Janssen Research & Development

Statistical Analysis Plan

Phase 3 Study Comparing Daratumumab, Lenalidomide, and Dexamethasone (DRd) vs Lenalidomide and Dexamethasone (Rd) in Subjects with Previously Untreated Multiple Myeloma who are Ineligible for High Dose Therapy

Protocol 54767414MMY3008 Amendment INT-4 Phase 3

JNJ-54767414 (daratumumab)

Status: Approved
Date: 3 October 2018
Prepared by: Janssen Research & Development
Document No.: EDMS-ERI-120250925, 3.0

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

TABLE OF CONTENTS

LIST OF IN-TEXT TABLES

AMENDMENT HISTORY

Unique Identifier for SAP Version	Date of SAP Version	Author	Changes from the Previous Version
1.0	24 February 2016	PPD	The initial version for the first interim analysis
2.0	3 October 2018	PPD PPD	1. Updated per the latest protocol 2. Added the efficacy part

ABBREVIATIONS

AE	adverse event
ALT	alanine aminotransferase
ASCT	Autologous Stem Cell Transplant
AST	aspartate aminotransferase
BSA	body surface area
CI	confidence interval
CR	complete response
CrCL	creatinine clearance
CRF	case report form
CSR	clinical study report
CTCAE	Common Terminology Criteria for Adverse Events
DOR	duration of response
DPS	data presentation specification
DRd	Daratumumab, lenalidomide, dexamethasone
ECG	electrocardiogram
ECOG	Eastern Cooperative Oncology Group
FISH	fluorescence in situ hybridization
FLC	free light chain
IA	interim analysis
ICF	informed consent form
IDMC	Independent Data Monitoring Committee
IMWG	International Myeloma Working Group
IRR	infusion-related reaction
ISS	International Staging System
ITT	intent-to-treat
IWRS	interactive web response system
MedDRA	Medical Dictionary for Regulatory Activities
M-protein	monoclonal protein, monoclonal paraprotein
MRD	minimal residual disease
NCI-CTCAE	National Cancer Institute Common Terminology Criteria for Adverse
ORR	overall response rate
OS	overall survival
PD	progressive disease
PBMC	peripheral blood mononuclear cell
PFS	progression-free survival
PP	per-protocol
PR	partial response
Rd	lenalidomide, dexamethasone
SAE	serious adverse event
SAP	statistical analysis plan
SC	subcutaneous
sCR	stringent complete response
SD	stable disease
SD	standard deviation
SOC	system organ class
SPEP	serum protein electrophoresis
TEAEs	treatment-emergent adverse events
TTR	time to response
TTP	time to disease progression
UPEP	urine protein electrophoresis
VGPR	very good partial response
WBC	white blood cells

1. INTRODUCTION

This statistical analysis plan (SAP) contains definitions of analysis sets, derived variables, and statistical methods for the planned analyses as specified in the protocol JNJ-54767414MMY3008, Amendment INT-4.

1.1. Overview of Trial Design

This is a randomized, open-label, active controlled, parallel-group, multicenter study in subjects at least 18 years of age with previously untreated multiple myeloma who are not candidates for high dose chemotherapy and ASCT. The primary objective is to evaluate if daratumumab in combination with Rd prolongs progression-free survival (PFS) compared with Rd alone. The secondary objectives are to compare the 2 treatment groups with respect to time to disease progression (TTP), complete response (CR) or better rate, MRD negativity rate, progression-free survival on next line of therapy (PFS2), overall survival (OS), time to next treatment, overall response rate (ORR), very good partial response (VGPR) or better rate, sCR rate, time to response, duration of response as well as to assess the safety and tolerability of daratumumab when administered in combination with Rd.

Approximately 730 subjects will be randomized in this study with 365 subjects planned per treatment arm (DRd or Rd). Randomization will be stratified by International Staging System (I vs II vs III), region (North America vs Other), and age (<75 vs ≥75). Within each stratum, subjects will be randomized using an equal allocation ratio of 1:1.

Subject participation will include a Screening Phase, a Treatment Phase, and a Follow-up Phase. The Screening Phase will be up to 21 days before Cycle 1 Day 1. The Treatment Phase will extend from Day 1 of Cycle 1 until discontinuation of all study treatment. For subjects assigned to DRd, daratumumab will be administered weekly for the first 8 weeks (Cycles 1-2) of treatment and then every other week for 16 weeks (Cycles 3-6), then every 4 weeks (from Cycle 7 and beyond) until disease progression or unacceptable toxicity. This will equate to 9 consecutive weeks of dosing at the start of the study and a total of 23 doses in the first year. Lenalidomide will be administered at a 25 mg or 10 mg orally (PO) on Days 1 through 21 of each 28-day cycle, and dexamethasone will be administered at a dose of 40 mg or 20 mg once a week for both treatment arms (per age, for more detail refer to protocol Section 6. Dosage and Administration). Those in the Rd arm will receive lenalidomide will be administered at a dose of 25 mg or 10 mg orally (PO) on Days 1 through 21 of each 28-day cycle, and dexamethasone will be administered at a dose of 40 mg or 20 mg once a week until disease progression or unacceptable toxicity. Subjects in both treatment arms will continue lenalidomide and dexamethasone until disease progression or unacceptable toxicity. In the DRd arm, subjects will continue on daratumumab until disease progression or unacceptable toxicity.

The Follow-up Phase will begin once a subject discontinues all study treatments. Subjects who discontinue for reasons other than disease progression must continue to have disease evaluations according to the Time and Events Schedule which will continue until confirmed progression disease (PD), death, lost to follow up, consent withdrawal, subsequent anticancer therapy, or

study end, whichever occurs first. After the clinical cut-off (see below for definition), data collection will be reduced as per protocol Section 9.1.4.

Two interim analyses are planned (for detail refer to Section 1.4).

The primary PFS analysis will occur when approximately 390 PFS events have been observed. The date established for the primary PFS analysis will serve as the clinical cut-off date, after which data collection in the study will be reduced. Investigators will be informed when each interim analysis is to occur. All available data prior to that time will be included in each of the respective analyses.

The end of the study will occur when 330 subjects have died, or 7 years after the last subject is randomized, whichever comes first. The sponsor will ensure that subjects benefiting from treatment with daratumumab will be able to continue treatment after the end of the study.

An Independent Data Monitoring Committee (IDMC) will be commissioned for this study to review efficacy and safety results at the planned interim analyses (more detail information including the timing refer to section 1.4). After the interim reviews, they will make recommendations regarding the continuation of the study. In addition, the IDMC may also review cumulative safety data every 6 months besides the 2 interim analyses.

Assessment of tumor response and disease progression will be conducted in accordance with the International Myeloma Working Group (IMWG) response criteria. An assessment of MRD will be conducted on bone marrow samples. Safety evaluations will include adverse event monitoring, physical examinations, electrocardiogram (ECG) monitoring, clinical laboratory parameters (hematology and chemistry), vital sign measurements, and Eastern Cooperative Oncology Group (ECOG) performance status. Blood samples will be drawn for assessment of pharmacokinetic parameters.

1.2. Statistical Hypotheses for Trial Objectives

The primary efficacy endpoint of this study is PFS. The null hypothesis is that there is no difference in PFS between daratumumab in combination with Rd and Rd alone in subjects with newly diagnosed multiple myeloma who are ineligible for high dose chemotherapy and autologous stem cell transplant.

The secondary endpoints such as ORR, VGPR or better rate, CR or better rate, MRD negativity rate, OS, TTP, PFS2, sCR rate, time to next treatment, time to response and duration of response will be evaluated as well.

1.3. Sample Size Determination

The sample size calculation is performed on the basis of the following assumption. Based on the published data, the median PFS for Rd arm is assumed to be approximately 24 months. Assuming the addition of daratumumab can reduce the risk of the disease progression or death by 25%, i.e., assuming the hazard ratio (DRd vs. Rd) of 0.75, a total of 390 PFS events is needed to achieve a power of 80% to detect this hazard ratio with a log-rank test (two-sided alpha is

0.05). With a 21-month accrual period and an additional 24-month follow-up, the total sample size needed for the study is approximately 730 (365/arm) subjects. The sample size calculation has taken into consideration an annual dropout rate of 5%.

Long-term survival follow-up will continue until 330 deaths have been observed or 7 years after the last subject is randomized, whichever is first. Therefore, this study will achieve approximately 80% power to detect a 27% reduction in the risk of death (hazard ratio = 0.73) with a log-rank test (two-sided alpha = 0.05) if 330 death events are observed at the study end.

1.4. Interim Analyses

Two interim analyses are planned. The first interim analysis, with a purpose to evaluate safety, has been performed after a total of approximately 100 subjects have been treated for at least 8 weeks or discontinued the study treatment. The second interim analysis will be performed when approximately 234 PFS events, which is 60% of the total planned events (390 PFS events), have been accumulated to evaluate the cumulative interim safety and efficacy of daratumumab in combination with Rd. The significance level at this interim analysis to establish the superiority of DRd over Rd with regard to PFS will be determined based on the observed number of PFS events at the interim analysis, using the O'Brien-Fleming boundaries as implemented by the Lan-DeMets alpha spending method. If the experimental arm (DRd) is numerically worse than the control arm in terms of PFS (observed hazard ratio >1 favoring the control arm), then the study may be terminated for futility.

More details about the conduct of these interim analyses can be found in a stand-alone IDMC charter.

2. GENERAL ANALYSIS DEFINITIONS

2.1. Visit Windows

For analyses of data by cycle, if data are collected by date (e.g., AE onset), the corresponding study evaluations will be assigned to actual sequential cycles, which are derived from the study treatment administration data. The start date of a particular cycle is defined as the date of the first scheduled dose of any component of the study treatment, and the end date of a cycle is the start date of the next cycle minus 1. For the last cycle, the end date is defined as the end of treatment visit date or the minimum of last study treatment date plus 30 days and subsequent antimyeloma therapy minus 1 day, if the end of treatment visit date is not available.

In general, if data (e.g., laboratory and vital sign etc.) are collected by cycle, the nominal cycle will be used to summarize data. However, due to possible cycle delays, assessment performed in the same cycle may not be well aligned in time scale for different subjects.

2.2. Pooling Algorithm for Analysis Centers

All participating centers in the study will be pooled together for analyses.

2.3. Study Treatment and Study Drug

In this study, study treatment refers to lenalidomide, dexamethasone, and/or daratumumab. Study drug refers to daratumumab.

2.4. Study Treatment Dosing Date

Study treatment dosing date is the date on which a subject actually receives study treatment (partial or complete) and will be recorded in the study treatment administration dataset.

For subjects who receive DRd treatment, the first study treatment date is defined as the earliest date of non-zero dose of the following administration: lenalidomide, dexamethasone or daratumumab. The last study treatment date is defined as the latest date of non-zero dose of the following administration: lenalidomide, dexamethasone or daratumumab.

For subjects who receive Rd treatment, the first study treatment date is defined as the earliest date of non-zero dose of the following administration: lenalidomide or dexamethasone. The last study treatment date is defined as the latest date of non-zero dose of the following administration: lenalidomide or dexamethasone.

2.5. Baseline Measurement

Baseline measurement is defined as the closest non-missing measurement taken on or prior to the first study treatment administration (including time if time is available, with exception of parameters associated with disease-related efficacy assessment such as SPEP, UPEP, kappa, lambda, kappa/lambda ratio, serum calcium, and albumin).

2.6. Unique Lab Value

In general, in instances when there are multiple records at a given visit date for lab parameters associated with disease assessment, the following rules will be applied to select the unique lab value for analysis: a) multiple records from both central and local lab, central lab value always takes precedence over local lab value; b) multiple records from central lab, select the latest value by time, visit number, or sequence number as the unique lab value; c) multiple records from local lab, select the latest lab value by time, visit number, or sequence number as the unique local lab value.

2.7. Imputation of Partial Dates

Unless specified otherwise, no data imputation will be applied for missing safety and efficacy evaluations. For analysis and reporting purpose, partial dates in adverse event (AE onset date; AE end date), concomitant therapies (start date; end date), MM diagnosis date, and start date of subsequent antimyeloma therapy will be imputed.

2.7.1. Missing/Partial Adverse Event Onset Date

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

- When month and year are present and the day is missing,
  - If the onset month and year are the same as the month and year of first study treatment, the day of first study treatment or the day-component of the AE end date (possibly imputed) is imputed, whichever is earlier
  - If the onset month and year are not the same as the month and year of first study treatment, then 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 first study treatment. If AE end date is available and is prior to first study treatment, the day and month of AE end date are imputed. Otherwise, the day and month of first study treatment are imputed
  - If the onset year is different from the year of first study treatment, the 1st of January is imputed
- If the onset date is completely missing, the earlier one of the date of first study treatment and the AE end date is imputed as the onset date.

No imputation will be done for partial or missing AE onset time.

2.7.2. Missing/Partial Adverse Event End Date

If the end date of an adverse event is missing completely or partially, the following imputation rules will be used.

- If month and year are present and the day of the month is missing, the last day of the month is imputed.
- If only a year is present, the 31st of December is used.
- If the imputed date is later than the date of death (if available), the date of death will be used as the imputed date instead.
- If the year of end date is missing, no imputation will be applied.

No imputation will be done for partial or missing AE end time.

2.7.3. Partial Concomitant Medication Start/End Date

In case of partially missing concomitant medication start/end dates, the following imputation rules will be applied. If the date is completely missing, no imputation will be performed.

- If only the day is missing, the 15th day of the month will be used
- If both the day and month are missing, the 30th of June will be used.

- If the medication was taken prior to study start, and the imputed start date is after first treatment date, further adjust of the imputed start date as the day prior to first dosing date; If the medication was taken after study start, and the imputed start date is prior to first dosing date, further adjust the imputed start date as first dosing date. Also adjust the imputed medication end date so that it is on or after first dosing date.

2.7.4. Partial Multiple Myeloma Diagnosis Date

For partial date of original multiple myeloma diagnosis, the following imputation rules will apply:

- If only the day is missing, set to 15th (as imputed date), and pick minimum of imputed date, date of collection and date of randomization
- If both the day and month are missing, set to January 1, and pick minimum of imputed date, date of collection and date of randomization
- If year is missing, no imputation will be applied.

If the imputed date of original diagnosis is after the randomization date, further adjust the imputed start date as the day before the randomization date.

2.7.5. Partial Subsequent Anticancer Therapy Start Date

If year or month of subsequent anticancer therapy start date is missing or no components of the start date are present, no imputation will be performed.

If only the day is missing, the following steps apply:

- If the month and year of the start date are the same as the month and year of last dosing date, the day of last dosing date or the day-component of the stop date of subsequent anticancer therapy is imputed, whichever is earlier.
- If the start month and year are not the same as the month and year of last dosing date, the first day of the month is imputed.

No imputation will be applied for missing or partial subsequent anticancer therapy end date.

2.8. General Analysis Method

In general, continuous variables will be summarized using descriptive statistics such as mean, standard deviation (SD), median and range. Categorical variables will be summarized using frequency and percentage. For time-to-event variables, which is defined as from randomization to the date of the event, the Kaplan-Meier method will be used for descriptive summaries. For the calculation of time-to-event and duration-of-event variables, the difference between the start date and the end date plus 1 day will be used.

2.9. Analysis Sets

The following analysis sets are defined.

- Intent-to-treat (ITT): defined as subjects who have been randomly assigned to the DRd or Rd group. Analyses of demographics, baseline characteristics and efficacy endpoints will be primarily analyzed based on this population.
- Safety population: defined as subjects who have received at least 1 administration of any study treatment (partial or complete). This population will be used for all safety analyses. The safety analyses grouping will be according to treatment actually received.
- Response-evaluable: defined as subjects who have a confirmed diagnosis of multiple myeloma and measurable disease at baseline or screening visit. In addition, subjects must have received at least one administration of study treatment and have adequate post-baseline disease assessments. Measurable disease is defined as follows:
  - IgG myeloma: Serum monoclonal paraprotein (M-protein) level ≥1.0 g/dL or urine M-protein level ≥200 mg/24 hours; or
  - IgA, IgD, IgE or IgM multiple myeloma: serum M-protein level ≥0.5 g/dL or urine M-protein level ≥200 mg/24 hours; or
  - Light chain multiple myeloma without measurable disease in serum or urine: Serum immunoglobulin free light chain (FLC) level ≥10 mg/dL and abnormal serum immunoglobulin kappa lambda free light chain ratio.

This population will be used as sensitivity analyses for selective response-related secondary endpoints such as CR or better rate.

- Per-protocol (PP) population: defined as subjects who are randomized and meet all eligibility criteria.
- Pharmacokinetics-evaluable: is defined as subjects assigned to DRd group who received at least 1 administration of daratumumab and have at least 1 pharmacokinetic sample concentration value after the first infusion. All pharmacokinetics analyses are based on the pharmacokinetic evaluable population.
- Immune response-evaluable: is defined as subjects assigned to DRd group who have at least 1 immunogenicity sample obtained after their first daratumumab administration.
- Molecular genetic evaluable: defined as subjects who meet one of the following biomarker criteria for risk assessment based on the next generation sequencing (NGS) data:
  - Standard risk: subjects that are negative (=molecular aberration absent) for all del17p, t(14;16), t(4;14).
  - High risk: subjects that are positive (=molecular aberration present) for any of del17p, t(14;16), t(4;14).
- Cytogenetic evaluable: defined as subjects who meet one of the following cytogenetic risk categories:
  - Standard risk: subjects that are negative for del17p, t(14;16), t(4;14) by FISH/Karyotype.
  - High risk: subjects that are positive for any of del17p, t(14;16), t(4;14) by FISH/Karyotype.

2.10. Definition of Subgroups

The following pre-specified subgroup analyses are to be performed at the 2nd IA for the efficacy and safety endpoints (Table 1). Additional subgroup analyses may be performed, if requested and deemed necessary by IDMC for making their decision.

Table 1: Subgroup Analyses of Efficacy and Safety Endpoints

Subgroup	Definition	Analysis Type
Sex	Male, Female	E, S
Age	<75 years, ≥75 years	E, S
Race	White, Others	E, S
Baseline renal function (CrCl)	E: >60 mL/min, ≤60 mL/min S: <30, 30 to <60, 60 to <90, ≥90 mL/min	E, S
Baseline hepatic function	Normal, Impaired [a]	E, S
Region	North America, Others	E, S
International Staging System (ISS)	I, II, III	E
Type of MM	IgG, Non-IgG	E
Cytogenetic risk	High risk [b], Standard risk	E
ECOG performance score	0, 1, ≥2	E

E: efficacy (PFS, ORR, CR or better rate); S: TEAE
[a] Includes mild (total bilirubin ≤ ULN and AST > ULN or ULN < total bilirubin ≤ 1.5×ULN), moderate (1.5×ULN < total bilirubin ≤ 3×ULN), severe (total bilirubin > 3×ULN).
[b]. High risk is defined as positive for any of del17p, t(14;16) or t(4;14) by FISH/Karyotype.

2.11. Other General Definitions

2.11.1. Treatment Emergent Adverse Events

Treatment emergent adverse events (TEAEs) are defined as any AE with onset date and time on or after that of the first dose of study drug infusion through 30 days after the last dose of study drug or the day prior to start of subsequent therapy, whichever is earlier; or any AE that is considered related to (probably, possibly, or very likely) study drug regardless of the start date of the event. AEs with missing or partial onset date and time will be considered as treatment-emergent unless the onset date and time of an AE can be determined as earlier than that of the first dose, or later than 30 days after last dose.

2.11.2. Linking of Treatment Emergent Adverse Events

All event records of the same preferred term from the same subject are to be linked by the onset date and the end date. If an event is followed by another event of the same preferred term with an onset date (or date/time) the same as or 1 day (or 1 minute if applicable) after the end date (or date/time) of the previous record and any features of the adverse event (i.e.: toxicity grades/seriousness/action taken) are different between these two records, these 2 records should be linked together and considered as 1 event, the maximal toxicity grade of the 2 records would be used for the analysis.

3. SUBJECT INFORMATION

3.1. Demographics and Baseline Characteristics

Unless specified otherwise, all demographic and baseline characteristics variables will be summarized for the ITT population. No statistical comparison between the 2 treatment groups is planned.

The distribution of subject enrollment will be presented for each treatment group according to region and country.

Subject demographic and baseline characteristic variables: age (<65 years, 65 to <70 years, 70 to <75 years, and ⩾75 years), sex, ethnicity, race, weight (kg), height (cm), body surface area (BSA, m^2) and ECOG performance status will be summarized by treatment group and overall. A listing of subject demographic and baseline characteristics will be provided as well.

Baseline disease characteristics including type of multiple myeloma (IgG, IgA, IgM, IgD, IgE, light chain only, biclonal, or negative immunofixation), type of measurable disease (Serum only (IgG, IgA, Other - IgD, IgM, IgE and biclonal), Serum and urine, Urine only, Serum FLC only, or not evaluable), ISS staging at baseline (I, II, III), time since initial MM diagnosis (months), number of lytic bone lesions (None, 1-3, 4-10, more than 10), presence of diffuse myeloma-related osteopenia (Yes, No), number of extramedullary plasmacytomas (0, ⩾1), presence of extramedullary plasmacytomas (Yes, No), bone marrow biopsy/aspirate % plasma cells (<10, 10 – 30, >30), bone marrow biopsy % plasma cells (<10, 10 – 30, >30), bone marrow aspirate % plasma cells (<10, 10 – 30, >30), standard-risk and high-risk cytogenetic abnormalities (del17p, t(4;14), t(14;16)), will be summarized and tabulated by treatment group and overall.

A descriptive summary of selected hematology and biochemistry laboratory analytes at baseline will be provided for each treatment group and overall. In addition, baseline toxicity grade of each selected laboratory analyte in hematology and chemistry panel will be summarized by treatment group using frequency based on the safety population.

Medical history collected at baseline or screening visit will be summarized by system-organ class and preferred term for each treatment group and overall.

A summary of stratification factors (ISS staging, region, and age) used in the randomization based on IWRS will be provided to evaluate whether or not randomization process was appropriately executed in the study. The stratified log-rank test and stratified cox model use these stratification factors in randomization.

3.2. Disposition Information

An overview of subject disposition in the study will be provided. The overview includes a summary of total number of subjects who are randomized to each treatment group, the number

and percentage of subjects who are randomized but not treated in each treatment group, and total number of subjects who are treated in each treatment group. For all treated subjects (defined as subjects who have received at least 1 administration of any study treatment), the number and percentage of subjects who discontinued treatment including reason for discontinuation as indicated by the investigators will be summarized. The similar summaries will be presented for all randomized subjects who discontinued from study.

A list of subjects who discontinued study treatment will be provided for safety subjects. This listing will include subject ID, treatment group, date of treatment discontinuation, study day of last dose, reason for discontinuation as well as the specific adverse events (MedDRA preferred term/verbatim term) if discontinuation due to AEs and primary cause of death if discontinuation due to death. A similar list of subjects who discontinued study will be provided for ITT subjects.

3.3. Extent of Exposure

Extent of exposure to study treatments will be summarized and presented based on the safety population.

The number and percentage of subjects treated within each cycle will be summarized by treatment group. The maximum number of treatment cycles received for each subject will be summarized by frequency and descriptive statistics.

Duration of study treatment, defined as the number of months from the date of the first administration of study treatment to the date of the last administration of study treatment, will be summarized.

The number of daratumumab administrations will be summarized for subjects treated with DRd. The total dose administered for daratumumab (mg/kg), lenalidomide (mg), and dexamethasone (mg) will be summarized overall, by cycle of daratumumab treatment.

The dose intensity, which is defined as the sum of total dose administered in all cycles divided by the number of treatment cycles, will be calculated for each study treatment and summarized accordingly. Additionally, the daratumumab dose intensity will be summarized by cycles of daratumumab treatment.

The relative dose intensity (%) defined as the ratio of total actually received dose and total planned dose (planned dose level times the number of administered infusions/medications) will be calculated for each study treatment and summarized by treatment group using descriptive statistics.

The number of subjects with treatment dose delay within the cycle, dose reduced, dose re-escalated per protocol, or dose skipped for lenalidomide and dexamethasone will be summarized for each treatment group. The reasons (AE or other) for treatment dose delay within the cycle, dose reduced, dose re-escalated per protocol, or dose skipped for lenalidomide and dexamethasone will be reported. In addition, a summary of study treatment dose modifications by cycle will be provided.

3.4. Protocol Deviations

Major protocol deviations will be summarized for the ITT population by the following types (not limited) of deviation for each treatment group:

- Developed withdrawal criteria but not withdrawn
- Entered but did not satisfy criteria
- Received a disallowed concomitant treatment
- Received wrong treatment or incorrect dose
- Efficacy assessment deviation
- Safety assessment deviation

A listing of subjects with major protocol deviations including subject ID, type of deviation, and reasons for deviation will be provided.

3.5. Prior, Concomitant and Subsequent Therapies

With the study population of newly diagnosed multiple myeloma subjects who are not candidates for high dose chemotherapy and ASCT, prior systemic use of corticosteroids is limited to a short course (equivalent of dexamethasone 40mg/day for 4 days) for purposes other than multiple myeloma. If any, a listing of all prior systemic use of corticosteroids will be provided.

Concomitant medications collected in the CRF page during the study will be summarized by therapeutic class, pharmacologic class, and drug name for each treatment group. A similar summary will be provided for subjects who received growth factor support, pre-infusion medication and post-infusion medication, respectively. In addition, systemic steroids as concomitant medication use during the study will be summarized. Additionally, prophylactic antiviral medication use will be tabulated.

The total number of subjects who received subsequent antimyeloma therapy will be reported for subjects in safety population in each treatment group. A summary of subsequent antimyeloma therapy will be presented by therapeutic class, pharmacologic class and drug name. In addition, for subjects who received subsequent antimyeloma therapy, their best response to the first subsequent antimyeloma therapy will be summarized.

4. EFFICACY

A validated computerized algorithm, which is based on the IMWG response criteria (Durie 2006, Rajkumar 2011)[1, 2] and has been used and validated by an independent review committee (IRC) in Study MMY2002, will be used to determine response and disease progression for each subject. As a sensitivity analysis, investigator assessment of response and disease progression using the IMWG response criteria will also be performed.

4.1. Analysis Specifications

4.1.1. Level of Significance

All statistical hypothesis tests and 95% confidence interval presented will be 2-sided.

The primary hypothesis is to be tested at the 0.05 significance level (overall). The exact significance level at the second interim analysis is to be determined by the observed number of events per the O'Brien-Fleming alpha spending function. Assuming 234 PFS events are observed at the second interim analysis, the alpha to be spent will be 0.0076 (2-sided) for the interim analysis and 0.0476 (2-sided) for the primary PFS analysis (390 PFS events occur). If the observed two-sided p-value is smaller than this significance level as specified above, the superiority of DRd versus Rd with respect to PFS will be established.

If the primary endpoint of PFS is statistically significant, the following secondary endpoints ordered below will be sequentially tested, each with an overall two-sided alpha of 0.05, by utilizing a hierarchical testing approach as proposed by Tang and Geller (1999) that strongly controls Type I error rate:

1) CR or better rate
2) VGPR or better rate
3) MRD negativity rate (10^{-5})
4) Overall response rate (ORR)
5) Overall survival (OS)

These secondary endpoints will be tested at the 2nd IA and the primary PFS analysis. The primary PFS analysis will be skipped if PFS is positive at the 2nd IA. If this is the case, the protocol will be amended to add an analysis for secondary endpoints at a similar timing as the planned primary PFS analysis (note that OS will also be tested at its protocol-specified final analysis). The significance level at the second interim and the primary PFS analyses will be determined by the alpha-spending function specific to endpoints:

- For ORR, VGPR or better rate, CR or better rate, and MRD negativity rate (10^{-5}), the information fraction is expected to be 80% at the 2nd IA. The O'Brien-Fleming alpha-spending function as implemented by the Lan-DeMets method will be used for alpha spending: 0.0244 (two-sided) at the 2nd IA and 0.0428 (two-sided) at the primary PFS analysis.
- For OS, a modified linear alpha spending function will be used to determine the alpha level at three looks (2nd IA, primary PFS analysis, and final OS analysis). The alpha level will be 0.0001 for the first OS look, and linear spending function will be used to determine the alpha level for the second and last looks. For example, if 59% targeted OS events are observed at the second look, the corresponding alpha level will be 0.0295.

If the null hypothesis for any of these endpoint fails to be rejected at the second interim analysis, then any subsequent endpoint(s) listed above will not be tested until the next analysis time point (e.g., primary PFS analysis), if applicable. If the null hypothesis for an endpoint is rejected at the

second interim analysis, it will remain being rejected and will not be re-tested at any subsequent time points, if any.

4.1.2. Data Handling Rules

There is no imputation planned for missing efficacy endpoint values.

4.2. Primary Efficacy Endpoint

The primary efficacy endpoint is progression-free survival (PFS) based on the computerized algorithm.

4.2.1. Definition

PFS is defined as the duration from the date of randomization to either progressive disease, according to the IMWG response criteria, or death, whichever occurs first. Subjects who start subsequent antimyeloma therapies for multiple myeloma without disease progression will be censored at the last disease assessment before the start of subsequent therapies. Subjects who withdrew consent from the study before disease progression will be censored at the last disease assessment before withdrawal of consent to study. Subjects who are lost to follow-up will be censored at the last disease assessment before subjects are lost to follow-up. Subjects who have not progressed and are still alive at the cutoff date for analysis will be censored at the last disease assessment. Subjects without any post-baseline disease assessment will be censored at the randomization.

Determination of dates of PFS event and dates for censoring is summarized in Table 2 as follows.

Table 2: PFS Event and Censoring Method

Situation	Date of Progression or Censoring	Outcome
Disease progression prior to start of subsequent antimyeloma therapy	Earliest date that indicates disease progression	PFS event
Death without subsequent antimyeloma therapy	Date of death	PFS event
No post-baseline disease assessment	Randomization	Censored
Other (e.g., withdrawal of consent to study participation, lost to follow-up, start of subsequent antimyeloma therapy etc.)	Date of last disease assessment prior to withdrawal of consent to study participation, lost to follow-up, or subsequent antimyeloma treatment	Censored

4.2.2. Analysis Methods

Analysis of PFS will be based on the ITT population. The Kaplan-Meier method will be used to estimate the distribution of overall PFS for each treatment group. The median PFS with 95% CI will be provided. In addition, the number and percentage of subjects who had a PFS event or were censored will be reported. The Kaplan-Meier PFS curve will also be plotted by treatment group.

The primary treatment comparison of the distribution of overall PFS will be based on a stratified log-rank test. The p-value from a stratified log-rank test will be reported. Hazard ratio and its 95% confidence interval will be estimated based on a stratified Cox's regression model with treatment as the sole explanatory variable. Stratification factors used in the analyses include ISS staging (I, II, III), region (North America vs other), and age (<75 years vs ⩾75 years).

In addition, 12, 24 and 36-months of PFS rate with 95% CI will be estimated by Kaplan-Meier method and reported for each treatment group.

Additionally, reasons for PFS event and censoring will be summarized for ITT population.

4.2.3. Sensitivity Analysis of PFS

The following sensitivity analysis is planned to evaluate the robustness of the primary endpoint analysis of PFS.

4.2.3.1. Progressive Disease Based on Investigator Assessment

A sensitivity analysis of PFS, in which progressive disease is based on investigator assessment according to the IMWG response criteria, will be performed in a similar manner as described in the Section 4.2.2.

The PFS definition used in the sensitivity analysis is similar to that defined in the Section 4.1.2, except for date of progressive disease and date of censoring. The date of progressive disease is the date of initial disease progression recorded in the Disease Progression CRF page or earliest date of confirmed progressive disease recorded in the Evaluation of Response CRF page, based on investigator assessment. Similarly, the censoring date is the latest date of disease response recorded in the Evaluation of Response CRF page, based on investigator assessment.

In addition, reasons for PFS and censoring based on investigator assessment will be summarized for ITT population.

4.2.3.2. Not Censored for Start of Subsequent Anticancer Therapies

A sensitivity analysis of PFS derived from the algorithm by not censoring data due to start of subsequent anticancer therapies for subjects, who have not developed a confirmed progressive disease, will be performed in a similar manner as described in Section 4.2.2.

The PFS definition used in the sensitivity analysis is similar to that defined in Section 4.2.1, except for censoring data due to start of subsequent anticancer therapies. Subjects who start subsequent anticancer therapies for multiple myeloma without disease progression will NOT be censored at the last disease assessment before the start of subsequent therapies. If there is no confirmed progressive disease, the subjects will be censored at the last disease assessment before subjects are lost to follow-up or withdrawal of consent to study.

Determination of dates of PFS event and dates for censoring is summarized in Table 3 as follows.

Table 3: PFS Event and Censoring Method

Situation	Date of Progression or Censoring	Outcome
Disease progression	Earliest date that indicates disease progression	PFS event
Death	Date of death	PFS event
No post-baseline disease assessment	Randomization	Censored
Other (e.g., withdrawal of consent to study participation, lost to follow-up, start of subsequent anticancer therapy etc.)	Date of last disease assessment	Censored

4.2.3.3. Censored for Death/PD after Missing More Than One Disease Evaluation

A sensitivity analysis of PFS derived from the algorithm by censoring for death or progression after missing consecutive evaluations will be performed in a similar manner as described in Section 4.2.2.

The PFS definition used in the sensitivity analysis is similar to that defined in Section 4.2.1, except for death or progression after missing more than one disease evaluation. For any PFS (death or progression) event identified by the computer algorithm, if the event date and the latest date of scheduled disease evaluation (includes serum M-protein, urine M-protein, serum FLC and corrected calcium only) immediately preceding the event differs more than 2.5 disease evaluation intervals, which indicates that subject missed at least one scheduled disease evaluation, then this event will not be considered as a PFS event in the sensitivity analysis. Instead, the subject will be censored at the date of last disease evaluation (includes serum M-protein, urine M-protein, serum FLC and corrected calcium only) prior to the PFS event originally identified.

4.2.3.4. Per-protocol Analysis of PFS

A sensitivity analysis of PFS derived from the algorithm based on Per-protocol population will be performed in a similar manner as described in Section 4.2.2.

4.2.3.5. Unstratified Analysis of PFS

A sensitivity analysis of PFS by using unstratified log-rank test and unstratified Cox's regression model will be performed in a similar manner as described in Section 4.2.2.

4.3. Secondary Endpoints

Secondary efficacy endpoints include overall response rate (ORR), VGPR or better rate, CR or better rate, MRD negativity rate, OS, TTP, PFS2, sCR rate, time to subsequent antimyeloma treatment, time to response and duration of response.

4.3.1. CR or Better Rate

4.3.1.1. Definition

CR or better rate is defined as the proportion of subjects with a response of CR or better (i.e., CR or sCR) based on the computerized algorithm, according to IMWG response criteria, during or after the study treatment.

4.3.1.2. Analysis Methods

CR or better rate will be calculated for each treatment group based on the ITT and response-evaluable population. The number and percentage of subjects in the following response categories will be presented by treatment group: stringent complete response (sCR), complete response (CR), sCR+CR, very good partial response (VGPR), VGPR or better (sCR+CR+VGPR), partial response (PR), overall response (sCR+CR+VGPR+PR), stable disease (SD), progressive disease (PD), and not evaluable (NE). The corresponding 95% exact CI will be provided.

Stratified CMH test will be used to test treatment difference in ORR, VGPR or better rate, CR or better rate and sCR rate. The CMH estimate of odds ratio and its 95% confidence interval and p-value for testing treatment difference will be reported. Stratification factors used in the analysis include ISS staging (I, II, III), region (North America vs. Other), and age ($<$75 years vs. $\geqslant$75 years).

A sensitivity analysis, which disease response is based on investigator assessment according to the IMWG response criteria, will be performed in a similar manner as described above.

4.3.2. VGPR or Better Rate

4.3.2.1. Definition

VGPR or better rate is defined as the proportion of subjects with a response of VGPR or better (i.e., VGPR, CR or sCR) based on the computerized algorithm, according to IMWG response criteria, during or after the study treatment.

4.3.2.2. Analysis Methods

Similar statistical methods will be applied as described in Section 4.3.1.2for CR or better rate analysis.

4.3.3. MRD Negativity Rate

For this study, three threshold values, 10^{-4}, 10^{-5} and 10^{-6}, will be used to evaluate MRD negativity status and its predictive value for PFS.

4.3.3.1. Definition

MRD negativity rate is defined as the proportion of subjects who have negative MRD at any time point after the date of randomization and prior to subsequent antimyeloma therapy. MRD

positive subjects include subjects of which all tested results were found to be MRD positive, or indeterminate, or unavailable (calibration failure or missing).

4.3.3.2. Analysis Methods

The MRD negativity rate will be calculated for each treatment group based on the ITT population. The corresponding 95% exact CI will be provided.

For each threshold value, Fisher's exact test will be used to test if the MRD negativity rate is the same between the two treatment groups in the previously untreated myeloma setting.

For the purpose of hierarchical testing (Section 4.1.1), the threshold value of 10^{-5} will be employed using ITT populations.

The durability of MRD negativity may be examined for each treatment group by the proportion of subjects remaining MRD negative 12 months after initial MRD negativity if data becomes available.

4.3.4. Overall Response Rate (ORR)

4.3.4.1. Definition

ORR is defined as the proportion of subjects who achieve a partial response or better (i.e., PR, VGPR, CR or sCR) based on the computerized algorithm, according to IMWG response criteria, during or after the study treatment.

4.3.4.2. Analysis Methods

Similar statistical methods will be applied as described in Section 4.3.1.2 for CR or better rate analysis.

4.3.5. Overall Survival

4.3.5.1. Definition

Overall survival (OS) is measured from the date of randomization to the date of death due to any cause. Subjects who are lost to follow-up will be censored at the time of lost to follow-up. Subjects who are still alive at the clinical cut-off date for the analysis will be censored at the last known alive date. The date of last known alive will be determined by the maximum collection/assessment date from among selected data domains within the clinical database.

4.3.5.2. Analysis Methods

OS, including 12, 24 and 36-months survival rate, will be analyzed for the ITT population. The Kaplan-Meier method will be used to estimate the distribution of OS for each treatment group. Median OS with 95% CI will be provided. In addition, the number and percentage of subjects who had died or were censored will be reported. Additionally, the survival rate with 95% CI at 12, 24 and 36-months will be estimated using Kaplan-Meier method for each treatment group.

Due to the expected small number of death events at the second interim analysis, the distribution of OS for the 2 treatment groups will be compared based on an un-stratified log-rank test. A p-value from an un-stratified log-rank test will be reported. Hazard ratio and its 95% confidence interval will be estimated based on an un-stratified Cox's regression model with treatment as the sole explanatory variable.

In addition, a summary of reasons for censoring of overall survival will be provided.

If the null hypothesis of OS is not rejected at the second interim analysis, for OS at the primary PFS analysis when 360 PFS events observed, the alpha to be spent will be determined by a linear alpha spending function based on the observed number of deaths at that time, i.e., the cumulative alpha to be spent will be the total alpha (0.05) multiplied by the proportion of the observed number of deaths out of the total planned number of deaths (330, expected at time of the final OS). See Section 4.1.1 for more details on alpha spending and type I error control.

4.3.6. Time to Disease Progression (TTP)

4.3.6.1. Definition

TTP is defined as the time between the date of randomization and the date of first documented evidence of confirmed PD, as defined in the IMWG response criteria, or death due to PD, whichever occurs first. Subjects who start subsequent antimyeloma therapies for multiple myeloma without disease progression will be censored at the last disease assessment before the start of subsequent therapies. Subjects who withdraw consent to study or are lost to follow-up or die without disease progression will be censored at the last disease assessment. Subjects who have not progressed at the cutoff date for analysis will be censored at the last disease assessment. Subjects without any post-baseline disease assessment will be censored at the randomization.

Determination of dates of TTP event and dates for censoring is summarized in Table 4 as follows.

Table 4: TTP Event and Censoring Method

Situation	Date of Progression or Censoring	Outcome
Disease progression prior to start of subsequent anticancer therapy	Earliest date that indicates disease progression	TTP event
Death due to disease progression prior to start of subsequent anticancer therapy	Date of death	TTP event
No post-baseline disease assessment	Randomization	Censored
Other (e.g., withdrawal of consent to study participation, lost to follow-up, start of subsequent anticancer therapy etc.)	Date of last disease assessment prior to subsequent anticancer treatment	Censored

4.3.6.2. Analysis Methods

Similar statistical methods will be applied as described in Section 4.2.2for PFS analysis, including descriptive and comparison of the distribution of overall TTP.

4.3.7. Progression-free Survival on Next Line of Therapy (PFS2)

4.3.7.1. Definition

Progression-free survival on next line of therapy (PFS2) is defined as the time from randomization to progression on next line of therapy or death, whichever comes first. Any deaths are considered as PFS2 events. Subjects who start next line of therapy without disease progression on study treatment will be censored at the last disease assessment before starting next line of therapy. For subjects who start next line of therapy after progression on study treatment, are still alive and not yet progress on next line of therapy, they will be censored on the last date of follow-up. Subjects without any post-baseline follow-up will be censored at the randomization.

Determination of dates of PFS2 event and dates for censoring is summarized in Table 5 as follows.

Table 5: PFS2 Event and Censoring Method

Situation	Date of Progression or Censoring	Outcome
No post-baseline disease assessment	Randomization	Censored
Alive and no disease progression on study treatment	Date of last disease assessment prior to start of 1st line on next therapy	Censored
Disease progression on study treatment and progress on the 1st line of next therapy or any death	Minimum of earliest date that indicates progression on the 1st line of next therapy and date of death	PFS2 event
Other	Minimum of start date of 2nd line of next therapy minus 1 and last date of follow-up	Censored

4.3.7.2. Analysis Methods

Similar statistical methods will be applied as described in Section 4.2.2 for PFS analysis.

4.3.8. sCR Rate

4.3.8.1. Definition

sCR rate is defined as the proportion of subjects with a response of sCR based on the computerized algorithm, according to IMWG response criteria, during or after the study treatment.

4.3.8.2. Analysis Methods

Similar statistical methods will be applied as described in Section 4.3.3.2 for CR or better rate analysis.

4.3.9. Time to Subsequent Antimyeloma Treatment

4.3.9.1. Definition

Time to subsequent antimyeloma treatment is defined as the time from randomization to the start of subsequent antimyeloma treatment. Death due to PD without start of subsequent

therapy will be considered as event. Subjects who withdrew consent to study or are lost to follow, or die due to causes other than disease progression will be censored at the date of death or the last date known to be alive.

4.3.9.2. Analysis Methods

The Kaplan-Meier method will be used to estimate the distribution of time to subsequent antimyeloma treatment for the ITT population. Median time to subsequent antimyeloma treatment with 95% CI will be tabulated for each treatment group. In addition, a Kaplan-Meier curve for time to subsequent antimyeloma treatment will be plotted. The hazards ratio and its 95% CI will be obtained through a stratified Cox's regression model with treatment as the sole explanatory variable. Treatment comparison will be made via a stratified log-rank test.

4.3.10. Time to Response

4.3.10.1. Definition

Time to response (i.e., time to first response) is defined as the time between the date of randomization and the first efficacy evaluation that the subject has met all criteria for PR or better based on the computerized algorithm for patients who had PR or better as their best response.

4.3.10.2. Analysis Methods

For subjects who achieve a confirmed response, descriptive statistics (n, mean, standard deviation, median, and range) will also be provided to summarize time to response, time to first response, time to VGPR or better response and time to CR or better response.

4.3.11. Duration of Response

4.3.11.1. Definition

Duration of response (DOR) is defined for subjects with a confirmed response (PR or better) as the time between first documentation of response and disease progression based on the computerized algorithm, according to IMWG response criteria, or death due to PD, whichever occurs first. Responders without disease progression will be censored at the censoring time point for TTP.

4.3.11.2. Analysis Methods

No formal statistical comparison of DOR between the 2 treatment groups is planned. Analysis of DOR will be based on subjects who achieved a confirmed response of PR or better. Median DOR with 95% CI will be estimated based on the Kaplan-Meier method for each treatment group. The Kaplan-Meier duration of response curve will be plotted by treatment group.

4.4. Subgroup Analysis of Efficacy Endpoints

For assessment of internal consistency and investigation of homogeneity of the treatment effect across subgroups, subgroup analyses of the primary and selected secondary efficacy endpoints will be conducted based on pre-specified subgroups defined in Section 2.10.

A forest plot of subgroup analysis on PFS and selected secondary endpoints (e.g. ORR) will be generated, respectively.

Additionally, the primary endpoint of PFS and secondary endpoint of TTP may be explored by responder vs non-responder for each treatment group.

4.5. Functional Status and Well-being

4.5.1. Definition

Functional status and well-being will be assessed using 2 PRO measures, the EORTC QLQ-C30 and the EQ-5D-5L.

The EORTC QLQ-C30 includes 30 items resulting in 5 functional scales (physical functioning, role functioning, emotional functioning, cognitive functioning, and social functioning), 1 Global Health Status scale, 3 symptom scales (fatigue, nausea and vomiting, and pain), and 6 single items (dyspnea, insomnia, appetite loss, constipation, diarrhea, and financial difficulties). The recall period is 1 week (the past week).

The instrument contains 28 items using a verbal rating scale with 4 response options: “Not at All,” “A Little,” “Quite a Bit,” and “Very Much” (scored 1 to 4). Two additional items use response options (1 to 7): 1 = Very Poor, to 7 = Excellent. All scale and item scores will be linearly transformed to be in the range from 0 to 100 according to the algorithm in EORTC QLQ-C30 scoring manual, version 3.0 (Fayers et al, 2001). A higher score represents a higher ("better") level of functioning, or a higher ("worse") level of symptoms.

The EQ-5D-5L is a generic measure of health status. For purposes of this study, the EQ-5D-5L will be used to generate utility scores for use in cost effective analyses. The EQ-5D-5L is a 5-item questionnaire that assesses 5 domains including mobility, self-care, usual activities, pain/discomfort and anxiety/depression plus a visual analog scale rating “health today” with anchors ranging from 0 (worst imaginable health state) to 100 (best imaginable health state). The scores for the 5 separate questions are categorical and cannot be analyzed as cardinal numbers. However, the scores for the 5 dimensions are used to compute a single utility score ranging from zero (0.0) to 1 (1.0) representing the general health status of the individual (but allows for values less than 0 by UK scoring algorithm).

4.5.2. Analysis Methods

Compliance rates for completion of EORTC QLQ-C30 and EQ-5D-5L at each time point will be generated based on number of expected.

Key PRO endpoints

- EORTC QLQ-C30 global health status/quality of life subscale
- EQ-5D-5L utility score
- EQ-5D-5L visual analog scale (VAS)

The change from baseline at each time point will be summarized descriptively by treatment group.

A distribution based method will be used to define worsening/improvement in scores, i.e., half standard deviation away from the mean score at baseline combining both treatment groups. Time to worsening and time to improvement will be derived. Death due to disease progression will be considered as worsening. Subjects who have not met the definition of worsening/improvement will be censored at the last PRO assessment. Subjects without baseline assessment or post-baseline assessment will be censored at date of randomization.

Time to improvement will be descriptively reported. Time to worsening will be estimated using Kaplan-Meier methods. The hazard ratio for DRd relative to Rd and its associated 95% confidence interval (CI) will be calculated based on the stratified Cox proportional hazards model by the stratification factor at randomization.

A mixed effects model with repeated measures analysis will be conducted estimating change from baseline at each time point between two treatments. ITT subjects who have a baseline value and at least one post-baseline value are included in the analysis. Change from baseline will be fitted to a mixed effects model including subjects as a random effect, and baseline value, treatment group, time in week, treatment-by-time interaction, and stratification factors as fixed effects. Line plot of LS mean of change from baseline with standard error over time will be displayed by treatment arm.

Secondary PRO endpoints

These may include other QLQ-C30 scales:

- functional scales: physical, role, cognitive, emotional, and social
- symptom scales: fatigue, pain, and nausea and vomiting
- single-item score: dyspnea, loss of appetite, insomnia, constipation, diarrhea, and financial difficulties

The change from baseline at each time point may be summarized descriptively by treatment group.

Time to worsening/improvement and mixed effect model analysis, as described for the key PRO endpoints, may be performed as appropriate. Line plot of LS mean of change from baseline with standard error over time may be displayed by treatment arm.

5. SAFETY

Safety assessment will be evaluated through AEs and clinical hematology and chemistry laboratory tests. Safety analyses will be based on the safety population and presented by the treatment actually received.

5.1. Adverse Events

All adverse events whether serious or non-serious, will be reported from the time a signed and dated informed consent form (ICF) is obtained until 30 days after the last dose of study treatment, until the subject withdraws consent for study participation, or until the subject starts subsequent antimyeloma therapy, whichever occurs first. AEs will be recorded in standard medical terminology and graded according to the National Cancer Institute Common Terminology Criteria for Adverse Events (NCI-CTCAE), Version 4.03. For AE reporting, the verbatim term used in the CRF by investigators to identify adverse events will be coded using the latest version of Medical Dictionary for Regulatory Activities (MedDRA) coding dictionary.

Unless otherwise specified, at each level (e.g., system organ class and/or preferred term) of subject summarization in reporting the incidence of the AE, a subject is counted once if one or more events were recorded. For summarizing new onset events, all event records of the same preferred term from the same subject are to be linked by the onset date and the end date. If an event is followed by another event of the same preferred term with an onset date (or date/time) the same as or 1 day (or 1 minute if applicable) after the end date (or date/time) of the previous record and any features of the adverse event (i.e.: toxicity grades/seriousness/action taken) are different between these two records, these 2 records should be linked together and considered as one event. A Grade 5 event will be linked to previous event of the same preferred term if the onset date of grade 5 record is the same or one day after the end date of previous record.

All summaries of AEs will be based on treatment-emergent adverse events (TEAEs), which are defined as any AE that occurs after start of the first study treatment through 30 days after the last study treatment; or the day prior to start of subsequent antimyeloma therapy, whichever is earlier; or any AE that is considered drug-related (very likely, probably, or possibly related) regardless of the start date of the event; or any AE that is present at baseline but worsens in toxicity grade or is subsequently considered drug-related by the investigator.

The incidence of TEAEs will be summarized overall, by MedDRA system organ class (SOC) and preferred term, by toxicity grade, and by relationship, to study treatment administration. Specifically, the following AE summaries will be presented by treatment group.

5.1.1. Overview of TEAEs

An overview of TEAEs reported through the study will be provided for each treatment group. The overview will include summaries of subjects with TEAEs, with TEAEs related to study treatment, with TEAEs of maximum toxicity grade of 1 to 5, SAEs, TEAEs leading to discontinuation of all study treatment.

5.1.2. All TEAEs

- Incidence of TEAEs by MedDRA SOC and preferred term.
- Most commonly reported (>10%) TEAE by MedDRA SOC and preferred term.

5.1.3. Toxicity Grade 3 or 4 TEAEs

- Incidence of toxicity Grade 3 or 4 TEAEs, by MedDRA SOC and preferred term
- List of subjects with any toxicity Grade 3 or 4 TEAEs
- Most commonly reported (>5%) Grade 3 or 4 TEAE by MedDRA SOC and preferred term

5.1.4. Study Treatment-Related TEAEs

- Incidence of TEAEs considered by the investigator to be related to study treatment, by MedDRA SOC, preferred term and relationship to study treatment
- Incidence of TEAEs with toxicity Grade 3 or 4 considered by the investigator to be related to study treatment, by MedDRA SOC and preferred term and relationship to study treatment

5.1.5. Serious Adverse Events (SAEs)

- Incidence of treatment-emergent SAEs, by MedDRA SOC and preferred term
- Incidence of treatment-emergent SAEs considered by the investigator to be related to study treatment, by MedDRA SOC, preferred term and relationship to treatment
- Most commonly reported (>2%) SAEs by MedDRA SOC and preferred term
- List of subjects with any treatment-emergent SAEs

5.1.6. TEAEs Leading to Dose Modifications

Incidence of TEAEs leading to dose modifications will be summarized by MedDRA SOC and preferred term. The summaries will be presented by all grades and Grade 3 or 4 for each treatment. This table will include TEAEs leading to cycle delays or dose modification of at least 1 of the study treatments, the dose modifications include dose delay within the cycle, dose reduced (not applicable to daratumumab), dose re-escalated per protocol (not applicable to daratumumab), or dose skipped.

5.1.7. TEAEs Leading to Discontinuation of Any Study Treatment

A summary of number of subjects who discontinued any study treatment because of 1 or more TEAEs by MedDRA system-organ class and preferred term will be provided. The summaries will be presented by all grades and Grade 3 or 4 for each treatment group. The AEs leading to discontinuation of any study treatment are based on AEs recorded in the AE CRF page with an action taken of drug withdrawal for any study treatment.

A listing of subjects who discontinued any study treatment because of TEAEs will be provided

5.1.8. TEAEs Leading to Discontinuation of All Study Treatment

A summary of number of subjects who discontinued all study treatment because of 1 or more TEAEs by MedDRA system-organ class and preferred term will be provided. A listing of subjects who discontinued all study treatment because of TEAEs will be provided, this listing includes TEAEs leading to discontinuation of all study treatment for those subjects indicated as having discontinued study treatment due to an adverse event on the end of treatment CRF page.

5.2. Deaths

5.2.1. Death Due to TEAEs

The number of subjects who died due to treatment-emergent adverse events will be summarized by preferred term and relationship to study treatment for each treatment group. The TEAEs included in this table are AEs with outcome death or toxicity grade of 5 recorded in the AE CRF page.

A listing of subjects who died due to treatment-emergent adverse events will be provided.

5.2.2. Deaths

A summary of all deaths and cause of death will be tabulated overall and by treatment group. Specifically, the number of subjects who died during the study will be summarized for the ITT population. The primary cause of death collected on the death information CRF page will be reported. If the primary cause of death is AE, the number of subjects who have related AE and unrelated AE will be further reported.

Subjects who died within 30 days of last study treatment dose and within 60 days of first study treatment dose, respectively, based on the safety population.

A listing of all deaths will be provided.

5.2.3. Adverse Events by Subgroups

The following subgroup analysis of adverse events will be performed based on subgroups specified in Section 2.10:

- Overview of TEAEs
- All TEAEs
- Toxicity Grade 3 or 4 TEAEs
- SAEs.

5.3. Adverse Events of Clinical Interest

The adverse events of clinical interest would be focus on below items (detail would be defined on the DPS or below sections),

- IRR
- Infections and Infestations
  - Opportunistic infections
  - Virus infections
- Hemorrhage events
- Interferences for blood typing
- Tumor lysis syndromes
- Intravascular hemolysis
- Cytopenia
  - Neutropenia events
  - Thrombocytopenia events
  - Anaemia
  - Lymphopenia
- Second Primary Malignancies

5.3.1. Infusion-Related Reactions (IRR)

Subjects with any IRR associated with daratumumab administration will be summarized by MedDRA system-organ class and preferred term. The summaries will be presented by all grades, Grade 3, 4, and 5. In addition, the total number of subjects with IRR in more than 1 infusion will be reported. Additionally, the timing of IRR associated with daratumumab administration will be evaluated through a summary of IRR by event onset time.

A listing of subjects with Grade 3 or higher treatment-emergent infusion-related reactions associate with daratumumab administration will be provided. In addition, subjects with treatment-emergent infusion-related reactions results in discontinuation of daratumumab will be listed.

5.3.2. Infections and infestations

Infections and infestations refer to adverse events with SOC of infections and infestations. A summary of number of subjects with 1 or more toxicity Grade 3 or 4 treatment-emergent infections and infestations by MedDRA preferred term and relationship to treatment will be provided. Additional summary analyses may include by onset time (i.e., $\leq$ 6 months vs. 6-$\leqslant$12 months vs. 12-$\leq$24 months vs. >24 months) and/or exposure-adjusted incidence rate of infections/infestations.

5.3.2.1. Opportunistic Infections

Opportunistic Infections to adverse events with PT terms which would be defined on the DPS. The summaries will be presented by all grades and Grade 3 or 4 for each treatment group.

5.3.3. Cytopenia

Cytopenia includes Neutropenia events, Thrombocytopenia events, Anaemia and Lymphopenia. The four types of Cytopenia to adverse events with PT terms which would be defined on the DPS. The summaries will be presented by all grades and Grade 3 or 4 for each treatment group.

5.3.4. Hemorrhage

Hemorrhage defined by Standardized MedDRA Queries (SMQ) with the first subcategory SMQ of hemorrhage terms (exclude laboratory terms). The summaries will be presented by all grades and Grade 3 or 4 for each treatment group.

5.3.5. Second Primary Malignancies

A listing of subjects who reported second primary malignancies during the study will be provided. This listing will include diagnosis, study day of diagnosis, recurrence of a prior existing malignancy (yes, no) and pathology diagnosis (biopsy, aspirate etc.) etc. information whenever a second primary malignancy is observed. In addition, cumulative study treatment exposure, the treatment for second primary malignancy and the outcome information will also be presented in the listing.

5.4. Clinical Laboratory Tests

The evaluation of clinical laboratory tests will focus on the following selected laboratory analytes:

Hematology panel:

- hemoglobin
- platelet count
- white blood cell (WBC) count with absolute neutrophils and lymphocytes

Blood chemistry panel:

- alkaline phosphatase
- creatinine
- uric acid
- glucose
- bilirubin
- aspartate aminotransferase (AST)
- alanine aminotransferase (ALT)

- sodium
- potassium
- corrected calcium
- phosphate
- creatinine clearance

Blood samples for serum hematology are taken at the screening visit, on Days 1, 8, 15, and 22 of Cycles 1-2, Days 1 and 15 of Cycles 3-6, Day 1 of Cycles 7+, and at the End-of-Treatment visit. Blood sample for biochemistry are taken at the screening visit, on Day 1 of each treatment cycle and at the End-of-Treatment visit.

Descriptive statistics (mean, standard deviation, median, range) will be used to summarize observed laboratory values and change from baseline in observed value at each scheduled visit for each treatment group. Line plot of mean with standard error for each laboratory analyte over time will be displayed by treatment group for hemoglobin, neutrophils, lymphocytes, platelets, WBC, AST, ALT, creatinine, and creatinine clearance.

The worst toxicity grade in hematology and chemistry during the treatment will be summarized by treatment group and toxicity grade. Shift tables from baseline to worst toxicity grade during the treatment will be provided for each laboratory analyte listed above. These tables will summarize the number of subjects with each baseline CTC grade and changes to the maximum CTC grade.

5.5. Vital Signs and Physical Examination Findings

Vital signs (systolic and diastolic blood pressure, heart rate, and temperature) values at baseline will be summarized.

Post baseline physical examination findings were collected as AEs, and therefore will not be summarized.

5.6. Electrocardiogram (ECG)

Electrocardiograms (ECG) will be performed at Screening, Day 1 of Cycle 3 and Cycle 6 immediately after end of daratumumab infusion, and End-of-Treatment visit.

The number and percentage of subjects with normal or abnormal 12-lead ECG results will be summarized.

5.7. ECOG Performance Score

ECOG performance status, which evaluates the effect of the disease status on the activities of daily living, will be assessed at Screening, Day 1 of Cycle 3, Cycle 6, Cycle 9 and Cycle 12 for the first year, every 6th cycle thereafter until EOT (PD), post-PD collected at Week 8 and 16. Descriptive statistics will be used to summarize ECOG performance status at baseline, Day 1 of each treatment cycle (including change from baseline), worst score during treatment (including

change from baseline), and End-of-Treatment visit (including change from baseline) for each treatment group. Shift table from baseline to worst score during the treatment will be provided.

6. PHARMACOKINETICS/IMMUNOGENICITY/PHARMACODYNAMICS

Unless specified otherwise, descriptive statistics (e.g., number of observations, mean, standard deviation, median, and range) will be used to summarize pharmacokinetics and pharmacodynamics data. In addition, coefficient variation and geometric mean will be provided in the pharmacokinetic concentration summary.

6.1. Pharmacokinetics

6.1.1. Sampling Timepoints

For subjects assigned to DRd, blood samples to assess serum concentration (pharmacokinetics) of daratumumab will be obtained at Day 1 of Cycle 1, 6 and 12, End-of-Treatment and Post-Treatment Week 8. On a daratumumab dosing day, blood samples need to be collected before (up to 2 hours but not after the start of infusion) and after (up to 2 hours but not before the end of infusion) daratumumab administration.

6.1.2. Pharmacokinetic Parameters

The pharmacokinetic parameters are defined as:

- Minimum observed concentration (C_{min}) - the concentration observed immediately before infusion.
- Maximum observed concentration (C_{max}) - the concentration observed after the end of infusion

For daratumumab, the pharmacokinetic evaluations include C_{min} and C_{max}.

6.1.3. Analysis Methods

Pharmacokinetic analyses will be performed on the pharmacokinetic-evaluable population. All serum concentrations below the lowest quantifiable concentration or missing data will be recorded as such in the concentration data set. Concentrations below the lowest quantifiable concentration will be treated as zero in the summary statistics. All subjects and samples excluded from the analysis will be clearly documented.

Descriptive statistics will be used to summarize daratumumab serum concentrations at each sampling time point. A plot of mean (±SD) daratumumab serum peak and trough concentrations over time will be provided.

If sufficient data are available, population pharmacokinetic analysis of serum concentration-time data of daratumumab may be performed using nonlinear mixed-effects modeling. If population pharmacokinetic analysis is conducted, it may include data from other clinical studies; details will be provided in a population pharmacokinetic analysis plan and results will be presented in a separate report.

6.2. Immunogenicity

6.2.1. Sampling Timepoints

Samples to assess the generation of antibodies to daratumumab (immunogenicity) will be obtained from all subjects in the DRd group at Cycle 1 Day 1 predose, End-of Treatment, and Post-Treatment Week 8. In addition, any time an infusion-related reaction is observed during the study, an unscheduled blood sample should be drawn as soon as possible after the reaction for potential immune response analysis.

6.2.2. Analysis Methods

The incidence of anti-daratumumab antibodies (immunogenicity) will be summarized for all subjects who receive a dose of daratumumab and have appropriate samples for detection of antibodies to daratumumab. In addition, subjects who are positive for antibodies to daratumumab will also be listed.

6.3. Pharmacokinetic/Pharmacodynamic Analyses

If sufficient data are available, other pharmacokinetic/pharmacodynamic modeling may be performed, including exploring the relationship between serum concentrations of daratumumab and endpoints of clinical efficacy. If analysis is conducted, details and results of the analysis will be presented in a separate report.

7. BIOMARKER

Biomarker studies are designed to identify markers predictive of response (or resistance) to daratumumab. Planned analyses are based on the availability of clinically valid assays and may be deferred if emerging study data show no likelihood of providing useful scientific information. Results of biomarker analyses may be presented in a separate report.

Blood samples will be drawn from all subjects in both treatment groups to better understand the mechanism of action and mechanism of resistance of daratumumab. Minimal residual disease (MRD) will be assessed from bone marrow aspirates for all subjects who achieve a CR/sCR. Molecular subtyping will be done to evaluate daratumumab response rates in high-risk multiple myeloma subpopulations.

7.1. Minimal Residual Disease (MRD)

Bone marrow aspirates will be collected to monitor MRD in those subjects who attain or suspect to have a CR/sCR. MRD will be monitored using next generation sequencing (NGS) on bone marrow aspirate.

7.1.1. Sampling Timepoints

Samples are requested at time of suspected CR/sCR and at 12, 18, 24 and 30 months post C1D1 (+/-1 month). If one of these time points occurs within 1 month of suspected CR, a repeat bone marrow will not be requested. These bone marrow tests will only be required if patient's response is near CR or better by blood and urine evaluations (for detail timepoint refer to protocol Section 9.2.1.5).

7.1.2. Analysis Methods

Details on MRD negativity rate analyses are described in Section 4.3.4.

In addition, to evaluate the relationship between MRD negativity and clinical endpoints (e.g. PFS) may be explored.

7.2. Molecular Subtyping

7.2.1. Sample collection and characterization

Viable, frozen CD38+ bone marrow aspirate samples were collected at baseline for DNA/RNA sequencing analysis. Peripheral blood mononuclear cell (PBMC) samples were collected at baseline to use as controls.

7.2.2. Molecular Subtypes

A portion of the bone marrow aspirate samples collected as specified in the Section 7.2.1 will be utilized for translocation/mutation/genomic analysis using exome-sequencing and RNA-sequencing to assess whether specific molecular subgroups having chromosomal aberrations such as del17p, t(4;14), t(14;16), are responsive to daratumumab treatment. PBMC samples will be sent for exome sequencing and used as controls for mutation and copy number analysis.

7.2.3. High-risk/Standard-risk Molecular Subgroup Analysis

High-risk and standard-risk molecular subgroups have been defined in Section 2.9.

To determine if daratumumab combined with Rd (DRd) will lead to improved clinical responses in high-risk molecular subgroups, the following exploratory analysis will be conducted by using the similar analysis methods specified in Sections 4.3.1.2 and 4.2.2.

- To evaluate ORR and MRD negativity rate for subjects in high-risk molecular subgroup and subjects with specific molecular subtyping such as del17p, t(14;16), t(4;14)
- To evaluate PFS and OS for subjects in high-risk molecular subgroup and subjects with specific molecular subtyping such as del17p, t(14;16), t(4;14)

If subgroup size allows, these results will also be compared with the subjects treated with DRd in the standard-risk molecular subgroup, to explore whether subjects treated with DRd in high risk molecular subgroup will have similar clinical outcome as subjects in standard risk subgroup.

8. MEDICAL RESOURCE UTILIZATION

Medical resource utilization (excluding study infusion administration) will be descriptively summarized by treatment group. Frequencies of hospitalization, outpatient visits, type of hospitalization or outpatient visit, reasons for hospitalization or outpatient visit, durations of hospitalization or outpatient visit will be calculated and tabulated.

REFERENCES

1. Durie BG, Harousseau JL, Miguel JS, et al. International uniform response criteria for multiple myeloma. Leukemia 2006;20:1467–1473. Corrigenda/Erratum in: Leukemia. 2007; 21:1134-1135.
2. Rajkumar SV, Harousseau J-L, Durie B, et al. Consensus recommendations for the uniform reporting of clinical trials: report of the International Myeloma Workshop Consensus Panel 1. Blood. 2011; 4691-4695.
3. Sonneveld P, Avet-Loiseau H, et al. Treatment of multiple myeloma with high-risk cytogenetics: a consensus of the International Myeloma Working Group. Blood. 2016 Jun 16;127(24):2955-62

ATTACHMENTS

ATTACHMENT 1: Additional Exploratory Analysis to Support HEMAR

1. DEFINITION OF SUBGROUPS

Subgroup analyses will be performed using the criteria listed below to determine whether the treatment effect is consistent among subgroups. Analyses will be conducted for the ITT population and for the following subgroups:

- For subjects who reached CR/sCR as their best response
- For subjects who reached VGPR as their best response
- For subjects who reached PR as their best response
- For subjects who reached PD as their best response
- For subjects who had baseline ECOG of 2
- For subjects who achieved MRD negativity (10^{-4}, 10^{-5} and 10^{-6})

Subgroup analyses will be performed if data warrants.

2. TIME-TO-EVENT ENDPOINTS FOR SUBGROUP ANALYZES

Kaplan-Meier estimates will be used to estimate distribution of time to event by treatment arm based on all ITT population. Data will be calculated and summarized with descriptive statistics. The following time-to-event endpoints will be analyzed by pre-defined subgroups as defined in section 2.10 and in section 1, Appendix:

- PFS
- TTP
- Time to subsequent antimyeloma treatment
- PFS2
- OS
- Time to best response
- DOR

2.1. Subgroup Analysis by Center for PFS, OS and EQ-5D and EORTC QLQ C30

3. EXPOSURE ADJUSTED INCIDENCE RATES (EAIR)

3.1. Restriction on the first event

The analysis restricts on the occurrence of the first event per patient and ignores the existence of later (multiple) events as these cannot be assumed to occur independent of previous events (e.g.: patients suffering from infections may have in general a higher risk of having other complications and may even have a higher risk of getting other infections). The occurrence of multiple events is subject to another analysis considering the absolute number of adverse events per patient.

For these reasons the EAIR should be interpreted as *'rate until the first event occurs'*. Rates estimated from several patients can be averaged on the level of a preferred term (PT), of a system organ class (SOC), or on a global level (see below).

The interpretation of EAIRs is simple and consistent on the preferred-term level only, and can be expressed as "*Average number of TEAEs per preferred-term emerging per person-month of exposure*".

The aforementioned considerations apply in the same way to EAIRs estimated on the global level: when EAIRs are collapsed into the global estimate (first analyses), the estimate can be interpreted as the "Average number of TEAEs emerging per person-month *and PT*", because estimation has been performed on a *'per PT'*-basis (per *average* or *typical* PT among all PTs).

Comparing EAIRs on the level of the SOC or on the global level involves data destruction because a patient's information is reduced to the first TEAE only (and possibly to a TEAE of marginal relevance among many TEAEs with higher clinical relevance).

The EAIR analysis focuses on the 'speed' by which TEAEs emerge. The analysis restricts on the first event of a patient because independence of TEAEs cannot be assumed. The necessity to restrict on the first event entails considerable data destruction when deriving SOC-specific EAIRs or the EAIR on a global level. To overcome this, the *'per PT'*-analysis, which is reported in both Tables identically, is preferable.

Comparing EAIRs between the analyses outlined below on a SOC-specific or a global level demonstrates that the *'per PT'*- method makes the interpretation of results more difficult. However, it can be suggested that this method provides a more robust approach when the two treatment arms are to be compared on a SOC-specific or global level. A t-Test like comparison of PT-specific estimates between the two treatment arms may provide a more robust, comprehensive and easy-to-communicate way of visualizing and comparing results.

3.2. Duration of exposure: censored & non-censored

The incidence rate for a patient is derived from the duration of exposure to treatment of that patient. When averaging incidence rates, a patient's duration of exposure is given either A) by the time when the event has occurred (non-censored data), or B) by the total duration of treatment in case the patient does not show the adverse event in question (censored data). Depending on whether a patient has an adverse event or not, the duration of exposure enters the denominator in its non-censored or censored form, respectively.

3.3. Incidence rate per patient

The incidence rate for a specific event of a patient *i* is the reciprocal of time *t* when the first event occurs:

$$EAIR_i = \frac{1}{t_i}.$$

3.4. Average EAIR

The *EAIR* averaged over all patients is

$$EAIR = \frac{\sum_{i=1}^{n} TEAE_i}{\sum_{i=1}^{n} t_i},$$

whereby

a) a TEAE enters the sum in the nominator unweighted ($TEAE_i$ =1, otherwise $TEAE_i$ =0), and
b) the duration of exposure enters the denominator as described before: $t_i = \begin{cases} time\ of\ TEAE \text{ if occurring (non-censored data)} \\ total\ duration\ of\ treatment \text{ if no event occurs (censored data)} \end{cases}$.

3.5. EAIRs on the level of a SOC and on the global level on a 'per-PT' basis

3.5.1. Average EAIR per PT

The *EAIR* for a specific PT is an average over all patients, i.e.

$$EAIR_{PT} = \frac{\sum_{i=1}^{n} TEAE_{PT,i}}{\sum_{i=1}^{n} t_{PT,i}},$$

whereby the number of TEAEs and durations of exposure enter the nominator and the denominator.

3.5.2. Average EAIR per SOC

The average *EAIR* per SOC considers the first event of each patient within the SOC. The denominator includes the exposure time of each adverse event of all PTs within the SOC, per patient, i.e.

$$EAIR_{SOC} = \sum_{i=1}^{n} TEAE_{SOC,i} \sum_{PT=1}^{n\ PTs\ per\ SOC} \frac{1}{t_{PT,i}},$$

where $TEAE_{SOC,i}$ is the first event per patient per SOC and $t_{PT,i}$ is the exposure time for a specific preferred term of a given patient.

Note: This *EAIR* is an incidence rate per *average (or typical)* preferred term in that SOC (cf. 3.6.1).

3.5.3. Average EAIR on a global level

The average *EAIR* on a global level only considers the first event per patient across all events. The denominator includes the exposure times of all PTs, i.e.

$$EAIR_{global} = \sum_{i=1}^{n} TEAE_i \sum_{PT=1}^{n\ PTs} \frac{1}{t_{PT,i}},$$

where $TEAE_i$ is the first event of a patient overall and the $t_{PT,i}$'s are PT-specific exposure times of that patient.

Note: This EAIR is an incidence rate *per average (or typical)* preferred term.

3.6. Second analyses

3.6.1. Average EAIR per PT

The *EAIR* for a specific PT is an average over all patients as described before, i.e.

$$EAIR_{PT} = \frac{\sum_{i=1}^{n} TEAE_{PT,i}}{\sum_{i=1}^{n} t_{PT,i}},$$

whereby the number of TEAEs and durations of exposure enter the nominator and the denominator.

3.6.2. Average EAIR per SOC

The average *EAIR* per SOC considers the first event per patient per SOC only, and only one (the corresponding) exposure time in the denominator (confer before, where the denominator in the $EAIR_{SOC}$ depends on the number of PTs per SOC):

$$EAIR_{SOC} = \frac{\sum_{i=1}^{n} TEAE_{SOC,i}}{\sum_{i=1}^{n} t_{SOC,i}},$$

Note: This EAIR is an incidence rate *per SOC*.

3.6.3. Average EAIR on a global level

The average *EAIR* on a global level considers the overall first event per patient only, and only one (the corresponding) exposure time in the denominator (confer before, where the denominator in the $EAIR_{SOC}$ depends on the overall number of PTs):

$$EAIR_{global} = \frac{\sum_{i=1}^{n} TEAE_{i}}{\sum_{i=1}^{n} t_{i}},$$

whereby $TEAE_i$ represents the first TEAE among all TEAEs of patient i and t_i as before (time when TEAE occurs (non-censored data) or total duration of treatment if no event occurs (censored data))

4. ADDITIONAL TIME TO EVENT ANALYSES

In case of different exposure times, time adjustment for AE is necessary. Hazard Ratio and Kaplan-Meier curves will be conducted including number of patients at risk for the following safety endpoints:

- Any TEAE
- Any Serious TEAE
- Any TEAE leading to death
- Any Grade 3 or 4 TEAE
- Any Grade 3 or higher TEAE
- Any TEAE leading to treatment discontinuation

Detailed description by preferred term:

- TEAEs by preferred term with prevalence>=10%/
- Grade 3 or 4 TEAEs preferred term with prevalence>=5%
- Grade 3 or higher TEAEs by preferred term prevalence>=5%
- Serious TEAEs preferred term with prevalence>=2%
- TEAEs leading to treatment discontinuation preferred term with prevalence>=1%
- TEAE leading to death preferred term without prevalence cut-off

ATTACHMENT 2: Multiple Myeloma IMWG Algorithm

PD and Response Algorithm
54767414MMY3008

The issues addressed by the program are:

1. Whether or not the disease progressed and date* of progression
2. The date* of censoring for subjects whose disease never progressed
3. Reason(s) for PD
4. The date* and category of best and first response
5. Duration of response

*Day post-randomization for randomized subjects.

1. Determination of PD and Relapse from CR

1.1. IMWG Criteria

1.1.1. Progressive Disease

PD is to be used for calculation of time to disease progression and progression-free survival for all subjects including those experiencing CR.

PD is defined as satisfying any one of the criteria listed below. These are identical to the IMWG criteria as described in the protocol. Further explanations (noted in superscript) pertaining to particular PD criteria are provided in Section 1.2, Clarifications and Modifications.

a. Increase (1.2.2) of ≥25% in the level of serum M-protein and absolute increase (1.2.2) must be ≥0.5 g/dL (5 g/L) (1.2.3; 1.2.4).

b. Increase (1.2.2, 1.2.18) of ≥25% in the 24-hour urinary light chain excretion (urine M-protein) and absolute increase (1.2.2) must be ≥200 mg/24 hours (1.2.3).

c. Only in subjects without measurable serum and urine M-protein levels: increase (1.2.2) of ≥25% in the difference between involved and uninvolved FLC levels (1.2.20) and absolute increase (1.2.2) must be >10 mg/dL (1.2.3).

d. Only in subjects without measurable serum and urine M-protein levels and without measurable disease by FLC levels: increase (1.2.2) of ≥25% in the level of bone marrow plasma cells percentage and absolute increase (1.2.2) must be ≥10% (1.2.3).

1

e. Definite increase [1.2.8] in the size of existing bone lesions [1.2.9] or soft tissue plasmacytomas [1.2.9; 1.2.10].

f. Definite development of new bone lesions [1.2.11] or soft tissue plasmacytomas [1.2.12, 1.2.13].

g. Development of hypercalcemia [1.2.14; 1.2.15; 1.2.16] (corrected serum calcium >2.875 mmol/L or >11.5 mg/dL) that can be attributed solely to the PC proliferative disorder.

1.1.2. Relapse from CR

Relapse from CR is to be used for calculation of disease-free survival for subjects experiencing CR.

Relapse from CR is defined as satisfying any one of the criteria listed below (a, b, or c). These are identical to the IMWG criteria as described in the protocols. Further explanations (noted in superscript) pertaining to particular relapse from CR criteria are provided in Section 1.2, Clarifications and Modifications.

a. Reappearance of serum or urinary M-protein by immunofixation or electrophoresis [1.2.3; 1.2.7].

b. Development of ≥5% plasma cells in the bone marrow.

c. Appearance of any other sign of progression (i.e., new plasmacytomas, lytic bone lesion, or hypercalcemia) [1.2.11; 1.2.12; 1.2.13; 1.2.14; 1.2.15].

1.2. Clarifications and Modifications

In order to allow these rules to be applied consistently and to be programmed, the Sponsor has added certain clarifications and modifications for using the IMWG criteria.

1.2.1. Measurable disease is defined in the protocol by at least one of the following measurement: (1) serum M-protein ≥ 1g/dL (≥ 10 g/L) or ≥ 0.5 g/dL (≥ 5 g/L) for subjects with IgA, IgD, IgE or IgM multiple myeloma; (2) urine M-protein ≥ 200 mg/24h; (3) serum FLC assay: involved FLC level ≥ 10 mg/dL (≥ 100 mg/L) provided serum FLC ratio is abnormal.

1.2.2. The reference point for calculating increase and % of increase for M-protein, FLC and bone marrow plasma cells will be the lowest response value and the "lowest response value" does not need to be a confirmed value.

1.2.3. Requires 2 consecutive (i.e., no intermediate values that do not meet the definition of PD) assessments made at any time before the institution of any new therapy (i.e., subsequent anti-cancer therapy). If a subject has an unconfirmed PD

2

followed by death due to PD within 30 days of unconfirmed PD, or PD (the same parameter as of initial observed PD) observed within 4 weeks after start of subsequent anti-cancer therapy, the unconfirmed PD will be considered as PD.

1.2.4. If nadir serum M-protein is ≥ 50g/L (≥5 g/dL), M-protein increases of ≥ 10g/L (1g/dL) is sufficient for progressive disease. It does not require meeting "increase of ≥ 25% in the level of serum M-protein".

1.2.5. Any 2 consecutive increase of serum M-protein ≥5 g/L (≥0.5 g/dL) is consistent with progressive disease, assuming that increase of ≥25% is met or not applicable, even if the serum M-protein level is below measurable disease threshold.

1.2.6. The baseline value for assessing disease progression is the last pretreatment measurement (it applies to SPEP, UPEP, FLC, plasmacytomas and bone lesions except as noticed in Section 1.2.10). For example, if there is a screening value and a Cycle 1 Day 1 value (prior to start of treatment) for M-protein, the program ignores the screening value and uses Cycle 1 Day 1 as the reference point for assessing PD for the first post-treatment results.

1.2.7. The program disregards serum M-protein and urine paraprotein measurements prior to Day 5 after dosing. These tests are considered too soon after dosing to be regarded as legitimate on-treatment values.

1.2.8. The program computes the date of progression as the earliest date of any of the tests listed in Section 1.1.1 (a, b, e, f or g for subjects with measurable serum or/and urine M-protein; or a, b, c, e, f or g for subjects without measurable serum and urine M-protein) that indicate PD. Similarity, the program computes the date of relapse from CR as the earliest date of any of the tests listed in Section 1.1.2 (a, b, or c) that indicate relapse from CR.

1.2.9. For PD due to bone lesions, the algorithm will rely on information collected on the eCRF regarding skeletal survey (i.e., increase in the size of lytic bone lesions or increase in the total number of lytic bone lesions) and other radiology reports. At any time, study sites may report progressive disease based on an increase in the size or number of lytic bone lesions. The algorithm accepts this determination as definitive.

1.2.10. Plasmacytomas/bone lesions that are first reported during the first week on study are treated as baseline plasmacytomas/bone lesions. The rationale for this rule is that plasmacytomas/bone lesions take time to develop, so anything reported within 1 week of first dose almost certainly existed before the start of study drug.

3

In the case that no pre-treatment value for plasmacytomas/bone lesion is available, the first post-treatment value is used as baseline.

1.2.11. For plasmacytomas, the Sponsor has defined "definite increase in size" as an increase of over 50% in the sum of the products of the two longest perpendicular diameters when available, using the smallest previous product as the reference point.

1.2.12. New post-baseline bone lesions are evidence of PD. If no baseline bone lesions are available, then any subsequent data that report a bone lesion will be considered as development of new bone lesions.

1.2.13. New post-baseline plasmacytomas are evidence of PD, even if the measurements are not available. If no baseline plasmacytoma data are available, then any subsequent data that report a plasmacytoma will be considered as a "new" plasmacytoma and will be considered as evidence of PD.

1.2.14. Hypercalcemia will be based on corrected calcium values as long as they are available (i.e., corrected calcium values are collected or serum calcium and albumin are collected). If corrected calcium values are not available, free-ionized calcium values will be used if they are available. Hypercalcemia based on free-ionized calcium will be defined as >1.5 mmol/L.

1.2.15. For subjects who have hypercalcemia at baseline, the program will never assess PD due to "development of hypercalcemia". A subject is considered to have hypercalcemia at baseline if he or she meets one of the following criteria:

a. For subjects with corrected calcium values, any corrected calcium value >ULN which occurs on or before Day 4 (relative to first dose).
Note: if subject had central and local assessments done on the same day, the central record takes precedence.

b. For subjects without corrected calcium values in all visits, but with free-ionized calcium values on or before Day 4, any free-ionized calcium value >ULN which occurs on or before Day 4 (relative to first dose).

1.2.16. For subjects with normal calcium at baseline, if the first PD criteria met is a post-baseline hypercalcemic value, then the following programming algorithm will be applied:

a. The program will first search for evidence of progression in other parameters within the following 56 days. If PD is determined in other parameters within the following 56 days, then the subject is assessed as PD. The date of PD is the date of the initial

4

hypercalcemic value and hypercalcemia is indicated as the reason or one of the reasons for PD.

b. If PD is not determined in other parameters within 56 days, the program will search for the next available corrected calcium/free-ionized calcium value. If the second value also meets the criteria for hypercalcemia, the subject is assessed as PD. The date of PD is the date of the initial hypercalcemic value and hypercalcemia is indicated as the reason or one of the reasons for PD.

c. If there is no more follow-up calcium value, then the subject is assessed as PD.

d. Any PD identified through a), b) or c) will be flagged and reviewed by clinical to ensure that hypercalcemia can be attributed solely to the plasma cell proliferative disorder. Hypercalcemia due to other reasons will not be considered as PD in the algorithm.

1.2.17. Descriptive (or non-numeric) results from bone marrow aspirate or bone marrow biopsy will be given a numeric interpretation based on the following conventions:

a. Percentage plasma cells reported as a range on the CRF page will be assigned the mean value (i.e., lower range plus upper range divide by 2) that will be analyzed by the programming algorithm (e.g., an amount reported as 10-15% would be interpreted as 12.5%).

b. Percent plasma cells reported as <x% will be interpreted as (x-1)% and percent plasma cells reported as >x% will be interpreted as x%.

c. Other conventions include:

Reported as:	Interpreted as:
"Not significantly present"	4%
"Sheets of plasma cells not seen"	4%

1.2.18. Imputation of UPEP and SPEP values: If the serum immunofixation result is "Not Detected" and the SPEP value is missing or not done, then SPEP value is treated as 0. If the urine immunofixation is "Not Detected" and the UPEP is missing or not done, then UPEP value is treated as 0.

1.2.19. a. Spurious UPEP values are not used in the determination. Determination of a spurious value is as follows: If at least 4 values exist in the neighborhood (+/- 94 days) of the value of interest, the mean and standard deviation of the values is computed (not including the value of interest in the calculation). The value of interest is compared to the mean +/- 3 standard deviations. If the value < mean-3*standard deviation or the value > mean+3*standard deviation then the value is considered spurious. The last two values available for a particular subject

5

are never considered as spurious. The baseline value for a subject is also never considered as spurious.

b. If the first PD criteria met is two consecutive UPEP values and subject has started subsequent anticancer therapy, this subject is assessed as PD and date of PD is the date of first UPEP assessment. Otherwise, if the first PD criteria met is two consecutive UPEP values and subject has not started subsequent anticancer therapy, then the following programming algorithm will be applied:

    1) The program will first search for evidence of progression in other parameters within the following 94 days. If PD is determined in other parameters within the following 94 days, then the subject is assessed as PD. The date of PD is the date of the initial UPEP value.
    2) If PD is not determined in other parameters within 94 days and no more UPEP values are available, then the subject is assessed as PD due to UPEP.
    3) Otherwise, the third UPEP value is considered. If the third value meets the criteria for PD, the subject is assessed as PD due to UPEP. If the third value does not meet the criteria for PD, but other parameters indicate PD within the next 94 days, the subject is assessed as PD due to UPEP at the time of the initial UPEP value.
    4) Otherwise, the fourth UPEP value is considered. If the fourth value meets the criteria for PD, the subject is assessed as PD due to UPEP. If the fourth value does not meet the criteria for PD, the subject is not assessed as PD due to UPEP. If no fourth UPEP value is available, the subject is not assessed as PD.

1.2.20. Difference between involved and uninvolved FLC level is defined as absolute value of kappa FLC level minus lambda FLC level in the serum.

1.2.21. Development of plasma cell leukemia is considered as disease progression. The date of PD is the date of event onset.

2. Determination of Date of Censoring and Reason for PD

The date of last post-baseline efficacy measure is used as the censoring date for all subjects without progressive disease. Subjects that have no post-baseline efficacy data are censored at the date of randomization for randomized subjects.

The reason(s) for PD is defined as the initial reason(s) that caused the program to indicate PD as well as any other criteria that were met by the time of confirmation of PD. Indicator variables for each reason (SPEP, UPEP, FLC, bone marrow, bone lesion [increase in number, increase in size], extramedullary plasmacytoma [new extramedullary plasmacytoma, increase in size], hypercalcemia and plasma cell leukemia) are created.

6

3. Determination of Response Category and Duration of Response

3.1. IMWG Criteria

According to IMWG criteria, response categories include complete response (CR), stringent complete response (sCR), very good partial response (VGPR), partial response (PR), stable disease (SD), and progressive disease (PD) (defined in Section 1). Although minimal response (MR) is not officially a response category in the IMWG criteria, consensus recommendations state that for subjects with relapsed and /or refractory myeloma, MR should be reported separately in clinical trials. Categories of sCR, CR, VGPR, PR, and SD are determined using the IMWG criteria and MR adopted from the EBMT criteria for subjects with relapsed refractory myeloma as outlined below.

Further explanations (noted in superscript) pertaining to particular response criteria are provided in Section 3.2, Clarifications and Modifications. The definition for duration of response is also covered in Section 3.2.18.

3.1.1. Definition of CR

Requires all of the following:

a. Negative immunofixation of serum and urine [3.2.4; 3.2.5; 3.2.6; 3.2.19].

b. Disappearance of any soft tissue plasmacytomas.

c. <5% plasma cells in the bone marrow [3.2.7].

3.1.2. Definition of sCR

Requires all of the following:

a. CR as defined above.

b. Normal FLC ratio [3.2.8].

c. Absence of clonal bone marrow plasma cell (PCs) by immunohistochemistry, immunofluorescence [3.2.9] or 2- to 4-color flow cytometry.

3.1.3. Definition of VGPR

Requires any of the following:

a. Serum and urine M-component detectable by immunofixation but not on electrophoresis [3.2.5; 3.2.11;3.2.12], or

7

b. ≥90% reduction [3.2.10] in serum M-protein plus urine M-protein <100 mg/24 hours [3.2.5; 3.2.11]

c. If the serum and urine M-protein are not measurable, a reduction [3.2.10] of >90% in the difference between involved and uninvolved FLC levels [3.2.5; 3.2.13] is required.

d. In addition to the above criteria, if present at baseline, ≥50% reduction [3.2.10] in the size of soft tissue plasmacytomas is also required.

3.1.4. Definition of PR

Requires all of the following:

a. ≥50% reduction [3.2.10] of serum M-protein [3.2.5] and reduction [3.2.10] in 24-hour urinary M-protein by ≥90% or to <200 mg/24 hours [3.2.5].

b. If the serum and urine M-protein are not measurable, a reduction [3.2.10] of ≥50% in the difference between involved and uninvolved FLC levels [3.2.5; 3.2.13] is required.

c. In addition to the above criteria, if present at baseline, ≥50% reduction [3.2.10] in the size of soft tissue plasmacytomas is also required.

3.1.5. Definition of SD

Not meeting the criteria for sCR, CR, VGPR, PR, or PD.

3.2. Clarifications and Modifications

As was the case with PD, developing a program to assess response requires adding certain clarifications, minor modifications and additions to the IMWG criteria.

3.2.1. Only subjects with measurable disease at baseline are eligible for assessment of response (i.e., considered in the response-evaluable population). Measurable disease is defined in Section 1.2.1; only legitimated on treatment serum M-protein and urine paraprotein measurements are used for assessment of response. The legitimated on treatment measurements is defined in Section 1.2.6.

3.2.2. CR, sCR, VGPR, PR, MR and SD response categories require no known evidence of progressive or new bone lesions if radiographic studies were performed. Once the program has determined PD for a subject, no subsequent response assessments are performed. For example, a subject who progresses at week 6 cannot have a first response or best response that starts after week 6.

8

3.2.3. Subjects with measurable disease (defined in Section 1.2.1) in serum (SPEP) and urine (UPEP) need to be followed by both SPEP and UPEP for response assessment and categorization; Except for assessment of CR or better, subjects with measurable disease restricted to the SPEP will need to be followed only by SPEP (i.e., urine M-protein need not show a reduction, but the available urine M-protein values must not meet the criteria for PD); correspondingly, subjects with measureable disease restricted to the UPEP will need to be followed only UPEP (i.e., serum M-protein need not show a reduction, but the available serum M-protein values must not meet the criteria for PD). For example, a subject who has baseline values of 0.1 g/dL of IgG M-protein and 300 mg/24 hrs of urine paraprotein and who subsequently maintains values of 0.1 g/dL and 120 mg/24 hrs will be regarded as achieving a PR; Subjects with measurable disease in either SPEP or UPEP or both will be assessed for response only based on these two tests and not by the FLC assay.

3.2.4. To be considered CR, both serum and urine immunofixation must be carried out and be negative regardless of the size of baseline M-protein in the serum or urine; subjects with negative UPEP values pretreatment still require UPEP testing to confirm CR.

3.2.5. Requires 2 consecutive (i.e., no intermediate values that do not meet the definition of response) assessments made at any time before the institution of any new therapy (i.e., subsequent anti-cancer therapy).

3.2.6. For coding CR in subjects in whom the only measurable disease is by serum FLC levels: it requires a normal FLC ratio (Kappa/Lambda) in addition to CR criteria. However, a normal FLC ratio is not required if the involved FLC level decrease to below detectable level (Kappa<0.5 mg/L or Kappa<0.67 mg/L based on new reagent lot; or lambda<0.6 mg/L). The default reference range of 0.26 to 1.65 will be used to determine normal FLC ratio, if the reference range from central/local laboratory is not available.

3.2.7. If all criteria for confirmed CR were met, except that bone marrow aspirate and biopsy were not performed, and baseline bone marrow evaluation showed <5% plasma cells, then the algorithm accepts this as a CR. If both bone marrow aspirate and bone marrow biopsy were performed at baseline, then both values must have <5% plasma cells for the rule to be applied. For subjects without measurable disease at baseline or subjects whose baseline bone marrow plasma cells percentage is >5%, bone marrow confirmation is required for CR response.

9

3.2.8. Normal FLC ratio is required for all subjects regardless of whether disease at baseline was measurable on serum, urine, both or neither. However, a normal FLC ratio is not required if the involved FLC level decrease to below detectable level (Kappa<0.5 mg/L or Kappa<0.67 mg/L based on new reagent lot; or lambda<0.6 mg/L). The default reference range of 0.26 to 1.65 will be used to determine normal FLC ratio (Kappa/Lambda), if the reference range from central/ local laboratory is not available.

3.2.9. Presence/absence of clonal cells is based upon the kappa/lambda ratio. An abnormal kappa/lambda ratio by immunohistochemistry or immunofluorescence requires a minimum of 100 plasma cells for analysis. An abnormal ratio reflecting presence of an abnormal clone is kappa/lambda of >4:1 or <1:2.

3.2.10. Reductions are based on changes from baseline.

3.2.11. VGPR categories require serum and urine studies regardless of whether disease at baseline was measurable on serum, urine, both or neither. For coding VGPR in subjects in whom the only measurable disease is by SPEP, it is not required that UPEP value must meet VGPR criteria, but UPEP test must be performed and not meeting PD criteria whenever SPEP meet VGPR criteria; correspondingly, for coding VGPR in subjects with measureable disease restricted to the UPEP, it is not required that SPEP value must meet VGPR criteria, but SPEP test must be performed and not meeting PD criteria whenever UPEP meet VGPR criteria.

3.2.12. Serum and urine M-protein via SPEP and UPEP must be reported as 0, not detected, or below level of quantification and positive serum or urine immunofixation.

3.2.13. Difference between involved and uninvolved FLC level is defined as absolute value of kappa FLC level minus lambda FLC level in the serum.

3.2.14. Skeletal survey is not required for assessment of response unless clinically indicated. However, if skeletal survey is performed, there should be no indication of disease progression before confirmation of response.

3.2.15. The date of first/best response is the earliest date that all available and required response criteria are met. The date of serum immunofixation response is the initial date of response, not the date of confirmation. The date of urine immunofixation response is the initial date of response, not the date of confirmation. The latest date of response (date at which all criteria were met) among all the relevant response criteria will also be stored in the analysis dataset.

10

3.2.16. Duration of response (DOR) applies to subjects achieving at least PR by IMWG criteria, and is measured from start of achieving PR (first observation of PR before confirmation) to the time of disease progression, with deaths owing to causes other than progression not counted, but censored. DOR is derived as (date of PD or date of censoring – date of first response + 1).

3.2.17. If the first two or more scheduled assessments are missing followed immediately by a confirmed PD (> study day 56), the best response should be SD instead of PD. If best response is assessed as SD, date of best response is not assigned.

3.2.18. Occasionally, the results of the urine immunofixation and UPEP may conflict. If the urine IFE is negative and the UPEP has any value less than 200 mg/24 hrs, the UPEP is disregarded and the urine results are treated as satisfying the requirement for a CR. Similarly, if serum IFE becomes and remains negative, appearance of low level of paraprotein (≤0.5 g/dL or 5 g/L) in SPEP will be disregarded.

3.2.19. A DIRA result of NEGATIVE is treated as equivalent to a negative immunofixation of serum. Since there is no repeat of the DIRA test, we consider a single DIRA test as equivalent of two consecutive immunofixation of serum tests if that indicates a CR.

3.2.20. Subjects with at least one post-baseline disease assessment corresponding to the type of measurable disease at baseline and also not falling into any response category or progressive disease are assigned as response category- stable disease (SD).

11

4. References

1. BGM Durie *et al.* International uniform response criteria for multiple myeloma. *Leukemia 2006*
2. Rajkumar *et al.* Consensus recommendations for the uniform reporting of clinical trials: report of the International Myeloma Workshop Consensus Panel 1. *Blood 2011*

12