Statistical Analysis Plan

Protocol Title:	**A Randomized, Open-label, Phase 3 Study Comparing Carfilzomib, Dexamethasone, and Daratumumab to Carfilzomib and Dexamethasone for the Treatment of Patients With Relapsed or Refractory Multiple Myeloma**
Short Protocol Title:	A phase 3 study comparing KdD vs Kd in RRMM
Protocol Number:	20160275
NCT Number:	03158688
Authors:	PPD PhD Biostatistics Senior Manager, GBS, Amgen Inc. PPD MS Biostatistician I, PAREXEL International
Sponsor:	Amgen Inc. One Amgen Center Drive, Thousand Oaks, CA 91320
SAP Date:	**Document Version** — **Date** **Original (v2.0)** — 15 July 2019

NCT Number: 03158688
This NCT number has been applied to the document
for purposes of posting on clinicaltrials.gov

Version Number	**Date (DDMMMYYYY)**	**Summary of Changes, including rationale for changes**
Original (v1.0)	31 July 2017	Original SAP
[Amendment 1 (v2.0)]	15 July 2019	Add interim OS analyses per Protocol Amendment 4 (17 May 2017)

Table of Contents

List of Table

List of Abbreviations and Definition of Terms

Abbreviation or Term	Definition/Explanation
ALT	alanine aminotransferase
AST	aspartate aminotransferase
BSA	body surface area
CD38	cluster differentiation antigen 38
CI	confidence interval
COA	clinical outcome assessment
CR	complete response
CrCl	creatinine clearance
CRF	case report form
CRR	complete response rate
CTCAE	Common Terminology Criteria for Adverse Events
DMC	Data Monitoring Committee
DOR	duration of response
ECG	electrocardiogram
ECOG	Eastern Cooperative Oncology Group
End of Study	defined as the date when the last subject is assessed or receives an intervention for evaluation in the study (ie, last subject last visit), following any additional parts in the study (eg, long-term follow-up), as applicable
EORTC	European Organisation for Research and Treatment of Cancer
EQ-5D-5L	EuroQol quality of life 5 dimensions 5 level version
FC	flow cytometry
FEV1	forced expiratory volume in 1 second
FISH	fluorescence in-situ hybridization
FVC	forced vital capacity
Heart rate	number of cardiac cycles per unit of time
HR	hazard ratio
ICH	International Council for Harmonisation
Interactive Voice Response (IVR)	telecommunication technology that is linked to a central computer in real time as an interface to collect and process information.
Interactive Web Response (IWR)	web based technology that is linked to a central computer in real time as an interface to collect and process information.
IRC	Independent Review Committee

Abbreviation or Term	Definition/Explanation
ISS	**International Staging System**
ITT	**Intention-to-Treat**
IMWG-URC	**International Myeloma Working Group-Uniform Response Criteria**
IxRS	**interactive voice/web response system**
KdD	**20/56 mg/m2 twice weekly carfilzomib, dexamethasone, and daratumumab**
Kd	**20/56 mg/m2 twice weekly carfilzomib and dexamethasone**
K-M	**Kaplan-Meier**
LTFU	**long-term follow-up**
LVEF	**left ventricular ejection fraction**
MRD[-]CR	**minimal residual disease negative-complete response; defined as achievement of CR (includes sCR) per International Myeloma Working Group-Uniform Response Criteria (IMWG-URC) and MRD[-] status as assessed by NGS (at a 10-5 level)**
NGS	**Next-Generation sequencing**
ORCA	**Onyx Response Computer Algorithm**
ORR	**overall response rate**
OS	**overall survival**
PD	**progressive disease**
PDn	**pharmacodynamics**
PK	**pharmacokinetic**
PFS	**progression free survival**
PR	**partial response**
PS	**performance status**
QLQ-C30	**quality of life questionnaire – core 30 items**
QLQ-MY20	**quality of life questionnaire - myeloma 20 items**
QoL	**quality of life**
QT interval	**QT interval is a measure of the time between the start of the Q wave and the end of the T wave in the heart's electrical cycle as measured by ECG**
QTc	**QT interval corrected for heart rate using accepted methodology**
R-ISS	**Revised International Staging System**
RRMM	**relapsed or refractory multiple myeloma**
SAP	**statistical analysis plan**
sCR	**stringent complete response**
SNPs	**single nucleotide polymorphisms**
VGPR	**very good partial response**
WBC	**white blood cells**

1. Introduction

The purpose of this Statistical Analysis Plan (SAP) is to provide details of the statistical analyses that have been outlined within the protocol amendment **4** for study 20160275, (carfilzomib) dated **17 May 2019**. The scope of this plan includes the primary **analysis for PFS, interim** and final analyses **for OS** that are planned and will be executed by the Amgen Global Biostatistical Science department unless otherwise specified.

2. Objectives, Endpoints and Hypotheses

2.1 Objectives and Endpoints

Objectives	**Endpoints**
Primary	
• To compare carfilzomib, dexamethasone, and daratumumab (KdD) to carfilzomib and dexamethasone (Kd) in terms of progression free survival (PFS) in patients with multiple myeloma who have relapsed after 1 to 3 prior therapies	• Progression free survival defined as time from randomization until disease progression or death from any cause. **Disease** progression **is** determined by a blinded Independent Review Committee (IRC).
Secondary	
Key secondary objectives: • To compare the overall response rate (ORR; defined as the proportion of best overall response of stringent complete response [sCR], complete response [CR], very good partial response [VGPR], and partial response [PR]) between two arms • To compare the rate of minimal residual disease negative-complete response (MRD[-]CR) in bone marrow aspirates at 12 months (± 4 weeks) as determined by Next-Generation sequencing (NGS) between two arms • To compare the overall survival (OS) between two arms Additional secondary objectives are to compare the following between the 2 arms: • safety and tolerability • duration of response (DOR) • time to next treatment **(TTNT)** • time to progression (TTP)	Key secondary endpoints: • ORR, defined as the proportion of **subjects with** best overall response **(BOR)** of **stringent complete response (**sCR**)**, **complete response (**CR**)**, **very good partial response (**VGPR**)**, and **partial response (**PR**)** by IRC • MRD[-]CR rate, defined as **proportion of subjects with BOR** of CR or better by IRC per International Myeloma Working Group Uniform Response Criteria (IMWG URC) and MRD[-] status as assessed by NGS (at a 10^{-5}level) at 12 months • OS, defined as time from randomization until death from any cause. Additional secondary endpoints: • DOR**, defined as the time from first evidence of PR or better per IMWG-URC to the earliest of disease progression or death due to any cause for subjects with a best response of PR or better**

Objectives	Endpoints
• time to response **(TTR)** • persistence of MRD[-]CR • complete response rate (CRR) • MRD[-] rate • quality of life	• **TTNT, defined as the time from randomization to the initiation of subsequent non-protocol anti-cancer treatment for multiple myeloma** • TTP**, defined as the time from randomization to documented disease progression** • **TTR, defined as the time from randomization to the earliest date a response of PR or better is first achieved and subsequently confirmed for subjects with a best response of PR or better** • sustained MRD[-]CR**,** defined as the proportion of subjects **who** maintain MRD[-]CR for 12 months or more after achieving MRD[-]CR status • CRR**,** defined as the proportion of **subjects with BOR** of sCR or CR • MRD[-] rate**, defined as the proportion of subjects with MRD[-] status as assessed by NGS (at a 10^{-5} level) at 12 months** • Global Health Status/Quality of Life (**GHS/**QoL) measured by European Organisation for Research and Treatment of Cancer (EORTC) QLQ-C30 version 3 questionnaire • subject incidence of treatment-emergent adverse events **(TEAE)** • safety laboratory values, left ventricular ejection fraction (LVEF), forced expiratory volume in 1 second (FEV1)/forced vital capacity (FVC) ratio, and vital signs at each scheduled assessment

Objectives	Endpoints
Exploratory	
CCI	

2.2 Hypotheses and/or Estimations

The KdD regimen will provide significant improvement in PFS over the Kd regimen. **The hypotheses will be tested using a fixed sequence hierarchical testing procedure to control the family-wise type I error rate at 1-sided 0.025 level.**

3. Study Overview

3.1 Study Design

This is a phase 3 multicenter, open-label, randomized study in subjects with relapsed or refractory multiple myeloma (RRMM) who have received 1 to 3 prior therapies.

Subjects will be randomized in a 2:1 ratio to 1 of 2 arms: Arm 1: KdD vs. Arm 2: Kd.

Randomization will be performed using an interactive voice/web response system (IxRS) and subjects will be stratified based on the following criteria: 1) International Staging System (ISS) stage (Stage 1 or 2 vs Stage 3) at screening, 2) prior proteasome inhibitor exposure (yes vs no), 3) number of prior lines of therapy (1 vs ≥ 2), and 4) prior cluster differentiation antigen 38 (CD38) antibody therapy (yes vs no).

Subjects will receive the treatment determined by randomization for a maximum of 4 years or until confirmed disease progression, unacceptable toxicity, withdrawal of consent, or death (whichever occurs first). No crossover between the treatment arms will be allowed. All subjects will be assessed for multiple myeloma disease response according to the IMWG URC (Kumar et al, 2016) using central laboratory test results every 28 ± 7 days. Disease response assessments will be performed every 28 ± 7 days until confirmed progressive disease (PD), irrespective of cycle duration including dose delays or treatment discontinuation.

Following progression or discontinuation of study drug(s), subjects will have 2 follow-up visits (30 days [+ 3] and 8 weeks [± 7 days] after last dose of all study drug[s]) and then remain in long term follow up (LTFU) where data on survival status **and subsequent antimyeloma therapy** will be gathered every 12 weeks ± 2 weeks.

3.2 Sample Size

One hundred eighty-eight PFS events are required to have at least 90% power to demonstrate superiority at an alternative HR of 0.6 (arm 1 vs arm 2), using a log rank test at 1-sided overall significance level of 0.025. CCI [redacted] CCI [redacted]. With 450 subjects randomized (300 in arm 1 vs 150 in arm 2), it is anticipated that 188 events will be accrued at approximately 27 months after the first subject is

randomized assuming events follow an exponential distribution. CCI [redacted] The description of sample size calculation for log-rank test factoring in accrual and dropout can be found in Lachin and Foulkes, 1986; Lakatos, 1988; and Chow et al, 2003. The actual timing of primary analysis will be determined by actual enrollment rate, dropout rate, and PFS event rate; hence, it is subject to change as these factors may vary in the study. The minimum detectable PFS HR that will achieve statistical significance is approximately 0.738 (corresponding to about 36% improvement in PFS) at the primary analysis. Amgen may choose to increase sample size upon observation of slower than expected PFS event rate.

CCI [redacted]

The sample size calculation was performed using East® software (Version 6.3 or above).

4. Covariates and Subgroups

4.1 Planned Covariates

Not applicable.

4.2 Subgroups

In addition to the stratification factors for randomization, the following covariates will be used to examine primary and selected secondary endpoints in subgroups as appropriate. When there is not a sufficient number of subjects in the subgroup (ie, less than 10% of subjects in a treatment arm), relevant subgroups may be combined.

- baseline demographics and characteristics:
  - age (<=75, >75)
  - sex (female, male)
  - race (white and other categories depending on frequency observed)
  - region (North America, Europe, Asia **Pacific**, Other)
- baseline organ function and comorbid conditions:
  - ECOG PS (0-1, 2)
  - baseline CrCl (15-30, 30-50, 50-80, ≥80 mL/min)

- baseline disease characteristics:
  - prior Lenalidomide exposure (yes vs no)
  - refractory to Lenalidomide (yes vs no)
  - prior Bortezomib or Ixazomib exposure (yes vs no)
  - refractory to Bortezomib or Ixazomib (yes vs no)
  - **prior IMiD exposure (yes vs no)**
  - **refractory to IMiD (yes vs no)**
  - revised ISS stage (Stage 1 or 2, Stage 3)
  - IgG vs non-IgG
  - determination of measurable disease at baseline (**based on SPEP only, based on UPEP only or both SPEP and UPEP**, based on SFLC only)
  - β2-microglobulin level (< 3.5, ≥3.5 and < 5.5, ≥5.5 mg/L)
  - risk group as determined by **genetic abnormality per IMWG** (high risk group, standard risk group)
  - presence of soft tissue plasmacytoma (yes, no)

5. Definitions

Baseline

Unless otherwise defined, baseline will be defined as the latest value measured on/before day1 of the first cycle of protocol-specified therapy. The protocol specifies that all study procedures on day 1 should be completed before the initiation of protocol-specified therapy, which will be the assumption in the analysis. If a day 1 value is not available, the latest value before the day of the start of protocol-specified therapy will be used. If a subject doesn't receive any protocol-specified therapy, then the latest value prior to or on randomization date will be used.

For plasmacytoma and skeletal survey, baseline will be defined as the latest assessment from screening period until 7 days after the Day 1 Cycle 1.

Body Surface Area (BSA)

BSA will be calculated using the Mosteller formula (Mosteller, 1987):

BSA (m^2) = ([Height(cm) x Weight(kg)]/3600)$^{1/2}$.

Complete Response Rate (CRR)

CRR is defined as the proportion of subjects whose best overall response is sCR or CR.

Death Date

For subjects who die during the study, the death date will be recorded on the end of study CRF **page** in the end of study date. If only the day of a death date is missing, death date will be imputed using the following rules: 1) Day 1 of the month will be used to impute if year and month indicate that death happened later than last known alive date; 2) One day after last known alive date will be used to impute if death happened in the same month and year as last known alive date.

The imputed death date will be used in calculation of duration of response, progression-free survival and overall survival.

Duration of Response (DOR)

DOR will be calculated only for subjects who achieve a best overall response of PR or better, ie, sCR, CR, VGPR, or PR. The duration will be calculated in months from the earliest date a response of PR or better is first achieved and subsequently confirmed to the earliest date of confirmed PD or death due to any cause:

DOR = (PD / death date - response start date + 1) / 30.4

Subjects will be censored using the same censoring rules for PFS as listed in Table 1 if applicable.

Genetic Risk Group

The **high-risk group consists of** subjects with **the following** genetic abnormalities t(4;14), t(14;16), and**/or** deletion 17p. The **standard-risk group consists of** subjects **who do not have any of the above genetic subtype**. **The unknown group consists of subjects who cannot be identified as high-risk nor standard-risk.**

International Staging System (ISS) Stage at Baseline

ISS stage at baseline will be calculated using serum beta-2 microglobulin and serum albumin values collected at baseline, based on the criteria published by the International Myeloma Working Group (Greipp 2005):

Stage I: Serum beta-2 microglobulin < 3.5 mg/L and serum albumin ≥ 3.5 g/dL

Stage II: Serum beta-2 microglobulin < 3.5 mg/L and serum albumin < 3.5 g/dL or Serum beta-2 microglobulin 3.5–<5.5 mg/L irrespective of the serum albumin

Stage III: Serum beta-2 microglobulin ≥ 5.5 mg/L

Revised International Staging System (R-ISS) Stage at Baseline

R-ISS stage at baseline will be calculated using ISS, risk group by FISH, and LDH, based on the criteria published by the International Myeloma Working Group (Palumbo, 2015):

Stage I: ISS stage I and standard risk group by FISH and normal LDH

Stage II: Not R-ISS stage I or III

Stage III: ISS stage III and {either high-risk group by FISH or high LDH}

Last Known Alive Date

Last known alive date is the latest date of the following dates before death date:

- Date of Randomization on Subject Enrollment CRF
- Date First Taken, Date Last Taken on Concomitant Medications CRF
- Date Performed on ECOG Performance Status, Vital Signs, Echocardiogram, Electrocardiogram, Transfusions, Surgery, Procedure CRFs
- Admission Date, Discharge Date on Hospitalizations, CRF Date of Examination on Physical Measurement CRF
- Date Collected on Reproductive Status and Pregnancy Test (Local Lab), Chemistry (Local Lab), Hematology (Local Lab), Coagulation (Local Lab) CRFs and in central lab data
- Start Date, Stop Date on Investigational Product Administration CRF
- Date Started and Date Ended or Resulted in Death on Events CRF
- Start date, Stop date on Concurrent Radiotherapy, Anti-Myeloma / Anti-Cancer Therapies
- Subject Status Date if status is Alive on Survival Status CRF
- Assessment Date on Skeletal Survey and Plasmacytoma Assessment CRF
- Date of Clinical Outcome Assessment

Minimal Residual Disease Negative-Complete Response (MRD[-]CR)

MRD[-]CR is defined as achievement of CR (includes sCR) per International Myeloma Working Group-Uniform Response Criteria (IMWG-URC) and MRD[-] status as assessed by NGS (at a 10^{-5} level) **at 12 months landmark**.

MRD[-]CR rate is defined as the proportion of subjects achieving MRD[-]CR at 12 months landmark.

Sustained MRD[-]CR is defined as maintaining MRD[-]CR for 12 months or more after achieving MRD[-]CR status.

Sustained MRD[-]CR rate is defined as the proportion of subjects achieving sustained MRD[-]CR.

MRD[-] rate is defined as the proportion of subjects that achieved MRD[-] status as assessed by NGS (at a 10^{-5} level) at 12 months landmark.

12-month landmark is defined as 12 months (-4 months / +1 month).

Onyx Response Computational Assessment (ORCA)

Responses that are derived using ORCA will be based on the International Myeloma Working Group Uniform Response Criteria (Durie et al. 2006, with corrections, Rajkumar et al. 2011, Kumar et al. 2016) (IMWG-URC). The detailed algorithm is documented in a separate document "SPECIFICATIONS FOR ONYX RESPONSE COMPUTATIONAL ASSESSMENT (ORCA) BASED ON IMWG UNIFORM RESPONSE CRITERIA".

Overall Response Rate (ORR)

ORR is the proportion of subjects whose best overall response is PR, VGPR, CR, or sCR.

Overall Survival (OS)

OS time in months will be calculated from time of randomization until death due to any cause:

$$OS = (\text{death date} - \text{randomization date} + 1) / 30.4$$

Subjects still alive will be censored at the date last known to be alive. If the date last known to be alive is after the date that triggers the analysis (ie, the data cutoff date), the subject will be censored at the analysis trigger date.

Progression-Free Survival (PFS)

PFS time will be calculated from the time of randomization (in months) until PD or death due to any cause, whichever occurs first:

$$PFS = (\text{PD / death date} - \text{randomization date} + 1) / 30.4$$

The duration of PFS will be right censored for subjects who meet any of the following conditions: 1) no baseline **/ post-baseline** disease assessments; 2) starting a new anti myeloma therapy before documentation of progressive disease or death; 3) progressive disease or death immediately after more than 70 days without disease assessment visit or; 4) alive without documentation of disease progression before the

analysis trigger date**; 5) lost to follow up or withdrawn consent**. The censoring rules for PFS primary analysis are described in Table 1. These rules are based on the May 2007 FDA Guidance for Industry, 'Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics' (https://www.fda.gov/ucm/groups/fdagov-public/@fdagov-drugs-gen/documents/document/ucm071590.pdf).

Table 1. Censoring Rules for Primary PFS Analysis

Situation	Date of Progression or Censoring	Outcome
No baseline / post-baseline disease assessments	Date of randomization	Censored
New anti-myeloma treatment started before documentation of PD or death	Date of last disease assessment prior to start of a new anti-myeloma treatment	Censored
Death or PD immediately after more than 1 consecutively missed disease assessment visit*	Date of last disease assessment visit before the first missed visit	Censored
Alive and without PD documentation	Date of last disease assessment	Censored
Lost to follow up or withdrawn consent	**Date of last disease assessment**	**Censored**
Death or PD between planned disease assessments	Date of death or first disease assessment showing PD, whichever occurs first	Progressed
Death before first disease assessment	Date of death	Progressed

* If death or PD is more than 70 days after previous disease assessment, or randomization date if there is no previous disease assessment

Randomization Date

Randomization Date is defined as the date the subject was allocated to a treatment group. Per protocol, study treatment will ideally commence on the day of randomization, but at least within 5 calendar days of randomization, unless approved by the Medical Monitor.

Refractory to prior multiple myeloma therapy

Subject is refractory to a drug of interest received in prior regimens if the data collected on prior multiple myeloma therapy CRF indicate that any of the following criteria is met:

- **Best overall response to any regimen containing the drug was no better than stable disease (SD) (ie, SD or progressive disease (PD).**

- **Reason the drug was stopped was progression in any regimen.**
- **Date of relapse/progression is after start date and within 60 days after stop date of the drug in any regimen.**

Relative Dose Intensity (RDI)

RDI reflects whether the dose intensity of a therapy was implemented as planned. It will be calculated as the ratio of actual dose intensity relative to planned dose intensity

$$\text{Relative Dose Intensity (RDI)} = \frac{\text{Actual dose intensity}}{\text{Planned dose intensity}}$$

For Carfilzomib **as Amgen investigational product**: Actual dose intensity is defined as the actual amount of drug in mg/m^2 delivered to a subject per week of treatment.

$$\text{Actual Dose Intensity (mg/m}^2\text{/week)} = \frac{\text{Actual cumulative dose (mg/ m}^2\text{)}}{\text{Number of weeks of actual treatment}}$$

Actual cumulative dose in (mg/m^2) is the sum of received doses (mg) divided by BSA (m^2) of the patient. **Each subject's first dose in (mg/m^2) of carfilzomib will be calculated based on baseline BSA using the Mosteller formula. BSA should not be revised unless the subject experiences a change in body weight of ≥ 20% in which case BSA will be recalculated and the new BSA will be used in dose (mg/m^2) calculation**. BSA will be capped at 2.2 for carfilzomib. Number of weeks of actual treatment will be calculated as (Last Dose Date of Carfilzomib – First Dose Date of Carfilzomib + i) / 7, where i = 7 if the last infusion is given on day 1 or 8 within the last cycle, i = 6 if the last infusion is given on day 2 or 9, i = 14 if the last infusion is given on day 15, i = 13 if the last infusion is given on day 16.

Planned dose intensity is defined as the planned amount of carfilzomib in mg/m^2 delivered to a subject per week of treatment.

$$\text{Planned Dose Intensity (mg/m}^2\text{/week)} = \frac{\text{Planned cumulative dose (mg/ m}^2\text{)}}{\text{Number of protocol specified treatment weeks}}$$

It will be calculated as the planned cumulative dose of carfilzomib in mg/m^2 divided by the planned number of weeks for the treatment per protocol based on the corresponding cycle and day of the last carfilzomib infusion. Per protocol, one cycle is 28 days (4 weeks), so the planned number of treatment weeks will be calculated as 4 x (c-1) + j, where c is the cycle in which the last carfilzomib infusion is given and j =1 if the last carfilzomib infusion is given on day 1 or 2 within the last cycle, j=2 if the last infusion is

given on day 8 or 9, j=4 if the last infusion is given on day 15 or 16. The planned cumulative dose of carfilzomib is the summation of planned carfilzomib dose (mg/m^2) per week as : [20x2, 56x2, 56x2, 0] in cycle 1 and [56x2, 56x2, 56x2, 0] in all other cycles.

For Dexamethasone **as non-Amgen non-investigational product**: Actual dose intensity is the actual amount of drug in mg delivered to a subject per week of treatment.

$$\text{Actual Dose Intensity (mg/week)} = \frac{\text{Actual cumulative dose (mg)}}{\text{Number of weeks of actual treatment}}$$

The cumulative dose in mg is the summation of total quantity administered (mg) over the study. Number of weeks of actual treatment will be calculated as (Last Dose Date of Dexamethasone – First Dose Date of Dexamethasone + i)/7, where i=7 if the last dexamethasone dose is given on day 1, 8, 15, and 22, i=6 if the last dexamethasone dose is given on day 2, 9, and 16.

Planned dose intensity (mg/week) is defined as the planned amount of dexamethasone in mg delivered to a subject per week of treatment.

$$\textbf{Planned Dose Intensity (mg/week)} = \frac{\textbf{Planned cumulative dose (mg)}}{\textbf{Number of protocol specified treatment weeks}}$$

It will be calculated as the planned cumulative dose of dexamethasone in mg divided by the planned number of weeks for the treatment per protocol based on the corresponding cycle and day of the last dexamethasone dose. Per protocol, one cycle is 28 days (4 weeks), so the planned number of treatment weeks will be calculated as 4 x (c-1) + j, where c is the cycle in which the last dexamethasone dose is taken and j =1 if the last dose is taken on day 1 or 2 within the last cycle, j=2 if the last dose is taken on day 8 or 9, j=3 if the last dose is taken on day 15 or 16, j=4 if the last dose is taken on day 22. The planned cumulative dose of dexamethasone is the summation of planned dexamethasone dose (mg) per week as: 40 mg/week in all cycles for subjects ≤ 75 years of age; [40 mg, 28 mg, 28 mg, 20 mg] in cycle 1, 20 mg/week starting from cycle 2 for subjects > 75 years of age in KdD group; 20 mg/week for subjects > 75 years of age in Kd group.

For Daratumumab **as non-Amgen investigational product**: Actual dose intensity is defined as the actual amount of drug in mg/kg delivered to a subject per week of treatment.

$$\text{Actual Dose Intensity (mg/kg/week)} = \frac{\text{Actual cumulative dose (mg/kg)}}{\text{Number of weeks of actual treatment}}$$

Actual cumulative dose in (mg/kg) is the sum of received doses (mg) divided by baseline weight (kg) of the patient. **Each subject's first dose in (mg/kg) of daratumumab will be calculated based on baseline weight. If the subject experiences a change in body weight of ≥ 10%, the new weight will be used in dose (mg/kg) calculation.** Number of weeks of actual treatment will be calculated as (Last Dose Date of Daratumumab – First Dose Date of Daratumumab + i)/7, where i=7 if the last infusion is given on day 1, 8, 15, 22 within the last cycle, i=6 if the last infusion is given on day 2.

Planned dose intensity is defined as the planned amount of Daratumumab in mg/kg delivered to a subject per week of treatment.

$$\text{Planned Dose Intensity (mg/kg/week)} = \frac{\text{Planned cumulative dose (mg/kg)}}{\text{Number of protocol specified treatment weeks}}$$

It will be calculated as the planned cumulative dose of **daratumumab** in mg/kg divided by the planned number of weeks for the treatment per protocol based on the corresponding cycle and day of the last **daratumumab** infusion. Per protocol, one cycle is 28 days (4 weeks), so the planned number of treatment weeks will be calculated as 4 x (c-1) + j, where c is the cycle in which the last daratumumab infusion is given and j =1 if the last daratumumab infusion is given on day 1 or 2 within the last cycle, j=2 if the last infusion is given on day 8, j=3 if the last infusion is given on day 15, j=4 if the last infusion is given on day 22. The planned cumulative dose of **daratumumab** is the summation of planned **daratumumab** dose (mg/m^2) per week as : [8x2, 16x1, 16x1, 16x1] in cycle 1; [16x1, 16x1, 16x1, 16x1] in cycle 2; [16x1, 0, 16x1, 0] in cycle 3-6; and [16x1, 0, 0, 0] in all other cycles.

Study Day 1

Study Day 1 is the first day that the protocol-specified investigational products are administered to the subject.

Time to Complete Response

Time to complete response will be calculated only for the subjects who achieve a best overall response of CR or better, ie, sCR or CR. It will be calculated in months from randomization date to the earliest date of response of CR or better is first achieved and subsequently confirmed:

Time to Complete Response = (response start date - randomization date + 1) / 30.4

Time to Next Treatment

Time to next treatment is defined as the time (in months) from randomization to the initiation of subsequent non-protocol anti-cancer treatment for multiple myeloma.

Time to **Progression** (TTP)

TTP is defined as the time (in months) from randomization to documented disease progression.

Time to Response

Time to response will be calculated only for subjects who achieve a best overall response of PR or better, ie, sCR, CR, VGPR, or PR. It will be calculated in months from randomization date to the earliest date a response of PR or better is first achieved and subsequently confirmed:

Time to Response = (response start date - randomization date + 1) / 30.4

Treatment-emergent Adverse Event

Treatment-emergent adverse events (TEAE) are events with an onset after the administration of the first dose of any **study treatment** and within **the end of study or** 30 days of the last dose of any **study treatment, whichever occurs earlier**.

6. Analysis Sets

6.1 Full Analysis Set (Intent-to-Treat (ITT) Population)

The full analysis set will include all randomized subjects. All subjects will be analyzed according to treatment to which they are randomized. Full analysis set will be used for the primary and key secondary endpoints

6.2 Safety Analysis Set

The safety population will include all randomized subjects who received at least 1 dose of any study treatment (ie, carfilzomib, dexamethasone, or daratumumab). Subjects in the analyses based on the safety analysis set will be analyzed according to the treatment group corresponding to the actual treatment received.

6.3 Per Protocol Set(s)

The per protocol analysis set is a subset of the full analysis set which includes subjects who do not have important protocol deviations that are considered to have an effect on efficacy outcomes. The list of important protocol deviations is maintained by the sponsor on an ongoing basis and will be finalized before the primary analysis of the study**. The Per-protocol Sets are different for OS analysis from PFS/ORR analysis.**

Subjects with the following protocol deviations will be excluded from the per protocol set for OS analysis:

- PPD **- Inclusion criteria of relapsed or progressive multiple myeloma**
- PPD **- Inclusion criteria of measurable disease**
- PPD **- Inclusion criteria of at least PR to at least 1 prior line of therapy**
- PPD **- Inclusion criteria of 1 - 3 prior therapies for multiple myeloma**
- PPD **- Inclusion criteria of prior exposure to carfilzomib therapy**
- PPD **- Inclusion criteria of prior exposure to anti-CD38 antibody therapy**
- PPD **- Exclusion criteria of history of other malignancy within the past 5 years**
- PPD **- Subject received therapy with a marketed or investigational anticancer therapeutic or radiation to large marrow reserves for either a therapeutic or palliative intent prior to confirmed progressive disease**
- PPD **- Subject received different treatment type from randomized treatment assignment; subject in arm 2 received daratumumab**
- PPD **– Incorrect carfilzomib dose (under dose)**

In addition to the deviations listed above, subjects who had the following deviations will also be excluded from per-protocol set for PFS and ORR analyses:

- PPD **- Subject permanently discontinued treatment due to progressive disease (PD) based on local labs not central labs, except in the case of disease progression based on hypercalcemia**
- PPD **- Screening lab disease assessment performed outside window and not reported prior to C1D1 dosing, such that the primary endpoint cannot be assessed**

- PPD **- Failure to obtain extramedullary plasmacytoma assessment and/or bone lesion assessment not carried out within 30 days prior to randomization or not obtained post-randomization and prior to C1D1 such that the primary endpoint cannot be assessed**
- PPD **- Failure to obtain all required response assessments per protocol schedule of assessments and IMWG URC specifications to assess PFS**

6.4 MRD Evaluable Analysis Set

The MRD evaluable analysis set is a subset of the full analysis set, which excludes subjects who don't have baseline MRD sample or don't have post-baseline MRD assessment due to technical issues.

6.5 12-month Landmark MRD Analysis Set

The 12-month landmark MRD analysis set is a subset of the full analysis set, which excludes subjects who do not have the opportunity to have MRD sample collected at month 12. The subjects who do not have the opportunity to have MRD sample collected at month 12 are defined as subjects who ended the treatment prior to month 12.

7. Planned Analyses

CCI

CCI

[redacted]

7.2 Primary Analysis

The timing for the primary analyses of PFS for arm 1 vs arm 2 will be event driven and will happen when approximately 188 PFS events are reached cumulatively.

CCI

[redacted]

CCI

[redacted]

Testing of the key secondary endpoints will be performed using a fixed sequence hierarchical testing procedure in the order of ORR, MRD[-]CR, and OS such that the overall Type I error rate is strongly controlled under 0.025 (1-sided). PFS, ORR, MRD[-]CR and OS will be tested at the PFS primary analysis.

The hierarchical testing procedure will be applied as following:

- **PFS (arm 1 vs arm 2) is tested at alpha level of 0.025 at the primary analysis**
- **If PFS is significant at the primary analysis, then ORR, MRD[-]CR are tested sequentially at alpha level of 0.025. Starting with the hypothesis of ORR, if any hypothesis in the sequence is rejected at a 1-sided significance level of 0.025, then the subsequent hypothesis will be tested. Otherwise, if any hypothesis failed to be rejected, then the subsequent hypotheses will not be tested.**
- **If PFS, ORR, and MRD[-]CR are all statistically significant, then OS will be tested multiple times with an overall alpha of 0.025. During the primary analysis, OS will be tested at a significance level of 0.001 per section 7.1.**

CCI

[redacted]

8. Data Screening and Acceptance

8.1 General Principles

The objective of the data screening is to assess the quantity, quality, and statistical characteristics of the data relative to the requirements of the planned analyses.

8.2 Data Handling and Electronic Transfer of Data

The Amgen Global Study Operations-Data Management (GSO-DM) department will provide all data to be used in the planned analyses. This study will use the RAVE database. Laboratory data will be collected by COVANCE Central Laboratory Services and transferred to Amgen GSO-DM periodically in cumulative files. Quality of life data

will be collected by ERT and transferred to Amgen GSO-DM periodically in cumulative files.

8.3 Handling of Missing and Incomplete Data

The descriptive statistics will identify the extent of missing data. Rules for handling missing data related to endpoints are described in the endpoint definitions or in the description of analyses. The handling of incomplete and partial dates for adverse events and concomitant medications are described in Appendix A.

8.4 Detection of Bias

If applicable the methods to detect bias are described in the analyses of particular endpoints.

8.5 Outliers

Pharmacokinetic (PK) concentration data will be evaluated for outliers by visual inspection, and decisions to re-assay individual samples will be made in accordance with standard pharmacokinetic evaluation practice.

8.6 Distributional Characteristics

If applicable the distributional characteristics will be explored for particular endpoints.

8.7 Validation of Statistical Analyses

Programs will be developed and maintained, and output will be verified in accordance with current risk-based quality control procedures.

Tables, figures, and listings will be produced with validated standard macro programs where standard macros can produce the specified outputs.

The production environment for statistical analyses consists of Amgen-supported versions of statistical analysis software; for example, the SAS System version 9.4 or later.

9. Statistical Methods of Analysis

9.1 General Considerations

The efficacy analyses of PFS and key secondary endpoints will be conducted on the full analysis set. Treatment effects in efficacy endpoints will be evaluated and compared between KdD vs Kd.

In principle, summary statistics including mean, standard deviation, median, first and third quartiles, will be provided for continuous variables. Frequency and percentage will be summarized by treatment arm for binary and categorical variables. Proportions and

the corresponding 95% CI will be based on **Clopper-Pearson Method** and the treatment comparison will be based on Cochran-Mantel-Haenszel test. Exact tests will be considered for subgroup analyses when the cell size is considered small.
Time to event endpoints will be estimated using the Kaplan-Meier (K-M) method. Stratified log-rank test statistics and associated p-values will also be calculated. Hazard ratios will be estimated using stratified Cox proportional hazards models. **Stratification factors include original International Staging System (ISS) stage (Stage 1 or 2 vs Stage 3) at screening; prior proteasome inhibitor exposure (yes vs no); number of prior lines of therapy (1 vs ≥ 2); prior cluster differentiation antigen 38 (CD38) antibody therapy (yes vs no). A stratification factor will not be used if any level of the factor contains less than 5% of the whole population.**

The adequacy of the proportional hazard assumption will be assessed using the plot of the logarithm of the estimated hazard function based on the Kaplan-Meier method against the logarithm of time-to-event endpoints. The scaled Schoenfeld residuals by time plot will be examined for evidence of a non-zero correlation, which indicates non-proportionality. In addition, a treatment-by-time(log) interaction test will be performed using a Cox model with stratification factors.

For PFS, response and disease progression will be determined by an IRC in a blinded manner. In addition, response and disease progression outcomes will be determined locally by investigators in an unblinded manner and centrally by the sponsor using a validated computer algorithm (Onyx Response Computer Algorithm, ORCA) in a blinded manner. The primary analysis of PFS will be based on IRC assessed outcomes; the timing will be event driven and will happen when approximately 188 PFS events are reached. The PFS outcomes assessed by the investigators as well as by ORCA will serve as supportive analyses of PFS.

The primary comparison of PFS will be tested using a log rank test stratified by the randomization stratification factors per IxRS at 1-sided significance level of 0.025.

9.2 Subject Accountability

The number and percent of subjects who were screened, randomized, received study treatment, entered long-term follow-up before disease progression and long-term follow up for survival will be summarized by treatment group. The number and percent of subjects who discontinued study treatment, long-term follow-up before disease progression, and study will be tabulated, along with the reason for discontinuation. The number and percent of subjects randomized will be tabulated by the stratification factors.

The number and percent of subjects randomized will be tabulated by study site. Key study dates for the first subject randomized, last subject randomized, and data cut-off date for analysis will be presented.

9.3 Important Protocol Deviations

Important Protocol Deviations (IPDs) categories are defined by the study team before the first subject's initial visit and updated during the IPD reviews throughout the study prior to database lock. These definitions of IPD categories, subcategory codes, and descriptions will be used during the course of the study.

9.4 Demographic and Baseline Characteristics

Demographic and baseline disease characteristics will be summarized by treatment group and overall using descriptive statistics for the Full Analysis Set. These include, but are not limited to the following.

- Baseline demographics and characteristics:
  - Age at randomization
    - As continuous variable
    - As categorical variable: <75, ≥ 75 years
    - **As categorical variable: 18 – 64, 65 – 74, 75 – 84, ≥ 85 years**
  - Sex (Male, Female)
  - Race (White and other categories depending on frequency observed)
  - Ethnicity (Hispanic or Latino, Not Hispanic or Latino)
  - **Height (cm)**
  - Weight (kg)
  - Body surface area (m^2) **(as continuous in m^2; as categorical variable: ≤ 2.2, > 2.2 m^2)**
  - Body mass index
  - Region (North America, Europe, Asia **Pacific**, Other)
  - Frailty status **(Fit, Intermediate Fitness, Frail, Not available)**
- Baseline organ function and comorbid conditions:
  - ECOG performance status (0-1 [0, 1], 2)
  - Hemoglobin (g/L) **(as continuous in g/L; as categorical variable: < 105, ≥ 105 g/L)**
  - Absolute Neutrophil Count (**10^9/L**) **(as continuous in 10^9/L; as categorical variable: < 1.5, ≥ 1.5 *10^9/L)**
  - Platelet count (10^9/L) **(as continuous in 10^9/L; as categorical variable: < 105, ≥ 105 *10^9/L)**

- Corrected calcium (mg/dL): calculated **by Covance (central laboratory)** as [serum calcium (mg/dL) + 0.8×(4 - serum albumin (g/dL))] **(as continuous in mg/dL; as categorical variable: ≤11.5, > 11.5 mg/dL)**
- Creatinine clearance (CrCl, **as continuous in** mL/min; as categorical variable: **< 15,** 15 - < 30, 30-<50, 50-<80, >=80 mL/min): Measured or calculated CrCl according to the Cockcroft-Gault formula **by Covance (central laboratory)**:

$$CrCL(mL/\min) = \frac{(140 - Age) \times Weight(kg)}{72 \times S_{Cr}(mg/dL)} \times (0.85\, female)$$

- **Left ventricular ejection fraction (LVEF) (%)**
- **Baseline ECG (Normal, Abnormal and not clinically significant, Abnormal and clinically significant, not evaluable)**
- Baseline hypertension history (yes/no)
- Baseline history of ischemic heart disease (yes/no)

- Baseline disease characteristics:
  - **Original** ISS stage per IXRS **at screening (I or II, III)**
  - **Original ISS stage at baseline (I, II, III)**
  - Revised ISS stage **at baseline (I, II, III)**
  - MM subtype **(IgG, IgA, IgD, IgE, IgM, None; Kappa, Lambda, Not detectable within each subtype)**
  - Determination of measurable disease at baseline (**based on SPEP only, based on UPEP only or both SPEP and UPEP**, based on SFLC only)
  - β2-microglobulin level (**as continuous in** mg/L; as categorical variable: < **3.5**, ≥ **3.5** and < 5.5, ≥ 5.5 mg/L)
  - **Albumin (as continuous in g/dL; as categorical variable: < 3.5, ≥ 3.5 g/dL)**
  - **SFLC Kappa/Lambda ratio (Normal, Abnormal)**
  - Presence of plasmacytoma (yes, no)
  - **Presence of bone lesion (Yes, No)**
  - **Plasma cell involvement in bone marrow (%) (as continuous in %; as categorical variable < 50%, ≥ 50%)**
  - Risk group as determined by **genetic abnormality per IMWG** (high-risk group, standard-risk group**, and unknown group**)
  - Time from Initial Multiple myeloma diagnosis to randomization (months)
  - Number of prior lines of therapy **(as continuous variable and as categorical variable 0, 1, 2, 3, and > 3)**
  - Prior Lenalidomide exposure (yes, no)
  - Refractory to Lenalidomide exposure (yes, no)
  - Prior Bortezomib exposure (yes, no)
  - **Refractory to Bortezomib exposure (yes, no)**
  - **Prior Bortezomib or Ixazomib exposure (yes, no)**

- **Refractory to Bortezomib or Ixazomib exposure (yes, no)**
- Prior Proteasome Inhibitor exposure (yes, no)
- **Refractory to Proteasome Inhibitor exposure (yes, no)**
- **Prior IMiD exposure (yes, no)**
- **Refractory to IMiD exposure (yes, no)**
- **Best overall response to last prior systemic therapy (sCR, CR, VGPR, PR, SD, PD)**
- Refractory to the last prior line of therapy (yes, no)
- Prior Transplant (yes (Autologous, Allogeneic), no)
- **Number of prior transplant (1, 2, > 2)**

9.5 Efficacy Analyses

The efficacy analyses will be based on the Full Analysis Set.

Table 3. Efficacy Endpoint Summary Table

Endpoint	Primary Summary and Analysis Method	Sensitivity Analysis
Primary Endpoint		
Progression-Free Survival	Based on IRC assessments: • KM summaries • 1-sided p-value from stratified log-rank test. Hazard ratio and 95% CI from stratified Cox regression.	• Investigator assessments: same as primary summary and analysis method based on investigator assessments. • Internal computational assessments: Same as primary summary and analysis method based on ORCA assessments. • Unstratified analyses: 1-sided p-value from unstratified log-rank test, hazard ratio and 95% CI from unstratified Cox regression. • Initiation of new anti-myeloma therapy treated as PFS Event: The data censoring rules are the same as those for the primary analysis of PFS except that the use of new anti-myeloma therapy will be treated as an event rather than a mechanism for censoring. The same analysis method as for primary analysis will be used. • Initiation of new anti-myeloma therapy treated as neither a PFS event nor a censoring event: The data censoring rules are the same as those for the primary analysis of PFS except that the initiation of new anti-cancer therapy will be excluded as a mechanism for censoring. The same analysis method as for primary analysis will be used.

Table 3. Efficacy Endpoint Summary Table

Endpoint	Primary Summary and Analysis Method	Sensitivity Analysis
Primary Endpoint (Continued)		
Progression-Free Survival (Continued)		• LTFU/Consent withdrew: The data censoring rules are the same as those for the primary analysis of PFS except that subjects who were lost to follow-up or withdrew consent without PD/death are treated as having an event at the next scheduled assessment time in both treatment arms • Per-Protocol subset: same as primary summary and analysis method for Per-Protocol subset. **This analysis might be performed only if the per-protocol population is less than 90% of the ITT population.** • Missing assessment: The data censoring rules are the same as those for the primary analysis of PFS except that subjects missed at least one disease assessment are treated as a mechanism for censoring at their last evaluation in the Kd arm and are treated as having an event at the next scheduled assessment time in KDd arm. Scheduled assessment dates: same as primary summary and analysis methods, except the analysis is based on the scheduled assessment dates instead of actual assessment dates.
Secondary Endpoints		
Overall Response Rate	Based on IRC assessments: • Point estimate of ORR and 95% CI by treatment group using the Clopper Pearson method. • 1-sided p-value from the Cochren-Mentel-Haenszel chi-squire test controlling for the randomization stratification factors per IXRS. • An estimate of the common odds ratio (95% CI) will be provided as a measure of the relative treatment effect	• Investigator assessments: Same as primary summary and analysis method based on investigator assessments. • **Internal** computational assessments: Same as primary summary and analysis method based on ORCA assessments. • Per-Protocol subset: Same as primary summary and analysis method for Per-Protocol subset. **This analysis might be performed only if the per-protocol population is less than 90% of the ITT population.**

Table 3. Efficacy Endpoint Summary Table

Endpoint	Primary Summary and Analysis Method	Sensitivity Analysis
Secondary Endpoints (Continued)		
MRD[-]CR (as assessed by NGS at a 10^{-5}level)	CR component based on IRC assessments: • Point estimate of MRD[-]CR and 95% CI by treatment group using the Clopper Pearson method. • 1-sided p-value from the Cochren-Mentel-Haenszel chi-squire test controlling for the randomization stratification factors per IXRS. • An estimate of the common odds ratio (95% CI) will be provided as a measure of the relative treatment effect	• **Investigator assessments: same as primary summary and analysis method based on investigator assessments**. • **Internal computational assessments: same as primary summary and analysis method based on ORCA assessments.** • **MRD evaluable population: same as primary summary and analysis method for MRD evaluable population. This analysis might be performed only if the MRD evaluable population is less than 90% of the ITT population.**
Overall Survival	• KM summaries • 1-sided p-value from stratified log-rank test. • Hazard ratio and 95% CI from stratified Cox regression.	• Unstratified analyses: 1-sided p-value from unstratified log-rank test, hazard ratio and 95% CI from unstratified Cox regression. • Per-Protocol subset: same as primary summary and analysis method for Per-Protocol subset. **This analysis might be performed only if the per-protocol population is less than 90% of the ITT population.**

Page 3 of 3

9.5.1 Analyses of Primary Efficacy Endpoint(s)

The distribution of PFS time including median and quartiles will be summarized descriptively using the Kaplan-Meier method. The corresponding 95% confidence intervals for the median and quartiles will be constructed using the method of Klein and Moeschberger (1997) with log-log transformation. PFS rates at selected landmark time points will be provided and the corresponding 95% confidence intervals will be calculated using the method of Kalbfleisch and Prentice (1980). The duration of the follow-up for PFS will be estimated by reverse Kaplan-Meier method (Schemper and Smith 1996).

The inferential comparison between treatment groups will use the log-rank test stratified by the randomization stratification factors per IXRS at level of 0.025 (1-sided). The HR

and its 95% CI will be estimated using a Cox proportional hazards model stratified by the same randomization stratification factors.

The primary endpoint of PFS will be analyzed within each of the subgroups listed in section 4. Specifically, to determine whether the treatment effect is consistent across subgroups**,** the estimate of the hazard ratios (with 95% CI) for PFS between the treatment groups will be provided. Additionally, a treatment-by-subgroup interaction test **may** be provided using a Cox proportional hazards model stratified by the stratification factors.

Piecewise Cox models may be explored given evidence of non-proportional hazards (Collett, 2003). This model will allow estimation of an overall weighted hazard ratio (weights equal to fraction of total events in each interval (Lu & Pajak, 2000)) as well as within interval treatment hazard ratio. Additional analysis may be performed to explore potential sources for non-proportionality by considering baseline prognostic factors and other potential confounding factors.

9.5.2 Analyses of Secondary Efficacy Endpoint(s)

Key secondary endpoints

If PFS is significant, the key secondary endpoints will be tested by sequential testing in the order of ORR, MRD[-]CR **rate** by NGS, and OS.

The analysis for ORR will be done if the primary analysis of PFS reaches statistical significance. **Subsequently, the analysis of MRD[-]CR rate will be done if ORR analysis reaches statistical significance.**

The inferential comparison between treatment groups for both the ORR and MRD[-]CR will be made using the Cochran Mantel Haenszel chi-square test controlling for the randomization stratification factors per IxRS. The ORR/MRD[-]CR will be calculated by treatment group and the associated 95% CI will be estimated using the Clopper-Pearson method. An estimate of the common odds ratio (95% CI) will be provided as a measure of the relative treatment effect. The odds ratio (and 95% CI) will be estimated using the Mantel-Haenszel method. The Kd arm will serve as the reference treatment group in the calculation of the odds ratio. The primary analysis of ORR will be based on IRC assessed outcomes. The analyses based on investigator-assessed and ORCA will serve as supportive analyses. Similarly, the CR portion of the MRD[-]CR endpoint will be based on IRC assessments. **The MRD assessment at month 12 is required for all subjects regardless of response. The subjects, who have progressed, might have**

withdrawn consent or stopped the study for other reasons prior to this time point, will be assessed as MRD positive and the reasons for not having an assessment will be summarized.

For subgroups listed in section 4, odds ratio (with 95% CI) will be provided for ORR and MRD[-]CR rate between the treatment groups. A treatment-by-subgroup interaction test **may** be provided using a logistic regression model stratified by the stratification factors.

Overall survival will be analyzed using the same method as described for the PFS endpoints after PFS, ORR, MRD[-]CR all reach statistical significance. In the case that the PFS results are not statistically significant at the primary PFS analysis, the sponsor may stop the study and if so, the subjects will not be followed for OS any further.

Subgroup analysis for OS will be performed using the same method described for PFS as appropriate. If there is evidence to support non-proportional hazards, a piecewise proportional hazard model may be explored as described for PFS.

Other secondary endpoints

Duration of response (DOR) is defined as the time (in months) from first evidence of PR or better per IMWG-URC to the earlier of disease progression or death due to any cause for subjects with a best response of PR or better. For those who are alive and have not experienced disease progression at the time of data cutoff for analysis, duration of response will be right-censored based on the censoring conventions defined previously for PFS (refer to Table 1). The distribution of DOR, including the median and quartiles and their corresponding 95% CIs, will be characterized using the Kaplan-Meier method based on the subjects who achieve a best response of PR or better. No inferential comparison between treatment arms will be made for duration of response.

Time to next treatment is defined as the time (in months) from randomization to the initiation of subsequent non-protocol anti-cancer treatment for multiple myeloma. Time to next treatment for subjects who do not start the subsequent treatment for multiple myeloma will be censored at the date when the subject's information is last available. Time to next treatment will be summarized descriptively using the Kaplan-Meier method. This analysis will be based on the ITT population.

Time to progression (TTP) is defined as the time (in months) from randomization to documented disease progression. Analysis of TTP will be the same as PFS (see Section 8.4.1.1) except that death will be treated as a censoring event for TTP and will use similar methods as those for the primary PFS analysis based on the ITT population.

For descriptive purposes, time to overall response (defined as the time [in months] from randomization to the earliest of sCR, CR, VGPR, or PR per IMWG-URC) will be summarized using the descriptive statistics for a continuous variable by treatment group for those who achieve a best response of sCR, CR, VGPR, or PR.

The rate of sustained MRD[-]CR (defined as the proportion of subjects that maintain MRD[-]CR for 12 months or more after achieving MRD[-]CR status), the CRR rate (defined as the proportion of best overall response of sCR or CR), and the MRD[-] rate at 12-month will be compared between treatment groups using the similar method for ORR/MRD[-]CR analysis.

The inferential analysis of the endpoint of QLQ-C30 Global Health Status/QOL scale is described in section 9.7.2.

CCI

CCI

[redacted]

9.6 Safety Analyses

9.6.1 Analyses of Primary Safety Endpoint(s)

Safety and tolerability will be assessed, where applicable, by incidence, severity, seriousness, and changes from baseline for all relevant parameters including AEs, deaths, laboratory tests, and vital signs.

All safety analyses will be based on the Safety Analysis Set (see section 6.2).

9.6.2 Adverse Events

The Medical Dictionary for Regulatory Activities (MedDRA) version 20.0 or later will be used to code all events categorized as adverse events to a system organ class and a preferred term.

The subject incidence of adverse events will be summarized for all treatment-emergent adverse events, grade 3 or higher TEAEs, serious adverse events, adverse events leading to withdrawal of investigational product, and fatal adverse events, and adverse events of interest (EOI). In the event that a subject experiences repeated episodes of the same AE, the subject will be counted once within each system organ class and similarly

counted once within each preferred term and the event with the highest severity grade and/or strongest causal relationship to each treatment will be used for purposes of incidence tabulations.

Subject incidence of all treatment-emergent adverse events, grade 3 or higher TEAEs, serious adverse events, adverse events leading to withdrawal of investigational product, and fatal adverse events will be tabulated by system organ class **in alphabetical order** and preferred term in **descending frequency order**.

Subject incidence of events of interest (standardized MedDRA queries and/or Amgen customized queries) will also be summarized according to their categories and preferred term in descending order of frequency. Time to onset and duration of select EOIs may also be summarized.

In addition, summaries of treatment-emergent AEs, grade 3 or higher TEAEs, serious adverse events, adverse events leading to withdrawal of investigational product, and fatal adverse events by preferred term in any treatment arm will be provided in descending order of frequency.

Summaries of treatment-emergent and serious adverse events will be tabulated by system organ class, preferred term, and grade. The fatal adverse events **will also be provided** by system organ class **in alphabetical order** and preferred term in descending order of frequency.

A summary of the number of deaths and the cause of death, classified by deaths within 30 days of last dose of study drug and deaths more than 30 days after the last dose, will be provided.

All AEs, including TEAEs, will be included in individual subject listings.

All on study deaths will be listed.

9.6.3 Laboratory Test Results

Laboratory parameters will be summarized using descriptive statistics, by postdose shifts relative to baseline, and a summary of subject incidence of clinically significant values.

For hematology, chemistry, and other laboratory parameters, the baseline values and changes from baseline values will be summarized descriptively.

For the summary of changes from baseline values, subjects without a baseline and/or post-baseline value will be excluded; values from both scheduled and unscheduled

assessments will be included. Laboratory results from samples taken > 30 days after the last administration of protocol therapy will be excluded from all laboratory summaries.

Laboratory test results will be graded using the NCI CTCAE (Version 4.03). Shifts in laboratory toxicity grades to outside the normal range will be evaluated for selected laboratory parameters (1. hematology analytes in decreasing direction: Hemoglobin, Lymphocyte, Absolute Neutrophil Count, Platelet, WBC; 2. chemistry analytes in increasing direction: Alanine Aminotransferase, Aspartate Aminotransferase, Total Bilirubin, Corrected Calcium, Serum Creatinine, Potassium, Sodium, Magnesium; 3. chemistry analytes in decreasing direction: Albumin, Corrected Calcium, Potassium, Magnesium, Phosphorus, Sodium) by assessing the maximum increase and/or decrease observed during the course of study treatment relative to the baseline toxicity grade.

The subject incidence of Grade 3 and 4 hematological laboratory abnormalities (including neutropenia, thrombocytopenia, and anemia) will be provided by treatment group.

The subject incidence of Grade 3 and 4 nonhematological toxicities (including liver function test [LFT], CrCl) will be provided by treatment group.

9.6.4 Vital Signs

Vital signs including systolic/diastolic blood pressure, pulse, respiratory rate, and temperature will be summarized by changes from baseline values for each treatment group using descriptive statistics.

For the summary of changes from baseline by visit, subjects without a baseline and/or post baseline value will be excluded; values from unscheduled assessments will be excluded. **Subjects with missing data for a scheduled assessment time point will be excluded from the graphical summary for that time point.** Vital sign results taken > 30 days after the last administration of protocol therapy will be excluded from all vital sign summaries.

9.6.5 Electrocardiogram

The electrocardiogram (ECG) measurements from this clinical study were performed as per standard of care for routine safety monitoring, rather than for purposes of assessment of potential QT interval corrected (QTc) effect. Summaries over time and/or changes from baseline over time will be provided by treatment group for all ECG parameters (PR, QRS, QT, QTc, and heart rate).

9.6.6 Antibody Formation

Serum from venous blood samples collected from all subjects in the KdD arm will be assessed for the generation of antibodies to daratumumab (immunogenicity) on day 1 of cycles 1, 7, and 12; as well as both follow-up visits. The number of subjects who test positive for anti-drug antibody will be summarized for each visit. Antibody titers and neutralizing antibodies measurements will be summarized descriptively by visit for these subjects.

9.6.7 Exposure to Investigational Product

Drug exposure (for carfilzomib, dexamethasone, daratumumab, respectively) including duration and intensity will be summarized descriptively for each treatment group.

Descriptive statistics will be produced to describe the exposure to investigational product by treatment group. The number of cycles of protocol-specified therapy administered will be summarized with an additional breakdown of the number of cycles started. In addition, the duration of therapy, the cumulative dose, and the average dose per administration and relative dose intensity will be summarized for each drug. The number and percent of subjects with dose modifications (eg, dose reductions, dose interruptions) and reason for modification will be summarized for both treatment groups.

9.6.8 Exposure to Concomitant Medication

The number and proportion of subjects receiving therapies of interest will be summarized by preferred term for each treatment group as coded by the World Health Organization Drug (WHO DRUG) dictionary. In addition, the number and proportion of subjects receiving anti-myeloma therapies **while on study** will be summarized by WHODRUG preferred term for each treatment group in the Full Analysis Set.

9.7 Other Analyses

9.7.1 Analyses of Pharmacokinetic or Pharmacokinetic/Pharmacodynamic Endpoints

PK samples will be obtained from all subjects to assess the serum concentration of carfilzomib and from all subjects in the KdD arm to assess the serum concentration of daratumumab. Data will be utilized to perform exploratory population PK modeling, but results will not be included in the clinical study report. Details regarding objectives, data handling, and methodology pertaining to any modeling activities will be provided in a separate population modeling analysis plan.

9.7.2 Analyses of Clinical Outcome Assessments

Detailed analyses of clinical outcome assessments are described in the separated Supplemental Statistical Analysis Plan (SSAP) for GHS.

9.7.3 Analyses of Biomarker Endpoints

Pharmacogenomics markers and potential cardiac risk single nucleotide polymorphisms (SNPs) will be explored.

9.7.4 Echocardiogram

Echocardiogram, including LVEF, RVEF, RV Functions and abnormal findings, will be summarized by visit using descriptive statistics for each treatment group based on the Safety Analysis Set. In particular, LVEF, RVEF and RV Functions will also be summarized by actual values and changes from baseline values for each treatment. Additionally, the unscheduled ECHO assessments (as clinically indicated, such as cardiac failure) will be summarized for each treatment group using descriptive statistics, based on the number of adverse events that trigger the unscheduled ECHO assessments.

9.7.5 Pulmonary Function Test

Pulmonary Function Tests, including Spirometry and DLCO, will be summarized by actual values and changes from baseline values by visit using descriptive statistics for each treatment group.

10. Changes From Protocol-specified Analyses

SAP contains some changes in subgroups from Protocol:

- **Added below subgroups:**
  - **prior Bortezomib or Ixazomib exposure (yes vs no)**
  - **refractory to Bortezomib or Ixazomib (yes vs no)**
  - **prior IMiD exposure (yes vs no)**
  - **refractory to IMiD (yes vs no)**
- **Changed risk groups determined by FISH to risk groups determined by genetic abnormality per IMWG**

11. Literature Citations / References

Anderson JR, Cain KC, Gelber RD. Analysis of survival by tumor response, Jon Clinical Oncology J Clin Oncol 1983;1:710-719

Clopper CJ and Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial, Biometrika. 1934; 26(4):404-413.

Collett D. Modelling Survival Data in Medical Research. 2nd edition. London, UK: Chapman & Hall/CRC; 2003.

Dimopoulos MA, Moreau P, Palumbo A, et al. Carfilzomib and dexamethasone versus bortezomib and dexamethasone for patients with relapsed or refractory multiple myeloma (ENDEAVOR): a randomized, phase 3, open-label, multicenter study. Lancet Oncol. 2016;17(1):27-38.

Durie BG, Harousseau JL, Miguel JS, et al. International Myeloma Working Group.

International uniform response criteria for multiple myeloma. Leukemia. 2006; 20(9):1467–73. Erratum in: Leukemia 2006; 20(12):2220. Leukemia 2007; 21(5):1134.

Giobbie-Hurder A, Gelber RD, Regan MM Challenges of guarantee- time bias

J Clin Oncol, 31 (2013), pp. 2963-2969

Greipp PR, San Miguel J, Durie BG, et al. International staging system for multiple myeloma. J Clin Oncol. 2005;23(15):3412–20.

Hryniuk W, Goodyear, M. The calculation of received dose intensity. Journal of Clinical

Oncology 8:1935–1937, 1990

Kalbfleisch, J. D. and Prentice, R. L. The Statistical Analysis of Failure Time Data, New

York: John Wiley & Sons; 1980

Klein, J. P. and Moeschberger, M. L. (1997), Survival Analysis: Techniques for Censored and Truncated Data, New York: Springer-Verlag.Longo D, Duffey P, DeVita V, Wesley M, Hubbard S, Young R. The calculation of actual or received dose intensity: A comparison of published methods. Journal of Clinical Oncology 9:2042–2051, 1991

Kumar S, Paiva B, Anderson KC, et al. International Myeloma Working Group consensus criteria for response and minimal residual disease assessment in multiple myeloma. *Lancet Oncol*. 2016;17(8):328-346

Lu, J., Pajak, T. F. Statistical power for a long-term survival trial with a time-dependent treatment effect. Control Clin Trials. 2000 Dec; 21(6): 561–573.

Mallinckrodt, C.H., Lane, P.W., Schnell, D., Peng Y, and Mancuso, J.P . Recommendations for the primary analysis of continuous endpoints in longitudinal clinical trials. Drug Information Journal, 2008; 42: 303-319.

Mosteller RD. Simplified calculation of body surface area. N Engl J Med 1987;317(17):1098 (letter).

Palumbo A, Avet-Loiseau H, Oliva S, et al. Revised International Staging System for Multiple Myeloma: a report from International Myeloma Working Group, J Clin Oncol, 33 (26) (2015), pp. 2863–2869

Rajkumar SV, et al. Consensus recommendations for the uniform reporting of clinical trials: report of the international myeloma workshop consensus panel 1. Blood 117:4691-4695.2011

Schemper M, Smith TL. A note on quantifying follow-up in studies of failure time.

Controlled Clinical Trials 17:343–346, 1996

Simon R, Makuch RW. A non-parametric graphical representation of the relationship between survival and the occurrence of an event: application to responder versus non-responder bias. Statistics in Medicine 3: 35-44, 1984

US Food & Drug Administration: Guidance for Industry. Clinical Trial Endpoints for the

Approval of Cancer Drugs and Biologics. www.fda.gov/cder/guidance/7478fnl.htm

Specifications for Onyx Response Computational Assessment (ORCA) Based on IMWG Uniform Response Criteria – Implementation for Study 20160275 (CANDOR)

12. Appendices

Appendix A. Technical Detail and Supplemental Information Regarding Statistical Procedures and Programs

A1. The following data will be imputed using the following algorithm:

Adverse Events

Concomitant Medications **(other than anti-cancer therapy)**

Table 4. Imputation Rules for Partial or Missing Start Dates

Start Date		Stop Date: Complete: *yyyymmdd*		Stop Date: Partial: *yyyymm*		Stop Date: Partial: *yyyy*		Stop Date: missing
		< 1st dose	**≥ 1st dose**	**< 1st dose *yyyymm***	**≥ 1st dose *yyyymm***	**< 1st dose *yyyy***	**≥ 1st dose *yyyy***	
Partial: *yyyymm*	= **1st dose** *yyyymm*	2	1	2	1	n/a	1	1
	≠ 1st dose *yyyymm*		2		2	2	2	2
Partial: *yyyy*	**= 1st dose *yyyy***	**3**	1	3	1	n/a	1	1
	≠ 1st dose *yyyy*		3		3	3	3	3
Missing		4	1	4	1	4	1	1

- 1 = Impute the date of first dose
- 2 = Impute the first day of the month
- 3 = Impute January 1 of the year
- 4 = Impute January 1 of the stop year

Note: If the start date imputation leads to a start date that is after the stop date, then do not impute the start date.

Imputation rules for partial or missing stop dates:

For partial stop date mmyyyy, impute the last day of the month.

For partial stop date yyyy, impute December 31 of the year.

For completely missing stop date, do not impute.

If the stop date imputation leads to a stop date that is after the death date, then impute the stop date as the death date.

If the stop date imputation leads to a stop date that is before the start date, then there is a data error and do not impute the stop date. (ie. set the stop date as missing).

A2. The anti-cancer therapy date will be imputed using the following algorithm:

If the start day of new anti-cancer therapy is missing and month and year are not the same as last dosing date of study treatment, it will be assumed to be the first day of the month. If the start day of new anti-cancer therapy is missing and month and year are same as last dosing date of study treatment, and the patient does not have protocol deviation of using excluded procedure while on study, the start date will be assumed as last dosing date of study treatment plus 1 day. In other situations, do not impute.

Appendix B. Code Fragment

```
CC
```

Appendix C. Reference Values/Toxicity Grades

Laboratory Values

Safety laboratory values below a distinct limit (eg. detection limit, documented as "< [limit]") will be substituted by half of the limit and values above a distinct limit (documented as "> [limit]") will be substituted by the limit itself for all analyses.

A Grade will be assigned to each laboratory based on CTCAE version 4.0 [v4.03: June 14, 2010], as detailed in **Table 5**. Depending on the toxicity definition, the same result may be assigned to two grading for deviations towards higher or lower values. Values not meeting any of the criteria will be assigned a grade 0.

Table 5. Grading of Select Laboratory Parameters

Laboratory Abnormality [Unit]	Grade 1	Grade 2	Grade 3	Grade 4
Decreased Hemoglobin [g/L]	100 - < LLN	80 - < 100	< 80	not defined
Decreased Lymphocytes [G/L]	0.8 - < LLN	0.5 - < 0.8	0.2 - < 0.5	< 0.2
Decreased Neutrophils [G/L]	1.5 - < LLN	1.0 - < 1.5	0.5 - < 1.0	< 0.5
Decreased Platelets [G/L]	75 - < LLN	50 - < 75	25 - < 50	< 25
Decreased WBC [G/L]	3.0 - < LLN	2.0 - < 3.0	1.0 - < 2.0	< 1.0
Increased AST	> ULN – 3×ULN	> 3×ULN – 5×ULN	> 5×ULN – 20×ULN	> 20×ULN
Increased ALT	> ULN – 3×ULN	> 3×ULN – 5×ULN	> 5×ULN – 20×ULN	> 20×ULN
Increased Total Bilirubin	> ULN – 1.5×ULN	> 1.5×ULN – 3×ULN	> 3×ULN – 10×ULN	> 10×ULN
Increased Corrected Calcium [mmol/L]	> ULN - 2.9	> 2.9 – 3.1	> 3.1 – 3.4	> 3.4
Increased Creatinine	> 1 - 1.5×BL or > ULN - 1.5×ULN	> 1.5 - 3.0×BL or > 1.5 - 3.0×ULN	> 3.0×BL or > 3.0-6.0× ULN	> 6.0×ULN
Increased Potassium [mmol/L]	5.5 - > ULN	> 5.5 – 6.0	> 6.0 – 7.0	> 7.0

Page 1 of 2

Footnotes defined on next page of table

Table 5. Grading of Select Laboratory Parameters

Laboratory Abnormality [Unit]	Grade 1	Grade 2	Grade 3	Grade 4
Increased Sodium[mmol/L]	> ULN – 150	> 150 – 155	> 155 – 160	> 160
Increased Magnesium [mmol/L]	> ULN – 1.23	not defined	> 1.23 – 3.30	> 3.30
Increased Uric Acid [mmol/L]^	> ULN – 0.59 without physiologic consequences	not defined	> ULN – 0.59 with physiologic consequences	> 0.59
Decreased Albumin [g/L]	30 – < LLN	20 – < 30	< 20	not defined
Decreased Corrected Calcium [mmol/L]	2.0 – < LLN	1.75 – < 2.0	1.5 – < 1.75	< 1.5
Decreased Potassium [mmol/L]*	3.0 – < LLN	3.0 – < LLN; symptomatic; intervention indicated	2.5 – < 3.0	< 2.5
Decreased Magnesium [mmol/L]	0.5 – < LLN	0.4 – < 0.5	0.3 – < 0.4	< 0.3
Decreased Phosphorus [mmol/L]	0.8 – < LLN	0.6 – <0.8	0.3 – < 0.6	< 0.3
Decreased Sodium [mmol/L]	130 – < LLN	not defined	120 – < 130	< 120

Page 2 of 2

BL: baseline value, LLN: Lower limit of normal, ULN: Upper limit of normal
Clinical criteria from CTCAE 4.0 grading were not considered in order to assign grades unless specified otherwise
*: Details will be recorded on Potassium CTC Details CRF for central lab values and Chemistry (Local Lab) CRF for local lab values.
^: Details will be recorded on Uric Acid CTC Details CRF for central lab values and Chemistry (Local Lab) CRF for local lab values.

Appendix D. Clinical Outcome Assessment Forms/Instruments

The following sections describe the scoring algorithms used for both the QLQ-C30 and QLQ-MY20 questionnaires. Scoring procedures are similar for both questionnaires and can be found in the EORTC QLQ-C30 Scoring Manual, ver. 3 (Fayers et al. 2001) and Cocks et al (2007). All scale scores range from 0 to 100.

QLQ-C30 Scoring

For all scales, calculate the raw score (RS) of a scale using the mean of the item scores in the scale as follows:

$$RS = (S1 + S2 + \ldots + Sn) / n$$

where Si: i=1, …, n, are the item scores and n is the number of items with valid scores, assuming the number of items with valid scores meets the minimum requirement as specified in Table 6 or this scale score will be assumed missing.

Use a linear transformation to standardize the raw score in order that scores will range from 0-100:

Global Health Status/QOL = {(RS-1)/range} * 100

Functional Scales = {1- (RS-1)/range} * 100

Symptom Scales = {(RS-1)/range} * 100

where range for each scale is defined in **Table 6**.

For the Global Health Status/QOL scale and functional scales in QLQ-C30 a higher score represents a better health state and for the symptom scores in QLQ-C30 a lower score represents a better health state.

QLQ-MY20 Scoring

For all scales, calculate the raw score (RS) of a scale using the mean of the item scores in the scale as follows:

$$RS = (S1 + S2 + \ldots + Sn) / n$$

where Si: i=1, …, n, are the item scores and n is the number of items with valid scores, assuming the number of items with valid scores meets the minimum requirement as specified in Table 6 or this scale score will be assumed missing.

Use a linear transformation to standardize the raw score in order that scores will range from 0-100:

Disease Symptom Scale (DS) = {(RS – 1) / range} * 100

Side Effects of Treatment Scale (SE) = {(RS – 1) / range} * 100

Future Perspective Scale (FP) = {1 - (RS – 1) / range} * 100

Body Image (BI) = {1 - (RS – 1) / range} * 100

where range for each scale is defined in **Table 6**.

Table 6. QLQ-C30 and QLQ-MY20 Scales and Scoring Details

	Number of Items	Item Range [a]	Item Numbers	Minimum Not Missing
QLQ-C30				
Global Health status/QOL	2	6	29,30	1
Functional Scales				
Physical Functioning	5	3	1-5	3
Role Functioning	2	3	6,7	1
Emotional Functioning	4	3	21-24	2
Cognitive Functioning	2	3	20,25	1
Social Functioning	2	3	26,27	1
Symptom Scales / Items				
Fatigue	3	3	10,12,18	2
Nausea/vomiting	2	3	14,15	1
Pain	2	3	9,19	1
Dyspnoea	1	3	8	N/A
Insomnia	1	3	11	N/A
Appetite Loss	1	3	13	N/A
Constipation	1	3	16	N/A
Diarrhoea	1	3	17	N/A
Financial Difficulties	1	3	28	N/A
QLQ-MY20				
Symptom Scales				
Disease Symptoms	6	3	31-36	3
Side Effects of Treatment	10	3	37-46	5
Functional Scales/Items				
Future Perspective	3	3	48-50	1
Body Image	1	3	47	N/A

[a] Range is the difference between the maximum possible value of the Raw Score and the minimum possible value.

For the Disease Symptoms and the Side Effects of Treatment scales a high score represents a high level of symptomatology / problems (symptom scale). For Body Image and Future Perspective scales a high score represents a high level of functioning (functional scale), ie, a better future perspective or body image.