Statistical Analysis Plan

Protocol Title:	A Randomized, Open-label, Phase 3 Study Comparing Once-weekly vs Twice-weekly Carfilzomib in Combination with Lenalidomide and Dexamethasone in Subjects with Relapsed or Refractory Multiple Myeloma (A.R.R.O.W.2)
Short Protocol Title:	A Study Comparing Once-weekly vs Twice-weekly Carfilzomib in Combination with Lenalidomide and Dexamethasone in Subjects with Relapsed or Refractory Multiple Myeloma
Protocol Number:	20180015
NCT Number:	03859427
Authors:	[redacted], Ph.D., MS Biostatistics Senior Manager, GBS, Amgen Inc. [redacted], Ph.D. Biostatistics Director, GBS, Amgen Inc.
Sponsor:	Amgen Inc. One Amgen Center Drive, Thousand Oaks, CA 91320, United States
SAP Date:	**Document Version** — **Date** Original (v 1.0) — 21 December 2018 Amendment 1 (v 2.0) — 09 December 2021 **Amendment 2 (v 3.0)** — **22 May 2023**

Version Number	Date (DDMMMYYYY)	Summary of Changes, including rationale for changes
Original (v1.0)	21 December 2018	
Amendment 1 (v2.0)	09 December 2021	1. Section 2.1 Objectives and Endpoints: • changed the wording for some endpoints: 'overall response rate' to 'overall response', "1-year PFS" to "PFS over the duration of the study", 'subject incidence' to 'incidence', 'MRD[-]CR rate' to 'MRD[-]CR', 'MRD[-] rate' to 'MRD[-] status', changed the time window for MRD[-] status from ± 2 weeks to ± 4 weeks per protocol amendment • added OS as the secondary endpoint per protocol amendment • [redacted] 2. Section 2.2 Hypotheses and/or Estimations: • updated the language to distinguish the hypothesis for overall response/PFS and patient-reported convenience 3. Section 3.1 Study Design: • added follow-up for survival • added "Following the safety follow up visit, all subjects with confirmed PD before 12 months from randomization will be followed for survival every 28 ± 7 days until 12 months after randomization, death, loss to follow-up, or withdrawal of full consent, whichever comes first." • Changed "the first 50% of the subjects have been randomized and had a best overall response (BOR) assessed" to "the first 230 subjects (50% of the planned total 460 subjects) have been randomized and had the opportunity to be followed for a best overall response (BOR) assessment" for interim analysis plan. This update is also applied to Section 3.2, 6.5, and 7.1 • added DMC review for Japan patients • updated Figure 3-1. Study Schema (adding long-term follow-up) per protocol 4. Section 3.2 Sample Size: • added the software used in calculation 5. Section 4.1 Planned Covariates: • added "in the stratified analyses" for the stratification factors for randomization 6. Section 4.2 Subgroups:

		• specified that the subgroups will be defined based on the data reported on CRF • added subgroup “prior bortezomib treatment (yes vs no)” • added subgroup “baseline creatinine clearance (<50, ≥50 mL/min)” 7. Section 5 Definitions: • added definition for baseline bone lesion and plasmacytoma assessment • moved actual cumulative dose of study treatment to relative dose intensity section • updated description of BSA calculation in carfilzomib dose calculation • removed End of Study Treatment Date • updated the EOS date for the individual subject to clarify the context • removed the list of CRFs for deriving last known alive date, which will be documented in a separate document • defined MRD[-]CR rate instead of MRD[-]CR; • changed the time window for MRD[-] status from ± 2 weeks to ± 4 weeks • added definition for OS • added the censoring situation: (5) lost to follow-up or withdrawn consent • updated the calculation for number of weeks of actual treatment for dexamethasone and lenalidomide • updated Table 5-3 Planned Dexamethasone Dose Schedule • updated the planned dose intensity calculation for dexamethasone • updated the calculation for the number of weeks of actual treatment in lenalidomide actual dose intensity • updated the description for Study Day • updated the wording for censoring in TTP • updated the description for TEAE to align with DES 8. Section 6.3 Per Protocol Set: • updated wording for the inclusion criteria 103, 106, 107 and exclusion criteria 214, 221 and 222 per protocol; added exclusion criteria 206, 215 and 244 • added corresponding IPD numbers to Major treatment non-compliance 9. Section 6.4 PK/PDn Analyses Set(s):

		• updated the description 10. Section 6.5 Interim Analysis Set: • removed the description for independent safety data review 11. Add Section 6.6 Modified Full Analysis Set(mFAS) 12. Section 7.1 Interim Analysis and Early Stopping Guidelines & Section 7.3 Final Analysis: • added the description for database snapshot • removed the corresponding RR stopping boundary • removed the corresponding RR stopping boundary 13. Section 9.2 Subject Accountability: • added the summary for COVID-19 impact • added the summary for subjects who completed the 12 cycles treatment for each study drug 14. Section 9.3 Important Protocol Deviations: • added the summary for COVID-19 impact 15. Section 9.4 Demographic and Baseline Characteristics: • added albumin (g/dL) • added the summary for analysis population and listing of randomization 16. Section 9.5 Efficacy Analyses Table 9-1 and Table 9-2: • changed the wording for some endpoints: 'overall response rate' to 'overall response', "1-year PFS" to "PFS over the duration of the study", 'MRD[-]CR rate' to 'MRD[-]CR', 'MRD[-] rate' to 'MRD[-] status' • added OS • added mFAS in sensitivity analysis for ORR, PFS and Patient-reported convenience • added a sensitivity analysis based on Per Protocol Set for Physical functioning and role functioning • added "The testing is done once at the primary (final) analysis on ITT analysis set" to key secondary endpoints

		• removed reporting P-value for MRD[-]CR and MRD[-] status 17. Section 9.5.1 Analyses of Primary Efficacy Endpoint(s) for ORR: • changed U_h from "KRd 27mg/m2 BIW vs Rd" to "Rd vs KRd 27mg/m2 BIW" and log (1.325) to log (0.755) to align with sample size section • updated the expression for H_0 and H_a to align with the U_h change. • added SAS code for calculating 1-sided p-value • updated the description for constancy assumption check • added sensitivity analysis based on mFAS 18. Section 9.5.2 Analyses of Secondary Efficacy Endpoint(s): For PFS: • added H_0 and H_a for PFS endpoint. • changed U_h from "KRd 27mg/m2 BIW vs Rd" to "Rd vs KRd 27mg/m2 BIW" and log (0.552) to log (1.812) • added SAS code for calculating 1-sided p-value • added "The subcategory of death with the primary reason of COVID-19 infection or COVID-19 pneumonia will be included in the PFS events" to the PFS summary • added sensitivity analysis based on mFAS For Patient-reported convenience: • specified that the descriptive summary will be based on all randomized subjects as well as all expected subjects at the scheduled visit • changed the Safety Analysis Set to ITT Analysis Set • added sensitivity analysis based on mFAS For Physical functioning and role functioning (EORTC QLQ-C30) over time: • changed Health-related Quality-of-Life Analysis Set to ITT Analysis Set • added descriptive summary for the scale scores, change from baseline, and the completion rate • added the multiple imputation for missing data for Physical functioning and role functioning scale scores

		• added a sensitivity analysis based on Per Protocol Set • removed reporting p-value For MRD[-]CR, MRD[-] status: • added sensitivity analysis based on ORCA assessed response • removed reporting p-value Added analysis for OS 19. Section 9.6.2 Adverse Events: • removed the listings for AE and death 20. Section 9.6.3 Laboratory Test Results: • added the summary for ALT/AST/Total Bilirubin and potential Hy's law cases 21. Section 9.6.8 Exposure to Investigational Product and Non-Investigational Products: • separated the summary of Dose Change/Withheld to Dose missed and Dose reduction • added the number (%) with COVID-19 control measures for reason of dose modification 22. Appendix A. Handling of Incomplete Dates and Missing Dates: • added imputation rules for new anti-myeloma therapy start date • added imputation rules for prior multiple myeloma therapy and relapse/progression to prior multiple myeloma therapy • updated the imputation rules for death date per AMGEN GBS Oncology Endpoint Guide v1.0
Amendment 2 (v3.0)	**22 May 2023**	**1. Section 5. Definitions:** • **modified the language for Cytogenetic risk group** • **added definition for patient-reported convenience** **2. Section 9.2. Subject Accountability** • **for key study dates, added last subject last dose of IP, last subject end of study** **3. Section 9.3. Important Protocol Deviations** • **added subject listing for inclusion/exclusion criteria deviations**

		4. Section 9.4. demographic and baseline disease characteristics: • **added PI refractory status, Anti-CD38 refractory status, Refractory to the last prior line of therapy, Immunoglobulin heavy and light chain types** **5. Section 9.5.1. Analyses of Primary Efficacy Endpoint(s):** • **added the non-inferiority margin for ORR** **6. Section 9.5.2. Analyses of Secondary Efficacy Endpoint(s):** • **added additional sensitivity analyses** • **added the analysis of restricted mean survival time (RMST) for PFS** • **added "the observed treatment effect retention rate will be reported" for PFS analysis** • **added subgroup analysis for PFS and OS** • **added assessments for the adequacy of proportional hazard assumption and piecewise Cox models for OS** • **specified that the p-value of the hypothesis test for patient-reported convenience will be reported after Cycle 4** • **added safety follow-up visit in the comparison analysis for patient-reported convenience, and QLQ-C30** • **for the model of CTSQ analysis, changed the independent variable "baseline score" to "scale score measured at Cycle 2 Day 1 visit"** **7. Section 10. Changes from Protocol-specified Analysis:** • **specified that the RMST and observed treatment effect retention rate are included as additional analyses for PFS; the non-inferiority margin for ORR is specified in Section 9.5**

Table of Contents

List of Table

List of Figures

List of Abbreviations and Definition of Terms

Abbreviation or Term	Definition/Explanation
AE(s)	adverse event(s)
ALT	alanine aminotransferase
ANC	absolute neutrophil count
ANCOVA	analysis of covariance
AST	aspartate aminotransferase
AUC	area under the curve
BIW	twice-weekly
BOR	best overall response
BSA	body surface area
CI	confidence interval
Cmax	maximum plasma concentration
CMH	Cochran-Mantel-Haenszel
COA	clinical outcome assessment
CR	complete response
CrCl	creatinine clearance
CRF	case report form
CSR	clinical study report
CTCAE	Common Terminology Criteria for Adverse Events
CTSQ	Cancer Therapy Satisfaction Questionnaire
DMC	data monitoring committee
DOR	duration of response
ECG	electrocardiogram
ECOG PS	Eastern Cooperative Oncology Group Performance Status
EOI	event of interest
EORTC	European Organization for Research and Treatment of Cancer
EORTC QLQ-C30	European Organization for Research and Treatment of Cancer Quality-of-life Questionnaire Core 30
[redacted]	

Abbreviation or Term	Definition/Explanation
FDA	Food and Drug Administration
FISH	fluorescence in situ hybridization
HR	hazard ratio
HRQOL	health-related quality-of-life
IMWG-URC	International Myeloma Working Group Uniform Response Criteria
Interactive Voice/Web Response System (IxRS)	telecommunication/web-based technology that is linked to a central computer in real time as an interface to collect and process information
IPD	important protocol deviation
IRC	Independent Review Committee
ISS	International Staging System
ITT	intent-to-treat
IV	intravenous
KM	Kaplan-Meier
KRd	carfilzomib in combination with lenalidomide and dexamethasone
LFT	liver function test
LVEF	left ventricular ejection fraction
MedDRA	Medical Dictionary for Regulatory Activities
mFAS	modified full analysis set
MI	multiple imputation
MM	multiple myeloma
MMRM	mixed model for repeated measures
MRD	minimal residual disease
MRD[-]	minimal residual disease negative
NCI	National Cancer Institute
NGS	next-generation sequencing
OR	odds ratio
ORCA	Onyx response computer algorithm
ORR	overall response rate
OS	overall survival
PD	progressive disease
PDn	pharmacodynamics
PFS	progression-free survival

Abbreviation or Term	Definition/Explanation
PH	proportional hazard
PI	proteasome inhibitor
PK	pharmacokinetics
PR	partial response
PT	preferred term
QW	once-weekly
Rd	lenalidomide with dexamethasone
RMST	**restricted mean survival time**
RR	relative risk
RRMM	relapsed or refractory multiple myeloma
SAP	statistical analysis plan
sCR	stringent complete response
STD	standard deviation
[redacted]	
SSAP	supplemental statistical analysis plan
SWT	Satisfaction with Therapy
TEAE(s)	treatment-emergent adverse event(s)
TTP	time to progression
TTR	time to response
VGPR	very good partial response

1. Introduction

The purpose of this Statistical Analysis Plan (SAP) is to provide details of the statistical analyses that have been outlined within the superseding protocol amendment 3 for study 20180015, Carfilzomib, dated 02 September 2021. The scope of this plan includes the interim futility analysis and the primary analysis (the final analysis) that are planned and will be executed by the Amgen Global Biostatistical Science department unless otherwise specified.

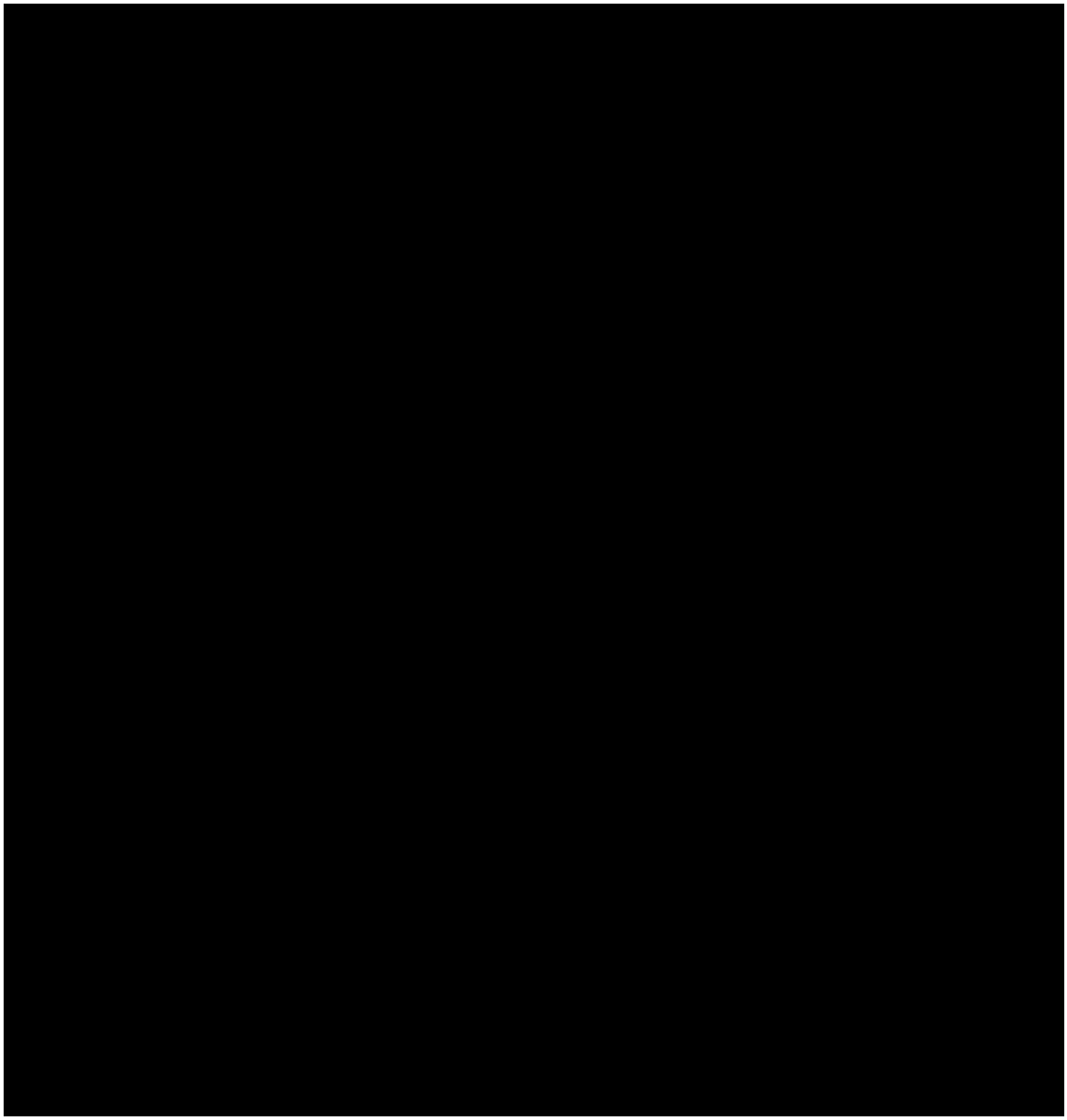
2. Objectives, Endpoints and Hypotheses

2.1 Objectives and Endpoints

Objectives	Endpoints
Primary	
• compare efficacy of 56 mg/m^2 carfilzomib administered once-weekly (QW) in combination with lenalidomide and dexamethasone (KRd 56 mg/m^2) to 27 mg/m^2 carfilzomib administered twice-weekly (BIW) in combination with lenalidomide and dexamethasone (KRd 27 mg/m^2) in subjects with RRMM with 1 to 3 prior lines of therapy	• overall response (defined as the best overall response of stringent complete response [sCR], complete response [CR], very good partial response [VGPR], and partial response [PR] per International Myeloma Working Group Uniform Response Criteria [IMWG-URC]) over the duration of the study
Key Secondary	
• compare progression-free survival (PFS) between treatment arms	• PFS over the duration of the study
• compare patient-reported convenience with carfilzomib-dosing schedule between treatment arms	• convenience as measured by the Patient-reported Convenience with Carfilzomib-dosing Schedule Question after cycle 4 of treatment
Secondary	
• describe safety and tolerability in treatment arms	• incidence of treatment-emergent adverse events
• compare additional efficacy parameters between treatment arms	• time to response (TTR) • duration of response (DOR) • time to progression (TTP)
• compare overall survival (OS) between treatment arms	• OS over the duration of the study
• compare rate of minimal residual disease negative (MRD[-]) in bone	• MRD[-]CR, defined as achievement of CR or better by Independent Review Committee (IRC) per IMWG-URC and

marrow aspirates between treatment arms	achievement of MRD negativity as assessed by next-generation sequencing (NGS) method at a 10^{-5} threshold over the duration of the study • MRD[-] status at 12 months, defined as achievement of MRD negativity at 12 months (± 4 weeks) from randomization, as assessed by NGS method at a 10^{-5} threshold
• compare patient-reported physical functioning and role functioning between treatment arms	• physical functioning and role functioning over time as measured by the Physical Functioning and Role Functioning scales of the European Organization for Research and Treatment of Cancer Quality-of-life Questionnaire Core 30 (EORTC QLQ-C30) over the duration of the study
• compare patient-reported treatment satisfaction between treatment arms	• treatment satisfaction as measured by the Satisfaction with Therapy (SWT) scale of the Cancer Therapy Satisfaction Questionnaire (CTSQ) after cycle 4 of treatment

Exploratory
[redacted]



2.2 Hypotheses and/or Estimations

KRd 56 mg/m^2 QW is non-inferior in terms of overall response and PFS, and is superior in terms of patient-reported convenience when compared with KRd 27 mg/m^2 BIW.

The hypotheses for the primary and key secondary objectives (ORR, PFS, and convenience after cycle 4 of treatment) 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 RRMM who have received 1 to 3 prior therapies.

Subjects will be randomized in a 1:1 ratio to 1 of 2 arms:

Arm 1: KRd using carfilzomib 56 mg/m^2 QW
Arm 2: KRd using carfilzomib 27 mg/m^2 BIW

Randomization will be performed using an interactive voice/web response system (IxRS) and subjects will be stratified based on the following criteria: original International Staging System (ISS) stage at study entry (stage 1 or 2 vs stage 3); prior lenalidomide treatment (yes vs no), prior proteasome inhibitor (PI) treatment (yes vs no), prior anti-CD38 exposure (yes vs no).

Subjects will receive the study drug(s) determined by randomization for a maximum of 12 cycles. No crossover between the treatment arms is allowed. After completion or discontinuation of all study drug(s), subjects will have a safety follow-up visit 30 (+3) days after the last dose of all study drug(s).

All subjects will be assessed for multiple myeloma disease response and disease progression by investigator and a blinded Independent Review Committee (IRC) according to the International Myeloma Working Group-Uniform Response Criteria (IMWG-URC) (Kumar et al, 2016; Rajkumar et al, 2011; Durie et al, 2006) using central laboratory test results every 28 ± 7 days from cycle 1 day 1 through the end of cycle 12 or disease progression until death, loss to follow-up, withdrawal of full consent, or first subsequent antimyeloma treatment (whichever occurs first), regardless of treatment cycle duration, dose delays or treatment discontinuation. The disease assessment schedule is independent of treatment schedules.

Following the safety follow-up visit, subjects who do not have confirmed progressive disease (PD) before 12 months from randomization are required to continue follow-up every 28 ± 7 days for survival, disease response assessments and report new antimyeloma treatment until 12 months after randomization, first subsequent antimyeloma treatment, death, loss to follow-up, withdrawal of full consent, or confirmed PD, whichever occurs first.

Following the safety follow up visit, all subjects with confirmed PD before 12 months from randomization will be followed for survival every 28 ± 7 days until 12 months after

randomization, death, loss to follow-up, or withdrawal of full consent, whichever comes first.

The response based on IRC assessment will be used for the primary analysis of efficacy endpoints.

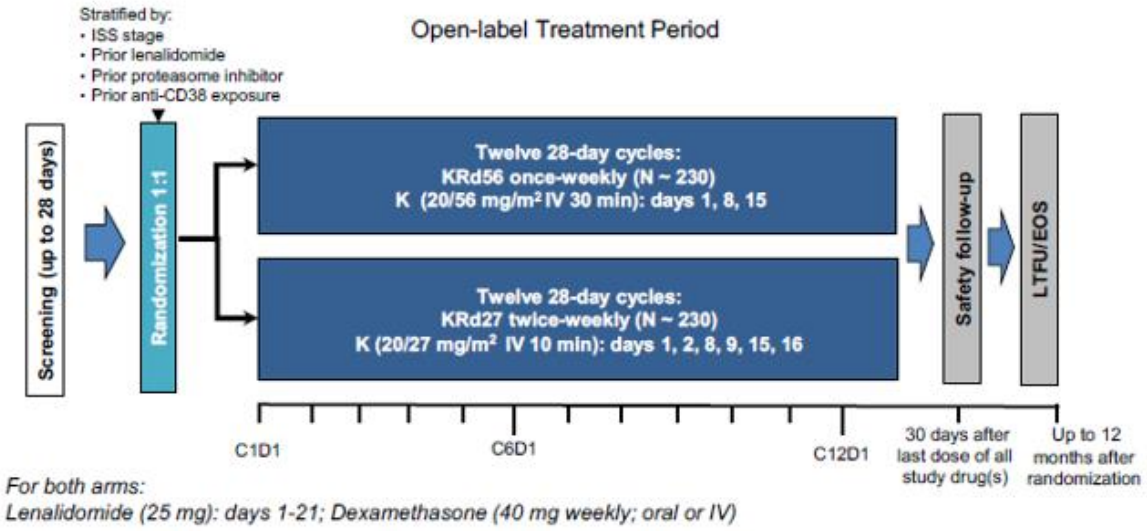
The independent data monitoring committee (DMC) will review safety data on a regular basis (approximately every 6 months) and review the efficacy data once for the interim futility analysis to provide recommendations relating to continuing, modifying, or stopping the study. The first planned DMC meeting for data review will take place when approximately 30 subjects (15 in each arm) have completed at least 1 cycle of treatment.

The primary purpose of the interim futility analysis is to assess the futility in terms of ORR. The interim futility analysis is planned to occur when the first 230 subjects (50% of the planned total 460 subjects) have been randomized and had the opportunity to be followed for a best overall response (BOR) assessment by the date when treatment was completed, confirmed PD or death occurred, subject was lost to follow-up, withdrew consent, or started new antimyeloma therapy, whichever occurred first.

The DMC will also perform a review per Japan-specific requirements to evaluate the tolerability of KRd 20/56 mg/m^2 QW on the first 3 to 6 Japanese subjects randomized to the KRd 20/56 mg/m^2 QW arm and received at least 1 cycle of study treatment. Details of the DMC's responsibilities will be described in the DMC Charter.

The overall study design is outlined in the study schema in Figure 3-1. The endpoints are defined in Section 2.1.

Figure 3-1. Study Schema



CxDx = cycle X day X; d = dexamethasone; EOS = End of Study; IV = intravenously; K = Kyprolis (carfilzomib); LTFU = long-term follow-up; R = lenalidomide.

3.2 Sample Size

The sample size was determined so that the primary objective could be tested via synthesis method at 1-sided 2.5% significance level with 80% power, including an interim futility analysis when the first 230 subjects (50% of the planned total 460 subjects) have been randomized and had the opportunity to be followed for the best overall response assessment by the date when treatment was completed, confirmed disease progression/death occurred, subject was lost to follow-up, withdrew consent, or started new antimyeloma therapy, whichever occurred first.

A sample size of approximately 460 subjects is needed to achieve 80% power for demonstrating that KRd 56 mg/m^2 QW preserves at least 60% of KRd 27 mg/m^2 BIW effect in terms of ORR at a 1-sided 2.5% significance level by the synthesis method. The stratified relative risk (RR) of ORR during the first 12 cycles of treatment in ASPIRE study (BIW Rd vs KRd 27 mg/m^2 RR and 95% CI: 0.755 [0.696, 0.818]) will be used as the historical reference for the test of non-inferiority. This calculation assumes a true RR of 1 with ORR = 86.6% for both arms (KRd 56 mg/m^2 QW vs KRd 27 mg/m^2 BIW), and an interim analysis for futility at 50% information fraction using an O'Brien-Fleming type beta-spending function. The reference ORR = 86.6% was determined based on the BOR observed by the end of 12 cycles of treatment in ASPIRE study (Amgen data on file).

The statistical software EAST 6 (version 6.4.1) and R (version 3.6.1) were used in the calculation for sample size determination.

4. Covariates and Subgroups

4.1 Planned Covariates

The planned covariates to be used for the primary analysis of the primary endpoint and the selected secondary endpoints in the stratified analyses are the stratification factors for randomization per IxRS: original ISS stage at study entry (stage 1 or 2 vs stage 3), prior lenalidomide treatment (yes vs no), prior PI treatment (yes vs no), prior anti-CD38 exposure (yes vs no).

4.2 Subgroups

The ORR, **PFS and OS** will be estimated for the following selected subgroups defined by the baseline data reported on CRF **or from the central lab**, as appropriate. When there is not a sufficient number of subjects in the subgroup, i.e., less than 5% of the whole population, relevant subgroups may be combined.

- original ISS stage at baseline (stage 1 or 2 vs stage 3)
- prior lenalidomide treatment (yes vs no)
- prior PI treatment (yes vs no)
- prior anti-CD38 exposure (yes vs no)
- prior bortezomib treatment (yes vs no)
- age (years) (<65, ≥65; 18 - <65, 65 - <75, ≥75)
- region (Europe vs Non-Europe)
- baseline creatinine clearance (CrCl, mL/min) (<50, ≥50)
- number of prior therapies (1 vs >1; 1 vs 2 vs >2)
- cytogenetic risk measured by FISH (high-risk (t(4;14), t(14;16), deletion 17p) vs standard risk)
- bortezomib refractory status (yes vs no)
- lenalidomide refractory status (yes vs no)
- prior transplant (yes vs no)

5. Definitions

Baseline

The baseline value is the latest value measured on/before day 1 of the first dose of any protocol-specified therapy. If a subject doesn't receive any protocol-specified therapy, then the latest value prior to or on randomization date will be used.

For bone lesion and plasmacytoma assessment, the baseline is defined as the most recent assessment within 45 days (bone lesion) or 28 days (plasmacytoma) prior to or on the randomization date, or up to 7 days (inclusive) after initiation of the study treatment if the patient has been dosed.

Best overall response by investigator assessment

Best overall response for a subject by investigator assessment is the best post baseline confirmed response by the analysis trigger date based on the responses by visit collected on myeloma response assessment CRF. The response assessments done after confirmed disease progression or initiation of new anti-myeloma therapy will be excluded from the analysis of primary endpoint.

Best overall response by IRC assessment

Best overall response for a subject by IRC as collected on IRC Evaluation CRF. Details of the IRC will be described in the IRC charter.

Cytogenetic Risk group as determined by Fluorescent in Situ Hybridization (FISH)

Cytogenetic risk group is defined based on the central laboratory FISH **analytes t(4;14), t(14;16) and deletion 17p regardless of any other FISH analyte test results.**
High risk group: Subjects who have **abnormal results in the tests of analytes** t(4;14) or t(14;16), and/or deletion 17p.
Standard risk group: Subjects who have **normal results in the tests of all the three analytes** t(4;14), t(14;16) and deletion 17p.
Missing: Subjects who cannot be identified as high **or** standard risk.

Death Date

Death date **for a subject is defined as the date** recorded on the End of Study CRF page **where the primary reason for ending the study is Death or the date on the Survival Status CRF page where the subject status is Dead**. Incomplete death dates will be imputed using the imputation rules as presented in Appendix A. The imputed death date will be used in calculation of duration of response, progression-free survival and overall survival.

Duration of Response (DOR)

For subjects with a PR or better, i.e., sCR, CR, VGPR, or PR, the DOR is defined as the time (months) from the earliest date when a PR or better is first achieved, and subsequently confirmed, to the earliest date of confirmed PD or death due to any cause.

For those who are alive and have not experienced PD by analysis time, DOR will be censored based on the same censoring rules for PFS as listed in Table 5-1 if applicable.

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

Duration of Study Treatment

Duration of treatment with carfilzomib, lenalidomide and dexamethasone will be defined as the time from the first start date of each drug to the last stop date of **each** drug. Duration of the whole study treatment will be from the earliest start date among the three study drugs to the latest stop date among the three study drugs.

Duration (week) = (last dose date of the drug – first dose date of the drug + 1) / 7

End of Study Date

For an individual subject, the end of study date is the date of withdrawal **of** full consent from the study, lost to follow-up, completeness of the final safety follow-up visit, or completeness of final long-term follow-up visit (whichever is later), decision by sponsor, or death. The end of study date will be recorded on the End of Study CRF page.

For the overall study, the end of study date is defined as the date when the last subject across all sites is assessed or receives an intervention for evaluation in the study (i.e., last subject last visit), following any additional parts in the study (e.g., long-term follow-up), as applicable.

First Dose Date of Study Treatment

It is the date on which a subject is administered the first dose of any study drug.

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, according to the criteria published by the International Myeloma Working Group (Greipp 2005):

Stage 1: Serum beta-2 microglobulin < 3.5 mg/L and serum albumin ≥ 3.5 g/dL

Stage 2: 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 3: Serum beta-2 microglobulin ≥ 5.5 mg/L

Investigational Product (IP)

IP for this study refers to Kyprolis® (carfilzomib).

Last Dose Date of Study Treatment

The Last Dose Date of Study Treatment for a subject is the last date when a non-zero dose of any study drug was administered.

Last Known Alive Date

Last Known Alive Date is the latest date before the death date, according to the dates recorded on relevant CRFs and in the data collected by the vendors (which will be specified in a separate document).

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

MRD[-]CR rate is defined as the proportion of subjects with achievement of CR or better by IRC per IMWG-URC and achievement of MRD negativity as assessed by next-generation sequencing (NGS) method at a 10^{-5} threshold over the duration of the study among all ITT subjects.

MRD[-] Rate at 12 Months

MRD[-] rate at 12 months is defined as the proportion of subjects with achievement of MRD negativity at 12 months (± 4 weeks) from randomization, as assessed by NGS method at a 10^{-5} threshold among all ITT subjects. Per protocol, the 12-month sample may be omitted if the MRD analysis with confirmed results was performed within 4 months prior to the scheduled test at 12 months from randomization, or if subject has started new antimyeloma therapy prior to 12-month landmark, or if disease progression is recorded. So, MRD negativity results from bone marrow samples obtained at 8 to 13 months from randomization and prior to new antimyeloma therapy or disease progression will be considered in the calculation.

Overall Response Rate (ORR)

ORR is the proportion of ITT subjects whose best overall response is sCR, CR, VGPR, or PR per IMWG-URC over the duration of the study.

Overall Survival (OS)

OS is defined as the time (in months) from randomization to the date of death due to any cause.

OS = (death date or censoring date - randomization date + 1) / 30.4

Subjects still alive or lost to follow-up or withdrawn consent from study by the analysis time will be censored at the date on which the subject is last known to be alive.

Progression-Free Survival (PFS)

PFS will be calculated from the date of randomization until the first documentation of PD or death due to any cause, whichever occurs first.

PFS (month) = (PD/death date or censoring date - randomization date + 1) / 30.4

The duration of PFS will be right censored for subjects who meet any one of the following conditions: (1) no baseline/no 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 1 consecutively missed disease assessment visit (that is, progressive disease or death immediately after more than **63** days without disease assessment visit); (4) alive without documentation of disease progression before the analysis trigger date; (5) lost to follow-up or withdrawn consent. These censoring rules for PFS primary analysis are **following the derivations used for historical ASPIRE study.**

Table 5-1. Censoring Rules for Primary PFS Analysis

Situation	Date of Progression or Censoring	Outcome
No baseline/no 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	Date of last disease assessment visit before the first missed visit	Censored

missed disease assessment visit*		
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 PD assessment	Date of death	Progressed

* If death or PD is more than **63** days after previous disease assessment **(63 days corresponds to approximately 2 cycles plus a 7-day window)**, or randomization date if there is no previous disease assessment.

Patient-reported Convenience

Patient-reported convenience is measured by the Patient-reported Convenience with Carfilzomib-dosing Schedule Question. The questionnaire is a carfilzomib-specific convenience single-item/question and will be collected on Day 1 of Cycle 2, Cycle 5 and Cycle 12 before dosing, and safety follow-up. The items in the questionnaire will be collapsed into two categories for analysis purpose: Convenient, and Inconvenient, where

- **Convenient = 4 (Very convenient), 3 (Convenient)**
- **Inconvenient = 2 (Inconvenient) or 1 (Very inconvenient)**

For the comparison analysis of key secondary endpoint (patient-reported convenience after Cycle 4 of treatment), subjects who reported at least one "Very convenient" or "Convenient" after Cycle 4 will be included in the Convenient category; subjects with missing response of patient-reported convenience at all visits after cycle 4 will be included in the Missing category; otherwise, subjects will be in the Inconvenient category.

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.

$$\text{Relative Dose Intensity (\%)} = 100 \times \frac{\text{Actual Dose Intensity}}{\text{Planned Dose Intensity}}$$

Carfilzomib:

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

$$\text{Actual Dose Intensity } (\text{mg}/m^2/\text{week}) = \frac{\text{Actual Cumulative Dose of Carfilzomib } (\text{mg}/\text{m}^2)}{\text{Number of Weeks of Actual Treatment}}$$

Actual cumulative dose of carfilzomib (mg/m^2) is the sum of received doses (mg) divided by body surface area (BSA) (m²) of the subject. BSA is to be determined by the Mosteller Formula (Mosteller 1987):

$$\text{BSA (m}^2\text{)} = (\text{[Height (cm)} \times \text{Weight (kg)]} / 3600)^{1/2}$$

BSA should be calculated at baseline and utilized to calculate required carfilzomib doses. BSA should be recalculated if weight changes by more than 20% (gain or loss from weight used in the previous BSA calculation), and the new recalculated BSA will be used for subsequent infusions until further weight changes by more than 20%. If BSA is > 2.2 m^2, then BSA will be capped at 2.2 for carfilzomib dose calculation.

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. 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.

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

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 cycle day 15 or 16.

The planned cumulative dose of carfilzomib is the sum of planned carfilzomib dose (mg/m^2) per week as specified in Table 5-2 across the planned treatment weeks.

Table 5-2. Planned Carfilzomib Dose Schedule

Arm	Cycle	Week	Protocol Specified Dose for Treatment Week (mg/m^2)
Carfilzomib 56 mg/m^2 QW with lenalidomide and dexamethasone	1	1st	20 (on Day 1)
	2 or later	1st	56 (on Day 1)
	All cycles	2nd	56 (on Day 8)
		3rd	56 (on Day 15)
		4th	0
Carfilzomib 27 mg/m^2 BIW with lenalidomide and dexamethasone	1	1st	40 (20 on each day of Day 1 & 2)
	2 or later	1st	54 (27 on each day of Day 1 & 2)
	All cycles	2nd	54 (27 on each day of Day 8 & 9)
		3rd	54 (27 on each day of Day 15 & 16)
		4th	0

Dexamethasone:

The actual dose intensity is the actual amount of dexamethasone in mg delivered to a subject per week of treatment.

$$\text{Actual Dose Intensity (mg/week)} = \frac{\text{Actual Cumulative Dose of Dexamethasone (mg)}}{\text{Number of Weeks of Actual Treatment}}$$

The actual cumulative dose of dexamethasone in mg is the sum 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 in cycle 1-9 or given on day 1, 8 in cycle 10 or later cycle, i = 6 if the last dose is given on day 2, 9, 16 in cycle 1-9, or on day 2, 9 in cycle 10 or later, i =14 if the last dose is given on day 15 in cycle 10 or later cycle, i = 13 if the last dose is given on day 16 in cycle 10 or later.

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

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

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 dexamethasone dose is taken on day 1 or 2, 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 in cycle 1-9, j=4 if the last dose is taken on day 22, j=4 if the last dose is taken day 15 or 16 in cycle 10-12.

The planned cumulative dose of dexamethasone is the sum of planned dexamethasone dose (mg) per week as specified in Table 5-3 (40 mg per week for week 1-4 in cycle 1-9, 40 mg per week for week 1-3 and 0 for week 4 in cycle 10 or later) across the planned treatment weeks.

Table 5-3. Planned Dexamethasone Dose Schedule

Cycle	Week	Day (Arm)	Protocol Specified Dose (mg)
All cycles	1st	Day 1 (KRd 56 mg/m^2 QW)	40
		Day 1 & 2 (KRd 27 mg/m^2 BIW)	40 on Day 1 (or 20 on each day of Day 1 & 2)
All cycles	2nd	Day 8 (KRd 56 mg/m^2 QW)	40
		Day 8 & 9 (KRd 27 mg/m^2 BIW)	40 on Day 8 (or 20 on each day of Day 8 & 9)
All cycles	3rd	Day 15 (KRd 56 mg/m^2 QW)	40
		Day 15 & 16 (KRd 27 mg/m^2 BIW)	40 on Day 15 (or 20 on each day of Day 15 & 16)
Cycle 1 – 9	4th	Day 22	40
Cycle 10-12	4th	Day 22	0

Lenalidomide:

The actual dose intensity is the actual amount of drug in mg delivered to a subject per week of treatment.

$$\text{Actual Dose Intensity (mg/week)} = \frac{\text{Actual Cumulative Dose of Lenalidomide (mg)}}{\text{Number of Weeks of Actual Treatment}}$$

The actual cumulative dose of lenalidomide in mg is the sum of total quantity administered (mg) over the study.

Number of weeks of actual treatment will be calculated as (Last Dose Date of lenalidomide – First Dose Date of lenalidomide + i) / 7, where the value of i depends on the last dose day in a cycle, which is shown in the table below.

Day # of last lenalidomide dose in a cycle	Value of i
Day 1 or 8	7
Day 2 or 9	6

Day 3 or 10	5
Day 4 or 11	4
Day 5 or 12	3
Day 6 or 13	2
Day 7 or 14	1
Day 15 to Day 21	14 - (Day # - 15)

Planned dose intensity (mg/week) is defined as the planned amount of lenalidomide in mg delivered to a subject per week of treatment. It will be calculated as follows.

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

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 lenalidomide dose is given, and j =1 if the last lenalidomide dose is given on day 1 - 7, j=2 if the last dose is given on day 8 -14, j=4 if the last dose is given on day 15 - 21.

The planned cumulative dose of lenalidomide is the sum of planned lenalidomide dose (mg) per week as specified in Table 5-4 (175 mg per week for week 1-3) across the planned treatment weeks.

Table 5-4. Planned Lenalidomide Dose Schedule

Cycle	Week	Day	Protocol Specified Dose (mg)
All cycles	1st	Day 1 - 7	175 (25 mg per day)
	2nd	Day 8 - 14	175 (25 mg per day)
	3rd	Day 15 - 21	175 (25 mg per day)
	4th	Day 22 - 28	0

Refractory to Prior Multiple Myeloma Therapy

Subject is refractory to a drug of interest received during prior regimens if the data collected on Prior Multiple Myeloma Therapy (CHEMOTHERAPY AND TRANSPLANT) CRF page indicates that any of the following criteria is met:

a. The best response reached during at least one regimen containing the drug of interest was stable disease or progressive disease

b. The reason that the drug of interest was stopped was progression in at least one regimen

c. The date of relapse/progression is after start date and within 60 days after stop date of the drug on interest in at least one regimen

Study Day 1

Study day 1 for a subject corresponds to the earliest date when any study drug (carfilzomib, lenalidomide, or dexamethasone) is administered. For subjects who were never treated, study day 1 corresponds to the randomization date.

Study Day

The number of days from the study day 1 to a date of interest, inclusive:

Study day = (date of interest – date of study day 1) + 1, where the date of interest is on or after the date of study day 1.

Study day = (date of interest – date of study day 1), where the date of interest is before the date of study day 1. The study day is negative 1 for the day before Study Day 1.

Time to Progression (TTP)

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

TTP (month) = (documented PD date or censoring date - randomization date + 1) / 30.4

The same censoring rule as per PFS described in Table 5-1 will be used for TTP except that death will be treated as a censoring event.

Time to Response (TTR)

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

TTR (month) = (confirmed response start date - randomization date + 1) / 30.4

Treatment-emergent Adverse Event (TEAE)

Treatment-emergent adverse events are defined as adverse events starting on or after the first dose of any study drug, and up to 30 days (inclusive) of the last dose of any study drug, excluding adverse events reported after End of Study date.

6. Analysis Sets

The analysis and reporting of the data from this study will be performed using the following analysis populations:

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

The ITT population constitutes all randomized subjects and will be the basis for the analyses of efficacy in this study. Subjects in the analyses based on the ITT population will be analyzed according to the treatment arm to which they were randomized.

6.2 Safety Analysis Set

The safety population includes all randomized subjects who receive at least 1 dose of any study treatment (carfilzomib, lenalidomide, or dexamethasone), and will be the basis for the analyses of safety. Subjects in the analyses based on the safety population will be analyzed according to the treatment arm corresponding to the actual treatment received.

6.3 Per Protocol Set

The per-protocol population will include all randomized subjects who do not have any major protocol deviations that might affect the interpretation of the analyses of the efficacy endpoints. Subjects with the following important protocol deviation (IPD) will be excluded from the per protocol set.

- Major inclusion criteria not met (103 - 109 in protocol):
    - Documented relapse or progression after the most recent myeloma treatment. Subjects refractory to the most recent line of therapy are eligible, unless the last treatment contained PI or lenalidomide and dexamethasone) (103)
    - Subjects must have at least PR to at least 1 line of prior therapy (104)
    - Subjects must have received at least 1 but not more than 3 prior lines of therapy for multiple myeloma (induction therapy followed by stem cell transplant and consolidation maintenance therapy will be considered as 1 line of therapy) (105)

- Inclusion criteria of prior therapy with PI (106)
- Inclusion criteria of prior therapy with a lenalidomide and dexamethasone containing therapy (107)
- Inclusion criteria of measurable disease (108)
- Eastern Cooperative Oncology Group Performance Status (ECOG PS) of $0 \leq 2$ (109)

- Major exclusion criteria not met

  Disease related (201 - 206 in protocol):

  - Waldenström macroglobulinemia (201)
  - Multiple myeloma of IgM subtype (202)
  - POEMS syndrome (polyneuropathy, organomegaly, endocrinopathy, monoclonal protein, and skin changes) (203)
  - Plasma cell leukemia (> 2.0×10^9/L circulating plasma cells by standard differential) (204)
  - Primary amyloidosis (patients with multiple myeloma with asymptomatic deposition of amyloid plaques found on biopsy would be eligible if all other criteria are met) (205)
  - Myelodysplastic syndrome (206)

  Other medical conditions (207 - 211 in protocol):

  - History of other malignancy within the past 5 years (other than protocol specified exceptions) (207)
  - Known HIV infection, or uncontrolled hepatitis B or C infection (subjects without sustained virologic response) (208)
  - Ongoing Graft versus host disease (209)
  - Acute active infection requiring systemic antibiotics, antifungal, antiviral agents (except antiviral therapy directed at hepatitis B) within 14 days prior to randomization (210)
  - Known cirrhosis (211)

  Cardiopulmonary considerations (214 - 215 in protocol):

  - Uncontrolled hypertension (214)
  - Active congestive heart failure (New York Heart Association Class III to IV) (215)

Prior or concomitant therapy (218 - 222, 228 in protocol):

- Immunotherapy, monoclonal antibody therapy or approved anti-cancer chemotherapy within 28 days prior to randomization (218, 219, 220)
- Glucocorticoid therapy exceeding a cumulative dose of 160mg within 14 days prior to randomization (221)
- Focal radiation therapy within 7 days prior to randomization, radiation to large marrow reserves within 28 days (i.e., prior radiation must have been to <30% of the bone marrow) (222)
- Currently receiving treatment in another investigational device or drug study within 28 days of randomization (228)

Organ Function Assessment:

- Calculated or measured creatinine clearance < 30 mL/min (calculation must be based on the Cockcroft and Gault formula) within 28 days prior to randomization (244)

- Major treatment non-compliance
  - Treatment received different from treatment randomized (IPD Criteria 500).
  - Incorrect carfilzomib dose: under-dosing that meets IPD Criteria 502.
  - Failure to obtain extramedullary plasmacytoma assessment and /or bone lesion assessment as part of baseline disease assessment such that the primary endpoint cannot be assessed (IPD Criteria 801).
  - Any C1D1 disease assessment or disease specific lab required at baseline is missed or collected in such a way that patient is not measurable as defined per protocol (IPD Criteria 802).
  - Failure to obtain 2 consecutive disease response assessments that may impact the assessment of the primary endpoint (IPD Criteria 803).

6.4 Pharmacokinetic/Pharmacodynamic (PK/PDn) Analyses Set(s)

The PK Analysis Set will include all subjects who have received at least 1 dose of carfilzomib and 1 post-dose PK sample collected, as defined by the Schedule of Assessments. The PDn Analysis Set will include a subset of subjects in the PK Analysis set who have consented to participate in the optional PK/PDn substudy. These subjects will be evaluated for pharmacokinetics and pharmacodynamics unless significant

protocol deviations affect data analysis or if key dosing, dosing interruption, or sampling information is missing.

6.5 Interim Analyses Set

The interim analysis set for interim futility analysis include the first 230 subjects (50% of the planned total 460 subjects) who have been randomized and had the opportunity to be followed for a BOR assessment by the date when treatment was completed, confirmed PD or death occurred, subject was lost to follow-up, withdrew consent, or started new antimyeloma therapy, whichever occurred first.

6.6 Modified Full Analysis Set

The modified full analysis set (mFAS) will be used in the sensitivity analysis for primary endpoint and key secondary endpoints in order to adjust for the COVID-19 impact on the ITT population. The mFAS is a subset of ITT Analysis Set excluding subjects who are impacted by COVID-19.

The COVID-19 impact refers to:

(1) any COVID-19 adverse events which are identified using the COVID-19 Standardized MedDRA Query (SMQ) narrow search strategy, and COVID-19 events collected on the Confirmed COVID-19 Status CRF;

(2) the reason for Investigational Product (IP)/Non-IP Dose Change/Withheld/Dose Delay is COVID-19 control measures recorded on the CRFs for IP Administration and Non-IP Administration;

(3) the reason for ending IP/Non-IP is COVID-19 control measures recorded on CRFs for End of IP Administration and End of Non-IP Administration;

(4) IPDs related to COVID-19 control measures;

(5) COVID-19 related protocol deviation: 940 series, 950 series and 960 series of protocol deviation codes in the study IPD list.

7. Planned Analyses

7.1 Interim Analysis and Early Stopping Guidelines

An Independent Biostatistics Group (IBG) will perform the interim analyses and provide the interim report to an independent Data Monitoring Committee (DMC). The initial assessment from this committee will be planned after 30 subjects (approximately 15 for the experimental arm and 15 for the control arm) have been enrolled and have finished the first cycle of treatment to ensure safety of all arms. Thereafter, the DMC will review

all available safety/efficacy data periodically (approximately every 6 months) and evaluate the efficacy analysis once for the non-binding interim analysis for futility. The primary purpose of the interim futility analysis is to assess the futility in terms of ORR. The non-binding interim futility analysis is planned to occur when the first 230 subjects (50% of the planned total 460 subjects) have been randomized and had the opportunity to be followed for a best overall response (BOR) assessment by the date when treatment was completed, confirmed PD or death occurred, subject was lost to follow-up, withdrew consent, or started new antimyeloma therapy, whichever occurred first. Using an O'Brien-Fleming type beta-spending function, the stopping boundary for futility in p-value scale is 0.289. The study is futile in terms of ORR if the observed p-value > 0.289. With such a stopping boundary, the trial has over 70% probability to stop for futility when the null hypothesis of inferiority is true.

The IBG and DMC will have access to subjects' individual treatment assignments. To minimize the potential introduction of bias to the conduct of the study, members of the DMC and IBG will not have any direct contact with study center personnel or subjects. The DMC will communicate major safety concerns and recommendations regarding study modification or termination based on the safety data and interim futility stopping criteria to Amgen in accordance with the DMC charter.

Records of all meetings will be maintained by the DMC for the duration of the study. Records of all meetings will be transferred and stored in the TMF (in accordance with SOP-427356) at the conclusion of the study.

Further details are provided in the DMC charter.

Data will be subject to ongoing checks for integrity, completeness, and accuracy in accordance with the Data Management Plan with the expectation that all outstanding data issues are resolved ahead of the snapshot. The data will be locked to prevent further changes, and a snapshot of the locked database will be used in the analysis.

7.2 Primary Analysis

The primary analysis corresponds to the final analysis.

7.3 Final Analysis

The final analysis will be performed after all subjects have completed the study. Final analysis will be based on a clean database lock.

The hypotheses for the primary and key secondary objectives (ORR, PFS, and convenience after cycle 4 of treatment) 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. The testing is ordered as follows: non-inferiority of ORR, non-inferiority of PFS, and superiority of patient-reported convenience after cycle 4 of treatment.

Starting with the hypothesis of ORR, if any null hypothesis in the sequence is rejected at a 1-sided significance level of 0.025, then the subsequent null hypothesis will be tested. Otherwise, if any null hypothesis failed to be rejected, then the subsequent hypotheses will not be tested.

For the primary endpoint, the stopping boundary is 0.025 in p-value scale in the final analysis.

The final analysis of efficacy endpoints will be based on the ITT analysis set, while the final analysis of safety endpoints will be based on the safety analysis set. Sensitivity analyses of efficacy endpoints based on the per protocol set might be performed only if the PP population was less than 90% of the ITT population.

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. The database will be subject to edit checks outlined in the data management plan by Amgen Clinical Data Management (CDM) department. Data inconsistencies and suspicious values will be reviewed and resolved before the database is locked.

8.2 Data Handling and Electronic Transfer of Data

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. The data handling and electronic transfer of data are described in the data management plan (DMP).

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,

concomitant medications, new antimyeloma therapy, death, prior multiple myeloma therapy, and relapse/progression to prior multiple myeloma therapy are described in Appendix A.

The calculation of scores and methods to deal with missing data will be handled according to the EORTC QLQ-C30 questionnaire's standard scoring guidelines.

No imputation will be done for the analysis of the primary and key secondary endpoints. The frequency of missing disease assessments and deviation of the actual disease assessment times from the scheduled assessment times will be summarized by treatment arms. Sensitivity analyses will be performed to assess the impact of missing any disease or response assessment on the analysis of ORR and PFS (i.e., analysis in Per Protocol Set). Similar analysis will be performed for QOL endpoints.

8.4 Detection of Bias

If applicable the methods to detect bias are described in the analyses of particular endpoints (Section 9).

8.5 Outliers

Any suspected outliers will be investigated by the study team and will be included in the database unless determined to be an error or there is supporting evidence or explanation to justify the exclusion. Any outliers excluded from the analysis will be discussed in the Clinical Study Report (CSR), including the reasons for exclusion and the impact of their exclusion on the study. Pharmacokinetic (PK) plasma 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 endpoints. The statistical assumptions for analysis methods will be assessed. If the assumptions for the distributional characteristics are not met, these will be described, and further analyses may be carried out using data transformations or alternative analysis methods. The use of transformations or alternative analysis methods will be justified in the final study report.

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

Where applicable, descriptive statistics will be provided. For continuous variables, the number of subjects with non-missing data (n), mean, standard deviation (STD), median, minimum, and maximum will be presented. For categorical variables, the frequency (n) and percentage will be summarized in each category. The denominator for percentages is the number of subjects in the analysis set of interest for the summary. The binomial proportions and the corresponding 95% confidence intervals (CIs) will be based on the exact distribution methods (Clopper-Pearson interval) and the treatment comparison will be based on Cochran-Mantel-Haenszel (CMH) method. Time to event endpoints will be estimated using the Kaplan-Meier (KM) method and will be summarized with the number of subjects with events or censored, and censoring reasons, KM quartiles (when estimable) and corresponding two-sided 95% CIs, KM proportions at select time points, inferential comparison between treatment arms with associated p-values, hazard ratios from stratified Cox proportional hazard (PH) model, and KM curves.

The analyses of efficacy and safety endpoints will be based on the analysis sets defined in Section 6. The primary (final) analyses of the efficacy endpoints and Patient-report outcomes will use the ITT population, while the Per Protocol Set will be used for sensitivity analyses, if applicable. The analyses of the safety endpoint will be based on the safety population.

The primary analysis of ORR and PFS will be based on IRC assessed outcomes. The synthesis approach will be used to show that KRd 56 mg/m^2 QW preserves at least 60% of KRd 27 mg/m^2 BIW effect vs Rd in terms of ORR and 50% in terms of PFS. The null hypotheses for the primary and key secondary objectives (ORR, PFS, and convenience after cycle 4 of treatment) 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. The testing is ordered as follows: non-inferiority of ORR, non-inferiority of PFS, and superiority of patient-reported convenience after cycle 4 of treatment. Starting with the hypothesis of ORR, if any null hypothesis in the sequence is rejected, then the subsequent hypothesis

will be tested. Otherwise, if any hypothesis failed to be rejected, then the subsequent hypotheses will not be tested. For all other endpoints, the significance testing, if performed, will be considered descriptive.

Subgroup analyses of the primary endpoint will be performed as exploratory analyses for selected baseline factors.

9.2 Subject Accountability

The following subject disposition information will be summarized by treatment arm.

- Number of subjects screened
- Number (%) of subjects screened but not randomized
- Number (%) of subjects randomized
- Number (%) of subjects randomized but not dosed, along with the reasons for not being dosed
- Number (%) of subjects who received any cycle of treatment for each study drug (carfilzomib, lenalidomide, dexamethasone)
- Number (%) of subjects who completed the 12 cycles of treatment for each study drug (carfilzomib, lenalidomide, dexamethasone)
- Number (%) of subjects who discontinued treatment for each study drug (carfilzomib, lenalidomide, dexamethasone), along with the reasons for discontinuation of treatment
- Number (%) of subjects who discontinued each study drug (carfilzomib, lenalidomide, dexamethasone) due to COVID-19 impact
- Number (%) of subjects who **completed the study, and who** discontinued the study, along with the reasons for discontinuation

Key study dates for the first subject enrollment, last subject enrollment**, last subject last dose of investigational product,** and **last subject end of study** will be presented.

The number (%) of subjects who were enrolled will be tabulated by region, country and investigator site and randomization stratification factors for each treatment arm in ITT Analysis Set.

The listing of unique manufacturing lot numbers and the subject listing of manufacturing lot numbers will also be generated.

9.3 Important Protocol Deviations

IPD 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. Eligibility deviations are defined in the protocol. The following information of IPD and protocol deviations will be summarized, where applicable, with respect to the following:

- Number (%) of subjects with IPDs and total number of IPDs will be summarized by category and sub-category by treatment arm in ITT Analysis Set
- Number (%) of subjects with IPDs related to COVID-19 control measures will be summarized by treatment arm in ITT Analysis Set
- Number (%) of subjects with COVID-19 protocol deviations by protocol deviation category (940 series, 950 series and 960 series of protocol deviations codes) and by treatment arm in ITT Analysis Set
- Subject listing of IPD in ITT Analysis Set, including COVID-19 related IPDs with the descriptions
- Subject listing of COVID-19 protocol deviations based on protocol deviations category (940 series, 950 series and 960 series of protocol deviations codes) in ITT Analysis Set
- **Subject listing of inclusion/exclusion criteria deviations**

9.4 Demographic and Baseline Characteristics

The following demographic and baseline disease characteristics will be summarized by treatment arm and overall using descriptive statistics for the ITT Analysis Set.

- Baseline demographics and characteristics:
  - Age (years) (as continuous variable, as categorical variable: <65, ≥65; 18 - <65, 65 - <75, ≥75; 18 - <65, 65 - <75, 75 - <85, ≥85, unknown)
  - Sex (female, male)
  - Race (White and other categories depending on frequency observed)
  - Ethnicity (Hispanic or Latino, Not Hispanic or Latino)
  - Region (Europe, Non-Europe)
  - Height (cm)
  - Weight (kg)

- BSA (m^2) (as continuous variable, as categorical variable: ≤2.2, >2.2)
- Body mass index (BMI) (kg/m^2)

- Baseline organ function and comorbid conditions:
  - ECOG PS (0-1, 2)
  - Hemoglobin (g/L)
  - Absolute Neutrophil Count (10^9/L)
  - Platelet Count (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)
  - Creatinine clearance (CrCl, mL/min) (as continuous variable, as categorical variable: <30, 30 - <50, 50 - <80, ≥80; <50, ≥50)
    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) (%)
  - Hypertension history (yes, no)
  - Cardiopulmonary history (yes <by diagnosis category>, no)
- Baseline disease characteristics:
  - Original ISS stage (1 or 2, 3) at study entry per IxRS
  - Original ISS stage at baseline (1 or 2, 3) (derived based on the definition in Section 5)
  - Prior lenalidomide treatment (yes vs no)
  - Prior PI treatment (yes vs no)
  - Prior anti-CD38 exposure (yes vs no)
  - Prior bortezomib treatment (yes vs no)
  - Number of prior therapies (1 vs >1; 1 vs 2 vs >2)
  - Cytogenetic risk measured by FISH (high risk (t(4;14), t(14;16), deletion 17p), standard risk, missing)
  - Bortezomib refractory status (yes vs no)
  - Lenalidomide refractory status (yes vs no)
  - **PI refractory status (yes vs no)**
  - **Anti-CD38 refractory status (yes vs no)**
  - **Refractory to the last prior line of therapy (yes vs no)**
  - Presence of plasmacytoma (yes, no)
  - Prior transplant (yes, no)
  - Prior tobacco use (yes, no)

- β2-microglobulin level (mg/L) (as continuous variable; as categorical variable: <3.5, ≥3.5 and <5.5, ≥5.5)
- Albumin (g/dL) (as continuous variable; as categorical variable: <3.5, ≥3.5)
- Time from initial multiple myeloma diagnosis to randomization (months)
- **Immunoglobulin heavy and light chain types (IgA, IgD, IgE, IgM, IgG, None; Kappa, Lambda, Not detectable within each heavy chain type)**
- Time since last relapse (months)

In addition, the analysis population in each analysis set will be summarized for each treatment arm. The listing of randomization will be also generated.

9.5 Efficacy Analyses

The main efficacy analyses will be based on the ITT Analysis Set. Other Analysis Sets might be used for various sensitivity analyses.

Table 9-1. Primary Efficacy Endpoint Summary Table

Endpoint	Primary Summary and Analysis Method	Sensitivity Analysis
Overall Response over the duration of the study	**Response** based on IRC assessments: • Point estimate of ORR and associated 95% CI (Clopper Pearson method) by treatment arm • Risk ratio and associated 95% CI using Cochran-Mantel-Haenszel (CMH) method controlling for randomization stratification factors as a measure of treatment effect • P-value (1-sided, 2.5% significance level in the final analysis) of non-inferiority test via synthesis approach to show that KRd 56 mg/m^2 QW preserves at least 60% of KRd 27 mg/m^2 BIW effect vs Rd	• **Response** based on investigator assessments: Same as primary summary and analysis method • **Response based on** internal computational assessments: Same as primary summary and analysis method based on Onyx response computer algorithm (ORCA) assessments • Unstratified analyses: risk ratio and associated 95% CI from unstratified CMH • Per Protocol Set: Same as primary summary and analysis method mFAS: Same as primary summary and analysis method

Table 9-2. Secondary Efficacy Endpoint Summary Table

Endpoint	Primary Summary and Analysis Method	Sensitivity Analysis
PFS over the duration of the study	**PD** based on IRC assessments: • KM estimates for PFS distribution by treatment arm and PFS rates with 95% CI at 6 and 12 months • Hazard ratio and 95% CI from stratified Cox proportional hazards (PH) model • P-value (1-sided, 2.5% significance level) of non-inferiority test via synthesis approach to show that KRd 56 mg/m^2 QW preserves at least 50% of KRd 27 mg/m^2 BIW effect vs Rd. The testing is done once at the primary (final) analysis on ITT analysis set. • **Restricted mean survival time (RMST) with 95% CI in each arm and the between-arm difference in RMST with 95% CI based on ITT analysis set** • **Observed treatment effect retention rate based on ITT analysis set**	• **PD** based on investigator assessments: Same as primary summary and analysis method • **PD based on** internal computational assessments (ORCA): Same as primary summary and analysis method • Unstratified analyses: hazard ratio and 95% CI from unstratified Cox PH model • Per Protocol Set: Same as primary summary and analysis method • mFAS: Same as primary summary and analysis method • **Initiation of new anti-myeloma therapy treated as PFS Event: The data censoring rules are 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 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**

Endpoint	Primary Summary and Analysis Method	Sensitivity Analysis
		analysis method as for primary analysis will be used • **Analysis based on scheduled assessment dates: same as primary summary and analysis methods, except that the analysis is based on the scheduled assessment dates instead of actual assessment dates** • **Analysis using interval censoring**: **PFS data will be treated as interval-censored (Section 9.5.2)**
Patient-reported Convenience with Carfilzomib-dosing Schedule Question after cycle 4 of treatment	• Proportion of patient-reported convenience and 95% CI (Clopper Pearson method) by treatment arm • Odds ratio (OR) with 95% CI from the CMH method stratified by the randomization stratification factors as a measure of treatment effect for whether carfilzomib dosing is reported as convenient or inconvenient • P-value (1-sided, 2.5% significance level) from the CMH chi-square test controlling for the randomization stratification factors. The testing is done once at the primary (final) analysis on ITT analysis set.	• Logistic regression model including the randomization stratification factors and treatment arm • mFAS: Same as primary summary and analysis method
Time to response	Response based on IRC assessments: • Descriptive statistics (mean, STD, median, minimum and maximum) among responders by treatment arm	• **Response** based on investigator assessments: Same as primary summary and analysis method
Duration of response (DOR)	Response based on IRC assessments: • KM estimates for DOR distribution by treatment arm	• **Response** based on investigator assessments: Same as primary summary and analysis method
Time to progression (TTP)	**PD** based on IRC assessments: • KM estimates by treatment arm	• **PD** based on investigator assessments: Same as

Endpoint	Primary Summary and Analysis Method	Sensitivity Analysis
	• Stratified Cox PH model	primary summary and analysis method
OS over the duration of the study	• KM estimates for OS distribution by treatment arm and OS rate with 95% CI at 6 and 12 months • Hazard ratio and 95% CI from stratified Cox proportional hazards (PH) model	• Unstratified analyses: hazard ratio and 95% CI from unstratified Cox PH model • Per Protocol Set: Same as primary summary and analysis method
MRD[-]CR	CR or better response component based on IRC assessments: • MRD[-]CR rate and 95% CI (Clopper Pearson method) by treatment arm • OR with 95% CI as a measure of treatment effect estimated by CMH method stratified by the randomization stratification factors	• **Response** based on investigator assessments for CR or better: Same as primary summary and analysis method • **Response based on** internal computational assessments (ORCA): Same as primary summary and analysis method
MRD[-] status at 12 months from randomization	• MRD[-] rate and 95% CI (Clopper Pearson method) by treatment arm • OR with 95% CI as a measure of treatment effect estimated by CMH method stratified by the randomization stratification factors	
Physical functioning and role functioning (scales of EORTC QLQ-C30) over time	• Repeated measures analysis of covariance (ANCOVA) adjusting for the baseline covariates (treatment arm, randomization stratification factors, baseline score and visit) for the comparison of mean score over time between treatment arms	• Per Protocol Set: Same as primary summary and analysis method • Restricted maximum likelihood-based mixed model for repeated measures (MMRM) for the comparison of mean score over time between treatment arms
Patient-reported treatment satisfaction (the SWT scale of CTSQ) after cycle 4 of treatment	• ANCOVA at the corresponding fixed time point	

Table 9-3. Exploratory Efficacy Endpoint Summary Table

Endpoint	Primary Summary and Analysis Method	Sensitivity Analysis
Described in Section 2.1 table	Will be described in a supplemental analysis plan (SSAP) finalized before database lock	Will be described in a SSAP finalized before database lock

9.5.1 Analyses of Primary Efficacy Endpoint(s)

The ORR will be calculated by treatment arm and the associated 95% CI will be estimated using the Clopper Pearson method (Clopper CJ and Pearson, 1934). As a measure of treatment effect, relative risk (risk ratio) and associated 95% CI will be estimated using CMH method controlling for randomization stratification factors. The KRd 27 mg/m^2 BIW arm will serve as the reference treatment arm in the calculation of the risk ratio. The non-inferiority comparison of ORR between treatment arms will be performed using the synthesis approach (FDA, 2016) to show that KRd 56 mg/m^2 QW preserves at least 60% of KRd 27 mg/m^2 BIW effect vs Rd under the constancy assumption (the effect of KRd 27 mg/m^2 BIW in the current study is consistent with the effect that was observed in the historical study ASPIRE). The 1-sided p-value from the non-inferiority test will be reported.

The null and alternative hypotheses are expressed as:

H_0: $U_n \le (1\text{-}r)*U_h$ versus H_a: $U_n > (1\text{-}r)*U_h$

The synthesis test statistic is calculated as:

$[U_n - (1-r)*U_h] / \sqrt{SE_n^2 + (1-r)^2 * SE_h^2}$, where

r = retention rate = 60%

U_h = log-relative risk of ORR of Rd vs KRd 27mg/m^2 BIW by cycle 12 from the historical study ASPIRE = log (0.755)

SE_h = standard error of U_h = 0.041

U_n = log-relative risk of ORR of KRd 56 mg/m^2 QW vs KRd 27 mg/m^2 BIW from this non-inferiority study

SE_n = standard error of U_n

In the primary (final) analysis, the H_0 will be rejected if the synthesis test statistic > $Z_{0.975}$, where $Z_{0.975}$ is the critical value corresponding to a test of 1-sided Type I error rate of 0.025.

SAS code for calculating 1-sided p-value for the non-inferiority test:

```
data test;
r = 0.6;
Uh = log(0.755);
SEh = 0.041;
RR = x.xxx /*relative risk of ORR (QW arm vs BIW arm), driven by the
             data from ARROW2 study*/
lowerCI = x.xxxx; upperCI = x.xxxx /*95% CI for RR, driven by the data
                                     from ARROW2 study*/
Un = log(RR);
SEn = (log(upperCI)-log(lowerCI))/(2*probit(1-0.05/2));
SE_syn = sqrt(SEn**2+((1-r)*SEh)**2);
z = (Un-(1-r)*Uh)/SE_syn;
pvalue_1sided = 1-probnorm(z);
run;
```

In order to check the adequacy of the constancy assumption, the summaries of baseline demographics and characteristics in KRd 27mg/m^2 BIW treatment arm from ARROW2 study will be clinically evaluated by the DMC against those in ASPIRE historical study for clinical judgement on the consistency of the two trial populations.

The primary analysis of ORR will be based on the IRC **response** assessments for the ITT Analysis Set. Several sensitivity analyses will be considered for ORR following the same method as the primary analysis: (1) analysis based on investigator assessment in ITT Analysis Set; (2) analysis based on ORCA assessments in ITT Analysis Set; (3) analysis based on the IRC assessments using unstratified model in ITT Analysis Set; (4) analysis based on the IRC assessments in Per Protocol Set; (5) analysis based on the IRC assessments in mFAS. The concordance rate between the results from the IRC, investigator and ORCA will be summarized.

In order to align with the European Medicines Agency (EMA) "Guidance on the Choice of the Non-inferiority Margin" for non-inferiority trials (EMA, https://www.ema.europa.eu/documents/scientific-guideline/guidelinechoice-non-inferioritymargin_en.pdf, 2005) and help with the interpretability of the results, 0.87 is considered as a margin for RR estimated from primary analysis of ORR. This value is outside the 95% CI for RR of response rates by Cycle 12 (Rd vs KRd 27 mg/m^2 BIW) from the historical trial ASPRE (0.70, 0.82), and it was chosen in corroboration with clinical considerations for this patient population.

Subgroup analyses will be performed to explore the consistency of the treatment effect on ORR for subgroups described in Section 4.2. The estimate of risk ratio with 95% CI **using CMH method controlling for randomization stratification factors** will be provided for ORR between the treatment arms. **A forest plot will be produced for the estimated risk ratio with 95% CI of each subgroup.**

9.5.2 Analyses of Secondary Efficacy Endpoint(s)

Key secondary efficacy endpoints

If the null hypothesis for ORR is rejected at a 1-sided significance level of 0.025, then the key secondary endpoints will be tested by sequential testing in the order of non-inferiority of PFS and superiority of patient-reported convenience after cycle 4 of treatment. Otherwise, if any null hypothesis failed to be rejected, then the subsequent hypotheses will not be tested.

PFS is defined in Section 5. The number of subjects with PFS events or censored, and censoring reasons will be presented by treatment arm. The subcategory of death with the primary reason of COVID-19 infection or COVID-19 pneumonia will be included in the PFS events. The distribution of PFS time including median and other quartiles will be summarized descriptively using the KM method (Klein and Moeschberger, 1997). The corresponding 95% confidence intervals for the median and other quartiles will be constructed using KM method with log-log transformation. PFS rates at 6 and 12 months will be estimated, and the corresponding 95% CIs 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). KM curves will also be presented. The HR and its 95% CI will be estimated using a Cox proportional hazards model stratified by the randomization stratification factors. The KRd 27 mg/m^2 BIW arm will serve as the reference treatment arm in the calculation of the HR.
The non-inferiority comparison of PFS between treatment arms will be performed using the synthesis approach to show that KRd 56 mg/m^2 QW preserves at least 50% of KRd 27 mg/m^2 BIW effect vs Rd. The 1-sided p-value from the non-inferiority test at 2.5% significance level will be reported. The synthesis test statistic and p-value will be calculated in the same way as for ORR analysis except that r (retention rate) = 50% and the HR will be used as the measure of relative risk. The testing is done once at the primary (final) analysis on ITT analysis set.

The null and alternative hypotheses are expressed as:

H_0: $U_n \geq (1\text{-}r)*U_h$ versus H_a: $U_n < (1\text{-}r)*U_h$

The synthesis test statistic is calculated as:

$[U_n - (1-r)*U_h] / \sqrt{SE_n^2 + (1-r)^2 * SE_h^2}$, where

r = retention rate = 50%

U_h = log-hazard ratio of PFS of Rd vs KRd 27mg/m^2 BIW by cycle 12 from the

historical study ASPIRE = log (1.812)

SE_h = standard error of U_h = 0.137

U_n = log-hazard ratio of PFS of KRd 56 mg/m^2 QW vs KRd 27 mg/m^2 BIW from this non-inferiority study

SE_n = standard error of U_n

The H_0 will be rejected if the synthesis test statistic < $Z_{0.025}$, where $Z_{0.025}$ is the critical value corresponding to a test of 1-sided Type I error rate of 0.025.

SAS code for calculating 1-sided p-value for the non-inferiority test:

```
data test;
r = 0.5;
Uh = log(1.812);
SEh = 0.137;
HR = x.xxx /*hazard ratio (QW arm vs BIW arm), driven by the data from
             ARROW2 study*/
Un = log(HR);
SEn = x.xxx; /*number x.xxx is driven by the data from ARROW2 study*/
SE_syn = sqrt(SEn**2+((1-r)*SEh)**2);
z = (Un-(1-r)*Uh)/SE_syn;
pvalue_1sided = probnorm(z)
run;
```

The primary analysis of PFS will be based on the IRC assessments **of PD** for the ITT Analysis Set. Similar to the primary analysis, several sensitivity analyses will be performed **including recommendations from the "Appendix 1 to the guideline on the evaluation of anticancer medicinal products in man" (EMA, https://www.ema.europa.eu/en/documents/scientific-guideline/appendix-1-guideline-evaluation-anticancer-medicinal-products-man-methodological-consideration-using_en.pdf, 2013):** (1) analysis based on investigator assessment in ITT Analysis Set; (2) analysis based on ORCA assessments in ITT Analysis Set; (3) analysis based on the IRC assessments using unstratified model in ITT Analysis Set; (4) analysis based on the IRC assessments in PP Analysis Set; (5) analysis based on the IRC assessments in mFAS; **(6) initiation of new anti-myeloma therapy treated as PFS Event in ITT Analysis Set; (7) initiation of new anti-myeloma therapy treated as neither a PFS event nor a censoring event in ITT Analysis Set; (8) analysis based on scheduled assessment dates instead of actual assessment dates in ITT Analysis Set; (9) analysis using interval censoring in ITT Analysis Set: PFS data will be treated as interval-censored. The interval will be constructed as follows: (i) if the PFS event is PD, then interval will be (date of last assessment before PD,**

date of the assessment indicating PD]; (ii) if the PFS event is death, then interval will be [date of death, date of death]; and (iii) if no PFS event is observed, then the interval will be (date of last assessment,]. These intervals will then be transformed from calendar time scale to time-since-randomization scale by subtracting individual randomization dates plus 1 day. The non-parametric maximum likelihood estimate of the survival curves will be computed based on the interval censored PFS data using the Expectation-Maximization iterative convex minorant algorithm (EM-ICM) (Wellner & Zhan, 1997).

The concordance in the assessment of progressive disease by the IRC, investigator, and ORCA will be summarized for ITT Analysis Set.

The observed treatment effect retention rate in ITT Analysis Set, calculated as (U_h – U_n) / U_h, where U_h and U_n are defined above, will be reported as the supportive information.

The adequacy of the proportional hazard assumption will be assessed using the plot of the logarithm of the estimated hazard function based on the KM 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, an interaction between treatment and the logarithm of the time to event will be tested using a Cox model stratified by randomization stratification factors to test for non-proportionality.

Piecewise Cox models may be explored **in ITT Analysis Set** given evidence of non-proportional hazards (Collett, 2003). This method 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.

Further, the restricted mean survival time (RMST) will be considered as an alternative measure of treatment effect (Uno, 2015; Weir, 2018). With an RMST boundary time of 13 months, the RMST (area under the survival curve) with 95% CI in each arm, and the between-arm difference in RMST (the area between the survival curves of the 2 arms) with its corresponding 95% CI (Wald) will be estimated in the ITT Analysis Set.

Subgroup analyses will be performed to explore the consistency of the treatment effect on PFS for subgroups described in Section 4.2. The estimate of hazard ratio with 95% CI using a stratified Cox model will be provided between the treatment arms. A forest plot will be produced for the estimated hazard ratio with 95% CI of each subgroup.

Patient-reported convenience **at each reporting cycle (2, 5,12, safety follow-up) and after Cycle 4 are defined in Section 5**.

The frequency and proportion of each category will be summarized by treatment arm at **Cycle 2, C**ycle 5, **C**ycle 12 and safety follow-up based on all randomized subjects (ITT Analysis Set) as well as all expected subjects at the scheduled visit (i.e., randomized subjects who are still alive and remaining on carfilzomib treatment at the scheduled visit, subjects who **have ended all study treatment and are remaining on study for** safety follow-up visit). **The 95% CI will be estimated using Clopper Pearson method for the proportion of the Convenient category based on all randomized subjects in each arm.**

For the key secondary endpoint analysis (Convenience after Cycle 4), Cochran-Mantel-Haenszel method stratified by the randomization stratification factors will be used to evaluate the treatment effect (OR with associated 95% CI) **and the 1-sided p-value** for whether carfilzomib dosing is reported as convenient **after cycle 4**. The superiority test significance level will be 2.5%. The comparison analyses will be done once at the primary **(final)** analysis using the data collected at **C**ycle 5, **C**ycle 12**, and safety follow-up** based on the ITT Analysis Set. Following the same method, a sensitivity analysis will be performed based on mFAS. A logistic regression model including the randomization stratification factors and treatment arm will be also considered for sensitivity analysis of the comparison between treatment arms in ITT Analyses Set. The missing data of outcome will not be imputed and will not be included in the analysis of the comparison between treatment arms.

Other secondary efficacy endpoints

TTR is defined in Section 5. TTR will be summarized descriptively by the non-missing sample size (n), mean, STD, median, minimum and maximum among responders by treatment arm based on IRC. Similar analysis will be performed using the Investigator assessment of response as a sensitivity analysis on ITT Analysis Set.

DOR is defined in Section 5. DOR will be summarized descriptively by treatment arm based on IRC assessment of response. The distribution of DOR, including the median and other quartiles and their corresponding 95% CIs, will be characterized using the KM method based on the subjects who achieve a best response of PR or better. No inferential comparison between treatment arms will be made for DOR. A similar analysis will be performed by using the Investigator disease assessment as a sensitivity analysis on ITT Analysis Set.

TTP is defined in Section 5. Analysis of TTP will be performed in the same way as the primary PFS analysis. The distribution of TTP will be characterized using the KM method. The HR and its 95% CI will be estimated using a Cox model stratified by the randomization stratification factors. A similar analysis will be performed by using the Investigator disease assessment as a sensitivity analysis on ITT Analysis Set.

OS is defined in Section 5. Analysis of OS will be performed in the same way as the primary PFS analysis based on the ITT Analysis Set. The number of subjects with OS events or censored, and censoring reasons will be presented by treatment arm. The subcategory of death with the primary reason of COVID-19 infection or COVID-19 pneumonia will be included in the OS events. The distribution of OS time including median and other quartiles will be summarized descriptively using the KM method (Klein and Moeschberger, 1997). The corresponding 95% CIs for the median and other quartiles will be constructed using KM method with log-log transformation. OS rates at 6 and 12 months will be estimated, and the corresponding 95% CIs will be calculated using the method of Kalbfleisch and Prentice (1980). The duration of the follow-up for OS will be estimated by reverse Kaplan-Meier method (Schemper and Smith 1996). KM curves will also be presented. The HR and its 95% CI will be estimated using a Cox proportional hazards model stratified by the randomization stratification factors. The KRd 27 mg/m^2 BIW arm will serve as the reference treatment arm in the calculation of the HR. **The adequacy of the proportional hazard assumption will be assessed using the same method as described for the PFS. Piecewise Cox models may be explored in ITT Analysis Set given evidence of non-proportional hazards (Collett, 2003).**

Similar to the primary analysis of OS, the sensitivity analyses will be performed based on unstratified method for the ITT Analysis Set, or for the Per Protocol Set.

Subgroup analyses will be performed to explore the consistency of the treatment effect on OS for subgroups described in Section 4.2. The estimate of hazard ratio

with 95% CI using a stratified Cox model will be provided between the treatment arms. A forest plot will be produced for the estimated hazard ratio with 95% CI of each subgroup.

MRD[-]CR is defined in Section 2.1. The MRD[-]CR rate (defined in Section 5) will be calculated by treatment arm and the associated 95% CI will be estimated using the Clopper Pearson method. As a measure of treatment effect, OR and associated 95% CI will be estimated using CMH method controlling for randomization stratification factors. The KRd 27 mg/m^2 BIW arm will serve as the reference treatment arm in the calculation of the OR. The analysis of MRD[-]CR will be based on the IRC assessments of response for the ITT Analysis Set where patients without MRD assessment will be considered as having MRD positive status.

Similar to the primary analysis of MRD[-]CR, the sensitivity analyses will be performed based on investigator assessments and ORCA assessments for CR or better.

MRD[-] status at 12 months is defined in Section 2.1. MRD[-] rate at 12 months (defined in Section 5) will be analyzed similarly to MRD[-]CR rate in the primary analysis of MRD[-]CR.

Physical functioning and role functioning (EORTC QLQ-C30) over time: Physical Functioning and Role Functioning are measured by the Physical Functioning and Role Functioning scales of the European Organization for Research and Treatment of Cancer Quality-of-life Questionnaire Core 30 (EORTC QLQ-C30) and will be administered before dosing on Day 1 of **C**ycle 1, 3, 5, 7, 9, 12, and safety follow-up. The scale scores are linearly converted to range from 0-100. Principles for scoring are detailed in Appendix B. A higher score for functional scales represents a better functional status.

The scale scores will be summarized descriptively with the non-missing data (n), mean, STD, median, minimum and maximum by treatment arm at each visit. Change from baseline at each visit will be also summarized using the same statistics.

The completion rate for scale score will be presented by treatment arm and by visit based on all randomized subjects (ITT Analysis Set) as well as all subjects expected to have an assessment at the scheduled visit (i.e., randomized subjects who are still alive and remaining on study treatment at the scheduled visit, subjects who **have ended all study treatment and are remaining on study for** safety follow-up visit).

In addition to descriptive analyses, the comparison of mean score over time between treatment arms will be performed using repeated measures analysis of covariance

(ANCOVA) adjusting for the baseline covariates in a multivariate model. The dependent variable of the models will be the scale scores measured at each post-baseline visit. The model will include treatment arm, randomization stratification factors, the baseline score (covariate, the scale score on cycle 1 day 1 before dosing) and visit as a repeated measure. The least square mean by treatment arm and the overall least square mean difference between treatment arms will be reported along with the corresponding 95% CI.

All the above analyses will be based on the ITT Analysis Set. A similar analysis will be performed on Per Protocol Set as a sensitivity analysis.

In the repeated measure ANCOVA, the missing scale scores will be imputed using multiple imputation (MI) method (Rubin DB, 1976; SAS Institute Inc., 2015) with the assumption of missing at random (MAR). The steps are as follows.

- Assuming the missing data pattern is arbitrary, the MI will be performed by treatment group using the fully conditional specification (FCS) method with linear regression models, which is implemented with SAS 9.4's PROC MI under the FCS statement. The imputation model will include all the variables specified in ANCOVA model (treatment arm, randomization stratification factors, the baseline scale score). There are 50 imputed datasets to be generated. The seed of the pseudorandom number generator used to randomly generate imputations for the missing values is specified as seed=54321. Prior to MI, the dataset needs to be converted from a long to the wide format (one record per subject).
- Repeated measures analysis of ANCOVA will be performed on each imputed dataset generated in MI. Prior to analysis of the completed dataset, the dataset needs to be restructured into a long format.
- The parameter estimates from the repeated measures analysis of ANCOVA based on each imputed dataset will be combined using SAS PROC MIANALYZE with MODELEFFECTS statement.

A restricted maximum likelihood-based mixed model for repeated measures (MMRM) under the assumption of MAR will be also considered as sensitivity analysis to test the difference between the overall least squares mean of the treatment arms. The dependent variable of the models will be the scale scores measured at each visit. The model will include treatment arm, visit (coded using integers representing cycle number), treatment-by-visit interaction, randomization stratification factors, the baseline score

(covariate), and baseline score-by-visit interaction as the fixed effect; subject as the random effect. An empirical structure will be assumed for the variance-covariance matrix of the fixed effect parameters and an unstructured covariance matrix will be used for the random effect. The least square mean by treatment arm and the least square mean difference between treatment arms will be reported along with the corresponding 95% CI.

Patient-reported treatment satisfaction (the Satisfaction with Therapy scale of CTSQ) after cycle 4 of treatment: Patient-reported treatment satisfaction are measured by the Satisfaction with Therapy (SWT) scale of the Cancer Therapy Satisfaction Questionnaire (CTSQ) and will be collected on Day 1 of Cycle 2, Cycle 5, Cycle 12 before dosing, and safety follow-up. The scale scores will be converted to range from 0-100, with a higher score associated with the best outcome on each domain. Principles for scoring are detailed in Appendix B.

Patient-reported treatment satisfaction (CTSQ) after cycle 4 of treatment will be analyzed at the corresponding fixed time point using ANCOVA method. The dependent variable of the models will be the scale scores measured at each visit (**Cycle 5 Day 1, Cycle 12 Day 1 and safety follow-up)**. The model will include treatment arm, randomization stratification factors and the **scale score measured at Cycle 2 Day 1**. The mean score difference with 95% CI between treatment arms will be reported. The analyses will be performed for the ITT Analysis Set.

9.5.3 Analyses of Exploratory Efficacy Endpoint(s)

Details regarding the analyses of exploratory endpoints of COA will be provided in a SSAP.

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, vital signs, electrocardiogram (ECG) and left ventricular ejection fraction (assessed by echocardiogram (ECHO)). All safety analyses will be based on the Safety Analysis Set.

9.6.2 Adverse Events

The Medical Dictionary for Regulatory Activities (MedDRA) version 24.1 or later will be used to code all adverse events (AEs) to a system organ class (SOC) and a preferred

term (PT). AEs will be graded for severity using the National Cancer Institute Common Terminology Criteria for Adverse Events (NCI-CTCAE) version 5.0. The events of interest (EOI) search strategies will be based on the standardized MedDRA query (SMQ) and/or Amgen customized MedDRA query (AMQ). Incomplete AE start dates will be imputed according to the specifications described in Appendix A.

The number and percentage of subjects experiencing AEs will be summarized for all treatment-emergent adverse events (TEAEs), grade 3 or higher TEAEs, serious TEAEs, treatment-related TEAEs, TEAEs leading to discontinuation of investigational product/non-investigational product, TEAEs leading to discontinuation of any study drug, TEAEs leading to dose reduction or interruption of investigational product/non-investigational product, TEAEs leading to dose reduction or interruption of any study drug, fatal TEAEs and treatment-emergent EOI. The subject incidence will be presented by treatment arm and tabulated by SOC (in alphabetical order) and/or PT (in descending order of frequency), and/or severity. If a subject experiences repeated episode of the same AE, the subject will be counted only once within each SOC and similarly counted once within each PT and the event with the highest severity grade will be used for purposes of incidence tabulations.

In addition, summaries of TEAEs and serious TEAEs occurring in at least 5% of the subjects by PT in any treatment arm will be provided 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 of study drug, will be provided.

9.6.3 Laboratory Test Results

Laboratory test results will be graded for severity using the NCI-CTCAE version 5.0 and will be summarized for each treatment arm using descriptive statistics for baseline values and changes from baseline values by cycle (per table 2-1 from protocol), and a summary of subject incidence of grade 3 and 4 laboratory abnormalities.

For the summary of changes from baseline values, subjects without a baseline and/or post-baseline value will be excluded; values from unscheduled assessments will not be included. Laboratory results from samples taken > 30 days after the last administration of protocol therapy will be excluded from all laboratory summaries.

Shifts in laboratory toxicity grades to outside the normal range will be evaluated for laboratory parameters by assessing the maximum increase and/or decrease observed

during the course of study treatment relative to the baseline toxicity grade. The following selected laboratory parameters may be considered in this analysis.

(1) hematology analytes in decreasing direction: Hemoglobin, Lymphocyte, Absolute Neutrophil Count (ANC), Platelet, White Blood Cell (WBC);

(2) chemistry analytes in increasing direction: Alanine Aminotransferase (ALT), Aspartate Aminotransferase (AST), Total Bilirubin, Corrected Calcium, Serum Creatinine, Potassium, Sodium, Magnesium, Uric Acid;

(3) chemistry analytes in decreasing direction: Albumin, Corrected Calcium, Potassium, Magnesium, Phosphorus, Sodium.

The subject incidence of Grade 3 and 4 hematological laboratory abnormalities and the subject incidence of Grade 3 and 4 nonhematological toxicities (including liver function test (LFT), creatinine) will be provided by treatment arm in the same table.

The summary table for ALT, AST, Total Bilirubin, and the potential Hy's Law cases will also be considered.

9.6.4 Vital Signs

Vital sign results (systolic/diastolic blood pressure, heart rate, respiratory rate, and temperature) will be summarized using descriptive statistics for baseline values and changes from baseline by cycle for each treatment arm.

For the summary of changes from baseline, subjects without a baseline and/or p**o**st-baseline value will be excluded; values from unscheduled assessments will not be included. Vital sign results taken > 30 days after the last administration of protocol therapy will be excluded from all vital sign summaries.

9.6.5 Physical Measurements

The baseline physical measurements (height (cm), weight (kg), BSA (m^2)) and the change from baseline of weight and BSA will be summarized by cycle for each treatment arm.

9.6.6 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 QTc effect. ECG data might be presented in listings.

9.6.7 Left Ventricular Ejection Fraction (LVEF)

LVEF (assessed by ECHO) and the change of LVEF from baseline will be summarized using descriptive statistics by treatment arm by visit.

9.6.8 Exposure to Investigational Product and Non-Investigational Products

Descriptive statistics will be produced to describe the exposure to all study drugs (carfilzomib, lenalidomide, and dexamethasone) by treatment arm for subjects in the Safety Analysis Set. The extent of exposure will be evaluated, where applicable, with respect to the following:

- Number of treatment cycles subject dosed

  For all study drugs, it is defined as the total number of treatment cycles in which at least one dose of any study drug is administered. For each study drug (carfilzomib, lenalidomide, dexamethasone), it is defined as the total number of treatment cycles in which at least one dose of carfilzomib/lenalidomide/dexamethasone, respectively, is administered.

- Number of subjects dosed in each cycle for all study drugs and each study drug
- Treatment duration (week): (last dose date – first dose date + 1) / 7

  For all study drugs, the last/first dose date refers to the last/first dose date of any study drug. For carfilzomib/lenalidomide/dexamethasone, the last/first dose date refers to the last/first dose date of carfilzomib/lenalidomide/dexamethasone, respectively.

- Number of doses administered (non-zero dose) of each study drug during the treatment period of the study
- Cumulative dose received (mg, mg/m^2) of each study drug during the treatment period of the study, defined in Section 5.
- Average dose per administration (mg, mg/m^2) of each study drug during the treatment period of the study, defined as the total cumulative dose received divided by the number of doses administered.
- Average dose per administration (mg, mg/m^2) excluding the 20 mg/m^2 of carfilzomib on cycle 1 day 1 and/or day 2 during the treatment period of the study, defined as the total cumulative dose received divided by the number of doses administered.
- Relative dose intensity of each study drug defined in Section 5.

- The number (%) of subjects with dose modifications of each study drug will be tabulated and the reasons for dose modification will also be summarized. If the reason for dose modification is COVID-19 control measures which is recorded in the specified field (Other) on the CRF, then the number (%) with COVID-19 control measures will also be presented.

  Dose modifications will include the following:

  − Dose missed (for each study drug, derived based on the Carfilzomib/ Lenalidomide/Dexamethasone Investigational Product Administration CRF)

  − Dose reduction (each study drug, derived based on the Carfilzomib/ Lenalidomide/Dexamethasone Investigational Product Administration CRF)

  − Dose delay (for carfilzomib only, as captured on the Investigational Product Administration (Carfilzomib) CRF)

  − Dose interruption (for carfilzomib only, as captured on the Investigational Product Administration (Carfilzomib) CRF)

The primary reason for study drug discontinuation will be summarized along with the summary of subject disposition (Section 9.2).

9.6.9 Exposure to Concomitant Medication

The number and proportion of subjects receiving concomitant medications from study day 1 through 30 days of the last dose of any study drug will be summarized by preferred term or category for each treatment arm as coded by the World Health Organization Drug (WHO DRUG) dictionary by treatment arm in the Safety Analysis Set. For the purpose of determining if a medication should be noted as a concomitant medication, the imputation rules stated in Appendix A will be used.

9.7 Other Analyses

[redacted]

9.7.2 Analyses of Clinical Outcome Assessments (COA)

The analyses of the secondary endpoints of COA (patient-reported convenience after cycle 4 of treatment, physical functioning and role functioning (EORTC QLQ-C30) over time, and patient-reported treatment satisfaction (CTSQ) after cycle 4 of treatment) are described in Section 9.5.2. Details regarding the analyses of exploratory endpoints of COA will be provided in a supplemental statistical analysis plan by the Department of Health Economics and Outcomes Research (HEOR).

9.7.3 Analyses of Biomarker Endpoints

Analyses of biomarker endpoints MRD[-] CR rate and MRD[-] rate are described in Section 9.5.2.

10. Changes from Protocol-specified Analyses

The RMST and observed treatment effect retention rate are included as additional analyses for PFS. The non-inferiority margin for ORR is specified in Section 9.5.

11. Literature Citations / References

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.

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.

European Medicines Agency (EMA). Guideline on the Choice of the Non-Inferiority Margin. July 2005.

European Medicines Agency (EMA). Appendix 1 to the guideline on the evaluation of anticancer medicinal products in man - methodological consideration for using progression-free survival (PFS) or disease-free survival (DFS) in confirmatory trials - Scientific guideline. July 2013.

FDA. Non-inferiority clinical trials to establish effectiveness. Guidance for Industry. November 2016.

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

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; 17:328-346.

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. 1991; 9:2042-2051.

Lu, J., and Pajak, T. F. Statistical Power for a Long-Term Survival Trial with a Time-Dependent Treatment Effect, Controlled Clinical Trials. 2000; 21, 561–573.

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 S, et al. Consensus recommendations for the uniform reporting of clinical trials: report of the International Myeloma Workshop Consensus Panel 1. Blood. 2011; 117:4691-4695.

Rubin DB. Inference and missing data. Biometrika. 1976; 63: 581-592.

SAS Institute I. Sas/stat 9.4 user's guide: the MI procedure. Cary (NC): SAS Institute Inc. 2015.

Schemper M, Smith TL. A note on quantifying follow-up in studies of failure time. Controlled Clinical Trials. 1996; 17:343–346.

Uno H, et al. Alternatives to hazard ratios for comparing the efficacy or safety of therapies in noninferiority studies. Ann Intern Med. 2015; 163:127-134.

Weir IR and Trinquart L. Design of non-inferiority randomized trials using the difference in restricted mean survival times. Clin Trials. 2018; 15(5): 499-508

Wellner JA and Zhan Y. A hybrid algorithm for computation of the NPMLE from censored data. J. Amer. Statist. Assoc. 1997; 92(439): 945-959

12. Appendices

Appendix A. Handling of Incomplete Dates and Missing Dates

A1. Imputation Rules for Adverse Events and Concomitant Medications (other than the new anti-myeloma therapy) dates

Imputation Rules for Partial or Missing Start Dates:

Start Date		Stop Date						
		Complete: *yyyymmdd*		Partial: *yyyymm*		Partial: *yyyy*		
		< 1st dose	**≥ 1st dose**	**< 1st dose *yyyymm***	**≥ 1st dose *yyyymm***	**< 1st dose *yyyy***	**≥ 1st dose *yyyy***	Missing
Partial: *yyyymm*	= **1st dose** *yyyymm*	2	1	**n/a**	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. **For subjects who were never treated (first dose date is missing), partial start dates will be set to the first day of the partial month or first day of year if month is also missing.**

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. (i.e., set the stop date as missing).

A2. Imputation Rules for New Antimyeloma Therapy Start Date

If the start day of new antimyeloma 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 antimyeloma therapy is missing and month and year are same as last dosing date of study treatment, the start date will be assumed as last dosing date of study treatment. In other situations, do not impute.

A3. Imputation Rules for Partial or Missing Death Dates:

1. If death year and month are available but day is missing:
    - If mmyyyy for last known alive date = mmyyyy for death date, set death date to the day after the last known alive date.
    - If mmyyyy for last known alive date < mmyyyy for death date, set death date to the first day of the death month.
    - If mmyyyy for last known alive date > mmyyyy for death date, data error and do not impute.
2. If death year is available but both month and day are missing for death date:
    - If yyyy for last known alive date = yyyy for death date, set death date to the day after the last known alive date.
    - If yyyy for last known alive date < yyyy for death date, set death date to the first day of the death year.
    - If yyyy for last known alive date > yyyy for death date, data error and do not impute.
3. If a death date is totally missing, do not impute.

The imputed death date will be used in calculation of duration of response, PFS **and OS**.

A4. Imputation Rules for Dates of Prior Multiple Myeloma Therapy and Relapse/progression to Prior Multiple Myeloma Therapy:

If the day of prior multiple myeloma therapy or relapse/progression to prior multiple myeloma therapy is missing but month and year are available, then impute the date to 15th of the month. **If the date imputation leads to a stop date that is before the start date, then do not impute the date.** If month or year is missing or the date is completely missing, do not impute.

Appendix B. Clinical Outcome Assessment Forms/Instruments

EORTC QLQ-C30 Functional Scale Scoring:

The following sections describe the scoring algorithms for functional scales used in EORTC QLQ-C30 questionnaire. Scoring procedures 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.

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 12-1 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. For the functional scales in QLQ-C30, a higher score represents a better health state.

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

where range for each scale is defined in Table 12-1.

Table 12-1. EORTC QLQ-C30 Functional Scales and Scoring Details

	Number of Items	Item Range [a]	Item Numbers	Minimum Not Missing
QLQ-C30				
Functional Scales				
Physical Functioning	5	3	1 to 5	3
Role Functioning	2	3	6, 7	1
Emotional Functioning	4	3	21 to 24	2
Cognitive Functioning	2	3	20, 25	1
Social Functioning	2	3	26, 27	1

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

Only Physical Functioning and Role Functioning will be included in the analysis.

CTSQ SWT Scale Scoring:

The following sections describe the scoring algorithms for SWT scale used in CTSQ questionnaire. Scoring procedures can be found in the CTSQ Administration & Scoring Guide v1.0 (2006). These procedures result in a score ranging from 0 to 100 for each CTSQ domain, with a higher score associated with the best outcome on each domain.

Table 12-2 provides a summary of the domain structure and critical information required for scoring SWT domain in CTSQ.

Table 12-2. Scoring information for SWT Domain in CTSQ

CTSQ Domain	Description of Content of Items in Domain	Item numbers*	Total # of items	Minimum # of completed items required to score
Satisfaction with Therapy (SWT)	Worth taking even with side effects, Think about stopping CT, How worthwhile was CT, Benefits meet expectations, Satisf. with form of CT, Satisf. with recent CT, Would you take this CT again	Q7, Q9R, Q10, Q12, Q14, Q15, Q16	7	5

* "R" following item number indicates that reverse-coded version of the item is used in calculating the domain score.

Step 1. Reverse-Coding Required for the Item

The first step in scoring the SWT is to create new variables containing the reverse-coded response values for the CTSQ items Q9. This is done by subtracting the initial (raw) response value for each of these items from 6:

```
Q9R = 6 - Q9;
```

Creating new variables containing the reverse-coded response values for the item ensures that the highest-coded value (5) for the item is associated with the best possible response (greater satisfaction with therapy), and the lowest-coded value (1) is associated with the worst possible response.

Step 2. Scoring Procedures

If the number of completed items is greater than or equal to the minimum number indicated in Table 12-2 the domain is scored using the formula:

Domain score = [(Sum of completed item responses / Number of completed items) - 1] x 100 / (Maximum possible item response value – Minimum possible item response value)

However, if fewer items are completed than the minimum number indicated in Table 12-2, then the domain is not scored (i.e. a missing value is assigned).

Since the maximum possible item response value is 5 and the minimum possible response value is 1 for all CTSQ items, a simpler way to represent the above formula for the CTSQ domains is:

CTSQ domain score = (Mean of completed item responses – 1) x 25

In terms of SAS programming code, the scoring procedure can be performed as follows:

```
if n(of Q7 Q9R Q10 Q12 Q14 Q15 Q16) >= 5 then SWT = (mean(of Q7 Q9R Q10
Q12 Q14 Q15 Q16)-1)*25;
else SWT = . ;
```