occurring at information time of 0.89 and 1, an average enrollment rate of 3 patient /month for the first 5 months, and **3-**5 patient/month thereafter, an additional 28 months follow up for all patients after last patient is enrolled, and approximately 10% drop out rate, a total of 248 patients will need to be randomized in a 1:1 ratio into 2 treatment arms in order to achieve 145 death events. The O'Brien-Fleming stopping boundary (Lan-DeMets method) will be used to assign alpha level to the second IA and FA.

The parameters used for hematologic response endpoint are a 2-sided test at the significance level of $\alpha =$ **0.04**, power of **90%**, a null hypothesis hematologic response rate **40%**, and an alternative hypothesis hematologic response rate **65%**. Approximately **176** patients are needed for two-sample test of the difference of proportions. Testing for hematologic response hypothesis will be performed when approximately **176** patients enrolled have had opportunity to either complete 6 cycles of treatment or discontinue treatment before receiving 6 cycles of treatment. This is the first IA for the study, also the FA for hematologic response for statistical testing purpose, with the opportunity to claim hematologic response rate benefit. If the test is

statistically significant in favor of the dexamethasone plus MLN9708 arm, **the second IA for organ deterioration and mortality will be tested at 2-sided α = 0.05 based on fallback approach, otherwise it'll be tested at 2-sided α = 0.01.**

The second IA will be performed when approximately 218 patients enrolled have had the opportunity to complete 2 years of treatment or discontinue treatment before receiving 2 years of treatment. This will be the FA for 2-year vital organ deterioration and mortality rate. With 218 patients, the endpoint of 2-year vital organ deterioration and mortality rate is powered at 90% at a 2-sided alpha level of 0.05 with the assumption of 80% deterioration and mortality rate in the control arm and 60% rate in the MLN9708 arm. **If first IA result is not significant, power of 2-year vital organ deterioration and mortality rate will be 74% at a 2-sided alpha level of 0.01.** If the test for 2-year vital organ deterioration and mortality rate is not significant, **there will not be any formal testing**; otherwise, OS will be tested. If the test for OS is statistically significant, the complete hematologic response rate will be tested at the same alpha level as that for OS and the study may be stopped for evidence of efficacy; if the test for OS is not statistically significant, the study will continue to the FA.

Sections that also contain this change are:

- Section 3.2, Statistical Decision Rules
- Section 4.1, Interim Analysis
- Section 5.8.1.1, Primary Efficacy Analysis

**Purpose:** Update the decision rules as 2-year vital organ deterioration and mortality rate is the second primary endpoint and OS is the first key secondary endpoint.

The primary change occurs in Section 5.8.1.1, Primary Efficacy Analysis:

Formerly read:

The unstratified CMH test will be used to compare hematologic response rate between the 2 treatment arms when ~~94~~ ITT patients have had the opportunity to receive 6 cycles of treatment, or discontinued treatment prior to 6 cycles. If the hematologic response rate is higher in the dexamethasone plus MLN9708 arm (arm A), and the 2-sided CMH test p value is ~~< 0.05~~, null hypothesis will be rejected. Therefore statistical significance will be claimed for dexamethasone plus MLN9708.

…

The unstratified CMH test will be used to make comparisons between the 2

treatment arms **at the second IA** when approximately 218 patients enrolled have had the opportunity to complete 2 years of treatment or have discontinued treatment before receiving 2 years of treatment. If the 2-year organ deterioration and mortality rate is lower in the dexamethasone plus MLN9708 arm (arm A), and the 2-sided CMH test p value is ~~< 0.05~~, null hypothesis will be rejected. Therefore statistical significance will be claimed for dexamethasone plus MLN9708.

---

Now reads:

The unstratified CMH test will be used to compare hematologic response rate between the 2 treatment arms when **176** ITT patients have had the opportunity to receive 6 cycles of treatment, or discontinued treatment prior to 6 cycles. If the hematologic response rate is higher in the dexamethasone plus MLN9708 arm (arm A), and the 2-sided CMH test p value is **≤ 0.04**, null hypothesis will be rejected. Therefore statistical significance will be claimed for dexamethasone plus MLN9708.

...

The unstratified CMH test will be used to make comparisons between the 2 treatment arms when approximately 218 patients enrolled have had the opportunity to complete 2 years of treatment or have discontinued treatment before receiving 2 years of treatment. If the 2-yaer organ deterioration and mortality rate is lower in the dexamethasone plus MLN9708 arm (arm A), and the 2-sided CMH test p value is **≤ 0.05 when hematologic response endpoint is significant or ≤ 0.01 when hematologic response endpoint is not significant,** null hypothesis will be rejected. Therefore statistical significance will be claimed for dexamethasone plus MLN9708.

Sections that also contain this change are:

- Section 3.2, Statistical Decision Rules
- Section 4.1,Interim Analysis
- Section 5.1,Sample Size Justification

---

**Purpose:** Reduce redundant information.

The primary change occurs in Section 2.3, Safety Population:

Deleted text:

The safety population is defined as all patients who receive at least 1 dose of any treatment drug. ~~Patients will be analyzed according to the treatment actually received. That is, those patients who are randomized to Arm B but~~

---

~~received the regimen in Arm A will be included in Arm A; those patients who are randomized to Arm A but received the regimen in Arm B will be included in Arm B for safety analyses.~~

---

**Purpose**: Update CR criteria based on C16011 Protocol amendment 4.

The primary change occurs in Section 8.1 Appendix 1

Deleted text:
**Complete Response (CR):**
Negative serum & urine immunofixation, normal kappa:lambda FLC ratio, ~~<5% plasma cells in bone marrow without clonal dominance~~

Section 5.6.3, Baseline Disease Status, also contain this change.

---

**Purpose**: Update CR criteria based on C16011 Protocol amendment 4.

The primary change occurs in 5.6.3 Baseline Disease Status:

Deleted text:
Baseline disease status will be summarized by the treatment arms in the ITT population, including time since initial diagnosis (months), type and number of organ involvement, type and quantity of involved free light chain, dFLC, serum κ/λ ratio, serum and urine m-protein quantity, serum and urine immunofixation results, ~~% plasma cell in bone marrow core biopsy and/or aspirate,~~ skeletal survey results, cardiac biomarker stage, Eastern Cooperative Oncology Group (ECOG) performance status, ß2-microglobulin, serum creatinine, serum creatinine clearance, total urine creatinine, urine creatinine clearance, serum albumin, alkaline phosphatase, ALT, AST, serum cardiac markers (NT-proBNP, BNP, troponin T), NYHA classification, and other parameters as appropriate. Separate by-patient listings will also be presented.

Section 8.1, Appendix 1, also contain this change.

---

**Purpose:** Correct typographical errors, punctuation, grammar, and formatting

These changes are not listed individually.

---

## 8.4 Rationale for Amendment 1

The main purpose of this amendment is to incorporate the statistically-relevant changes in Protocol C16011, amendment 1, dated 01 June 2012.

Purposes for amendment 1 are to:

- Change the second primary endpoint from overall survival (OS) to 2-year vital organ (that is, heart or kidney) deterioration and mortality rate
- Include OS as a key secondary endpoint

- Remove liver from the list of vital organs with amyloid involvement required at study entry
- Revise definitions of relapsed and refractory disease to clarify 60-day period
- Replace reference to Medical Resource Utilization (MRU) with Health Utilization (HU) to reflect the Sponsor's updated terminology
- Replace reference to Independent Review Committee with Adjudication Committee to accurately reflect how hematologic and organ response and vital organ deterioration will be assessed
- Add hypothesis for 2-year vital organ deterioration and mortality rate; remove hypothesis for organ response and stabilization rate
- Update the decision rules as 2-year vital organ deterioration and mortality rate is the second primary endpoint and OS as the first key secondary endpoint
- Provide distribution of strong control of type I error control for key secondary endpoints and critical value calculation for complete hematologic response rate once OS is statistically significant at the final analysis
- Update sample size calculation on OS due to different information fraction between second interim analysis and final analysis, and different assumption on enrollment rate
- Provide sample size calculation on 2-year vital organ deterioration and mortality rate; Add one non-inferential analysis for hematologic response rate at the second interim analysis
- Clarify the definition of missing data and statistical methods to deal with missing data in primary analysis
- Provide the formula to calculate creatinine clearance
- Provide the formula to calculate months since diagnosis
- Add the populations for analysis at interim analyses
- Provide primary efficacy analysis for 2-year vital organ deterioration and mortality rate
- Add 2 additional sensitivity analyses for primary endpoints
- Update sensitivity analyses for key secondary efficacy analysis
- Add time to vital organ deterioration or death, vital organ response rate, organ response rate, and organ PFS as additional regular secondary efficacy endpoints and analysis
- Add Appendix 2 for the proof of strong control of Type I error rate for both secondary endpoints
- Correct typographical errors, punctuation, grammar, and formatting

## ELECTRONIC SIGNATURES

| Signed by | Meaning of Signature | Server Date (dd-MMM-yyyy HH:mm) |
|---|---|---|
| [redacted] | Clinical Approval | 15-Jul-2015 17:37 |
| [redacted] | Biostatistics Approval | 15-Jul-2015 17:46 |
| [redacted] | Statistical Approval | 15-Jul-2015 17:51 |
| [redacted] | Clinical Science Approval | 15-Jul-2015 19:30 |
| [redacted] | Biostatistics Approval | 15-Jul-2015 20:41 |