Statistical Analysis Plan

Study ID: 207499-JPN
Study Official Title: A Phase III, Multicenter, Open-Label, Randomized Study to Evaluate the Efficacy and Safety of Belantamab Mafodotin in Combination With Pomalidomide and Dexamethasone (B-Pd) Versus Pomalidomide Plus Bortezomib and Dexamethasone (PVd) in Participants With Relapsed/Refractory Multiple Myeloma (DREAMM 8)
NCT ID: NCT06956170
Date of Document: 13-Feb-2024

Note: The global study, 207499 (DREAMM-8), registered under NCT04484623, is a Multi-Regional Clinical Trial (MRCT), that includes a Japanese expansion cohort registered separately under NCT06956170. In the Japan SAP addendum on page 1, the global study NCT ID is referenced, as the development of this SAP preceded the availability of the Japan cohort NCT ID.

Additionally, for ease of access, the Global Study SAP has been appended within the same document for reference.

Information Type:	Statistical Analysis Plan (SAP)

TITLE PAGE

Protocol Title: Statistical Analysis Plan for: A Phase III, Multicenter, Open-Label, Randomized Study to Evaluate the Efficacy and Safety of Belantamab Mafodotin in Combination with Pomalidomide and Dexamethasone (B-Pd) versus Pomalidomide plus Bortezomib and Dexamethasone (PVd) in Participants with Relapsed/Refractory Multiple Myeloma (DREAMM 8)

This is the supplemental Statistical Analysis Plan (SAP) for Japan and North-East Asia (NEA) sub-population analysis and Japan expansion cohort analysis which will be produced to support reporting requirement for an expected regulatory submission in Japan.

Study Number: 207499

Compound Number: GSK2857916

Abbreviated Title: A Phase III Study of Belantamab Mafodotin plus Pomalidomide and Dexamethasone vs. Pomalidomide, Bortezomib and Dexamethasone in Participants with RRMM

Acronym: DREAMM8

Sponsor Name: GlaxoSmithKline Research & Development Limited

Regulatory Agency Identifier Number(s)

Registry	**ID**
Clinicaltrials.gov	NCT04484623
EudraCT	2018-004354-21
EU CT	2023-506877-37-00
IND	119333

Statistical Analysis Plan (SAP) Template v3.0 14 September 2022

TABLE OF CONTENTS

PAGE

LIST OF TABLES

PAGE

VERSION HISTORY

SAP Version	Approval Date	Protocol Version (Date) on which SAP is Based	Change	Rationale
SAP for Japan subgroup	24 May 2023	Protocol Amendment 3 (23 February 2023)	Not Applicable	Original version
SAP for Japan subgroup Amendment 1	13 Feb 2024	Protocol Amendment 4 (28 September 2023)	Changed the analysis timing of NEA analysis and Japan expansion cohort analysis. Added some analyses to be aligned with the OPS addendum.	Aligned with the protocol amendment. To meet the requirement for Japan submission.

1. INTRODUCTION

The purpose of this SAP addendum is to describe the planned analyses for Japan sub-population, North-East Asia (NEA) sub-population (NEA pooled region) and Japan expansion cohort, which will be produced to support the reporting requirement for an expected regulatory submission in Japan.

NEA sub-population analysis is planned with reference to ICH E17 guideline to assess the consistency between overall population and NEA sub-population.

This SAP addendum is conjunction with the SAP amendment for study 207499 approved in 11 October 2023 (original SAP). Unless specified in this SAP addendum, any analysis rules and definitions in Japan sub-population analysis, NEA sub-population and Japan expansion cohort analysis will be aligned with the ones described in the original SAP.

Additional details with regards to data handling conventions and the specification of data displays will be provided in the separate Output and Programming Specification (OPS) document.

2. STATISTICAL ANALYSES

2.1. General Considerations

2.1.1. General Methodology

Any considerations for data analyses, data handling conventions and methodology of analyses will follow the original SAP, unless otherwise specified.

Intent-to-Treat (ITT) analysis set will be used for all study population analyses and efficacy analyses, unless otherwise specified. Safety analysis set will be used for all safety analyses. PK analysis set will be used for all PK analyses. Analysis populations are defined in Section 2.3.

Unless otherwise specified, continuous data will be summarized using descriptive statistics: n, mean, standard deviation (std), median, minimum, and maximum. Categorical data will be summarized as the number and percentage of participants in each category.

Confidence intervals (CIs) will use 95% confidence levels unless otherwise specified.

2.1.2. Definition of sub-population and All Japan population

The Japan and Korea sub-populations will be defined as the population who enrolled in the D8 main study from Japan and Korea as country respectively. NEA sub-population will be defined as the population who enrolled in the D8 main study from Japan, China and Korea as country. Japan or NEA sub-population does not include the participants who enrolled in Japan expansion cohort.

All Japan population will include participants who enrolled in either D8 main study from Japan or Japan expansion cohort.

2.2. Planned Analysis

This supplemental analysis for Japan / NEA sub-population and all Japanese population will be conducted as shown in Table 1 below.

- Japan sub-population analysis will be conducted at the timing of IA2 (if PFS demonstrate statistical significance at IA2 in the main study) or the primary PFS analysis (otherwise) of D8 main study. This analysis is considered as the primary for Japanese regulatory submission.
- NEA sub-population analysis will be also conducted at the timing of IA2 (if PFS demonstrate statistical significance at IA2 in the main study) or the primary PFS analysis (otherwise) of D8 main study. This analysis is considered as the supportive analysis for Japanese submission.
- All Japan population analysis will be conducted at the timing of having enough number of PFS events in all Japanese population (i.e., approximately 10 events) as described in the Japan specific study protocol. If PFS does not demonstrate statistical significance at IA2, the analysis for all Japan population will not be conducted until the primary PFS analysis even if enough number of PFS events have been observed. Also, if PFS does not demonstrate statistical significance at both IA2 and the primary PFS analysis, the analysis for all Japan population will be conducted at the same timing as the primary PFS analysis. The analysis for all Japan population may be conducted before observing approximately 10 PFS events depending on the outcome of communications with health authorities based on the result of main study. This analysis is considered as the supportive analysis for Japanese submission. The analysis for all Japan population will be also conducted at the same timing/ after the timing for the final Overall Survival (OS) analysis of D8 study. Not only for the subjects in Japan expansion cohort but also for the subjects in the main study the data observed until the data cut-off for All Japan population analyses will be used.

Table 1 Summary of cohort and sub-population analyses

Cohort/Sub-population	At the timing of IA2 or the primary PFS analysis of D8 main study	At the timing that Japan expansion cohort has enough PFS events*	At the timing of the final OS analysis of D8 main study
Japan sub-population	Y (as primary for Japan submission)	N	N
NEA sub-population	Y	N	N
All Japan population	N	Y	Y

*: The analysis for all Japan population may be conducted before observing approximately 10 PFS events depending on the outcome of communications with health authorities based on the result of main study.

2.3. Analysis Sets

Analysis sets for Japan sub-population and NEA sub-populations will be aligned with original SAP.

For All Japan population, analysis sets will be defined as below.

Table 2 Analysis Sets for All Japan population

Analysis Set	Definition / Criteria	Analyses Evaluated
All Screened (All Japan)	• The All Screened Population will consist of all participants who sign the ICF to participate in the clinical trial. Participants in this population will be used for screen failure summary. • This population includes only the participants who were screened in either D8 main study from Japan or Japan expansion cohort.	Study Population
Enrolled (All Japan)	• The Enrolled population is defined as all subjects that have entered the study (e.g. subjects that are identified on the Screen Failure form as non-screen failures). • This population includes only the participants who enrolled in either D8 main study from Japan or Japan expansion cohort.	Study Population
Safety (All Japan)	• All randomized participants who take at least 1 dose of study treatment. Participants will be analyzed according to the treatment they actually received. • This population includes only the participants who were randomized in either D8 main study from Japan or Japan expansion cohort.	Safety
Intent-to-Treat (All Japan) (ITT All Japan)	• ITT Population will consist of all randomized participants whether or not randomized treatment was administered. This population will be based on the treatment to which the participant was randomized and will be the primary population for the analysis of efficacy data. Any participant who receives a treatment randomization number will be considered to have been randomized. • This population includes only the participants who were randomized in either D8 main study from Japan or Japan expansion cohort.	Efficacy Study Population
Pharmacokinetic (All Japan)	• The Pharmacokinetic Population will consist of those participants in the Safety Population from whom at least 1 PK sample has been obtained and analyzed. This population will be the primary population for PK analyses. • Data should be reported according to the actual treatment. • This population includes only the participants who were randomized in either D8 main study from Japan or Japan expansion cohort.	PK

2.4. Study Population Analyses

Unless otherwise specified, the study population analyses will be based on the ITT Analysis Set.

Within the summaries planed in the original SAP, the following data will be summarized for Japan sub-population, NEA sub-population and All Japan population.

- Subject disposition and reasons for study withdrawal
- Treatment status and reasons for treatment discontinuation
- Demographic characteristics
- Disease characteristics at Screening
- Duration of follow-up

For the analysis using All Japan population, the displays related to following data planned in the original SAP will be also created.

- Screening status and reasons for screen failure
- Important protocol deviation
- Study populations
- Concomitant medications
- Anti-myeloma therapy

2.5. Primary Endpoint Analyses

2.5.1. Definition of Endpoint

PFS is the primary endpoint of this study; it is defined as the time from randomization until the earliest date of Progressive disease (PD), or death due to any cause. The analyses of PFS will be based on the ITT Analysis Set, unless otherwise specified. PFS analysis will be based on responses per IMWG 2016 according to the Independent Review Committee (IRC) assessment.

2.5.2. Main Analytical Approach

The non-parametric Kaplan-Meier method will be used to estimate the survival curves for PFS. Kaplan-Meier plots of PFS will be presented by treatment arm for Japan sub-population, NEA sub-population and All Japan population. Kaplan-Meier estimates for the median PFS, the first and third quartiles and 6, 12 and 18 months PFS rate will be presented, along with 95% CIs. CIs for quartiles will be estimated using Brookmeyer-Crowley method [Brookmeyer, 1982].

The hazard ratio with its corresponding 95% CI for the Japan sub-population, NEA sub-population and All Japan population will be estimated respectively using unstratified Cox proportional hazard model with treatment arm only as the explanatory variable.

Only primary analysis of primary estimand will be conducted for the Japan sub-population, NEA sub-population and All Japan population. The definition of estimand is referred to the original SAP.

2.6. Secondary Endpoint(s) Analyses

Secondary endpoints will be analyzed based on the ITT Analysis Set, unless otherwise specified. Due to the small sample size, adjustment of any stratification factors and covariates will not be considered for Japan sub-population, NEA sub-population and All Japan population. Some of secondary endpoints defined in protocol of study 207499 will not be analyzed due to the small sample size.

For OS (Overall Survival), Kaplan-Meier estimates for the median, the first and third quartiles and 6, 12 and 18 months survival rate will be presented, along with 95% CIs for the NEA sub-population and All Japan population. CIs for quartiles will be estimated using Brookmeyer-Crowley method. The hazard ratio with its corresponding 95% CI for the NEA sub-population and All Japan population will be estimated respectively using unstratified Cox proportional hazard model with treatment arm only as the explanatory variable.

For DoR (Duration of Response), the non-parametric Kaplan-Meier method will be used to estimate the survival curves. Kaplan-Meier plots for DoR will be presented by treatment arm for the NEA sub-population. Kaplan-Meier estimates for the median, the first and third quartiles rate will be presented, along with 95% CIs for Japan sub-population, NEA sub-population and All Japan population. CIs for quartiles will be estimated using Brookmeyer-Crowley method.

For MRD (Minimal Residual Disease) negativity rate, the corresponding exact 95% CI, if applicable, will be provided by treatment arm and Best Response for All Japan population. Participants with unknown or missing responses will be treated as non-responders, i.e., these participants will be included in the denominator when calculating percentages of response.

For ORR (Overall Response Rate) based on IRC-assessment per IMWG as the Best Overall Response (BOR), the corresponding exact 95% CI will be provided by treatment arm and the exact 95% CI for the difference between treatment arm will be calculated for Japan sub-population, NEA sub-population and All Japan population.

TTR (Time to Response) will be summarized descriptively by treatment arm for Japan sub-population, NEA sub-population and All Japan population.

DoR and TTR will be assessed among participants with a confirmed PR or better as the BOR. DoR, MRD, TTR and ORR for Best Response will be analyzed based on IRC-assessment.

For the analysis using All Japan population, the displays for other efficacy endpoints planned in the original SAP (e.g. Time to Progression, Time to Best Response) will be also created. The displays will be identified in the OPS (TOC) for this SAP.

2.7. Safety Analyses

The safety analyses will be based on the Safety Analysis Set, unless otherwise specified.

The exposure related data will be summarized depending on treatment arm for Japan sub-population, NEA sub-population and All Japan population. The summaries to be created will be shown in OPS (TOC) for this SAP.

Within the summaries for adverse events planned in the original SAP, the followings will be summarized by treatment arm for Japan sub-population, NEA sub-population and All Japan population.

- Adverse event overview
- All adverse events
- Adverse events of maximum grade 3 or higher
- Drug-related adverse events
- Belantamab Mafodotin-related adverse events
- Death (not summarized for Japan sub-population)
- Serious adverse events
- Adverse events leading to permanent discontinuation of study treatment
- Adverse events leading to dose reduction
- Adverse events leading to dose interruption/delay

Regarding AESI, the following events will be summarized by treatment arm for Japan sub-population, NEA sub-population and All Japan population.

- Thrombocytopenia
- Thrombocytopenia and bleeding events
- Infusion-related reactions
- Corneal adverse events

For ocular findings, the KVA (Keratopathy Visual Acuity) scale events overview will be summarized for Japan sub-population, NEA sub-population and All Japan population.

Anti-GSK2857916 anti body will be summarized by treatment and planned time for Japanese sub-population, NEA sub-population and All Japan population.

For the analysis using All Japan population, the further displays for safety (e.g. other summary related to adverse events, laboratory data and ECOG performance status) planned in the original SAP will be also created. The displays will be identified in the OPS (TOC) for this SAP.

2.8. Pharmacokinetic Analyses

The pharmacokinetic analyses will be based on the Pharmacokinetic Analysis Set, unless otherwise specified.

Within the summaries planned in the original SAP, the followings will be summarized. The population for each summary is also described below.

- Plasma GSK2857916 ADC PK concentration: Japan sub-population, NEA sub-population, All Japan population, Korea sub-population
- Derived GSK2857916 ADC PK parameters: Japan sub-population, NEA sub-population, All Japan population, non-Japan sub-population, non-NEA sub-population, Korea sub-population
- Plasma cys-mcMMAF PK concentration: Japan sub-population, NEA sub-population, All Japan population, Korea sub-population
- Derived cys-mcMMAF PK parameters: Japan sub-population, NEA sub-population, All Japan population, non-Japan sub-population, non-NEA sub-population, Korea sub-population

The following figures will be created. The population for each plot is also described below.

- Median Plasma GSK2857916 (ADC) Concentration: Japan sub-population, NEA sub-population, All Japan population, Korea sub-population
- Median Plasma GSK2857916 (cys-mcMMAF) Concentration: Japan sub-population, NEA sub-population, All Japan population, Korea sub-population
- Median Pomalidomide PK Concentration: Japan sub-population

2.9. Exploratory Endpoint(s) Analyses

Any exploratory endpoints except for pharmacokinetic analyses will not be analyzed for Japan sub-population, NEA sub-population and All Japan population due to small sample size.

3. REFERENCES

Brookmeyer R, Crowley J. A confidence interval for the median survival time. Biometrics. 1982 Mar 1:29-41.

Information Type:	Statistical Analysis Plan (SAP)

TITLE PAGE

Protocol Title:	Statistical Analysis Plan for**:** A Phase III, Multicenter, Open-Label, Randomized Study to Evaluate the Efficacy and Safety of Belantamab Mafodotin in Combination with Pomalidomide and Dexamethasone (B-Pd) versus Pomalidomide plus Bortezomib and Dexamethasone (PVd) in Participants with Relapsed/Refractory Multiple Myeloma (DREAMM 8)
Study Number:	207499
Compound Number:	GSK2857916
Abbreviated Title:	A Phase III Study of Belantamab Mafodotin plus Pomalidomide and Dexamethasone vs. Pomalidomide, Bortezomib and Dexamethasone in Participants with RRMM
Acronym:	DREAMM 8

Sponsor Name: GlaxoSmithKline Research & Development Limited

Regulatory Agency Identifier Number(s)

Registry	**ID**
Clinicaltrials.gov	NCT04484623
EudraCT	2018-004354-21
EU CT	2023-506877-37-00
IND	119333

Statistical Analysis Plan (SAP) Template v3.0 14 September 2022

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

VERSION HISTORY

SAP Version	Approval Date	Protocol Version (Date) on which SAP is Based	Change	Rationale
SAP Version 3	02 Feb 2024	Protocol Amendment 4 (28 Sep 2023)	• Section 2, Section 4.7, Section 5 and throughout: Revised multiplicity strategy. Equivalent to assigning 0 weight to DoR in the weighted Bonferroni procedure defined within the protocol, a hierarchical procedure will be used such that OS is tested at 2.5% level conditional on successful rejection of the null hypothesis associated with PFS only, and MRD Negativity is tested at 2.5% level conditional on successful rejection of the null hypothesis associated with OS. DoR will remain key secondary but will not be formally tested (descriptive only). • Other changes: minor clarifications, restructuring of content and administrative updates. Revisions to supportive listings, to provide more focused outputs. • Section 4.1.2: added baseline definition for immunogenicity endpoint analyses. • Section 4.2.2 and Section 4.3.2.1: Clarified that an adequate assessment is defined as an assessment where the *confirmed* response is sCR, CR, PR, VGPR, MR or SD. In Section 4.2.2 also added details of the IDMC decision process for IA2, which will consider both PFS statistical significance and OS HR<1 for recommendation of early stopping due to efficacy. • Section 4.3.1.1: clarified the OS last contact date derivation used for censoring. • Section 4.3.1.2: Clarification on SAS code options due to changes in SAS versions. • Section 4.3.2.1: Updated confirmed response algorithm for alignment with IMWG, in particular, when unconfirmed PDs occur and to ensure confirmed response is present for final assessment.	Acknowledging the importance of OS to patients. Administrative updates to add clarification and/or remove discrepancies. Clarifications and additional details added for planned analyses. Revised approach to summaries and analysis based on emerging data and prioritization.

SAP Version	Approval Date	Protocol Version (Date) on which SAP is Based	Change	Rationale
			• Section 4.3.5.3: Added OSDI compliance, clarified the analysis sets to be used for compliance and that compliance will only be reported until treatment discontinuation. • Section 4.4.3.4: Included summary of worst-case post-baseline for FACT-GP5. • Section 4.5.1: updated planned dose intensity definition, with additional supportive relative dose intensity calculation added for belantamab mafodotin and bortezomib. Sensitivity analysis based on derived delays added. For duration of follow-up, clarified the last contact date used for duration of follow-up, and added summary for ongoing participants. • Section 4.5.2: Added a benefit-risk plot and summary of grade 3+ AEs by SOC and PT. • Section 4.5.2.4: Added visit-slotting for ocular data. Added outputs for unilateral and bilateral worsening in vision, clarified logic for corneal exam finding indicators if overall exam findings are normal. Added and removed ocular related outputs following a deepened understanding of important data for analysis. • Section 4.6.1: Added subgroup 'prior exposure to lenalidomide and anti-CD38 mAb' and redefined the 1 prior line with relapse subgroup. • Section 4.8: Specified changes to protocol defined analyses - added clarification on adequate assessment and updated multiplicity strategy. • Section 6.1: Added details and clarifications on study population analyses, including baseline disease characteristics of interest, focused prior and concomitant medication and prior and subsequent anti-MM therapy summary tables, to avoid redundancies. • Section 6.2.9.1: restructured and clarified the derivation of sponsor assessed KVA grade	

SAP Version	Approval Date	Protocol Version (Date) on which SAP is Based	Change	Rationale
SAP Version 2	11 Oct 2023	Protocol Amendment 4 (28 Sep 2023)	• Updates made to align with protocol changes throughout: Key Secondary endpoint/estimand ordering updated to align with multiplicity strategy, sequence/timing of planned analyses clarified, endpoint definitions updated, multiplicity strategy revised and further details added, second line analysis removed, unblinded data access for population PK dataset creation specified. • Other changes: minor clarifications, restructuring of content and administrative updates. • Section 4.1.1 General Methodology: Central vs local laboratory data use clarified. Pooling strategy for stratification clarified. • Section 4.2.4 Sensitivity Analyses: details of IRC and Investigator agreement evaluation included and pooling strategy for sensitivity analyses clarified. • Section 4.3.1 Key Secondary Endpoint(s): OS definition clarified in consideration of deaths obtained beyond DCO and beyond study discontinuation/withdrawal. Ratio of RMDOR included. OS analysis using IPCW method moved to supplementary SAP. Supportive summary of MRD negativity rate by best overall response included. • Section 4.3.2 Supportive Secondary Endpoints: BOR assessment clarified for participants without measurable disease at baseline. TTP endpoint rates at a fixed time removed for TTP and added for PFS2. • Section 4.3.5 Secondary Patient Reported Outcome Analyses: Visit slotting rules added. • Section 4.4.3Exploratory Patient Reported Outcome Analyses: Visit slotting rules added. Time to onset of event and duration of events summaries added. Planned analysis of driving and reading questions changed to a worst-case post-baseline shift summary.	Requirement for increased OS data maturity at the time of Primary PFS analysis and acknowledgment of the importance of OS endpoint. Resulted in: Addition of interim analyses for PFS and OS; Order of key secondary endpoints changed; Multiplicity adjustment strategy detailed; Increased targeted number of PFS events. Administrative updates to add clarification and/or remove discrepancies. Revised approach to supportive summaries based on emerging data and prioritization.

SAP Version	Approval Date	Protocol Version (Date) on which SAP is Based	Change	Rationale
			• Section 4.4.4 Exploratory MRD Negativity Endpoints: Visit window for sustained MRD negativity defined. Supportive plots for SPEP and UPEP removed. • Section 4.5.1 Extent of Exposure: Duration of treatment plot removed. Dose intensity calculation modified. Data capture and reporting of dose delays clarified. Patient profile plot of responders with extended dose delay removed. • Section 4.5.2 Adverse Events: Added and removed supportive adverse event/ocular summaries. Planned KVA analyses clarified. • Section 4.5.3.1 Laboratory Data: Removed supportive plots of laboratory data. • Section 4.7 Interim Analyses: Revised to align with protocol. Stopping boundaries for PFS and OS efficacy analyses added. • Section 5 Sample Size Determination: Updated to include an interim analysis and revised targeted number of events. Clarified sample-size re-estimation performed prior to decision to revise targeted PFS events. • Section 6.2 Appendix 2 Data Derivation Rules: Revised assessment window for PROs. Removed images of PRO scoring.	
SAP Original Version	13 Mar 2023	Protocol Amendment 3 (23 Feb 2023)	Not Applicable	Original version

1. INTRODUCTION

The purpose of this SAP (Version 3) is to describe the planned analyses to be included in the CSR for Study 207499. Additional detail with regards to data handling conventions and the specification of data displays will be provided in the Output and Programming Specification (OPS) document.

1.1. Objectives, Estimands and Endpoints

1.1.1. Objectives and Endpoints

Objectives	Endpoints
Primary	
To compare the efficacy of B-Pd with that of PVd in participants with RRMM	• Progression-Free Survival (PFS), defined as the time from randomization until the earliest date of PD based on IRC-assessment per IMWG criteria, or death due to any cause.
Key Secondary	
To further compare the efficacy of B-Pd with that of PVd in participants with RRMM	• Overall Survival (OS), defined as the interval of time from randomization to the date of death due to any cause.
	• Duration of Response (DoR), defined as the time from first documented evidence of PR or better until progressive disease (PD) or death due to any cause. Response will be based on IRC-assessment per IMWG criteria.
	• MRD negativity rate, defined as the percentage of participants who achieve MRD negative status (as assessed by NGS at 10^{-5} threshold) at least once during the time of confirmed CR or better response based on IRC-assessment per IMWG.

Objectives	Endpoints
Secondary	
To further assess the efficacy of B-Pd in terms of other efficacy outcomes in participants with RRMM	• Overall Response Rate (ORR), defined as the percentage of participants with a confirmed partial response (PR) or better (i.e., PR, VGPR, CR, and sCR) based on IRC-assessment per IMWG criteria. • Complete Response Rate (CRR), defined as the percentage of participants with a confirmed complete response (CR) or better (i.e., CR and stringent complete response (sCR)) based on IRC-assessment per IMWG criteria. • Very Good Partial Response (VGPR) or better rate defined as the percentage of participants with a confirmed VGPR or better (i.e., VGPR, CR, and sCR) based on IRC-assessment per IMWG criteria. • Time to Best Response (TTBR), defined as the interval of time between the date of randomization and the earliest date of achieving best response among participants with a confirmed PR or better based on IRC-assessment per IMWG. • Time to Response (TTR), defined as the time between the date of randomization and the first documented evidence of response (PR or better) among participants who achieve a response (i.e., confirmed PR or better) based on IRC-assessment per IMWG. • Time to Progression (TTP), defined as the time from the date of randomization until the earliest date of documented PD based on IRC-assessment per IMWG criteria, or death due to PD. • Progression-Free Survival following initiation of new anti-myeloma therapy (PFS2), defined as time from randomization to disease progression (investigator-assessed) after initiation of new anti-myeloma therapy or death from any cause, whichever is earlier. If disease progression after new anti-myeloma therapy cannot be measured, a PFS event is defined as the date of discontinuation of new anti-myeloma therapy, or death from any cause, whichever is earlier
To evaluate the safety and tolerability of B-Pd	• Incidence of AEs and changes in laboratory parameters • Ocular findings on ophthalmic exam
To describe the exposure to belantamab mafodotin after infusion	• Plasma concentrations of belantamab mafodotin, and cys-mcMMAF
To evaluate the PK of pomalidomide in combination with belantamab mafodotin and dexamethasone, in a subset of participants	• Derived PK parameter values, as data permit
To assess ADAs against belantamab mafodotin	• Incidence and titers of ADAs against belantamab mafodotin

Objectives	Endpoints
To evaluate the safety and tolerability of belantamab mafodotin based on self-reported symptomatic adverse effects when administered in combination with pomalidomide and dexamethasone	• Maximum post-baseline PRO-CTCAE score for each item attribute
To evaluate and compare changes in symptoms and HRQoL	• Change from baseline in HRQoL as measured by EORTC QLQC30, EORTC QLQ-MY20* and EORTC IL52*
Exploratory	
To further evaluate the safety and tolerability of belantamab mafodotin when administered in combination with pomalidomide and dexamethasone	• Changes in safety assessments, including vital signs
To further characterize the PK profile of belantamab mafodotin when administered in combination with pomalidomide and dexamethasone	• Derived PK parameter values for belantamab mafodotin and cys-mcMMAF, as data permit
To evaluate self-reported ocular symptomatic AEs of belantamab mafodotin when administered in combination with pomalidomide and dexamethasone	• Changes from baseline in symptoms and related impacts as measured by OSDI
To further evaluate and compare changes in HRQoL and symptoms	• Change from baseline in HRQoL as measured by EQ-5D-3L • Change from baseline in PGIS and change over time in PGIC
To further evaluate the impact of side effects on QoL	• Change from baseline in FACT-GP5
To further explore the efficacy in terms of MRD-negativity	• Sustained MRD negativity rate: defined as the percentage of participants who achieve MRD negative status assessed by NGS at 10^{-5} threshold at least twice, a minimum of 12 months apart and with no MRD positive (or indeterminate) result in between, during the time of confirmed CR or better response per IRC-assessment according to IMWG. • Imaging plus MRD-negativity rate, defined as the percentage of participants who achieve MRD negative status assessed by NGS at 10^{-5} threshold and have no evidence of disease on PET-CT at least once during the time of confirmed CR or better response per IRC-assessment according to IMWG.
To evaluate and compare nonprotocol specified HCRU	• Out-patient visits by physician specialty • Emergency room visits • Home healthcare visits • Inpatient hospitalizations (including duration by wards (intensive care unit vs. general ward)
To explore the exposure-response relationship between belantamab mafodotin exposure and clinical endpoints in participants treated with B-Pd	• Belantamab mafodotin exposure (e.g., concentration, Cmax, or AUC) vs. efficacy and safety endpoints (e.g., PFS, ORR, CRR, corneal events)

Objectives	Endpoints
To explore the relationship between clinical response and biologic characteristics including, but not limited to, BCMA expression on tumor cells and sBCMA concentrations	• Assess various tumor and blood-based biomarkers at baseline and on-treatment, by analysis of DNA, RNA and/or protein, including but not limited to evaluating baseline BCMA expression and/or immune status in tumor and tumor microenvironment and/or serum soluble BCMA levels, and their relationship to response to belantamab mafodotin
To explore the effect of host genetic variation on the response to belantamab mafodotin and disease under study as well as related drug classes and diseases	• Effect of host genetic variation in 1 or more candidate genes or across the genome on response to belantamab mafodotin and disease under study as well as related drug classes and diseases

Abbreviations: ADA=Anti-drug antibody; AE=adverse event; B-Pd=Belantamab mafodotin in combination with pomalidomide and dexamethasone; CR=complete response; CRR=complete response rate; cys-mcMMAF=Cysteine maleimidocaproyl monomethyl auristatin F; DoR=duration of response; EORTC IL52=European Organisation for Research and Treatment of Cancer Item Library 52; EORTC QLQ C30=European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire 30 item core module; EORTC QLQ MY20=European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire 20-item Multiple Myeloma Module; HRQoL=Health-related Quality of Life; IRC= Independent Review Committee; mAb=monoclonal antibody; MRD=Minimal Residual Disease; NGS=next generation sequencing; ORR=Overall Response Rate; OS=Overall Survival; OSDI=Ocular Surface Disease Index; PFS=Progression-free Survival; PFS2=progression-free survival on subsequent line of therapy; PR=partial response; PRO CTCAE=Patient Reported Outcomes version of the Common Terminology Criteria for Adverse Events; PK=Pharmacokinetic(s); PVd=pomalidomide plus bortezomib and dexamethasone; RRMM=Relapsed/refractory multiple myeloma; sCR=Stringent Complete Response; TTBR=time to best response; TTP=Time to Disease Progression; TTR=Time to Response; VGPR=Very Good Partial Response.
*EORTC IL52 (disease symptoms from the EORTC QLQ-MY20) applies to participants enrolled under the original protocol; EORTC QLQ-MY20 applies to participants enrolled under protocol amendment 1.

1.1.2. Estimands

Objective	Estimand Category	Estimand			
		Variable/ Endpoint	Analysis Set	Intercurrent Event Strategy	Population Level Summary Measure
Primary Objective: To demonstrate the superiority of B-Pd compared to PVd in PFS in participants with relapsed/refractory multiple myeloma (RRMM) [1]	Primary	PFS	ITT, mITT	• Disease assessments between scheduled visits: treatment policy • New anti-myeloma therapy: while on treatment • Extended loss to follow-up: while on treatment • Treatment discontinuation: treatment policy • Death: composite	Hazard ratio for B-Pd vs PVd
	Supplementary 1 (S1)	PFS	ITT	• Disease assessments between scheduled visits: hypothetical • New anti-myeloma therapy: while on treatment • Extended loss to follow-up: while on treatment • Treatment discontinuation: treatment policy • Death: composite	Hazard ratio for B-Pd vs PVd
	Supplementary 2 (S2)	PFS	ITT	• Disease assessments between scheduled visits: treatment policy • New anti-myeloma therapy: composite • Extended loss to follow-up: while on treatment • Treatment discontinuation: treatment policy • Death: composite	Hazard ratio for B-Pd vs PVd
	Supplementary 3 (S3)	PFS	ITT	• Disease assessments between scheduled visits: treatment policy • New anti-myeloma therapy: while on treatment • Extended loss to follow-up: treatment policy • Treatment discontinuation: treatment policy • Death: composite	Hazard ratio for B-Pd vs PVd
	Supplementary 4 (S4)	PFS	ITT	• Disease assessments between scheduled visits: treatment policy • New anti-myeloma therapy: while on treatment • Extended loss to follow-up: while on treatment • Treatment discontinuation: composite	Hazard ratio for B-Pd vs PVd

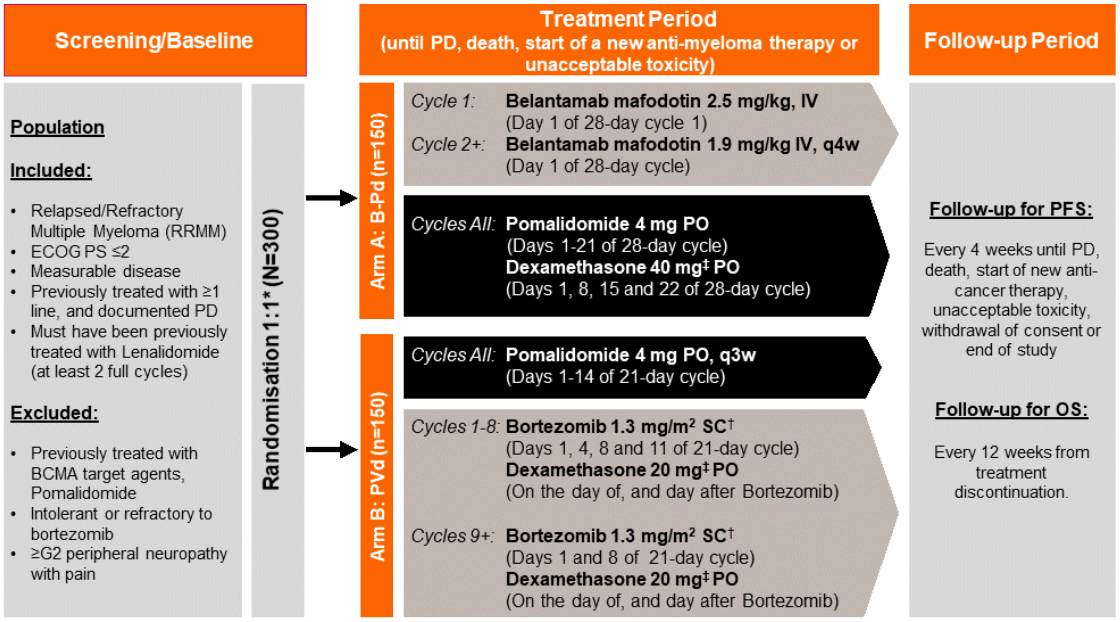
Objective	Estimand Category	Estimand			Population Level Summary Measure
		Variable/ Endpoint	Analysis Set	Intercurrent Event Strategy	
				• Death: composite	
	COVID-19 Supplementary	PFS	ITT	• Disease assessments between scheduled visits: treatment policy • New anti-myeloma therapy: while on treatment • Extended loss to follow-up: while on treatment • Treatment discontinuation: treatment policy • Death (not COVID-19 related): composite • Death (COVID-19 related): hypothetical	Hazard ratio for B-Pd vs PVd
Key Secondary Objectives: Superiority of B-Pd compared to PVd in OS, DoR and in MRD negativity in participants with relapsed/refractory multiple myeloma (RRMM) [1]	Primary	OS	ITT	• New anti-myeloma treatment: treatment policy • Treatment discontinuation: treatment policy	Hazard ratio for B-Pd vs PVd
	Primary	DoR	ITT	• Disease assessments between scheduled visits: treatment policy • New anti-myeloma therapy: while on treatment • Extended loss to follow-up: while on treatment • Treatment discontinuation: treatment policy • Death due to non-PD: composite • Death due to PD: composite	Difference in the restricted mean duration of response (RMDOR) for B-Pd vs PVd
	Responder Supplementary 1	DoR	Participants with a confirmed PR or better in the ITT	• Disease assessments between scheduled visits: treatment policy • New anti-myeloma treatment: while on treatment • Extended loss to follow-up: while on treatment • Treatment discontinuation: treatment policy • Death due to non-PD: while on treatment • Death due to PD: composite	Median DoR, summarized using the Kaplan-Meier method by treatment arm
	Primary	MRD negativity	ITT	• New anti-myeloma treatment: while on treatment • Treatment discontinuation: treatment policy	MRD Negativity Rate by treatment arm
Secondary Objectives (Efficacy):	Primary	ORR	ITT	• New anti-myeloma treatment: while on treatment • Treatment discontinuation: treatment policy	≥PR percentage by treatment arm

Objective	Estimand Category	Estimand			
		Variable/ Endpoint	Analysis Set	Intercurrent Event Strategy	Population Level Summary Measure
To demonstrate the superiority of B-Pd vs PVd in ORR/ CRR/ VGPR+/ TTBR/ TTR/ TTP/ PFS2 in participants with relapsed/refractory multiple myeloma (RRMM) [1]		CRR	ITT	• New anti-myeloma treatment: while on treatment • Treatment discontinuation: treatment policy	≥CR percentage by treatment arm
		VGPR+	ITT	• New anti-myeloma treatment: while on treatment • Treatment discontinuation: treatment policy	≥VGPR percentage by treatment arm
		TTBR	Participants with a confirmed PR or better in the ITT	• New anti-myeloma treatment: while on treatment • Treatment discontinuation: treatment policy	Descriptive summary of median TTBR by treatment arm
		TTR	Participants with a confirmed PR or better in the ITT	• New anti-myeloma treatment: while on treatment • Treatment discontinuation: treatment policy	Descriptive summary of median TTR by treatment arm
		TTP	ITT	• Disease assessments between scheduled visits: treatment policy • New anti-myeloma treatment: while on treatment • Extended loss to follow-up: while on treatment • Treatment discontinuation: treatment policy • Death due to non-PD: while on treatment • Death due to PD: composite	Hazard ratio for B-Pd vs PVd
		PFS2	ITT	• Disease assessments between scheduled visits: treatment policy • New anti-myeloma treatment: treatment policy • Extended loss to follow-up: treatment policy • Treatment discontinuation: treatment policy • Death: composite	Median PFS, summarized using the Kaplan-Meier method by treatment arm

[1] Have been previously treated with at least 1 prior line of MM therapy including a lenalidomide-containing regimen; see inclusion/ exclusion criteria for details

1.2. Study Design

Overview of Study Design and Key Features



Abbreviations: PD=progressive disease; RRMM=relapsed/refractory multiple myeloma.
* Stratification: Prior lines of treatment (1 vs. 2 / 3 vs. ≥4), prior bortezomib treatment (yes or no) and prior anti-CD38 treatment (yes or no). No more than 50% of participants with 2 or more prior lines of treatment will be enrolled. It is anticipated that no more than 15% of participants will be enrolled with 4 or more prior lines of treatment. No cross-over will be allowed.
† SC administration of bortezomib only
‡ Reduce the dose level of dexamethasone by half if age >75 years or have comorbidities or are intolerant to 40 mg dose in Arm A or 20 mg dose in Arm B, respectively.
Note: Prior to protocol amendment 1, ISS status (I vs II/III) was included as a stratification/randomization factor instead of prior anti-CD38 treatment (yes or no).

Design Features	**Overall Design:** This study will evaluate the efficacy and safety of belantamab mafodotin in combination with pomalidomide and dexamethasone (B-Pd) compared with pomalidomide, bortezomib and dexamethasone (PVd) in participants with RRMM previously treated with lenalidomide and at least 1 prior line of therapy. **Disclosure Statement:** This study is a parallel group study with 2 treatment arms and no masking. **Number of Participants:** Approximately 375 participants in Phase III will be screened to achieve approximately 300 participants randomized in a 1:1 ratio between the 2 arms. If the number of participants required by local regulatory agencies are not recruited within the planned recruitment target, enrollment may continue in separate cohorts until the country enrollment requirements are met. Additional participants that are enrolled in separate cohorts will not be included in the analysis portion of the study planned for the marketing application. However, these additional participants will be included in country-specific supplemental analyses, as detailed in the country-specific SAP. In these countries, respective regulatory

Overview of Study Design and Key Features			
	authorities require a sufficient representation of their population to be included in marketing authorizations. **Intervention Groups and Duration:** Following Screening, participants will be stratified based on the number of prior lines of therapy (1 vs. 2 /3 vs. ≥4), prior bortezomib treatment (yes or no) and prior anti-CD38 treatment (yes or no), and centrally randomized in a 1:1 ratio to Treatment Arm A or Treatment Arm B. No more than 50% of participants with 2 or more prior lines of treatment will be enrolled. It is anticipated that no more than 15% of participants will be enrolled in with 4 or more prior lines of treatment. No cross-over between 2 study arms will be allowed. Note: Prior to protocol amendment 1, ISS status (I vs II/III) was included as a stratification/randomization factor instead of prior anti-CD38 treatment (yes or no) in protocol amendment 2 or later.		
Study intervention	• **Treatment Arm A (B-Pd): belantamab mafodotin 2.5 mg/kg in C1 and 1.9 mg/kg in C2+ (IV), pomalidomide 4 mg, and dexamethasone 40 mg, q4w** In Treatment Arm A, belantamab mafodotin will be administered intravenously (IV) over at least 30 minutes at a single dose of 2.5 mg/kg on Day 1 (D1) of Cycle 1 and 1.9 mg/kg in Cycle 2 and beyond (2+) of every 28-day cycle (q4w). Pomalidomide will be taken orally 4 mg per day on Days 1-21 of each 28-day cycle. Dexamethasone will be administered orally at a dose of 40 mg per day on Days 1, 8, 15, and 22 of each 28-day cycle. For participants who are >75 years old or have comorbidities or are intolerant to dexamethasone 40 mg, dexamethasone may be administered at the lower dose of 20 mg in Arm A at the discretion of the investigator. • **Treatment Arm B (PVd): Pomalidomide 4 mg, bortezomib 1.3 mg/m^2, and dexamethasone 20 mg, q3w** In Treatment Arm B, pomalidomide will be administered PO at 4 mg daily on Days 1 to 14 of each 21-day cycle (i.e., q3w), with bortezomib injected SC at 1.3 mg/m^2 on Days 1, 4, 8, and 11 of each 21-day cycle for Cycles 1 through 8 and on Days 1 and 8 of each 21-day cycle for Cycles 9 and beyond (Cycles 9+). Dexamethasone will be administered PO at a dose of 20 mg on the day of and day after bortezomib, q3w or on Days 1, 2, 4, 5, 8, 9, 11, and 12 for Cycles 1 through 8, and then on Days 1, 2, 8, and 9 for Cycles 9+. For participants who are >75 years old or have comorbidities or are intolerant to dexamethasone 20 mg, dexamethasone may be administered at the lower dose of 10 mg on the day of and day after bortezomib in Arm B at the discretion of the investigator. Treatment will continue in both arms until progressive disease (PD), death, unacceptable toxicity, start of a new anti-myeloma therapy, withdrawal of consent, or end of the study, whichever occurs first. Dose delays or reductions may be required following potential drug-associated toxicities. Participants will be followed for PD and overall survival (OS).		
Study intervention assignment	All participants will be centrally randomized using a central Interactive Response Technology (IRT) system. Randomization list will be done centrally using a randomization schedule generated by the Contract Research Organization, which will assign participants in a 1:1 ratio to Treatment Arm A and Treatment Arm B. As this is an open-label study, no blinding of treatment identity is needed for either Treatment Arm A or Treatment Arm B.		
Analyses	**Analyses / Timing**	**Endpoints for analyses**	**Data to be used**
	Safety review by IDMC/ Reviewed periodically starting from when ~60 participants have been followed for 8 weeks, and then every 6 months or as requested by the IDMC thereafter	Key safety (AEs, SAEs, AESIs, deaths, ocular, exposure, dose modifications, laboratory parameters), descriptive efficacy summaries (e.g., response rates, counts of PFS/OS events) and study population summaries.	All data available at the time of the data cut

Overview of Study Design and Key Features			
	Interim analysis for harm (IA1) based on PFS and potential sample-size re-estimation / ~35 PFS events (~25% PFS information fraction)	Key safety, study population and PFS. Additional analyses may be performed to support decision making if requested by IDMC.	All data available at the time of the data cut
	IA2 ~145 PFS events (~84% information fraction)	Minimally, key safety, study population and PFS. Additional analyses may be performed to support decision making if requested by IDMC. All endpoints may be included if PFS is statistically significant.	All data available at the time of the data cut
	Primary PFS analysis/ IA3 ~173 PFS events (100% PFS information fraction) if PFS does not demonstrate statistical significance at IA2 OR alternatively, when: ~130 OS events (~60% OS information fraction) if PFS demonstrates statistical significance at IA2	All endpoints. A reduced set of outputs may be produced if PFS is significant at IA2. PFS will be descriptively analyzed but not formally tested if statistical significance is demonstrated at IA2.	All data available at the time of the data cut
	IA4 ~163 OS events (~75% OS information fraction)	Minimally, updated key safety, study population summaries and OS analysis.	All data available at the time of the data cut
	Final analysis ~217 OS events (100% OS information fraction)	Minimally, updated key safety, study population summaries and OS analysis.	All data

2. STATISTICAL HYPOTHESES

Details of the multiplicity adjustment are detailed in Section 2.1.

Primary Endpoint PFS

The following primary hypothesis will be tested, comparing the distribution of PFS between the two treatment groups:

$$H_0: \theta_1 \geq 1 \qquad VS. \qquad H_1: \theta_1 < 1$$

where, θ_1 is the PFS HR for B-Pd vs. PVd.

Key secondary endpoints

a) OS

The key secondary OS analysis will be the comparison of the distribution of OS between the two treatment groups. The following statistical hypothesis will be tested:

$$H_0: \theta \geq 1 \qquad VS. \qquad H_1: \theta < 1$$

where, θ is the OS HR for B-Pd vs. PVd.

b) MRD Negativity

The following statistical hypothesis will be tested to compare the proportion of participants with MRD negativity between the two treatment groups:

$$H_0: P_1 \leq P_0 \quad VS. \quad H_1: P_1 > P_0$$

where, P_0 =proportion of participants with MRD negativity Arm B (PVd) and P_1 =proportion of participants with MRD negativity Arm A (B-Pd).

c) DoR

The following statistical hypothesis will NOT be formally tested:

$$H_0: \mu_1 - \mu_0 \leq 0 \qquad VS. \qquad H_1: \mu_1 - \mu_0 > 0$$

where, μ_1 is the restricted mean duration of response (RMDOR) for participants in Arm A (B-Pd) and μ_0 is the RMDOR for participants in Arm B (PVd).

2.1. Multiplicity Adjustment

The global family-wise type I error (FWER) for this study is strongly controlled at 2.5% (one-sided).

Evaluation of primary and key secondary endpoints will be structured in terms of two families of hypotheses. The first family will be based on the primary endpoint PFS, and the second family will be based on two key secondary endpoints OS and MRD Negativity. Testing of the second family of hypotheses is conditional on the successful rejection of the null hypothesis for the first family. If successful, the full alpha will be propagated to the second family of hypotheses. OS will be tested first. Testing of MRD will be conditional on the successful rejection of the null hypothesis for OS. This testing procedure is aligned with a step-down (or hierarchical) testing procedure [Bretz, 2009; Lan, 1983; Li, 2017]. The multiple testing strategy (in relation to alpha-spending) is illustrated in Figure 1. Let H_i denote the one-sided null hypothesis for the primary and key secondary endpoints as defined by H_0 in Section 2, and let $i = 1, 2, 3$ denote the index indicating PFS, OS and MRD negativity rate, respectively.

PFS testing

PFS will be tested across 3 planned analyses: an analysis for harm (IA1), an analysis for efficacy (IA2) and the primary PFS analysis/IA3. A gamma beta-spending function with parameter of -3 is used to define a non-binding futility boundary for IA1, no alpha is allocated to this analysis. The Lan DeMets approach, that approximates the O'Brien and Fleming spending function [Lan, 1983], will be used to maintain an overall one-sided 2.5% type I error when testing PFS across IA2 and the primary PFS analysis/IA3, since these analyses provide the opportunity to make a claim of efficacy. All boundaries (see Section 4.7) will be adjusted based on the actual number of PFS events observed at the time of analysis.

Testing of key secondary endpoints: OS and MRD Negativity

Testing of H_2(OS) will be conditional on rejection of H_1(PFS).

Note that if H_1(PFS) fails to be rejected at IA2 but is later rejected at Primary PFS/IA3, then the full alpha will be propagated so that H_2(OS) will be tested at the 2.5% level.

OS will be tested across 4 planned analyses: IA2, primary PFS analysis/IA3, IA4 and at the OS final analysis. The Lan DeMets approach that approximates the O'Brien and Fleming spending function [Lan, 1983] will be used. The efficacy boundaries will be adjusted based on the actual number of OS events observed at the time of analysis and the alpha allocated.

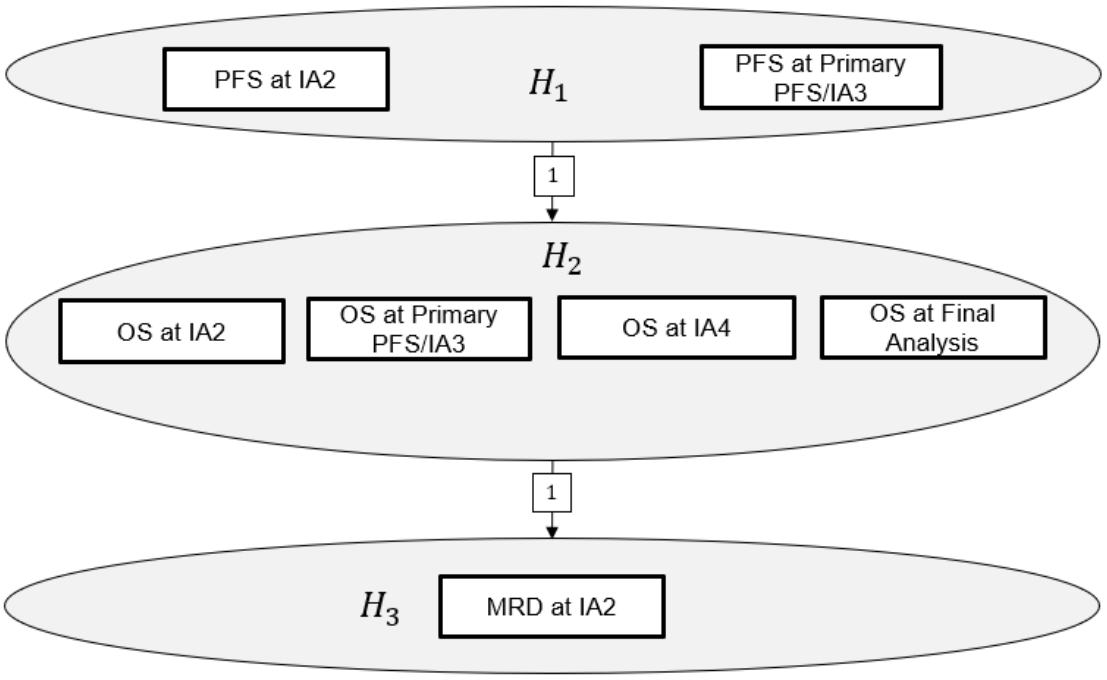
H_3(MRD) testing will be conditional on rejection of H_2(OS). Regardless of the timing of rejection of H_2(OS):

1. H_3(MRD) will only be tested using data available at IA2.

2. The full alpha allocated to OS (2.5% conditional on successful rejection of H_1(PFS)) will be propagated.

The remaining secondary efficacy endpoint DoR and other secondary endpoints will be analyzed without alpha adjustment.

Figure 1 Multiple Testing Strategy



Abbreviations: IA=Interim Analysis; MRD=MRD Negativity Rate; PFS=Progression-Free Survival; OS=Overall Survival. $\boldsymbol{H_i}$ denotes the one-sided null hypothesis for the primary and key secondary endpoints, where $\boldsymbol{i = 1, 2, 3}$ denotes the index indicating PFS, OS and MRD negativity rate, respectively.

Upon successful rejection of the hypothesis and regardless of the timing of rejection, the full alpha allocated to testing the hypothesis can be propagated. Arrows indicate the direction and proportion of alpha re-allocation. $\boldsymbol{H_1}$ will be tested at the one-sided 2.5% significance level. All other hypotheses will have an initial alpha of 0% assigned.

The number of rectangular boxes indicates the number of planned analyses with alpha allocation for a given hypothesis, with text indicating the corresponding endpoint and timepoint of data extraction to be tested. Alpha will be adjusted to account for multiple testing of an endpoint across timepoints using the Lan DeMets approach that approximates the O'Brien and Fleming spending function [Lan, 1983]. The efficacy boundaries will be adjusted based on the observed number of events at the time of analysis.

3. ANALYSIS SETS

Analysis Set[1]	Definition / Criteria	Analyses Evaluated
All Screened	The All Screened Population will consist of all participants who sign the ICF to participate in the clinical trial. Participants in this population will be used for screen failure summary.	Study Population
Enrolled	The Enrolled population is defined as all participants that have entered the study (e.g., participants that are identified on the Screen Failure form as non-screen failures).	Study Population
Safety	All randomized participants who receive at least 1 dose of study treatment (any component). Participants will be analyzed according to the treatment they actually received. **For Arm A**: B-Pd, if participants are incorrectly dosed with bortezomib at >50% of dosing visits then they will be assigned to Arm B: PVd as their actual treatment. Similarly, **for Arm B**: PVd, if participants are incorrectly dosed with belantamab mafodotin at >50% of dosing visits then they will be assigned to Arm A: B-Pd as their actual treatment. Data should be reported according to the actual treatment.	Safety Population
COVID-19	All participants in the Safety set who had a confirmed, probable, or suspected COVID-19 case diagnosis. Data should be reported according to the actual treatment.	Baseline Characteristics, Medical History and Laboratory Data
Intent-to-Treat (ITT)	ITT Population will consist of all randomized participants whether or not randomized treatment was administered. This population will be based on the treatment to which the participant was randomized and will be the primary population for the analysis of efficacy data. Any participant who receives a treatment randomization number will be considered to have been randomized.	Study Population Efficacy
Modified ITT (mITT)	Participants who met all criteria below will be included: • Have received at least 1 line of prior therapy including a lenalidomide-based therapy • With measurable disease at baseline[2] • Randomized and received at least one dose of planned study treatment (belantamab mafodotin or bortezomib) o Participants randomized to the belantamab mafodotin arm that received bortezomib will be excluded and vice versa. o Participants randomized but never treated will be excluded.	Efficacy (sensitivity analysis of primary endpoint and key secondary endpoint)
Belantamab mafodotin Pharmacokinetic (PK)	The belantamab mafodotin Pharmacokinetic Population will consist of those participants in the Safety Population from whom at least 1 belantamab mafodotin PK sample was obtained, analyzed and was measurable (Non-Quantifiable [NQ] values will be considered as non-missing values). Data should be reported according to the actual treatment.	PK analyses related to belantamab mafodotin (non-Pomalidomide)
Pomalidomide PK (Pom PK)	The Pomalidomide Pharmacokinetic Population will consist of those participants in the Safety Population from whom at least 1 pomalidomide PK sample was obtained, analyzed and was measurable (Non-Quantifiable [NQ] values will be	Pomalidomide PK analyses

Analysis Set[1]	Definition / Criteria	Analyses Evaluated
	considered as non-missing values). Data should be reported according to the actual treatment.	

Abbreviations: ICF=Informed Consent Form; ITT=Intent-to-Treat; PK=pharmacokinetic(s).

1. Analysis Set and population will be used interchangeably for analysis purposes
2. Measurable disease at baseline is defined as: a patient has at least one of the following measurements: a. Serum M protein ≥0.5 g/dL (≥5 g/L) or b. Urine M-protein ≥200 mg/24h or c. Serum FLC assay: Involved FLC level ≥10 mg/dL (≥100 mg/L) and an abnormal serum free light chain ratio (<0.26 or >1.65)

4. STATISTICAL ANALYSES

4.1. General Considerations

4.1.1. General Methodology

The Intent-to-Treat (ITT) population will be used for all study population analyses and efficacy analyses, unless otherwise specified, and Safety population will be used for all safety analyses.

Unless otherwise specified, the stratification factors entered for randomization will be used in the primary analysis. If there is any mis-stratification, a sensitivity analysis will be performed using the stratification data based on the clinical database for primary and key secondary endpoints, as appropriate.

All confidence intervals will be 2-sided at the 95% confidence levels unless otherwise specified.

Unless otherwise specified, continuous data will be summarized using descriptive statistics: n, mean, standard deviation (std), median, minimum and maximum. Categorical data will be summarized as the number and percentage of participants in each category.

For laboratory data used in safety summaries, central laboratory data will be used over local results if a participant has multiple non-missing results within the same visit and date. If central laboratory data is not available, local laboratory data will be used unless otherwise specified.

For efficacy analyses, only central lab values will be used, except for bone marrow data for plasma cells where local data is preferred to align with protocol. Where multiple assessments have been performed for bone marrow (e.g., use of central and local labs, aspirate and biopsy) the biopsy result should be used over the aspirate and local lab used over the central for plasma cells. MRD assessment will be based on central lab values.

For endpoint derivations dependent upon response assessments per IMWG, IRC-assessed response (as opposed to investigator-assessed response) will be used unless otherwise specified. Unless otherwise specified, response (including progression) requires confirmation for all efficacy analyses (Table 4).

Data from all participating centres will be integrated and no controlling for centre-effect will be considered in the statistical analyses. It is anticipated that patient accrual will be spread thinly across centres and summaries of data by centre is unlikely to be informative and will not be provided.

Only the assessments from the start of treatment up to the earlier of confirmed disease progression or the start of new anti-myeloma therapy will be considered in efficacy analyses of response data. If assessments are collected beyond this, they may be listed. Only new systemic anti-myeloma therapy taken are considered as anti-myeloma therapy (local radiotherapy and surgeries are not considered as systemic anti-myeloma therapy for the purpose of the efficacy analyses).

Based on Amendment 01 as depicted in Table 1 below, the study would have two stratification cohorts per randomization, with the first having stratification according to A*B*C for 12 strata and the second having stratification according to A*B*D for 12 strata; and so in all, it has 24 strata since Cohort 1 vs. Cohort 2 is also a stratification factor as a consequence of Cohort 2 having a revised structure for stratification that differs from the initial structure for stratification for Cohort 1.

Table 1 Stratification Factors by Protocol Amendment 01

Prior to Amendment 01 [Stratification *Cohort 1*]	After Amendment 01 [Stratification *Cohort 2*]
A: number of prior lines of therapy (1 vs. 2/3 vs. ≥4)	**A:** number of prior lines of therapy (1 vs. 2/3 vs. ≥4)
B: prior bortezomib treatment (yes or no)	**B:** prior bortezomib treatment (yes or no)
C: ISS status (I vs II/III)	**D:** Prior anti-CD38 treatment (yes or no)

Since stratification produces balance of the randomized treatment groups for the corresponding factors for stratification, there is no bias to analysis from its ignoring of a factor for stratification. Also, adjustment for all strata can lead to some strata being entirely non-informative by having 0 events for an endpoint like PFS or only including participants from one of the two treatment groups. The strata are at least minimally informative by when each stratum include at least one participant for each of the two treatment groups and at least one participant with an event and at least one participant with no event and follow-up at least as long as at least one participant with the event. Usually, the strata should be somewhat more informative than minimally informative, with this implying that each stratum should have approximately 10 participants and approximately 5 participants with a PFS event (or event appropriate per the endpoint).

Based on the above, primary analyses for all stratified analyses (e.g., stratified log-rank test and stratified cox proportional hazards model) will be stratified by two randomization factors; number of prior lines of therapy and prior bortezomib treatment. As appropriate, sensitivity analyses will be performed at the time of Primary PFS analysis and/or IA2 (if PFS is statistically significant at IA2) considering all 4 randomization factors as possible stratification factors, using a prespecified pooling of strata so that each stratum has approximately 10 participants and approximately 5 participants with a PFS event (or event appropriate per the endpoint). For the primary endpoint of PFS, an additional

supportive analysis will be performed (HR and corresponding 95% CI will be estimated from Cox proportional hazard model stratified by number prior lines of therapy and prior bortezomib use with treatment, ISS status and prior anti-CD38 treatment as explanatory variables).

Other considerations for data analyses and data handling conventions are outlined in the appendices and the Output Programming Specifications (OPS) document.

4.1.2. Baseline Definition

For all endpoints, unless otherwise specified, the baseline value will be the latest pre-dose assessment with a non-missing value, including those from unscheduled visits. If time is not collected, Day 1 assessments are assumed to be taken prior to first dose and used as baseline. For participants who did not receive study treatment during the study, baseline will be defined as the latest, non-missing collected value.

For laboratory data, baseline will be the latest non-missing pre-dose value from central lab. If no central lab value is available, the latest non-missing pre-dose value from local lab will be used.

For efficacy lab tests, in contrast to above, only central lab values will be used with the exception of bone marrow data where local data is preferred to align with protocol. Where multiple assessments have been performed for bone marrow (e.g., use of central and local labs, aspirate and biopsy) the biopsy result should be used over the aspirate and local lab used over the central.

For immunogenicity, to derive the baseline, consider only the belantamab mafodotin dosing as the first dose date/time.

Unless otherwise stated, if baseline data is missing no derivation will be performed and baseline will be set to missing.

4.2. Primary Endpoint(s) Analyses

4.2.1. Definition of Progression-Free Survival (PFS)

PFS is the primary endpoint of this study; it is defined as the time (months) from randomization until the earliest date of disease progression (PD) per IMWG [Kumar, 2016], or death due to any cause. The analyses of PFS will be based on the ITT Analysis Set, unless otherwise specified, and will use IRC assessment.

Determination of dates of PFS events and dates for censoring are described in Table 2.

4.2.2. Planned Analyses of PFS

Analyses of PFS will be based on response per IMWG [Kumar, 2016] according to the Independent Review Committee (IRC) assessment. Section 1.1.2 describes how intercurrent events will be handled whilst Table 2 below lists the censoring rules.

The following sets of analyses will be conducted:

1. Primary analysis of primary estimand (IRC-assessed response + primary censoring rules)
2. Primary analysis of supplementary estimand 1 [S1] (IRC-assessed response + alternative censoring rules 1)
3. Primary analysis of supplementary estimand 2 [S2] (IRC-assessed response + alternative censoring rules 2)
4. Primary analysis of supplementary estimand 3 [S3] (IRC-assessed response + alternative censoring rules 3)
5. Primary analysis of supplementary estimand 4 [S4] (IRC-assessed response + alternative censoring rules 4)
6. Primary analysis of COVID-9 supplementary estimand (IRC-assessed response + COVID-19 censoring rules)

Table 2 Assignments for Primary and Alternative Progression and Censoring Dates for PFS Analysis

Situation	Date of Event (Progression/Death) or Censoring	Outcome Event (Progression/Death) Or Censored
No (or inadequate) baseline assessments [1] and the participant has not died (if the participant has died follow the rules for death indicated at the bottom of the table)	Randomization	Censored
No adequate post-baseline assessments and the participant has not died (if the participant has died follow the rules for death indicated at the bottom of the table)	Randomization	Censored
Progression documented at scheduled visits and Progression documented without extended loss-to-follow-up time [4]	Date of assessment of progression	Event
Progression documented between scheduled visits and Progression documented without extended loss-to-follow-up time [4]	Date of assessment of progression	Event
	(S1) min (Date of next scheduled visit, date of death)	**(S1) Event**
With post-baseline assessment but no progression (or death)	Date of last 'adequate' assessment of response [2]	Censored
No adequate post-baseline assessment before start of new anti-myeloma therapy (prior to documented disease progression or death)	Randomization	Censored
	(S2) Date of starting new anti-myeloma therapy	**(S2) Event**
With adequate post-baseline assessment and new anti-myeloma treatment started (prior to documented disease progression or death) [3].	Date of last 'adequate' assessment of response [2] *(on or prior to starting anti-myeloma treatment)*	Censored
	(S2) Date of starting new anti-myeloma therapy	**(S2) Event**
Death before first scheduled assessment (or death at Baseline or without any adequate assessments)	Date of death	Event

Situation	Date of Event (Progression/Death) or Censoring	Outcome Event (Progression/Death) Or Censored
Death between adequate assessment visits	Date of death	Event
Death without extended loss-to-follow-up time [4]	Date of death	Event
Death or progression after an extended loss-to-follow-up time [4]	Date of randomization if no post-baseline assessments, or date of last 'adequate' assessment of response [2] prior to PD/death (prior to missed assessments): since disease assessment is every 4 weeks, a window of 63 days (8 weeks + 7-day window) will be used to determine whether there is extended time without adequate assessment. If the time difference between PD/death and max (last adequate disease assessment, randomization) is more than 63 days, PFS will be censored at the last adequate disease assessment prior to PD/death.	Censored
	(S3) Date of death or progression	**(S3) Event**
(S4) Treatment discontinuation due to clinical PD[5] before PD or death	**(S4) Date of treatment discontinuation**	**(S4) Event**

Abbreviations: CR=Complete Response; FLC=Free Light Chain; MR=Minimal Response; PD=Progressive Disease; PR=Partial Response; sCR=Stringent Complete Response; SD=Stable Disease; VGPR=Very Good Partial Response.

Note: (S1) (S2) (S3) (S4) Rules to be Applied for PFS Supplementary Analysis.

Event or censored are based on confirmed responses.

1. Adequate baseline assessment is defined as at baseline, a patient has at least one of the following measurements: a. Serum M-protein ≥0.5 g/dL (≥5 g/L) or b. Urine M-protein ≥200 mg/24h or c. Serum FLC assay: Involved FLC level ≥10 mg/dL (≥100 mg/L) and an abnormal serum free light chain ratio (<0.26 or >1.65)
2. An adequate assessment is defined as an assessment where the confirmed response is sCR, CR, VGPR, PR, MR, or SD. If the adequate assessment occurred on the same date as new anti-myeloma therapy, it is assumed that the assessment occurred first.
3. If PD or death and new anti-myeloma therapy occur on the same day assume the outcome is progression or death, and the date is the date of the assessment of progression or death. If anti-myeloma therapy is started prior to any adequate assessments, censoring date should be the date of randomization.
4. Extended loss-to-follow-up time = 8 weeks + 7-day window = 63-day window; without extended loss-to-follow-up time is defined as: ≤ 63 days; after an extended loss-to-follow-up time is defined as: >63 days. 5. Treatment discontinuation of any component due to physician decision = clinical relapse or where physician decision indicates clinical progression.

Refer to Table 4 for information regarding the derivation of confirmed response.

Interim PFS analysis (IA2)

An interim PFS analysis will be conducted when approximately 145 PFS events (~84% information fraction) are observed. Minimal safety and efficacy outputs will be produced in order for the Independent Data Monitoring Committee (IDMC) to assess the benefit:risk profile and make recommendations to continue the study or stop for efficacy and unblind (further details to be provided in the IDMC charter and OPS document). If

PFS demonstrates statistical significance at the IA2 using the primary estimand (see Section 4.7 for details on boundaries), then:

- PFS will be further analysed using all the sets of analyses as described above along with all other endpoints.

- PFS will only be descriptively analysed and not formally re-tested in the subsequent analyses. IA3 analysis will be driven by the OS events instead and a reduced set of outputs vs those planned for Primary PFS analysis will be produced.

As described in the IDMC charter, an additional safeguard will be considered by the IDMC when recommending whether the study should continue as planned to the primary PFS analysis or to unblind at IA2. In addition to statistical significance for PFS, the OS HR must be less than 1. Should the PFS threshold be met, but OS HR≥1, the study will continue as planned; the testing of all other endpoints, in addition to all other planned analyses, will be performed in alignment with PFS statistical significance not being achieved at IA2. Therefore, throughout the SAP, reference to PFS statistical significance at IA2 also implies that OS HR<1 criteria needs to be met.

Primary PFS analysis

If PFS at IA2 is not statistically significant, the primary PFS analysis will be conducted after observing approximately 173 PFS events in the randomized participants contributing to the analysis. Assuming successful PFS, OS will be tested at the 2.5% alpha level (see Section 4.7 for details on boundaries). Key secondary endpoints DoR and MRD will be analyzed descriptively without formally being tested based on the data available at the primary PFS analysis data cut-off. Regardless of timing of PFS statistical significance, formal testing (if applicable) of MRD negativity will be based on IA2 data.

4.2.3. Main Analytical Approach

The distribution of PFS for each treatment arm, at each planned analysis, will be estimated using the non-parametric Kaplan-Meier method. Kaplan-Meier plots of PFS will be presented by treatment arm. The median, 25th and 75th percentiles of PFS will be estimated and corresponding 95% confidence intervals will be estimated using the Brookmeyer-Crowley method [Brookmeyer, 1982]. A listing of participants PFS status will be produced.

The treatment relative effect in PFS will be compared by the one-sided stratified log-rank test. The stratified log-rank test (stratified by applicable randomization factors) will only be performed for the primary analysis of primary estimand of PFS (i.e., based on IRC-assessed response and primary event and censoring rules) based on the ITT Analysis Set.

The hazard ratio (HR) and its corresponding 95% CI will be estimated from a Cox proportional hazard model stratified by applicable randomization factors with treatment arm as the sole explanatory variable. Cox models will be fitted using SAS PROC PHREG with the Efron method to control for ties.

The type of events (progressions, deaths) and censoring reasons will be summarised. Depending on data maturity, PFS rate at 6, 12, and 18 months with corresponding 95% CI will also be estimated from the Kaplan-Meier analysis.

Stratification factors entered for randomization in the interactive voice recognition system (IVRS) will be used in the primary analysis. If there is any mis-stratification, sensitivity analyses will be performed using the stratification data based on the clinical database (eCRF/vendor data).

Statistical Methodology Specification

Endpoint / Variables
• PFS
Model Specification
• PFS will be analyzed across treatment arms using Kaplan-Meier analysis (PROC LIFETEST). • 95% Confidence intervals will be estimated using the Brookmeyer-Crowley method [Brookmeyer, 1982]. • The treatment relative effect in PFS will be tested by the stratified log-rank test (stratified by two randomization factor(s); number prior lines of therapy and prior bortezomib use). • A stratified Cox proportional hazard model (same stratification factors as above) with Efron's method of tie handling and treatment arm as the sole explanatory variable will be used to assess the magnitude of the treatment difference (i.e., the hazard ratio) in PFS between the treatment arms.
Model Checking & Diagnostics

• The proportional hazards assumption will be assessed using the following methods:
  ○ Kaplan-Meier plot by treatment arm
  ○ Plot of log(time) against log(-log[survival]) by treatment arm
  ○ Plot of Schoenfeld residuals for treatment
  ○ Evaluation of time-dependency of treatment effect by adding an interaction term of treatment and time in the Cox model. If the interaction term is significant (p< [0.10]), it is considered that the proportional hazards assumption is violated.
• If one or more of the procedures above demonstrates clear violation of the proportional hazards assumption in PFS, it is considered the proportional hazards assumption does not hold. Hazard ratio and corresponding 95% CI estimated from the Cox model will still be reported.
• More details for handling possible non-proportional hazards effect are provided in Section 4.2.4.1.

Model Results Presentation
• Kaplan-Meier estimates for the median PFS and the first and third quartiles will be presented, along with 95% CIs. • The p-value from the one-sided stratified log-rank test will be reported. Note: interpretation will be based on one-sided p-value. The critical value will be determined according to the assigned alpha level, in line with the multiplicity strategy. • Hazard ratio and corresponding 95% confidence interval from the Cox model will be reported.

4.2.4. Sensitivity Analyses

All PFS sensitivity and supportive/supplementary analyses will be performed at the time that statistically significant PFS (based on the primary estimand) is observed. If this is at the time of IA2, analyses may be repeated at the primary PFS analysis/IA3, if appropriate.

4.2.4.1. Non-Proportional Hazards Effect

If there is evidence (see diagnostics in 4.2.3) of non-proportional hazards effect in PFS, the Restricted Mean Survival Time (RMST) method [Uno, 2015] may be implemented if appropriate; the hazard ratio (HR) and corresponding 95% CI based on Cox proportional hazard model will still be reported.

Restricted Mean Survival Time (RMST)

RMST method may be conducted to account for the possible non-proportional hazards effect. The RMST is the expected survival time restricted to a specific time horizon t^*. *The cutoff t* for determining the RMST will be the smallest value among the largest observed time across study interventions.*

Statistical Methodology Specification

Endpoint / Variables
• PFS
Model Specification
• Additional analysis based on RMST will be conducted if the proportional hazard assumption does not hold. • RMST at *t** will be estimated from the Kaplan-Meier curve for each treatment arm: $$\mu_{t^*} = \int_0^{t^*} S(t)\, dt$$ • RMST difference at *t** ($\hat{\Delta}_{t^*}$) between treatment arms will be estimated as: $$\hat{\Delta}_{t^*} = \int_0^{t^*} \left[\hat{S}_T(t) - \hat{S}_C(t)\right] dt$$ • 95% CI for RMST difference and the p-value will be estimated using the following formula under normal approximation [Klein, 2005]: $$Var\left(\hat{\Delta}_{t^*}\right) = V[\hat{\mu}_{t^*}(T)] + V[\hat{\mu}_{t^*}(C)]$$ $$V[\hat{\mu}_{t^*}] = \sum_{i=1}^{D} \left[\int_{t_i}^{t^*} \hat{S}(t)dt\right]^2 \frac{d_i}{Y_i(Y_i - d_i)}$$ where d_i is the number of events and Y_i is number of participants at risk at t_i.
SAS Procedure
• SAS/STAT 15.1 will be used for the statistical analysis. • Proc LIFETEST will be used with RMST option to obtain the RMST in both the treatment groups. • Proc RMSTREG will be used to obtain the RMST difference between the groups and corresponding 95% CI. The option link=linear and loglink will be specified. "Mean Plot" with "CLBAND" option will be used to generate the RMST plot with confidence bands.
Model Results Presentation
If the proportional hazard assumption does not hold: • the p-value based on the RMST test will also be reported. • RMST difference at *t**, and the corresponding 95% confidence interval will be presented. • RMST ratio at *t**, and the corresponding 95% confidence interval will additionally be presented. • A plot of RMST up to *t** and the corresponding 95% simultaneous confidence bands will be generated.

4.2.4.2. Sensitivity Analysis of PFS Primary Estimand on Investigator-Assessed Response

This sensitivity analysis will include only the primary estimand of PFS (i.e., handling of intercurrent events based on primary event and censoring rules) and will be based on the investigator-assessed response. This analysis will only use the Kaplan-Meier estimates, Cox proportional hazards model stratified by randomization factors as described in Section 4.2.3.

The agreement between the IRC and Investigator-assessment of PD (including timing and occurrence) within and across treatment arms will be evaluated using the PhRMA method [Amit, 2011]. The agreement between the investigator and the IRC within a study intervention is represented in a tabular form as shown in Table 3.

The timing of investigator and IRC will be considered to agree if they occur within ±3 days of each other, aligned with the protocol-specified window for tumor assessments. Otherwise, progression by the investigator is considered earlier than IRC when progression is declared by investigator but not by IRC or IRC progression is declared after investigator progression; progression by the investigator is considered later than IRC when progression is declared by IRC but not by the investigator or the investigator progression is declared after the IRC progression. When summarized, a further breakdown may be provided versus the below table:

- PD
    - Complete agreement on timing and occurrence of PD (as per table)
    - Investigator PD declared later than IRC PD
    - Investigator PD declared earlier than IRC PD
- No PD

The early discrepancy rate (EDR) and late discrepancy rate (LDR) are defined as

$$EDR = \frac{b + a3}{a + b}$$

$$LDR = \frac{c+a2}{b+c+a2+a3}.$$

The EDR represents the positive predictive value of investigator assessment and quantifies the frequency with which the investigator declares progression early relative to IRC within each arm as a proportion of the total number of investigator assessed PD's. The LDR quantifies the frequency that investigator declares progression later than IRC as a proportion of the total number of discrepancies within the arm. If the distribution of discrepancies is similar between the study interventions, then this suggests the absence of evaluation bias favoring a particular study intervention.

The EDR and LDR will be calculated for each study intervention and the differential discordance around each measure will be summarized as the rate on the experimental arm minus the rate on the control arm. A negative differential discordance for the EDR and/or positive differential discordance for the LDR is suggestive of a bias in the investigator favoring the experimental arm.

Table 3 Agreement Between Investigator and IRC

	IRC	
Investigator	PD	No PD
PD	a = a1+a2+a3	b
No PD	c	d

a1: number of agreements on timing and occurrence of PD
a2: number of times investigator PD declared later than IRC PD
a3: number of times investigator PD declared earlier than IRC PD

A listing of participants with differing IRC and Investigator-assessed response will also be produced. All visits will be included for participants with any differing response assessments.

4.2.4.3. Sensitivity Analyses of PFS Primary Estimand Considering the Stratification Factors

All 4 Stratification Factors

A sensitivity analysis will be performed considering stratification by all 4 randomization factors (including all factors used prior to and following protocol amendment 1). However, a prespecified pooling of strata strategy will be applied so that each stratum minimally aligns with the stratification used for the primary analysis and any further stratification has informative information (e.g. minimally, approximately 10 participants and approximately 5 participants with a PFS event):

1. First construct strata according to A*B, with this producing 6 strata in all.
2. If a stratum according to A*B does has at least 30 participants and at least 10 events for Cohort 1, further stratify those participants according to C.
3. If a stratum according to A*B has at least 30 participants and at least 10 events for Cohort 2, further stratify those participants according to D.
4. if any of the (up to 24) stratum constructed in steps 1 and 2 has <10 participants and <5 events then:
    a. if any stratum exists for A*B alone with the same levels of A and B as the stratum with insufficient participants and/or events then combine with this stratum (i.e. combine the stratum with participants from the other cohort with the same levels of A and B).
    b. Otherwise, remove the stratification by C or D (as appropriate) for all stratum with the same combination of A*B within the affected Cohort (i.e. within the cohort, stratum with the same levels of A and B will be combined).

Note: if any stratum according to A*B does not have at least 30 patients and at least 10 events for Cohort 1 or 2, these strata will not be combined with any other stratum.

As an additional supportive analysis, HR and corresponding 95% CI will be estimated from Cox proportional hazard model stratified by number of prior lines of therapy and prior bortezomib use with treatment, ISS status and prior anti-CD38 treatment as explanatory variables.

Based on Stratification Data from the Clinical Database

If there is any mis-stratification for stratification factors entered for randomization, the following supplementary analysis will be performed using the stratification data based on the clinical database.

1. Primary analysis of primary estimand (IRC-assessed response + primary censoring rules)

The analytical approach is as follows: Cox proportional hazards model stratified by randomization factors (based on data from the clinical database).

4.2.4.4. Sensitivity Analysis of PFS Primary Estimand on mITT Analysis Set

The following sets of analyses will also be conducted based on the mITT Analysis Set using IRC-assessed response:

1. Primary analysis of primary estimand (IRC-assessed response + primary censoring rules)

The analytical approach for each analysis above is as follows: Kaplan-Meier estimates and Cox proportional hazards model stratified by randomization factors as described in Section 4.2.3.

4.2.5. Additional Estimands

Additional analyses of the supplementary estimands 1, 2, 3 and 4, as well as the COVID-19 supplementary estimand (i.e., handling of intercurrent events based on alternative event and censoring rules) of PFS will be conducted based on IRC-assessed response. For S1-S4, the associated censoring rules are defined in Section 4.2.2.

These additional analyses will only use the Kaplan-Meier estimates and Cox proportional hazards model stratified by randomization factors described in Section 4.2.3.

COVID-19 Supplementary Estimand

If the number of COVID-19 related deaths is considered to be high (approximately 3% of randomized participants or more), a PFS supplementary analysis may be performed. This will be similar to the primary analysis of the primary estimand, however, COVID-19 related deaths (where primary cause of death is COVID-19 related and secondary cause is

not related to the disease under study) will be censored, (instead of treated as an event) in order to approximate a COVID-19 post-pandemic treatment effect. Additional intercurrent events may be considered based on review of the blinded data, prior to database lock.

This study was designed in the absence of a COVID-19 pandemic. The study objectives were defined to inform clinical practice in a world without COVID-19 or in a world post-pandemic. It is expected that the pandemic will be temporary, where in the future, effective treatment for and prevention of infection by severe acute respiratory syndrome coronavirus-2 (SARS COV-2), the virus that causes COVID-19, will be available.

The primary analysis methods do not account for the impact of the COVID-19 pandemic. However, with few intercurrent events related to COVID-19, the estimated treatment effect will approximate the treatment effect in the absence of relevant intercurrent events related to COVID-19, in alignment with the study objectives.

This may be performed as part of the IDMC interim analysis review, if requested by IDMC. A sensitivity for the COVID-19 supplementary estimand may also be performed using investigator-assessed response.

4.3. Secondary Endpoint(s) Analyses

4.3.1. Key Secondary Endpoint(s)

4.3.1.1. Definition of Endpoints

Overall Survival (OS) is defined as the interval of time from randomization to the date of death due to any cause. Participants who are alive will be censored at the date last known alive.

Note: attempts to obtain survival status may occur following data cut off and prior to data extract. If participants are confirmed to be alive or if the death date is after the data cut off, then the participant will be censored at the date of data cut off. Survival status may be obtained from public records, if applicable per local laws. Survival status captured while on study and those retrieved following study discontinuation/withdrawal will be considered.

The last contact date/last known alive date will be determined by the latest collection/assessment date from among selected data domains within the clinical database that are indicative of participants last known alive date. This will include survival status data captured beyond data cut off and beyond study discontinuation/withdrawal. For participants with last contact date/last known alive date or death beyond the date of data cut off, the date of data cut off will be used as the last contact date/last known alive date. Details of the last contact date/ last known alive date derivation will be provided in a separate Output and Programming Specification (OPS) document.

When calculating overall survival, all deaths following subsequent anti-myeloma therapy will be included. This is the primary estimand of OS, and there is no supplementary estimand of OS.

Duration of response (DoR) is defined as the time from first documented evidence of PR or better until the earliest date of disease progression (PD), or **death due to any cause**.

Minimal Residual Disease (MRD) Negativity Rate is defined as the percentage of participants who achieve MRD negative status (as assessed by NGS at 10^{-5} threshold) at least once during the time of confirmed CR or better response based on IRC-assessment per IMWG [Kumar, 2016]. For analysis purposes, participants with a confirmed CR or better response who do not achieve MRD negative status (including missing/inconclusive assessment(s)) and participants without a confirmed CR or better response will be considered as having non-negative MRD.

4.3.1.2. Main Analytical Approach

OS: refer to Section 4.2.3 (i.e., Kaplan-Meier estimates, stratified log-rank test, Cox proportional hazards model stratified by randomization factors, and examination of non-proportional hazards effect). The analyses of OS will be based on the ITT Analysis Set, unless otherwise specified. In addition, pending on maturity of data, the survival probability at 6, 12 and 18 months with 95% CI will be estimated using Kaplan-Meier method. Kaplan-Meier plots of OS will be presented by treatment arm. A listing of participants OS status will be produced.

For **DoR**: For the primary analysis of DoR, all participants will be included in the analysis regardless of response status, to enable a valid statistical comparison between the two arms. Response will be based on IRC-assessment per IMWG criteria [Kumar, 2016]. DoR will be analyzed based on the restricted mean DoR (RMDOR) using a non-parametric approach [Huang, 2022]. Using this approach, non-responders will have an observed DoR of zero. The approach accounts for TTR, ORR and DoR where the summary measure is the time from response to progression or death. The RMDOR for a treatment arm is the difference between the KM curves of PFS and response/progression-free survival (RPFS). The RMDOR and the corresponding 95% confidence interval will be calculated for each arm. The difference in the RMDOR and the associated 95% CI and one-sided p-value (descriptive only) will be provided. Additionally, the ratio of the RMDORs (Arm A/Arm B) and associated 95% CI will be calculated. A listing of duration of response will be produced.

MRD Negativity Rate: The number and percentage of participants who are MRD negative will be summarized by treatment arms. The corresponding exact 95% CI for MRD negativity rate and associated p-value(s) will also be provided. Information of MRD will be included in the listing of response. Intercurrent event strategy is described on Section 1.1.2.

The primary analysis and formal testing of MRD negativity will be based on data available at the time of IA2, regardless of the timing of PFS statistical significance. At

the time of primary PFS analysis, data will be analyzed descriptively without formally being tested based on the data available at the data cut-off.

Statistical Methodology Specification

Endpoint / Variables
• MRD Negativity Rate
Model Specification
• N/A
SAS Procedure
• SAS/STAT 15.1 will be used for the statistical analysis. • Proc FREQ will be used with CMH option to produce the Cochran-Mantel-Haenszel statistics at the one-sided 0.025 alpha level • Proc FREQ will be used with binomial exact option to obtain fisher's exact test at the one-sided 0.025 alpha level (supportive nominal p-value) [METHOD=NOSCORE will be specified when using SAS/STAT 15.1 – if using SAS/STAT 14.2 or earlier for other deliverables, this is not required]
Model Results Presentation
• The MRD negativity rate and corresponding 2-sided 95% exact CIs will be summarized by treatment arm. • The p-value will be obtained using the *Cochran Mantel Haenszel* (CMH) test stratified by the two randomization factors (number of prior lines of therapy and prior bortezomib use) at the one-sided 0.025 alpha level. A supportive one-sided p-value will be calculated also from fisher's exact test. • Note: MRD interpretation will be based on one-sided p-value obtained using the CMH test. The critical value will be determined according to the assigned alpha level, in line with the multiplicity strategy.

4.3.1.3. Sensitivity Analyses

OS

The analyses in this section may be performed at each OS planned analysis assuming sufficient number of OS events have occurred. These analyses may be performed as required based on the specifications below:

- **RMST***:* If there is possible non-proportional hazards effect, refer to Section 4.2.4.1 for RMST method.
- **Analysis of OS based on Stratification Data from the Clinical Database:** If there is any mis-stratification for stratification factors entered for randomization, the following sensitivity analysis will be performed using the stratification data based on the clinical database. The analytical approach is Cox proportional hazards model stratified by randomization factors (based on data from the clinical database).

DoR

DoR sensitivity analyses will be conducted at the time of at the time of PFS IA2 (conditional upon PFS statistical significance at IA2) as well as PFS primary analysis. DoR will additionally be analysed as follows:

- Using investigator-assessed response according to IMWG (based on ITT, RMDOR)
- **Conventional DoR analysis in responders:** Using IRC-assessed response according to IMWG (based on ITT) but among participants who achieve a response (i.e., confirmed PR or better). This will be repeated where DoR includes deaths due to any cause and separately for deaths due to disease progression only.

A conventional DoR analysis will be performed, where responders without disease progression will be censored at the censoring time point for TTP, however, death due to causes other than PD will be handled the same as death due to PD. Distribution of DoR will be summarized using the Kaplan-Meier method by treatment arm. The median, 25th and 75th percentiles of DoR will be estimated and corresponding 95% confidence intervals will be estimated using the Brookmeyer-Crowley method [Brookmeyer, 1982]. Refer to Section 4.2.3 (i.e., definitions of required Kaplan-Meier estimates, including rates of DoR of 6, 12 and 18 months with corresponding 95% CIs, Cox proportional hazards model stratified by randomization factors, HR and 95% CIs). P-values will not be produced. Kaplan-Meier plots of DoR will be presented by treatment arm.

MRD Negativity

MRD additional analyses, as described in Section 4.3.1.2, will also be repeated as follows at the time of PFS IA2 (conditional upon PFS statistical significance at IA2) as well as PFS primary analysis:

- **Using investigator-assessed response** according to IMWG and based on ITT Analysis Set
- On the ITT Analysis Set but based on **participants with VGPR or better**, using
    - IRC-assessed response and
    - investigator-assessed
- Using the stratified Cochran Mantel Haenszel test, **considering stratification by all 4 randomization factors** (see Section 4.2.4.3 for details on strata and pre-specified pooling strategy).
    - **If there is any mis-stratification for stratification factors entered for randomization**, an additional analysis will be performed using the stratification data from the clinical database.

A supportive summary of MRD Negativity Rate by Best Overall Response will be provided in order to examine the breakdown of MRD Negative participants. MRD negative rates and associated 95% exact CIs will be summarized by treatment arm.

4.3.1.4. Additional Estimands

The sensitivity analysis of **conventional DoR analysis in responders** will be repeated for the responder supplementary estimand 1 at the time of PFS IA2 (conditional upon PFS statistical significance at IA2) and primary PFS analysis using the ITT Analysis Set.

DoR will be defined as the time from first documented evidence of PR or better until the earliest date of PD, or **death due to PD**, among participants who achieve a response (i.e., confirmed PR or better) based on IRC-assessment per IMWG criteria [Kumar, 2016]. Responders without disease progression will be censored at the censoring time point for TTP.

4.3.2. Supportive Secondary Endpoints

Primary analysis of supportive secondary efficacy endpoints will be based on IRC-assessed response and will be based on the ITT Analysis Set, unless otherwise specified.

All secondary efficacy endpoints will be analysed at the primary PFS analysis only, unless PFS demonstrates statistical significance at the IA2 or required for IDMC review of the benefit:risk. No subsequent analyses are planned. This includes all subgroup, sensitivity and supportive/supplementary analyses.

4.3.2.1. Definition of Endpoints

- **Overall response rate (ORR)** is defined as the percentage of participants with a confirmed PR or better (i.e., PR, VGPR, CR, and sCR) based on IRC-assessment per IMWG as the Best Overall Response (BOR).

The earliest date of the two consecutive assessments will be used as the date of the confirmed response. BOR is defined as the best confirmed response (stringent Complete Response [sCR] > Complete Response [CR] > Very Good Partial Response [VGPR] > Partial Response [PR] > Minimal Response [MR] > Stable Disease [SD] > Progressive Disease [PD] > Not Evaluable [NE]) from treatment start date until disease progression or initiation of new anti-myeloma therapy, whichever is earlier, based on IRC-assessed response per IMWG [Kumar, 2016] (see Table 4 for details).

Additionally, per IMWG [Kumar, 2016], if participants do not have measurable disease at baseline, they can only be assessed for at least a complete response (i.e. CR or sCR) or progressive disease. Therefore, in these cases BOR can only be with BOR assessed as SD, MR, PR or VGPR will be assigned a BOR of NE in alignment with IMWG criteria. Participants with only assessments of Not Evaluable or missing response will be treated as non-responders, i.e., they will be included in the denominator when calculating the percentage.

- **Complete response rate (CRR)** is defined as the percentage of participants with a confirmed complete response or better (i.e., CR and sCR) based on IRC-assessment per IMWG as the BOR.

- **Very good partial response rate (VGPR+)** is defined as the percentage of participants with a confirmed Very Good Partial Response (VGPR) or better (i.e., VGPR, CR, and sCR) based on IRC-assessment per IMWG.
- **Time to best response (TTBR)** is defined as the time (in months) between the date of randomization and the date of achieving BOR among participants with a confirmed PR or better (i.e., time to sCR if sCR achieved, if not then time to CR, if CR not achieved then time to PR) based on IRC-assessment per IMWG.
- **Time to response (TTR)** is defined as the time (in months) between the date of randomization and the first documented evidence of response (PR or better), among participants who achieve a response (i.e., PR or better) based on IRC-assessment per IMWG.
- **Time to progression (TTP)** is defined as the time from randomization until the earliest date of PD based on IRC-assessment per IMWG or death due to PD (equivocally or unequivocally). Determination of dates of TTP event and dates for censoring are described in Table 5.
- **Progression-free survival on subsequent line of therapy (PFS2**) is defined as time from randomization (in months) to disease progression after initiation of the first new anti-myeloma therapy or death from any cause, whichever is earlier. If progression after starting new anti-myeloma therapy cannot be measured, a PFS event is defined as the date of discontinuation of first new anti-myeloma therapy, or death from any cause, whichever is earlier. Determination of dates of PFS2 events and dates for censoring are described in Table 6.
- For the PFS2 analysis, progression (after anti-myeloma therapy) will be based on investigator-assessed response per IMWG.

Table 4 Response Confirmation Algorithm

#	Response at any given visit	Response at Subsequent Disease Assessment[1]	Confirmed Response at the given visit
1	sCR	sCR	sCR
2*	sCR	CR	CR
3	CR	sCR/CR	
4*	sCR/CR	VGPR	VGPR
5	VGPR	sCR/CR/VGPR	
6*	sCR/CR/VGPR	PR	PR
7	PR	sCR/CR/VGPR/PR	
8*	sCR/CR/VGPR/PR	MR	MR
9	MR	sCR/CR/VGPR/PR/MR	
10*	sCR/CR/VGPR/PR/MR	SD	SD
11*	sCR/CR/VGPR/PR/MR	PD (any reason) OR No subsequent disease assessment: **participant died or discontinued study or started new anti-myeloma therapy** before further	Last confirmed response. If no prior confirmed response exists then SD.

#	Response at any given visit	Response at Subsequent Disease Assessment[1]	Confirmed Response at the given visit
		adequate disease assessment	
12	PD (due to reasons other than imaging, i.e., plasmacytoma or bone lesion)	PD (any reason) including PD after initiation of new anti-myeloma therapy OR [2] No subsequent disease assessment: participant **died due to PD** before further adequate disease assessment and within 63 days of PD at First Time Point (including death due to PD after initiation of new anti-myeloma therapy)	PD
13	PD (due to reasons other than imaging, i.e., plasmacytoma or bone lesion)	sCR/CR/VGPR/PR/MR/SD OR No subsequent disease assessment: participant died **due to reasons other than PD** OR participant **died due to PD** after 63 days (8 weeks + 7-day window) of PD at first time point before further adequate disease assessment, OR No subsequent disease assessment: participant **discontinued study** before further adequate disease assessment	Last confirmed response. If no prior confirmed response exists then NE.
14	sCR/CR/VGPR/PR/MR/PD (due to reasons other than imaging, i.e., plasmacytoma or bone lesion)	No subsequent disease assessment: participant has not died, not discontinued from study or (except for PD), not started new anti-myeloma therapy; but as yet has no further adequate disease assessments	Last confirmed response. If no prior confirmed response exists then NE.
15	SD	Any OR No subsequent disease assessment	SD
16	PD due to imaging (plasmacytoma or bone lesion)	Any OR No subsequent disease assessment	PD
17	NE or missing	Any OR No subsequent disease assessment	NE

[1] Subsequent disease assessment is defined as the next non-missing or NE disease assessment following the given visit, before (or on the same date of) start of new anti-myeloma therapy except for confirmation of PD, for which PD or death due to PD after new anti-myeloma therapy are considered for confirmation of PD. No minimal time interval is required for the subsequent disease assessment, but a different sample is required for confirmation.
[2] Additional clinical consideration for confirmation of PD (not per IMWG)
Notes:

- SD does not need to be confirmed.
- PD due to imaging (i.e., plasmacytoma or bone lesion) does not need to be confirmed.
- Where criteria are not mutually exclusive, take the first that applies.
- Scenarios represented in lines 2, 4, 6, 8, 10, and 11, per IMWG criteria, should only occur if there is no previous confirmed response. Otherwise, *confirmed response at the given visit* should be the last confirmed response category. Downgrades in response are not expected per IMWG but the logic in these lines handles initial response assessments and unconfirmed upgrades (which may be entered to reduce site burden due to retrospective data entry updates), so that the confirmed response is per IMWG criteria. For the scenarios represented in lines 13, 14, and 17, in most cases this scenario will not apply, as *confirmed response at the given visit* should be the last confirmed response category per IMWG. Also note "NE" is not an IMWG response category. NE is used to characterize "Not Evaluable," as in a response category (per IMWG) cannot be determined. In case of data entry issues where downgrades in response are entered, and to handle unconfirmed PDs where no subsequent assessment exists, additional programming logic is implemented in lines 13 and 14 to ensure that confirmed response assessment aligns with IMWG criteria.

"Death due to PD" will be defined as a death equivocally or unequivocally due to the disease under study.

Table 5 Assignments for Progression and Censoring Dates for TTP Analysis

Situation	Date of Event (Progression/Death due to PD) or Censored	Outcome Event (Progression/Death due to PD) Or Censored
No (or inadequate) baseline tumor assessment [1] and the participant has not died due to PD (if the participant has died due to PD follow the rules for death indicated at the bottom of the table)	Randomization	Censored
No adequate post-baseline assessments and the participant has not died due to PD (if the participant has died due to PD follow the rules for death due to PD indicated at the bottom of the table)	Randomization	Censored
Progression documented at or between scheduled visits, without extended loss-to-follow-up time[2]	Date of progression	Event
With post-baseline assessment but no progression (or death due to PD)	Date of last 'adequate' assessment of response[3]	Censored
No adequate post-baseline assessment before start of new anti-myeloma therapy (prior to documented disease progression)	Randomization	Censored

Situation	Date of Event (Progression/Death due to PD) or Censored	Outcome Event (Progression/Death due to PD) Or Censored
With adequate post-baseline assessment and new anti-myeloma treatment started (prior to documented disease progression)[4].	Date of last 'adequate' assessment of response[3] (on or prior to starting anti-myeloma therapy)	Censored
Death due to PD before first scheduled assessment (or at baseline and without any adequate assessments)	Date of death	Event
Death due to PD between adequate assessment visits	Date of death	Event
Death from causes other than PD without extended loss-to-follow-up time[2]	Date of death	Censored
Death due to PD or progression after missing two or more scheduled assessments	• Date of randomization if no post-baseline assessments, OR • Date of last 'adequate' assessment of response [3] (prior to missed assessments): since disease assessment is every 4 weeks, a window of 63 days (8 weeks + 7-day window) will be used to determine whether there is extended time without adequate assessment. If the time difference between PD/death due to PD and max (last adequate disease assessment, randomization) is more than 63 days, TTP will be censored at the last adequate disease assessment prior to PD/death due to PD.	Censored

[1]Adequate baseline assessment is defined as at baseline, a patient has at least one of the following measurements: a. Serum M-protein ≥0.5 g/dL (≥5 g/L) or b. Urine M-protein ≥200 mg/24h or c. Serum FLC assay: Involved FLC level ≥10 mg/dL (≥100 mg/L) and an abnormal serum free light chain ratio (<0.26 or >1.65)

[2]Extended loss-to-follow-up time = 8 weeks + 7-day window = 63-day window; without extended loss-to-follow-up time is defined as: ≤63 days; after an extended loss-to-follow-up time is defined as: >63 days. Note that deaths due to causes other than PD will be handled similarly to death due to PD for the derivation of DoR endpoint.

[3]An adequate response assessment is defined as an assessment where the confirmed response is sCR, CR, VGPR, PR, MR, or SD. If the adequate assessment occurred on the same date as new anti-myeloma therapy, it is assumed that the assessment occurred first.

[4]If PD and New anti-myeloma therapy occur on the same day, assume the progression was documented first e.g., outcome is progression and the date is the date of the assessment of progression.

Table 6 Assignments for Progression and Censoring Dates for PFS2 Analysis

Scenario	Event or censored	Date
Death before starting any new line of anti-myeloma therapy	Event	Date of death
PD2[1] is observed	Event	Date of PD2
No PD2[1] is observed and patient died after starting the 1st new line of anti-myeloma therapy	Event	min (end date of the 1st new line of anti-myeloma therapy[2,3], date of death)
No PD2[1] or death is observed **AND** the 1st new line of anti-myeloma therapy ended (if 1st new line of anti-myeloma therapy is intended to be treated until PD)	Event	End date of the 1st new line of anti-myeloma therapy[2]
No PD2[1] or death is observed **AND** the 1st new line of anti-myeloma therapy is completed (if the 1st new line of anti-myeloma therapy is intended to be treated for a fixed number of doses, e.g., cell therapy) **AND** the 2nd new line of anti-myeloma therapy started	Event	Start date of 2nd new line of anti-myeloma therapy – 1 day
Otherwise censored	Censored	Last date known alive

[1]PD2: PD after the 1st new line of anti-myeloma therapy started and before the 2nd new line of anti-myeloma therapy started

[2]Start date of 2nd new line of anti-myeloma therapy – 1 day will be used if end date for 1st new line of anti-myeloma therapy is missing and the 2nd new line of anti-myeloma therapy started

[3]Start date of 2nd new line of anti-myeloma therapy – 1 day will be used instead if the 1st new line of anti-myeloma therapy is treated for a fixed number of doses, e.g., cell therapy

Note: Start date of new lines of anti-myeloma therapy will be defined as the earliest start date of any component within the line. Similarly, the end date of a line of anti-myeloma therapy will be defined as the latest end date of any component within the line.

4.3.2.2. Main Analytical Approach

- **ORR**: The number and percentage of participants with BOR in the following categories will be summarized by treatment arm: sCR, CR, VGPR, PR, overall response (sCR+CR+VGPR+PR), minimal response (MR), stable disease (SD), progressive disease (PD), and not evaluable (NE). The corresponding exact 95% CI for ORR will also be provided. Participants with unknown or missing responses will be treated as non-responders, i.e., these participants will be included in the denominator when calculating percentages of response. The difference in ORR between treatment arms and associated exact 95% CI for the difference will also be calculated.
- **CRR**: summaries of CRR (sCR, CR) by treatment arms will be provided in the same way as ORR.

- **VGPR+**: summaries of VPPR+ (i.e., VGPR or better including sCR, CR, VGPR) by treatment arms will be provided in the same way as ORR.
- **TTBR:** TTBR will be summarized descriptively by treatment arm using median and quartiles in the subset of participants with a confirmed response of PR or better as the Best Overall Response (BOR).
- **TTR**: same as TTBR.
- **TTP**: The distribution of TTP will be summarized using the Kaplan-Meier method by treatment arm. The median, 25th and 75th percentiles of TTP will be estimated and corresponding 95% confidence intervals will be estimated using the Brookmeyer-Crowley method [Brookmeyer, 1982]. Kaplan-Meier plots of TTP will be presented by treatment arm. TTP analysis will also be conducted using Cox proportional hazards model stratified by applicable randomization factors. Refer to Section 4.2.3 for details of the analytical approaches.
- **PFS2**: same as TTP assuming sufficient number of events are observed. In addition, pending on maturity of data, the survival probability at 6, 12 and 18 months with 95% CI will be estimated using Kaplan-Meier method.

4.3.2.3. Sensitivity analyses

All sensitivity/supportive analyses will be performed at the time of primary PFS analysis unless PFS demonstrates statistical significance at the IA2 or required for IDMC review of the benefit:risk.

With the exception of TTBR, TTP and PFS2, all secondary efficacy endpoint analyses will be repeated for the primary estimand but instead using the investigator-assessed response.

Supportive analyses will be provided, evaluating the agreement between the investigator and IRC-assessed response with confirmation, and providing the concordance between best responses for VGPR+, PR+ and CR+, where concordance is calculated as the percent agreement for responders and non-responders.

4.3.3. Pharmacokinetic Analyses

Pomalidomide analyses will be performed on the Pomalidomide PK population. All other pharmacokinetic analyses will be based on the Pharmacokinetic population, unless otherwise specified.

Belantamab Mafodotin Drug Concentration Measures / Concentration-time Data

Linear and semi-logarithmic individual concentration-time profiles and mean and/or median profiles (when appropriate) may be plotted for belantamab mafodotin (ADC) and cys-mcMMAF. Concentrations of belantamab mafodotin (ADC) and cys-mcMMAF will be listed for each participant and summarized (when appropriate) by planned time point.

Linear and semi-logarithmic individual concentration-time profiles and mean and/or median profiles (when appropriate) will be plotted for pomalidomide. Concentrations of pomalidomide will be listed for each participant and summarized (when appropriate) by planned time point.

Pomalidomide PK parameters

Pharmacokinetic parameters, described in Table 7, may be determined for pomalidomide, as data permit, for participants who undergo pomalidomide pharmacokinetic sampling.

Pomalidomide PK parameters may be generated using standard noncompartmental methods using WinNonlin, data permitting, or using a published population PK model [Li, 2015].

Calculations will be based on the actual sampling times.

Table 7 Derived Pomalidomide Pharmacokinetic Parameters

Parameter	Parameter Description
AUC(0-tlast)	Area under the concentration-time curve from time zero to the time of the last quantifiable concentration (C(tlast)) will be calculated using the linear trapezoidal rule for each incremental trapezoid and the log trapezoidal rule for each decremental trapezoid
AUC(0-τ)	Area under the concentration-time curve during the dosing interval
Cmax	Maximum observed concentration, determined directly from the concentration-time data for each cycle. .
tmax	Time to reach Cmax, determined directly from the concentration-time data for each cycle
Cτ, Ctrough	Trough concentration prior to the next dose for each cycle
tlast	Time of last observed quantifiable concentration

Pomalidomide pharmacokinetic parameters will be listed and summarized descriptively (mean, standard deviation, median, minimum, maximum, geometric mean, and the standard deviation, CV%, and 95% CI of log-transformed parameters) by cycle. These may be graphically presented, where appropriate.

Pomalidomide concentration-time data will be displayed similarly to belantamab mafodotin in order to support the Pomalidomide PK parameters.

Details of the planned displays are provided in the OPS and will be based on GSK Data Standards and statistical principles.

4.3.4. Immunogenicity (Anti-Drug Antibody) Analyses

For each participant, the anti-belantamab mafodotin (drug) antibody results, titers, and neutralizing antibody assay results will be listed for each assessment time point. The frequency and percentage of participants with positive and negative anti-drug antibody

and neutralizing antibody assay results will be summarized for each assessment time and overall, for each participant by treatment group. The immunogenicity analyses will be based on the Safety Analysis Set.

4.3.5. Secondary Patient Reported Outcome Analyses

The EORTC QLQ-C30 (version 3.0), EORTC QLQ-IL52 (disease symptom domain of EORTC QLQ-MY20), EORTC QLQ-MY20 and the PRO-CTCAE are three oncology-specific Health-Related Quality-of-Life (HRQoL) assessments that will be analysed in this study as supportive secondary endpoints.

Prior to protocol amendment 1, participants completed the EORTC QLQ-IL52 only. Following PA1, newly enrolled participants completed the EORTC QLQ-MY20. For the EORTC QLQ-IL52, the disease symptom domain of the EORTC QLQ-MY20 will be included in analyses.

The analysis of EORTC QLQ-C30, EORTC QLQ-MY20 and EORTC QLQ-IL52 will be based on the ITT Analysis Set while the analysis of PRO-CTCAE will be based on the Safety Analysis Set.

All questionnaires will be scored according to published scoring guidelines or the developer's guidelines if published guidelines are not available.

Visit-Slotting of PRO data will be implemented to accurately reflect visit schedule from treatment start date, as per protocol Schedule of Activities. Visit-Slotting details will be provided in the Output and Programming Specification (OPS) document.

4.3.5.1. Patient Reported Outcome Version of the Common Term Criteria for Adverse Events (PRO-CTCAE)

The Patient-Reported Outcomes Version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE) is a patient-reported outcome measure developed to evaluate symptomatic toxicity in participants on cancer clinical trials [Basch, 2014]. The PRO-CTCAE was designed to be used as a companion to the Common Terminology Criteria for Adverse Events (CTCAE), the standard lexicon for adverse event reporting in cancer trials. The PRO-CTCAE includes an item library of 124 items representing 78 symptomatic toxicities drawn from the CTCAE. PRO-CTCAE provides a systematic yet flexible tool for descriptive reporting of symptomatic treatment side effects in cancer clinical trials. In the present study, a subset of items selected from the PRO-CTCAE Version 1.0 Item library will be administered.

The levels and related code values for PRO-CTCAE are shown below.

Table 8 PRO-CTCAE Levels and Related Code Values

	Levels and related code values				
Response scale	0	1	2	3	4

	Levels and related code values				
Frequency	Never	Rarely	Occasionally	Frequently	Almost Constantly
Severity	None	Mild	Moderate	Severe	Very severe
Interference	Not at all	A little bit	Somewhat	Quite a bit	Very much
Present/Absence	No	Yes			

For each selected item from the library: proportion of PRO-CTCAE scores for attributes (frequency, severity and/or interference) will be presented with horizontally stacked bar charts by visit for each treatment group, side-by-side in the form of a butterfly plot. Maximum PRO-CTCAE score post-baseline for each item attribute will be summarized by counts and proportions. Proportion of participants with a maximum score of 3 or 4 for each item attribute (severe or very severe, frequently or almost constantly, quite a bit or very much) will also be reported. Proportions will be based on the number of participants with available data and participant with missing response will be excluded from analysis.

4.3.5.2. European Organization for Research and Treatment of Cancer Quality of Life Questionnaire 30-item Core Module (EORTC QLQ-C30)

The EORTC QLQ-C30 is a 30-item questionnaire containing both single- and multi-item measures [Aaronson,1993]. These include five functional scales (Physical, Role, Cognitive, Emotional, and Social Functioning), three symptom scales (Fatigue, Pain, and Nausea/Vomiting), a Global Health Status/QoL scale, and six single items (Constipation, Diarrhea, Insomnia, Dyspnea, Appetite Loss, and Financial Difficulties). Scores for each scale and single-item measure are averaged and transformed linearly to a score ranging from 0–100. Details of deriving domain scores (9 scales and 6 single items) and summary score can be found in Section 6.2.8 and more details will be provided in the OPS document.

- A high score for functional scales and for Global Health Status/QoL and summary score represent better functioning ability or Health-Related Qualify of Life (HRQoL) (higher score indicates improvement)
- whereas a high score for symptom scales and single items represents significant symptomatology [Proskorovsky, 2014] (lower score indicates improvement)

Descriptive summaries (mean, SD, median, min and max) of the actual value and change from baseline at selected time points will be provided for EORTC QLQ-C30 domain and symptom scores, including Global health status/QoL. Time points include but are not limited to worst-case post-baseline, end of treatment and last follow-up visit. The number and percentage of participants with post-baseline score improved by ≥10, and ≥5 points, respectively from baseline score will be summarized at selected time points. The number and percentage will be provided for summary score and each domain score. Should new thresholds be available at the time of the analysis (i.e., from ongoing EORTC group work) these modified thresholds will be used and specified in OPS.

Plots of mean change from baseline (including baseline) and 95% confidence interval over time by visit, and at end of study treatment, last follow-up, and worst-case post-

baseline for selected domain and symptom scores (fatigue, pain and physical functioning) and Global health status/QoL scores will also be provided.

Longitudinal changes from baseline by treatment group for selected EORTC QLQ-C30 domain and symptom scores (fatigue, pain and physical functioning) and Global health status/QoL scores will be explored using a restricted maximum likelihood-based mixed model for repeated measures (MMRM) to compare between-treatment difference adjusting for correlations across multiple time points within a patient and controlling for the baseline value and other variables. Adjusted mean difference and 95% CIs will be presented to illustrate the effect of treatment and associated plots of the least square means and 95% CIs will be provided.

The MMRM model will include patient, treatment, analysis visit, and treatment-by-visit interaction as explanatory variables, the baseline value as a covariate along with the baseline-by-visit interaction. Treatment, visit, and treatment-by-visit interactions will be fixed effects in the model; participant will be treated as a random effect. An unstructured covariance matrix will be used to model the within-participant variance and the Kenward-Roger approximation [Kenward, 2009] will be used to estimate the degrees of freedom. Restricted maximum likelihood (REML) estimation will be used. If the fit of the unstructured covariance structure fails to converge, the following covariance structures will be used in order until convergence is reached: toeplitz with heterogeneity (TOEPH), autoregressive with heterogeneity (ARH[1]), Toeplitz (TOEP), and autoregressive (AR[1]). If there are still issues with the fit of the model or estimation of the treatment effects, participant will be treated as a fixed effect.

4.3.5.3. EORTC QLQ-MY20 and EORTC IL52

The EORTC Quality of Life Questionnaire 20-item Multiple Myeloma module (QLQ-MY20) is a supplement to the QLQ-C30 instrument used in patients with multiple myeloma [Aaronson, 1993; Cocks, 2007]. The module comprises 20 questions that address four myeloma specific HRQoL domains: Disease Symptoms (EORTC IL52), Side Effects of Treatment, Future Perspective, and Body Image. Three of the four QLQ-MY20 domains are multi-item scales: Disease Symptoms (includes bone aches or pain, back pain, hip pain, arm or shoulder pain, chest pain, and pain increasing with activity); Side Effects of Treatment (includes drowsiness, thirst, feeling ill, dry mouth, hair loss, upset by hair loss, tingling hands or feet, restlessness/agitation, acid indigestion/heartburn, and burning or sore eyes); and Future Perspective (includes worry about death and health in the future, and thinking about illness). The Body Image scale is a single-item scale that addresses physical attractiveness. As with the QLQ-C30, QLQ-MY20 domain scores are averaged and transformed linearly to a score ranging from 0–100. A high score for Disease Symptoms and Side Effects of Treatment represents a high level of symptomatology or problems [Proskorovsky, 2014] (lower score indicates improvement), whereas a high score for Future Perspective and Body Image represents better outcomes (higher score indicates improvement). Details of deriving domain score can be found in Section 6.2.8.1. and more details will be provided in OPS document.

Descriptive summaries (mean, SD, median, min and max) of the actual value and change from baseline at selected time points will be provided for each domain score, for the

EORTC QLQ-IL52 and the EORTC QLQ-MY20, separately. The number and percentage of participants with post-baseline score improved by ≥10, and ≥5 points, respectively from baseline score will be summarized at selected time points. Should new thresholds be available at the time of the analysis (i.e., from ongoing EORTC group work) these modified thresholds will be used and specified in the OPS. In addition, the summary will also be provided in the subgroup for participants achieving a confirmed partial response (PR) or better based on the IRC-assessed response for the EORTC QLQ-IL52 disease symptom domain scores only.

Only participants enrolled following PA1 will be included in the EORTC QLQ-MY20 analyses, i.e., only those who were able to complete the EORTC QLQ-MY20 at baseline.

Plots of mean change from baseline (including baseline) and 95% confidence interval over time by visit, and at end of study treatment, last follow-up, and worst-case post-baseline will also be provided for the EORTC QLQ-IL52 disease symptom domain. In addition, the plot will also be provided in the subgroup for participants achieving a confirmed partial response (PR) or better based on the IRC-assessed response.

Longitudinal changes from baseline by treatment group for EORTC QLQ-IL52 and EORTC QLQ-MY20 domain scores will be explored using a restricted maximum likelihood-based mixed model for repeated measures (MMRM), using the same approach described for the EORTC QLQ-C30 analysis. Associated plots of least square means and 95% CIs will be provided for the EORTC QLQ-IL52 disease symptom domain score only.

Compliance of OSDI, PRO-CTCAE, EORTC QLQ-C30, EORTC QLQ-MY20 and EORTC QLQ-IL52

For each of the PROs OSDI, PRO-CTCAE, EORTC QLQ-C30, EORTC QLQ-MY20 and EORTC IL52, overall compliance and compliance by visit will be summarized, based on the following definitions.

- Number of participants expected to complete PRO form: Date of study discontinuation and/or date of death will be used to determine the last visit at which a patient is still expected under PRO follow-up.
- Evaluable forms:
  - OSDI: with at least one non-missing total score or subscale score
  - PRO-CTCAE: with at least one non-missing item score
  - EORTC QLQ-C30, EORTC QLQ-MY20: with at least one non-missing scale/domain score
  - EORTC IL52: with non-missing EORTC IL52 scale/domain score

The overall compliance rate is defined as the number of participants with an evaluable baseline form and at least one evaluable post-baseline form, divided by the number of participants expected to complete the baseline form.

Compliance by visit will be calculated as the number of participants with an evaluable form at that visit, divided by the number of participants expected to complete the form at that visit.

Compliance summaries will only include data up until treatment discontinuation; data collected at EOT and during PFS and OS follow-up will be excluded. The summaries will be based on the analysis set used for primary analysis for each PRO.

4.4. Exploratory Endpoint Analyses

Exploratory endpoints will be analysed at the primary PFS analysis only, unless PFS demonstrates statistical significance at the IA2 or required for IDMC review of the benefit:risk. No additional analyses at subsequent analyses are planned. This includes all subgroup, sensitivity and supportive/supplementary analyses.

4.4.1. Exploratory Pharmacokinetic Analyses

The analyses will be based on the PK Analysis Set, unless otherwise specified.

Concentration-time data from the participants with enhanced PK schedule may be analyzed using a standard non-compartmental approach according to current working practices and using Phoenix WinNonlin, version 6.3 or later, as data permit, to generate the following parameters:

- For belantamab mafodotin, as data permit:
  - For Cycle 1: Maximum observed plasma concentration (Cmax), time to Cmax (tmax), area under the plasma concentration-time curve from zero to the end of the dosing interval, tau, AUC(0-tau), and last time point where the concentration is above the limit of quantification (tlast).
  - For the first 6 doses of belantamab mafodotin (regardless of cycle in which dose occurred): concentration at the end of infusion (C-EOI), and predose plasma concentration (Ctrough)
- For cys-mcMMAF, as data permit:
  - Cmax, tmax, C-EOI, and AUC(0-168h) and tlast will be computed at Cycle 1.

Calculations will be based on the actual sampling times recorded during the study.

Derived PK Parameters listed in Table 9 will be listed and summarized descriptively (mean, standard deviation, median, minimum, maximum, geometric mean, and the standard deviation, CV%, and 95% CI of log-transformed parameters) by cycle/dose (as specified above). These may be graphically presented, where appropriate.

Table 9 Derived Belantamab Mafodotin Pharmacokinetic Parameters

Parameter	Parameter Description
AUC(0-t')	Area under the concentration-time curve to a fixed time t' (i.e., cys-mcMMAF AUC0-168h)
AUC(0-τ)	Area under the concentration-time curve during the dosing interval (i.e., ADC AUC0-504h)
Cmax	Maximum observed concentration, determined directly from the concentration-time data for each cycle. Cmax will not be derived when only predose and EOI samples were collected.
tmax	Time to reach Cmax, determined directly from the concentration-time data for each cycle
Cτ, Ctrough	Trough concentration prior to the next dose for each cycle
C-EOI	Observed plasma concentration at the end of infusion
tlast	Time of last observed quantifiable concentration
CL	Clearance (only from population PK analysis)
Vss	Volume of distribution at steady state (only from population PK analysis)
λz, lambda_z	Terminal phase rate constant (only from population PK analysis)

Plasma belantamab mafodotin and/or cys-mcMMAF concentration-time data from this study may be combined with data from other studies and may be analyzed using a population pharmacokinetic approach. The initial analysis, if performed, will use the most current population pharmacokinetic model at the time of the analysis to generate post hoc belantamab mafodotin pharmacokinetic parameter estimates for the individual participants in Treatment Arm A (B-Pd). Based on the individual post hoc parameter values, dosing information, and sample collection times, belantamab mafodotin plasma concentrations at the time of sample collection will be predicted for each participant. Model evaluation will consist of comparison of model-predicted and observed concentrations. If necessary, model estimation or refinement will be performed. Summary exposure measures (e.g., C_{max}, AUC) will be computed. The results of the population PK analysis, if performed, would include computation of systemic clearance (CL), volume of distribution, and/or terminal phase half-life (t1/2z).

Details of these population pharmacokinetic analyses may be provided under a separate data analysis plan and results may be provided in a separate report.

CPMS analysts or delegate(s) not involved in the study conduct will have access to a blinded population PK dataset (including, but not limited to, concentration, actual dosing information, demographics, and some vital sign and laboratory information, but excluding adverse event and efficacy information) at several time points (e.g., prior to primary PFS analysis) throughout the trial for population PK model development/refinement. Additionally, designated representatives not involved with study conduct may be unblinded for performing population PK and PKPD dataset preparation in support of planned analyses and PK display review. All other personnel will remain blinded to aggregate data by treatment group until database lock.

Details of the planned displays are provided in the OPS and will be based on GSK Data Standards and statistical principles.

4.4.2. Exposure-Response for Efficacy and Safety Endpoints

If deemed appropriate and data permit, exposure-response relationships between belantamab mafodotin exposure (e.g., concentration, Cmax, or AUC) vs. efficacy and safety endpoints (e.g., PFS, ORR, CRR, corneal events) may be explored using population methods. If data permit, the effects of covariates may be explored. Details of these analyses will be reported under a separate SAP, and the results of this analysis will be provided in a separate report.

4.4.3. Exploratory Patient Reported Outcome Analyses

EQ-5D-3L, OSDI and FACT GP5 are the exploratory Health-Related Quality-of-Life (HRQoL) assessments that will be analyzed in this study. EQ-5D-3L analyses will be based on the ITT Analysis Set, OSDI and FACT GP5 will be based on the Safety Analysis Set.

Visit-Slotting of PRO data will be implemented to accurately reflect visit schedule from treatment start date, as per protocol Schedule of Activities. Visit-Slotting details will be provided in the Output and Programming Specification (OPS) document.

4.4.3.1. OSDI

The impact of potential ocular toxicity on function and health-related quality of life will be assessed with the use of the Ocular Surface Disease Index (OSDI). The OSDI is a 12-item questionnaire designed to assess both the frequency of dry eye symptoms and their impact on vision-related functioning [Dougherty, 2011; Schiffman, 2000]. The OSDI has demonstrated good reliability, validity, sensitivity, and specificity, and can be used as a complement to other clinical and subjective measures of dry eye disease by providing a quantifiable assessment of dry eye symptom frequency and the impact of these symptoms on vision-related functioning.

For the OSDI, the total score will be calculated as well as scores for the three subscales (ocular symptom: item 1-3; vision-related function: item 4-9; and environmental triggers: item 10-12).

The total OSDI score = ([sum of scores for all questions answered × 100] / [total number of questions answered ×4]). Subscale scores are computed similarly with only the questions from each subscale used to generate its own score. Subscale scores are computed similarly with only the questions from each subscale used to generate its own score. A score of 100 corresponds to complete disability (a response of “all of the time” to all questions answered), while a score of 0 corresponds to no disability (a response of “none of the time” to all questions answered). Therefore, decrease in score from baseline means improvement.

For total score and each of the three sub-scales, the descriptive summary of the actual value and change from baseline at selected time points will be provided. Plots of mean change from baseline (including baseline) and 95% confidence interval over time by visit, and at end of study treatment, last follow-up, and worst-case post-baseline for

individual domains will also be provided. Plots will be provided for the total score and visual related functioning subscale only. 95% confidence intervals will only be produced if at least 3 values are non-missing at a visit, for each treatment arm.

Additionally, the number and percentage of participants with post-baseline score worsening/deterioration of ≥ minimal clinically important difference (MCID) from baseline score will be summarized at selected time points for Total score, Ocular Symptoms subscale and Vision-related Function subscale (higher score indicates worsening). MCIDs for total score and each sub-scale are listed in Table 10 below [Eliason, 2020]. Should new MCIDs be available at the time of the analysis, these modified thresholds will be used and specified in the OPS. As well as worsening/deterioration (≥MCID score increase from baseline), the number and percentage of participants with post-baseline score demonstrating a meaningful improvement (≥MCID score decrease from highest [worst] score at or following the first deterioration from baseline in OSDI) and resolution (deterioration that returns to baseline) will be summarized. The time to onset of first occurrence of a worsening/deterioration event, duration of first worsening/deterioration event until meaningful improvement and the duration of first deterioration event until resolution will be summarized.

Table 10 MCID for OSDI

Score	Total score	Ocular Symptoms	Vision-related Function
MCID	14.58	16.67	12.5

OSDI compliance will be summarized similarly to the secondary PRO endpoints.

In order to support the OSDI summary, additional details on driving and reading were reported by the site:

- At the time of this visit, the patient:
  - Is currently able to drive with little or no difficulty
  - Is able to drive but with some difficulty mainly due to eyesight issues
  - Stopped driving mainly due to eyesight issues
  - Stopped driving due to other reasons
  - Never drove
- At the time of this visit, the patient:
  - Is currently able to read with little or no difficulty
  - Is able to read but with some difficulty mainly due to eyesight issues
  - Stopped reading mainly due to eyesight issues
  - Stopped reading due to other reasons
  - Never drove

A summary of shifts in response from baseline to worst-case post-baseline will be produced separately, for reading and driving questions.

4.4.3.2. EuroQol Questionnaire (EQ-5D-3L)

The EQ-5D-3L descriptive system comprises the following 5 dimensions: mobility, self-care, usual activities pain/discomfort and anxiety/depression. Each dimension has 3 levels: no problems, some problems, extreme problems. The participant is asked to indicate their health state by selecting the most appropriate statement in each of the 5 dimensions. The EQ VAS records the participant's self-rated health on a vertical, visual analogue scale where the endpoints are labelled 'Best imaginable health state' and 'Worst imaginable health state'. The value of EQ ranges from 0 (worst) to 100 (best). Analysis of this data will be specified separately in another SAP.

4.4.3.3. Patient Global Impression Items

The Patient Global Impression of Severity (PGIS) assesses global impression of symptoms severity at baseline and subsequent timepoints. The second question, the Patient Global Impression of Change (PGIC) serves to rate the global change in symptoms at subsequent time points. In addition to evaluating symptom severity and change, these questions serve as anchors to establish thresholds of clinically meaningful change for the questionnaires in the study [Guy, 1976]. Analysis of this data will be specified separately in another SAP.

4.4.3.4. Functional Assessment of Cancer Therapy – General Population (FACT-GP5)

FACT GP5 is a single item from the FACT-G, which assesses how bothersome the side of effects of treatment are for participants. This item is being included to assess the overall tolerability of treatment from the patient's perspective.

The number and percentage will be reported for each category of FACT GP5 from 0 = Not at all, 1=A little bit, 2 = Somewhat, 3 = Quite a bit, 4 = Very much, 3+4, Any scale >0, by visit and treatment arm. Worst-case post-baseline will also be summarized by treatment arm.

4.4.4. Exploratory MRD Negativity Endpoints

These analyses will be based on the ITT Analysis Set.

4.4.4.1. Sustained MRD Negativity Rate

Sustained MRD negativity rate is defined as the percentage of participants who achieve MRD negative status as assessed by NGS at 10^{-5} threshold at least twice, a minimum of 12 months apart and with no MRD positive (or indeterminate) result in between, during the time of confirmed CR or better response based on IRC-assessed response according

to IMWG. A 1-month window will be considered (i.e. a minimum of 11 months apart), due to the protocol permitting a 1 month window for data collection. MRD samples missing between two valid MRD results, or MRD samples failed the test QC between two valid MRD results, will be excluded from the analysis. Participants who do not achieve sustained MRD negative status and participants without a confirmed CR or better response will be considered as not achieving sustained MRD negativity.

The number and percentage of participants who have sustained MRD negativity will be summarized descriptively by treatment arm, and the corresponding 95% exact CIs will be provided.

4.4.4.2. Imaging plus MRD Negativity Rate

Imaging plus MRD-negativity rate, defined as the percentage of participants who achieve MRD negative status as assessed by NGS at 10^{-5} threshold and have no evidence of disease on PET-CT at least once during the time of confirmed CR or better response based on IRC-assessed response according to IMWG. Participants who do not meet the criteria will be considered as non-imaging plus MRD-negative, i.e., participants meeting any of the following:

- do not achieve MRD negative status (including missing/inconclusive assessment) at least once during the time of confirmed CR or better response, or
- do not have **"**no evidence of disease on PET-CT at least once during the time of confirmed CR or better response"**,** or
- participants without a confirmed CR or better response.

Imaging plus MRD negativity rate will be analysed similarly to sustained MRD negativity rate.

If data are available, imaging-based assessment of MRD (i.e., PET-CT) will also be included in the listing of MRD Negativity Rate data.

4.4.5. Medical Resource Utilization and Health Economics

Healthcare Resource Utilization (HRU) data are collected, the analysis will be described in a separate analysis plan.

4.4.6. Pharmacodynamic and Biomarker Analyses

Pharmacodynamic and Biomarker analyses may be specified within a separate biomarker SAP, which may explore actual change and percent change of free-BCMA expression level from baseline, circulating-free DNA assessments at baseline, during response, and at end of treatment; the relationship between clinical response and other biologic characteristics, including BCMA expression on tumour cells, and sBCMA concentrations. If warranted, the results of these additional analyses will be provided in a separate report.

4.5. Safety Analyses

The safety analyses will be based on the Safety Analysis Set, unless otherwise specified.

4.5.1. Extent of Exposure

Since each component has different dosing schedules that also depend on treatment arm, duration of exposure and dose intensity will be summarized overall and separately for Cycles 1, 2-8 and 9+ to align with the protocol-defined changes in dosing regimen.

Unless otherwise specified, exposure will be summarized by treatment arms.

Duration of exposure

The treatment duration (days) for each component (overall and per cycle block) is defined as:

Treatment duration = [(last date of the study drug) – (first dose date of the study drug) +1]

Where, first dose date of the study drug is defined as the first dose of study drug within the period (overall or cycle block, as appropriate). This is usually Cycle X Day 1 visit but, if the visit is missing, this may be a later day within the cycle. Unscheduled visits should also be considered.

The last date of the study drug is defined as follows:

- If the last dose does not occur within the period (i.e., not the last cycle) then take the start date of the next cycle block -1
- Otherwise, take the last non-zero/non-missing dose date + number of days in the first scheduled off dose period immediately after the last non-zero/non-missing dose, regardless of date of death (if death occurs).

For example:

- If the last non-zero/non-missing belantamab mafodotin dose is at Cycle 2 Day 1 then the last date for belantamab mafodotin is Cycle 2 Day 1 plus 27 days
- If the last non-zero/non-missing pomalidomide dose is at Cycle 2 Day 4 then the last date for pomalidomide is Cycle 2 Day 4 plus 7 days (regardless of treatment arm)
- If the last non-zero/non-missing bortezomib dose is at Cycle 4 Day 8 then the last date for bortezomib is Cycle 4 Day 8 plus 2 days (but for Cycle 10 Day 8 would be plus 13 days)
- If the last non-zero/non-missing dexamethasone dose is at Cycle 1 Day 8 then the last date for dexamethasone is Cycle 1 Day 8 plus 6 days if participant is in Arm A and Cycle 1 Day 8 plus 1 day if participant is in Arm B

The treatment duration (days) for each component (overall and per cycle block) based on date of decision to discontinue treatment will also be derived as follows:

Treatment duration (based on date of decision to discontinue treatment) = [(latest of[last date of the study drug, date of decision to discontinue the study drug]) – (first dose date of the study drug) +1]

The total duration of exposure (days) across all components (overall and per cycle block [1, 2-8, 9+]) is defined as:

Total duration exposure=[(Overall study treatment end date) – (Overall study treatment start date) + 1]

The start date of the overall study treatment is defined as the first dose date of belantamab mafodotin, bortezomib, pomalidomide or dexamethasone, whichever is earlier (i.e., the first study drug start date). The overall study treatment end date (for the purpose of deriving total duration exposure) is the latest last date of the component across all components, as derived above.

Total duration of exposure based on date of decision to discontinue treatment across all components will also be derived as follows:

Total duration exposure (based on date of decision to discontinue treatment) =[(latest of [overall study treatment end date, latest date of decision to discontinue treatment]) – (Overall study treatment start date) + 1]

Treatment duration (days) and dose intensity (units/cycle) will be summarized for each component using mean, median, standard deviation, minimum, and maximum. The total duration of exposure across all components will also be summarized.

Dose intensity

The dose intensity (units/cycle) calculation (overall and per cycle block [1, 2-8, 9+]) is described below:

- **Cumulative actual dose divided by treatment duration in 3 weeks** (treatment duration in days / 21); will be used for pomalidomide and dexamethasone for treatment **Arm B (PVd)** and for bortezomib for both arms (although not expected for Arm A: B-Pd, included in case of dosing errors)
    - This will be repeated using treatment duration (based on date of decision to discontinue treatment) in the denominator
- **Cumulative actual dose divided by treatment duration in 4 weeks** (treatment duration in days / 28); will be used for pomalidomide and dexamethasone for treatment **Arm A (B-Pd)** and for belantamab mafodotin for both arms (although not expected for Arm B: PVd, included in case of dosing errors)
    - This will be repeated using treatment duration (based on date of decision to discontinue treatment) in the denominator

Notes:

1. Dose intensity units will depend on treatment being summarized (belantamab mafodotin, pomalidomide, bortezomib, dexamethasone).
2. For bortezomib, baseline body surface area (BSA) in m^2 will be used to convert bortezomib dose in mg to mg/m^2 using the following BSA formula [Haycock, 1978]:

   Baseline BSA (m^2) = 0.024265 x Baseline Height(cm)$^{0.3964}$ x Baseline Weight(kg)$^{0.5378}$
3. If baseline height or weight are missing, we will use the closest height/weight date to baseline.
4. For Belantamab mafodotin, if units are collected in mg, units are converted to mg/kg based on baseline weight. If the change of body weight is greater than 10%, the dose should be re-calculated based on the actual body weight at the time of dosing.

The following will also be summarised for each component:

- **Number of cycles** (non-zero, non-missing)
- **Cumulative dose** = sum of all actual (non-zero, non-missing) doses taken across the treatment.
- **Dose exposure** = total number of days on the study drug during the treatment phase, periods of dose break per protocol or dose interruptions will be excluded. Only non-zero, non-missing dosing days will be considered.
- **Average daily dose** = cumulative dose / dose exposure (unit/day).
- **Relative dose intensity** = dose intensity / planned dose intensity, where the planned dose intensity is defined as the expected dose intensity per protocol, given the number of actual dosing cycles i.e. for a participant receiving 4 cycles of treatment (considering all components), the planned dose intensity of belantamab mafodotin is *= [2.5 + (3 x 1.9)]/4= 2.05mg/kg/cycle.* For the overall record for belantamab mafodotin and bortezomib only, the relative dose intensity will also be calculated such that the planned dose intensity is defined as the cumulative planned dose up to the last date of the study drug divided by treatment duration (in 3 or 4 weeks, dependent upon treatment arm). The reduced dose of dexamethasone for participants >75 years will be considered.

Descriptive statistics will be produced similarly to dose intensity for number of cycles, cumulative dose, dose exposure, average daily dose and relative dose intensity. With the exception of the number of cycles, these will be summarized overall and by cycle block, separately, for each component. The number of cycles will be summarized overall.

A listing of exposure will be produced. A separate supportive listing of overall exposure and dose modifications will also be included.

Summaries of Dose Modifications

Summaries of dose modifications will be provided. All the dose reductions, infusion interruptions, and dose delays will be summarized and/or listed by component. Separate listings of dose reductions, dose interruptions, dose delays and dose escalations will be produced.

Dose Delays

Dose Delays will be collected in the exposure eCRF. Delays are expected to be captured at every visit where dose was skipped/missed/no dose was administered when planned up until the decision to discontinue treatment. For the purpose of analysis, this means that successive delays captured in the eCRF reflect a single prolonged delay. Therefore, a single dose delay will be reported when consecutive planned dosing dates are missed/skipped/no dose is administered when planned. It also means that delays beyond the last dose will be captured, to reflect the dose holds ongoing when decision is made to discontinue treatment.

Dose delays will be summarised by number of delays (i.e., how many periods of delayed dosing occurred), number of cycles with at least 1 delay, reasons for the delays (where multiple reasons may be reflected for one delay), and delay duration (days). The number and percentage of the delays for intervals of 1-21, 22-28, 29-42, 43-56 and >56, will be computed. For bortezomib and dexamethasone the delay intervals will be defined as 1-7, 8-14, 15-21, and >21 days.

Primary reasons for dose reductions and dose delays will also be summarized by visit.

Duration of delays is defined as period from the expected start date of dose to subsequent actual dosing date following dose delay. Duration of delay will be missing if dosing did not resume prior to data cut off, i.e., if the delay was ongoing or there was a subsequent decision to discontinue treatment.

A sensitivity analysis will be performed for dose delay. This summary will calculate dose delays for belantamab mafodotin and bortezomib only, by deriving the dose delays as follows:

1. If time between first dose of each cycle is more than xx days, then count as a delay:
    a) Belantamab mafodotin: xx days = 31 days (Q4W+3 days)
    b) Bortezomib: xx days = 24 days (Q3W+3 days)
2. Count an addition delay from a participants last dose of belantamab mafodotin/bortezomib to “end of study” if this is more than xx days. For “end of study”, consider the following:
    a) Date of death
    b) Date of decision to discontinue treatment

c) Treatment discontinuation date (EOT visit date)

d) Start of new anti-myeloma therapy

e) Last contact date

This calculated dose delay summary will include the number of participants with any dose delay, total number of dose delays, number of dose delays categories, and delay duration categories. Additional details will be described in the OPS.

Note: for the calculated dose delays, the protocol defined reduced dose levels of belantamab with Q8W dosing will be counted as a delay.

Dose Reductions

Dose Reductions are recorded on the exposure eCRF on every visit that a new reduction to a lower dose level (e.g., from Dose level 0 to Dose level -1; from Dose level -1 to Dose level -2, etc) or reduced frequency occurred. Protocol-mandated reduction (e.g., reduction of belantamab mafodotin at Cycle 2 Day 1) will not be counted as reduction. Number of dose reductions and reasons for reductions will be summarized. Additionally, the number of cycles with at least 1 reduction will be summarized.

If dose reductions are reflected at subsequent visits from the initial reduction, then a sensitivity analysis may be performed where the first reduction recorded on the eCRF will be considered and any subsequent reductions will be considered only if a further reduction (80% or less of previous non-missing, non-zero dose) was applied. Note: reductions to a dose of zero and reductions in frequency will not be considered, unless they are the first reported reduction.

To understand the extent of dose modifications for the combination therapy, the following will be added to the summary of total duration of exposure across all components:

- Number of cycles across all components (i.e., where at least one component was administered)
- Number of cycles with at least one dose reduction of any component
- Number of cycles with at least one dose delay of any component
- Number of cycles with at least one dose modification of any component

Additionally, **Duration of Follow-Up** will be summarized and is defined as the time from randomization to last known alive date or death. For participants with last known alive date or death beyond the date of data cut off, the date of data cut off will be used. Duration of follow-up will be summarized for all participants and separately for ongoing participants, both categorically and as a continuous variable.

4.5.2. Adverse Events

The safety analyses will be based on the Safety population, unless otherwise specified.

Unless otherwise specified, all adverse events whether serious or non-serious, will be reported from the start of treatment until at least 70 days after the last dose of study treatment, until the participant withdraws consent for study participation, or until the participant starts subsequent anti-myeloma therapy, whichever occurs first (i.e., treatment emergent).

Adverse events analyses including the analysis of adverse events (AEs), Serious AEs (SAEs) and other significant AEs will be based on GSK Core Data Standards. Separate listings will be produced for all adverse events, a list of Subject IDs associated with each individual adverse event and a listing of reasons for considering as a serious adverse event.

Unless otherwise specified, AEs will be summarized by treatment arms. AEs related to any component or AEs related to all components will be considered related to study treatment.

An overview summary of AEs will be presented, including counts and percentages of participants with:

- any AE
- AEs related to study treatment
- Grade 3&4 AEs
- Grade 3&4 AEs related to study treatment
- AEs leading to permanent discontinuation of study treatment
- AEs related to study treatment and leading to permanent discontinuation of study treatment
- AE leading to dose reductions
- AEs leading to dose interruption/delays
- SAEs
- SAEs related to study treatment
- fatal SAEs, and
- fatal SAEs related to study treatment

A summary of non-serious AEs that occurred in 5% of the participants or above will be provided (no rounding for the percentage will be used in terms of 5% threshold, e.g., event with 4.9% incidence rate should not be included in this table). The summary will be displayed by System Organ Class (SOC) and Preferred Term (PT).

Adverse events will be coded using the latest version of the standard Medical Dictionary for Regulatory Affairs (MedDRA dictionary) and graded by the investigator according to the NCI-CTCAE [CTCAE, v5.0] or the KVA scale, as appropriate.

A summary of number and percentage of participants with any adverse events by maximum grade will be produced. AEs will be sorted by PT in descending order. The summary will use the following algorithms for counting the participant:

- **Preferred term row**: participants experiencing the same AE preferred term several times with different grades will only be counted once with the maximum grade.
- **Any event row***:* Each participant with at least one adverse event will be counted only once at the maximum grade no matter how many events they have.

The frequency and percentage of AEs (all grades) will be summarized and displayed in three ways: 1) in descending order by maximum grade and PT 2) in descending order by SOC and PT and 3) in descending order by PT only. In the SOC row, the number of participants with multiple events under the same SOC will be counted once.

A listing of AEs occurring between the start of new anti-myeloma therapy and study treatment stop day + 70 days will be produced.

A separate summary will be provided for study treatment-related AEs. A study treatment-related AE is defined as an AE for which the investigator classifies the relationship to study treatment as “Yes”. A worst-case scenario approach will be taken to handle missing relatedness data, i.e., the summary table will include events with the relationship to study treatment as ‘Yes’ or missing. The summary table will be displayed in two ways: 1) by maximum grade sorted by PT in descending order and 2) in descending order by PT only.

In addition, AEs of maximum grade of 3 or higher will be summarized separately by PT and by SOC and PT.

A summary of common AEs (≥10% in either treatment group) will also be produced presenting the number and percentage of participants with the AE in both arms, sorted by relative risk and presenting the relative risk, associated 95% Wald CIs and forest plot (on the log scale) will be produced.

All AEs will be listed which will include subject IDs for each individual AE.

Details of the planned displays are provided in the OPS.

A figure outlining key benefit and risk endpoints will be produced. This will be produced as a forest plot and will include the number of participants experiencing each event, along with hazard ratios, RMST ratios, treatment difference, risk ratios (or other relevant summary statistics) and associated 95% confidence intervals. PFS, OS, DoR and MRD negativity will be included to summarize the benefit and Grade 3+ Thrombocytopenia, Grade 3+ Infections (Infections and Infestations SOC), Grade 3+ Pneumonia, Grade 3+ Neutropenia, Grade 3+ AEs and AEs leading to treatment discontinuation of all three

components will be included to evaluate the risk. Additional endpoints may be considered as clinically appropriate at the time of analysis.

4.5.2.1. Adverse Events of Special Interest

Adverse events of special interest (AESI) for belantamab mafodotin are corneal events (CTCAE), thrombocytopenia and infusion-related reactions (IRRs). In addition to events identified and collected in eCRF, a comprehensive list of MedDRA terms based on clinical review will be used to identify each type of event. Specifically for IRR, events identified by the comprehensive list of MedDRA terms based on clinical review would only be considered IRR if the event onset was reported on an infusion day after the start of infusion or within 24 hours following end of infusion AND led to a temporary interruption or prolongation of infusion time or treatment withdrawal. Changes to the MedDRA dictionary could occur between the start of the study and the time of reporting and/or emerging data from on-going studies may highlight additional adverse events of interest, therefore the list of terms to be used for each event of interest and the specific events of interest will be based on the SRT agreements in place at the time of reporting.

All symptomatic ocular AEs are to be reported based on NCI-CTCAE v5.0 criteria for the study duration. Belantamab mafodotin-related ocular exam findings will be reported based on NCI-CTCAE v5.0 criteria for eye disorders prior to consenting to protocol amendment 1 and will be graded according to the KVA scale after consenting to protocol amendment 1. Severity of all other AESIs will be graded using National Cancer Institute-Common Toxicity Criteria for Adverse Events [CTCAE, v5.0]. Guidelines for dose modifications and interruptions for management of common toxicities associated with the study treatment(s) are provided in the protocol. Dose modifications for belantamab mafodotin corneal events (following re-consent to protocol amendment 1) will be based on grading of corneal events according to the guidelines of the Keratopathy Visual Acuity (KVA) Scale.

Summaries of the number and percentage of participants with these AESIs will be provided for each category of AESI separately by preferred term and maximum grade. The time to onset and duration of first occurrence for each type of AESI will be summarized using summary statistics mean, standard deviation, median, minimum, and maximum. The number and percentage of participants who have time to onset of first occurrence (1-21, 22-28, 29-42, 43-56, 57-63, 64-84, >84 days) will be reported for corneal events (CTCAE) and thrombocytopenia. The number and percentage of participants who have time to onset of first occurrence in categories of (0-6, >6-12, >12-18, >18-24, >24 hours) will be reported for infusion-related reactions. The number and percentage of participants who have duration of first occurrence (1-21, 22-28, 29-42, 43-56, 57-63, 64-84, >84 days) will be reported for corneal events (CTCAE) and thrombocytopenia. The number and percentage of participants who have duration of first occurrence in categories of (0-12, >12-24, >24 hours) will be reported for infusion-related reactions. For an AESI which is based on a single adverse event term, the onset and duration will be calculated based on the start and end dates of the single term. For an AESI which is based on multiple adverse event terms, the onset and duration will be calculated by looking across all terms for the AESIs. The derived start date is identified as the onset of any term defined as the AESI. The derived end date is identified as last

end date for any terms once all concurrent terms for the AESI have resolved, i.e., the first time a participant is free of any adverse event term defined as the AESI.

The summary of event characteristics will be provided for each AESI respectively, including number of participants with any event, number of events, number of participants with any event that is serious, number of participants with any event that is related to study intervention, number of occurrences (1, 2, 3 or more), maximum grade, maximum grade for events related to study intervention, outcomes and the action taken for the event. The percentage will be calculated in two ways, one with number of participants with event as the denominator and the other with total number of participants as the denominator. The worst-case approach will be applied at participant level for the maximum grade, i.e., a participant will only be counted once as the worst case from all the events experienced by the participant. For action taken to an event, a participant will be counted once under each action, e.g., if a participant has an event leading to both study intervention discontinuation and dose reduction, the participants will be counted once under both actions.

The Summary of Characteristics II of Keratopathy Visual Acuity (KVA) Scale Events (Grades 2+) (See Section 4.5.2.4) will also be repeated for corneal events (CTCAE) including all grades.

For each of these AESI, a summary of cumulative incidence by number of cycles (≤1, ≤2, ≤4, ≤6, ≤8, ≤10, Any) received at first occurrence will be provided.

For thrombocytopenia, a separate summary of thrombocytopenia and bleeding event will be produced. This will include:

- number and percentage of participants with grade 3 or 4 platelet count decreased based on AE data
- number and percentage of participants with grade 2 or above bleeding event based on AE data
- number and percentage of participants with concomitant grade 3 or 4 platelet count decrease (based on AE data, or lab data collected following treatment start date) and grade 2 or above bleeding event (based on AE data where site indicates bleeding is associated with the thrombocytopenia AESI). A bleeding event will be considered as concomitant only if the start date is within ± 3 days of the onset of the platelet count decrease event.

The summary of event characteristics display produced for AESIs will be repeated for the other important risk, neutropenia. Additionally, a summary of neutropenia and infection events will be provided, summarizing:

- number and percentage of participants with grade 3 or 4 neutrophil count decreased based on AE data
- number and percentage of participants with infection event based on AE data

- number and percentage of participants with concomitant grade 3 or 4 neutrophil count decrease and infection event. Infections will be considered concomitant only if started within +/-7 days of the neutrophil count decrease.

A comprehensive list of MedDRA preferred terms for neutropenia and infections based on clinical review will be used to identify neutropenia and infection events.

Details of the planned displays are provided in the OPS.

For all AESIs and other important risks, changes to the MedDRA dictionary may occur between the start of the study and the time of reporting and/or emerging data from on-going studies may highlight additional risks; therefore, the list of terms to be used for each event and the specific events will be based on the safety review team (SRT) agreements in place at the time of reporting.

4.5.2.2. Death and Serious Adverse Events

All deaths will be summarised based on the number and percentage of participants. This summary will classify participants by time of death relative to the last dose of (any component of) treatment (>30 days or ≤30 days) and primary cause of death as indicated in the eCRF. Deaths related to COVID-19 may also be summarised. For summaries of death, both deaths captured while on study and those retrieved following study discontinuation/withdrawal will be included. A supportive listing will be generated to provide participant-specific details on participants who died.

All SAEs will be tabulated based on the number and percentage of participants who experienced the event. Separate summaries will also be provided for study treatment-related SAEs. The summary tables will be displayed in descending order by PT. The summary of all SAEs will also be created by SOC and PT, including the number of occurrences.

A study treatment-related SAE is defined as an SAE for which the investigator classifies the relationship to study treatment as “Yes”. A worst-case scenario approach will be taken to handle missing data, i.e., the summary table will include events with the relationship to study treatment as ‘Yes’ or missing. Study treatment-related SAEs will further be summarized by fatality: fatal vs non-fatal. A separate summary will be produced for non-serious treatment-related AEs.

SAEs are included in the listing of all adverse events. A separate supportive listing with participant-level details will be generated for Non-Fatal SAEs. Fatal SAEs will be summarized.

4.5.2.3. Adverse Events Leading to Discontinuation and Dose Modification

The following categories of AEs will be summarized separately by PT:

- AEs Leading to Permanent Discontinuation of Study Treatment
- AEs Leading to Dose Interruption/Delays

- AEs Leadings to Dose Reductions

Other dose modifications (e.g., dose escalations) are not expected per protocol. If these occur, they will be identified in the listing of all adverse events.

A summary of corneal events (CTCAE) leading to permanent discontinuation of belantamab mafodotin will also be produced.

4.5.2.4. Ocular Findings from Ophthalmic Exam

Visit-Slotting of ocular data will be implemented due to changes in visit labels during data collection, i.e. changed from capturing as Week X to Cycle X Day Y. Visit-Slotting details will be provided in the Output and Programming Specification (OPS) document.

Ocular Exam and Visual Acuity

As outlined in study protocol, ophthalmic exams are scheduled at screening, while on treatment, and at end of treatment for participants in both treatment arms. Ophthalmic exams in follow-up period (if needed) will only be conducted for B-Pd treatment arm. The ocular findings from ophthalmic exams will be summarized descriptively and from baseline to last follow-up, the following analyses will be performed:

Visual Acuity

The best corrected Visual Acuity (BCVA) summary will be based on the Logarithm of the Minimum Angle of Resolution (logMAR score), where:

$$\text{logMAR score} = -\log_{10}(\text{Snellen Acuity Score})$$

The following categories of logMAR score changes from baseline are defined:

- No change/improved vision is defined as a change from baseline <0.12;
- Possible worsened vision is defined as a change from baseline ≥0.12 to <0.3;
- Definite worsened vision is defined as a change from baseline ≥0.3.

- A summary of Best Corrected Visual Acuity (BCVA) logMAR score at selected timepoints will be presented.
- A summary of Worst Change Post-Baseline (by eye [Right and Left] and also based on participant's best and worst eye at baseline) in BCVA Score (logMAR Score) will be provided for categories "increase ≥0.12 to <0.3", "increase ≥0.3 to <0.6", "increase ≥0.6".
    - An additional summary of Worst Post-Baseline Change in Best Corrected Visual Acuity *Snellen-equivalent* will be provided.

- A summary of characteristics (II) of Definitely Worsened Best Corrected Visual Acuity (logMAR score change from baseline ≥0.3 in either eye) will include the following:
    - time to onset of first occurrence: summary statistics and frequency/percentage in categories (1-21, 22-28, 29-42, 43-56, 57-63, 64-84, >84 days);
    - outcome of first occurrence;
    - duration of first occurrence: summary statistics and frequency/percentage in categories (1-21, 22-28, 29-42, 43-56, 57-63, 64-84, >84days);
    - number of occurrences based on participants with the event i.e., worsened vision;
    - outcome post treatment exposure (i.e., for those participants with ongoing event as of last follow up);
    - duration of occurrence post treatment exposure;
    - outcome of last event;
    - time to last ocular exam date since the last dose (days; summary statistics and ≤80 or >80 days).
    - duration of occurrence for participants who resolved for last event
    - outcome of last event in participants who discontinued from study treatment.

Duration of an occurrence is defined as the time from onset of any worsened vision (change from baseline logMAR score ≥0.3 in either eye) to the first time the participant is free of worsened vision (i.e., < 0.3 logMAR score change from baseline in both eyes). It requires at least one day gap between the resolution of all worsened vision from first occurrence to the onset of second occurrence.

Summaries of unilateral and bilateral worsening in BCVA to 20/50 or worse will be produced. A unilateral event is defined as a period of time in which at least one eye has Snellen score of 20/50 or worse. The other eye must be better than 20/50 or missing but cannot be equal to or worse than 20/50. A period of unilateral worsened vision may involve a period where the worse eye changes from right to left or left to right. A bilateral event is defined a period of time in which both eyes have a Snellen score of 20/50 or worse. A participant may only be considered unilateral *or* bilateral at a visit, not both. The summaries will be repeated for worsening to 20/200. For the bilateral summaries (both 20/50 and 20/200 worsening), only participants with a Snellen score of 20/50 or better in at least one eye at baseline will be included. The summary will include the same descriptive statistics as included for the summary of characteristics (II) of Definitely Worsened Best Corrected Visual Acuity, defined above.

Corneal Exam

- Shift table from baseline to worst case post-baseline by eye (R/L) and worst eye for corneal epithelium findings:

- Corneal epithelium (Normal to Abnormal),
- Microcyst-like deposits (No to Yes).
- Subepithelial haze (No to Yes)
- Stromal opacity (No to Yes)
- Epithelial edema (No to Yes)
- Corneal epithelia defect (No to Yes)
    - Corneal erosion (No to Yes)
    - Corneal ulcer (No to Yes)
- Corneal neovascularization (No to Yes)
- Superficial punctate keratopathy severity (None vs Mild vs Moderate vs Severe)
- Stippled peripheral corneal staining ±vortex/whorl staining pattern (No to Yes)

Missing categories will be presented where relevant. If corneal epithelium exam results are normal, then all exam indicators will be considered as "No" and Superficial punctate keratopathy severity will be considered as None. Similarly, if corneal epithelia defect is No, then the corneal erosion and corneal ulcer indicators will be considered as No.

Lens

Supportive listings may be provided as required, e.g.

- Listing of participants who fall into each of the two categories of change from baseline in BCVA: "possible worsened vision", "definite worsened vision" with sub-categories "increase ≥0.3 to <0.6", and "increase ≥0.6".
- Listing of participants who fall into each of the two categories of decline in BCVA to 'light perception' (LP) or 'no light perception' (NLP) anytime post-baseline.
- Listings of participants with cataracts at baseline, pseudophakia at Baseline, who developed cataracts post-baseline, and who underwent cataract surgery post-baseline will be provided.
- Listing of impact on driving and reading if data are available
- Listing of corneal exam results
- Listing of cataract lens findings
- Listing of visual acuity and abnormal corneal exam results

Corneal Events Based on Keratopathy and Visual Acuity Scale (KVA Scale)

For ocular exam visits based on the ocular worksheet under the original protocol, KVA grade is not expected to be collected. For ocular exam visits based on the ocular

worksheet under the protocol amendment 1, KVA grade is expected to be collected for Arm A (B-Pd). To perform KVA analysis including ocular visits based on both the original and protocol amendment 1, the following methods will be used:

- **Investigator-Reported KVA grade**

  All data where Investigator-Reported KVA Grade is available will be reported. Missing KVA Grade (e.g., for assessments prior to protocol amendment 1 or not collected) will not be replaced. Summaries reporting Investigator-Reported KVA grade will be created for Arm A only.

- **Sponsor-Assessed KVA grade**

  At ocular exam visits where Investigator-Reported KVA is present and also where it is missing (e.g., for assessments prior to protocol amendment 1 or not collected), if data permit, KVA grade will be based on medical review, assisted by a programming algorithm. Investigator-Reported KVA Grade will not be used. Summaries reporting Sponsor-Assessed KVA grade will be created for Arm A only, with the exception of the Summary of Characteristics of Sponsor-Assessed Keratopathy Visual Acuity (KVA) Scale (overall, and for the visual acuity and corneal exam findings subcomponents, separately) which will be summarized for both treatment arms.

Unless otherwise specified, for the following analyses, KVA scale events will be summarized for Arm A (B-Pd), based at participant level, and separately by Sponsor-Assessed KVA grade and Investigator-Reported KVA grade. Analyses will include:

- **Keratopathy Visual Acuity (KVA) Scale Events Overview**

  Participants with any event, grade 3/4 events, events leading to permanent discontinuation of study treatment, events leading to dose reduction, events leading to dose interruption/delay. This will be produced for investigator-reported KVA only.

- **Summary of Characteristics of Keratopathy Visual Acuity (KVA) Scale**

  Participants with any event, events characteristics (study treatment related), maximum grade and action taken (study treatment withdrawn, dose reduced, dose not changed, dose interrupted/delayed) will be included. Percentages will be calculated based on all participants and also based on participants with an event.

  Note: for the sponsor-assessed displays, dose modification information will not be presented (as dose modification was performed based on investigator assessment).

  In addition to the overall KVA grade, the display will be repeated for the visual acuity and corneal exam findings subcomponents.

  Investigator-reported KVA grade data entry is only expected for data entry after consenting to protocol amendment 1 or later. Therefore, this display will be repeated for the subgroup of participants who had KVA data entry for all post-baseline assessments (based on Investigator-reported KVA grade). This subgroup will be

identified as all participants with date of randomization + 31 days (4 weeks + 3-day SoA window) on or after earliest consent date to protocol amendment 1 or later.

- **Summary of Characteristics II of Keratopathy Visual Acuity (KVA) Scale Events (Grades 2+)**

  A more detailed summary which includes time to onset of first occurrence, outcome of first occurrence, duration of first occurrence, number of occurrences, outcome post-treatment exposure, duration of event post-treatment exposure, outcome of last event, time to last ocular exam date since last dose, duration of event for participants who resolved for last event, outcome of last event in participants who discontinued from study treatment. Duration will be defined as time from onset of any grade 2+ event until the event is resolved (i.e., grade 1 or better).

  Note: for the sponsor-assessed displays, dose modification information will not be presented (as dose modification was performed based on investigator assessment).

  In addition to the overall KVA grade, the display will be repeated for the visual acuity and corneal exam findings subcomponents.

  This display will be repeated for the subgroup of participants who had KVA data entry for all post-baseline assessments (based on Investigator-reported KVA grade).

- **Summary of Cumulative Incidence of Keratopathy Visual Acuity (KVA) Scale Events by KVA Grade and Number of Doses Received at First Occurrence**

  A summary by KVA grade (1, 2, 3, 4, any) and number of doses of belantamab mafodotin (≤1, ≤2, ≤4, ≤6, ≤8, ≤10, any) received at first occurrence will be provided for Investigator-Reported KVA Grade only.

A table summarising the concordance between the Investigator-Reported and Sponsor-Assessed KVA grade will be produced for visits where Investigator-Reported KVA grade is available. The Sponsor-Assessed KVA grade will be calculated where Investigator-Reported KVA grade is present. This analysis will identify whether investigators have assessed KVA grade in line with sponsor expectations. The concordance will be calculated based on the worst case assessment and will also be summarized by visit.

The end of treatment exposure (or any summary measure with reference to dosing) for all KVA related outputs will be defined in relation to belantamab mafodotin only. Refer to Section 4.5.1.

Additionally, a corneal events display will be created combining AE and KVA source data. The summary will include the number and percentage of participants with, as well as the number of occurrences of:

- any corneal AE
- any corneal events by KVA scale
- corneal AE OR corneal events by KVA scale

- corneal AE AND corneal events by KVA scale
- corneal AE only
- corneal events by KVA scale only

The summary of Corneal Events (CTCAE and Investigator-Reported KVA) will be repeated for the subgroup of participants who had KVA data entry for all post-baseline assessments (based on Investigator-reported KVA grade).

CTCAE: Changes in Best Corrected Visual Acuity (BCVA) are converted from Snellen charts and converted to logMar score for assessment of visual acuity, coded using the standard MedDRA and graded by the investigator according to the NCI-CTCAE v5.0. Worst post-baseline BCVA change will be calculated using CTCAE grades. Only grade 2+ changes will be summarized.

For participants experiencing an investigator-reported KVA event of grade 2+, the grade of first occurrence and subsequent second occurrence (grade 2, 3, 4 or N/A for those not experiencing a subsequent grade 2+ event) will be summarized in the form of a shift table.

The total number of all Investigator-reported KVA events will be summarized on the event level, and a summary of the duration of all occurrences will be provided.

Dose Modifications

Additionally, a dose modification display will be created combining AE and KVA source data. Dose modifications (reduction, interruption / delay, treatment discontinuation) will be summarized at the participant level by the categories of reasons that lead to the dose modification, including any AE, non-corneal AE, corneal AE, corneal AE or corneal events by KVA scale, non-corneal AE or corneal events by KVA scale, any AE or corneal events by KVA scale.

A summary of Investigator-Reported KVA events of grade 2 or above leading to action taken with study treatment will be produced. The summary will be repeated, where one will summarize the number and percentage of participants experiencing any dose modification and the grade of the event at the time of dose modification, and the other display will consider the maximum grade of event, where any dose modification that occurred during the event or prior to subsequent event will be included.

4.5.3. Additional Safety Assessments (if applicable)

The analyses of non-laboratory safety test results including ECGs and vital signs will be based on GSK Core Data Standards, unless otherwise specified. Details of the planned displays are provided in OPS.

Unless otherwise specified, the denominator in percentage calculation at each scheduled visit will be based on the number of participants with non-missing value at each particular visit.

4.5.3.1. Laboratory Data

Laboratory evaluations including the analyses of Chemistry laboratory tests, Hematology laboratory tests, Urinalysis, and liver function tests will be based on GSK Core Data Standards. All laboratory data will be listed (Hematology, Chemistry and Urinalysis). Separate listings will be produced for participants with any value of potential critical importance outside of the normal range.

Summary of change from baseline by scheduled visits using mean, median, standard deviation, minimum and maximum will be provided.

Summaries of worst-case grade increase from baseline grade will be provided for all the lab tests that are gradable by CTCAE v5.0. These summaries will display the number and percentage of participants with a maximum post-baseline grade increasing from their baseline grade. Any increase in grade from baseline will be summarized along with any increase to a maximum grade of 3 and any increase to a maximum grade of 4. Missing baseline grade will be assumed as grade 0. For some laboratory tests, baseline is not gradeable and the shift from baseline is accounted for in the post-baseline grades – for these tests, the worst grade post-baseline will be summarized. For laboratory tests that are graded for both low and high values, summaries will be done separately and labelled by direction, e.g., sodium will be summarized as hyponatremia and hypernatremia.

For lab tests that are not gradable by CTCAE v5.0, summaries of worst-case changes from baseline with respect to normal range will be generated. Decreases to low, changes to normal or no changes from baseline, and increases to high will be summarized for the worst-case post-baseline. If a participant has a decrease to low and an increase to high during the same time interval, then the participant is counted in both the “Decrease to Low” categories and the “Increase to High” categories.

Separate summary tables for hematology, and chemistry laboratory tests will be produced. Liver function laboratory tests will be included with chemistry laboratory tests.

For spot urine albumin/creatinine ratio (mg/g), a shift table from baseline to worst post-baseline will be provided.

Detailed derivation of baseline assessment is specified in Section 4.1.2.

Unless otherwise specified, the denominator in percentage calculation at each scheduled visit will be based on the number of participants with non-missing value at each particular visit.

Summaries of hepatobiliary laboratory events including possible Hy’s law cases will be provided in addition to what has been described above. Possible Hy’s law cases are defined as any elevated alanine aminotransferase (ALT)>3×upper limit of normal (ULN), total bilirubin≥2×ULN and alkaline phosphatase (ALP)<2×ULN/missing. Total bilirubin≥2×ULN can be within 28 days following the ALT elevation and if direct bilirubin is available on the same day, it must be ≥ 35% of total bilirubin. ALP<2×ULN/missing means it is satisfied unless the ALP is ≥2xULN at the time of bilirubin elevation. The summary will be produced for worst case post baseline only.

An e-DISH plot of maximum post baseline total bilirubin versus maximum post baseline ALT will be created.

A plot of maximum post baseline ALT versus baseline ALT will also be provided.

A summary of Liver Monitoring/Stopping Event Reporting will be provided. Participants with liver monitoring/stopping events will be listed.

An additional summary of liver restarts/re-challenges will also be produced.

4.5.3.2. Vital Signs

The change from baseline of vital signs (temperature, systolic and diastolic blood pressure, heart rate) will be summarized by scheduled visit using mean, median, standard deviation, minimum and maximum. A supportive listing will also be produced.

A summary of changes in heart rate and temperature comparing the baseline value to the worst-case post baseline value will be provided. Heart rate will be categorized into 'Decrease to <60', 'Change to Normal or No Change' and 'Increase to >100'. The determination of the worst-case post baseline considers both scheduled and unscheduled assessments. If a participant has a decrease to low and an increase to high, then the participant is counted in both the "Decrease to <60" categories and the "Increase to >100" categories. Similarly, temperature will be categorized based on normal ranges.

In addition, summaries of grade increase in systolic blood pressure (SBP) and diastolic blood pressure (DBP) will be provided separately. These summaries will display the number and percentage of participants with any grade increase, increase to Grade 2 and increase to Grade 3 for worst-case post-baseline only. The grade definition for SBP (mmHg) is: Grade 0 (<120), Grade 1 (120-139), Grade 2 (140-159), Grade 3 (≥160). The grade definition for DBP is: Grade 0 (<80), Grade 1 (80-89), Grade 2 (90-99), Grade 3 (≥100). The summaries will be produced for worst-case post baseline only.

4.5.3.3. ECG

A 12-lead electrocardiogram (ECG) is obtained at screening as specified in the Schedule of Activities. The heart rate, PR, QRS, QT, and corrected QT (QTc) intervals according to Fridericia's formula (QTcF) will be obtained. No further ECGs are required but may be obtained as part of medical care.

A listing of QTc values of potential clinical importance will be provided using the collected values based on Fridericia's formula. All ECG findings will also be listed.

4.5.3.4. Pregnancies

While pregnancy itself is not considered to be an AE or SAE, any pregnancy complication or elective termination of a pregnancy for medical reasons will be recorded as an AE or SAE as described in the protocol. If participants or participants' partner

become pregnant while on the study, the information will be included in the narratives and no separate table or listing will be produced.

4.5.3.5. Performance Status

ECOG performance status will be summarized at baseline and each post-baseline scheduled visit. Summaries will use frequency and percentage of participants at each planned assessment time. A summary of change from baseline by scheduled visits will be performed, as well as the worst-case post-baseline and the best-case post-baseline changes during the study (improved, no change, deteriorated).

4.6. Other Analyses

4.6.1. Subgroup Analyses

The list of subgroups may be used in descriptive summaries and statistical analyses. Additional subgroups of clinical interest may also be considered.

- If the percentage of participants is small within a particular subgroup, then the subgroup categories may be refined prior to unblinding the trial.
- If the category cannot be refined further, then descriptive rather than statistical comparisons may be performed for the particular subgroup.

Due to the expected low number of events per strata, subgroup analyses will not be stratified and analysis models will not include stratification factors as covariates. Otherwise, subgroup analyses will be performed similarly to the primary analysis method including only the participants within the relevant subgroup category. P-values will not be presented. All subgroup analyses will be based on the clinical database using eCRF or vendor data (and not IVRS strata).

The following subgroup analyses (see Table 11) will be performed to compare the primary estimand of PFS between treatments, based on IRC-assessed response, as well as the primary estimand of OS between treatments, if data permit.

Table 11 Subgroup Analyses

Subgroup	Categories
Age	<65 years, 65-<75 years, ≥75 years
Gender	Female, Male
Ethnicity	Hispanic, non-Hispanic
Race	White, Black or African American, Other
Region	North America, Europe, North East Asia [Japan, China and Republic of Korea], Rest of World (ROW)
Number of prior lines of therapy	1 vs. >1
Time to relapse after initiation of 1L treatment[1]	Relapse ≤12 months vs >12 months

Subgroup	Categories
High risk cytogenetics[2]	High Risk vs. Standard Risk
ISS stage at Screening	I vs II/III
Prior anti-CD38 treatment	Yes vs No
Prior bortezomib use	Yes vs No
Baseline ECOG	0 vs 1 or 2
Prior stem cell transplant	Yes vs No
Refractory to lenalidomide	Refractory vs Non-refractory
Refractory to anti-CD38 treatment	Refractory vs Non-refractory
EMD at baseline	Yes vs No
Triple-exposed (PI, Immunomodulator, anti-CD38)	Yes vs No
Prior exposure to lenalidomide and anti-CD38 mAb	Yes vs No

EMD=Extramedullary disease; PI=Proteasome Inhibitor.
[1] Relapse is defined as time from start date of 1L treatment to date of randomization for participants with 1 prior line and from start date of 1L treatment to start date of 2L treatment for participants with >1 prior line.
[2] High risk is defined as at least one high-risk abnormality—del(17p), t(4;14), or t(14;16). Standard risk is defined as negative results for all three high-risk abnormalities—del(17p), t(4;14), or t(14;16). All other cases will be considered as missing or not evaluable.

All subgroup analyses will be performed at the time of primary PFS analysis, unless PFS demonstrates statistical significance at the IA2. For OS, subgroup analyses may be performed on more mature data at a later planned analysis timepoint, as clinically relevant. Subgroup analyses of other key secondary endpoints may be performed using the above subgroups, as clinically relevant.

4.6.2. Other Variables and/or Parameters

Randomization factors changed during protocol amendment 1 as defined in Table 1.

4.7. Interim Analyses

Several interim analyses are planned for the study, details are provided below.

Table 12 Summary of Planned Interim Analyses

Analyses	Main Purpose	Timing
Interim Analysis for harm (IA1)	Harm (inferior efficacy) and potential sample-size re-estimation	~35 PFS events (~25% PFS information fraction)
Interim Analysis 2 (IA2)	Early Efficacy PFS	~145 PFS events (~84% PFS information fraction)
Primary PFS analysis/Interim Analysis 3 (IA3)	Primary PFS analysis or Early Efficacy OS (if PFS significant at IA2)	~173 PFS events (100% PFS information fraction) if PFS does not demonstrate statistical significance at IA2 OR

Analyses	Main Purpose	Timing
		If PFS demonstrates stat. significance at IA2 then the trigger for IA3 is ~130 OS events (~60% OS information fraction)
Interim Analysis 4 (IA4)	Early Efficacy OS	~163 OS events (~75% OS information fraction)

Abbreviations: IA=interim analysis; OS=Overall Survival; PFS=progression-free survival.

The interim analysis for harm (IA1) is planned at the time of approximately 35 of the 139 targeted PFS events (25% information) have been observed (expected around 8 months from the date of first participant randomized in the study under H_1). The interim analysis will allow for stopping early for harm (inferior efficacy). It is expected that 212 participants have been enrolled at the time of IA. A non-binding gamma spending function with parameter -3 will be used to calculate the boundary based on actual observed number of PFS events.

Table 13 shows the stopping boundary for the interim analysis for harm (IA1) assuming 35 PFS events are observed. The stopping boundaries will be revised based on the observed PFS events included in the IA data. Further details of the interim analysis, if necessary, will be provided in the IDMC Charter.

Table 14 provides a summary of boundary crossing probabilities for harm under a range of underlying true hazard ratios.

Table 15 presents the PFS efficacy stopping boundaries while Table 16 presents the OS efficacy stopping boundaries. All stopping boundaries will be adjusted at the time of each analysis based on the actual number of events observed. Further details of the interim analysis, if necessary, will be provided in the IDMC Charter.

In addition, safety data will be reviewed periodically starting from when ~60 participants have been followed for 8 weeks, and then every 6 months or as requested by the IDMC thereafter. Descriptive summaries of efficacy (response rates, number of PFS/OS events) may be included in the safety reviews to support the benefit:risk assessment. Further details are provided in the IDMC charter. An IDMC consisting of at least 2 physicians and one statistician as defined in the IDMC Charter will review data.

GSK CPMS analysts or delegate(s) not involved in the study conduct will have access to a blinded population PK dataset (including, but not limited to, concentration, actual dosing information, demographics, and some vital sign and laboratory information, but excluding adverse event and efficacy information) at selected time points throughout the trial for population PK model development/refinement. Additionally, designated representatives not involved with study conduct may be unblinded for performing population PK and PKPD dataset preparation in support of planned analyses and PK display review. All other personnel will remain blinded to aggregate data by treatment group until database lock.

Table 13 Stopping Boundaries for Interim PFS Analyses for Harm (based on 139 targeted PFS events)

Information Fraction	N of Events	Cum. α Spent	Cum. β Spent	Boundary (p-value)	Boundary (HR)	Boundaries Crossing Probabilities for (Harm) (Incremental)	
						Under H0	Under H1
0.25	35	0	0.009	0.805	1.337	0.195	0.009
1	139	0.025	0.148	0.025	0.717	0.78	0.141

Abbreviations: HR=hazard ratio; N/A=not applicable. Per protocol, null hypothesis H0: θ≥1 vs alternative hypothesis H1: θ<1 where θ is the PFS HR (B-Pd vs PVd). To calculate the boundary crossing probabilities, it has been assumed that under H0: HR=1 and under H1: HR=0.6.

Table 14 Boundary Crossing Probabilities for Harm at the Interim Analysis Under a Range of Underlying True Hazard Ratios (based on 139 targeted PFS events)

Information Fraction	N of Events	Boundary (p-value)	Boundary (HR)	Underlying True HR	Boundaries Crossing Probabilities (Harm)
0.25	35	0.805	1.337	0.6	0.9%
				0.7	2.8%
				0.8	6.3%
				0.9	11.7%
				1	19.1%
				1.1	27.4%
				1.2	36.4%
				1.3	46.0%
				1.4	54.6%

Abbreviations: HR=hazard ratio; N/A=not applicable.

Table 15 Stopping Boundaries for Interim Analyses for PFS Efficacy (based on 173 targeted PFS events)

Information Fraction	N of Events	Cum. α Spent	Boundary (p-value)	Boundary (HR)	Boundaries Crossing Probabilities (Incremental)	
					Under H0	Under H1
0.838	145	0.014	0.014	0.695	0.014	0.812
1	173	0.025	0.021	0.734	0.01	0.1

Abbreviations: HR=hazard ratio; N/A=not applicable. Per protocol, null hypothesis H0: θ≥1 vs alternative hypothesis H1: θ<1 where θ is the PFS HR (B-Pd vs PVd). To calculate the boundary crossing probabilities, it has been assumed that under H0: HR=1 and under H1: HR=0.6.

Table 16 Stopping Boundaries for Interim Analyses for OS Efficacy based on 2.5% alpha allocation

Information Fraction	N of Events	Cum. α Spent	Boundary (p-value)	Boundary (HR)	Boundaries Crossing Probabilities (Incremental)	
					Under H0	Under H1
0.502	109	0.002	0.002	0.568	0.002	0.194
0.599	130	0.004	0.003	0.621	0.002	0.149
0.751	163	0.01	0.008	0.688	0.006	0.235
1	217	0.025	0.022	0.76	0.015	0.254

Abbreviations: HR=hazard ratio; N/A=not applicable. Per protocol, null hypothesis H0: $\theta \geq 1$ vs alternative hypothesis H1: $\theta < 1$ where θ is the OS HR (B-Pd vs PVd). To calculate the boundary crossing probabilities, it has been assumed that under H0: HR=1 and under H1: HR=0.67.

4.7.1. Sequence of Interim and Other Planned Analyses

All planned analyses are listed in Table 17 below.

If requested by the applicable regulatory authorities, additional analyses may be performed for country-specific supplemental analyses. The analysis requirements and expected timing will be detailed in the country-specific SAP, if applicable.

Additional analyses of OS may be performed upon requests or to provide updated data to the health authorities. The details of these analyses including the associated alpha adjustment, if any, will be described in an updated SAP.

Table 17 Details of Planned Analyses

Analyses	Main Purpose	Timing	Endpoints included	Alpha adjustment for Primary and Key Secondary Endpoints [1] [2]		
Safety review by IDMC	Safety review	Periodically starting from when ~60 participants have been followed for 8 weeks, and then every 6 months or as requested by the IDMC thereafter	Key safety (AEs, SAEs, AESIs, deaths, ocular, exposure, dose modifications, laboratory parameters), descriptive efficacy summaries (e.g., response rates, counts of PFS/OS events) and study population summaries.	Only for safety review - no alpha adjustment		
Interim Analysis for harm (IA1)	Harm PFS (inferior efficacy) and potential sample size re-estimation	~35 PFS events (25% PFS information fraction)	Key safety, study population and PFS. Additional analyses may be performed to support decision making if requested by IDMC.	Only for harm - no alpha adjustment		
Interim Analysis 2 (IA2)	Early Efficacy PFS	~145 PFS events (~84% PFS information fraction)	Minimally, key safety, study population and PFS. Additional analyses may be performed to support decision making if requested by IDMC. All endpoints may be included if PFS is statistically significant.	PFS a=0.014	If PFS alpha ≤0.014 (PFS stat. significant), test OS at overall alpha 0.025 (across all looks) so OS alpha=0.002 in this look If OS significant test MRD at alpha=0.025 DoR and supportive secondary endpoints will be analyzed but not tested	If PFS alpha >0.014 (not stat. significant): No further statistical testing at this analysis; Supportive secondary endpoints will not be analyzed.
Primary PFS analysis/ Interim	Primary PFS analysis or Early Efficacy OS (if PFS significant at IA2)	~173 PFS events (100% PFS information fraction).	All endpoints. A reduced set of outputs may be produced if PFS is significant at IA2. An endpoint will not be re-tested once statistically	If PFS at IA2 is significant, no further PFS testing	a) Test OS alpha=0.003 If OS significant test MRD (based on data available at IA2) at alpha=0.025	

Analyses	Main Purpose	Timing	Endpoints included	Alpha adjustment for Primary and Key Secondary Endpoints [1] [2]	
Analysis 3 (IA3)		If PFS demonstrates stat. significance at IA2 then the trigger for IA3 is ~130 OS events (~60% OS information fraction)	significant. For these endpoints, updates (without formal hypothesis testing) will be provided.		PFS, DoR, MRD and supportive secondary endpoints may be descriptively analyzed using data available at time of analysis but not tested
				If PFS at IA2 not significant then test PFS a=0.021	If PFS alpha≤ 0.021 (PFS stat. significant): 1) Test OS at overall alpha 0.025 (across all looks) so OS alpha=0.003 in this look If OS significant test MRD (based on data available at IA2) at alpha=0.025 Supportive secondary endpoints will not be analyzed. DoR, MRD and supportive secondary endpoints may be descriptively analyzed using data available at time of analysis but not tested If PFS alpha> 0.021 (not stat. significant): No further statistical testing at this analysis or subsequent analyses;
Interim Analysis 4 (IA4)	Early Efficacy OS	~163 OS events (~75% OS information fraction)	Minimally, updated key safety, study population summaries and OS.	If PFS significant (at any previous analysis) test OS alpha=0.008 If OS significant test MRD (based on data available at IA2) at alpha=0.025	
Final Analysis	Final OS analysis	~217 OS events (100% OS information fraction)	Minimally, updated key safety, study population summaries and OS.	If PFS significant (at any previous analysis)test OS alpha=0.022 If OS significant test MRD (using data available at IA2) at alpha=0.025	

Abbreviations: DoR=Duration of Response; IA=Interim Analysis; MRD=MRD Negativity Rate; PFS=Progression-Free Survival; OS=Overall Survival.
[1] Upon successful rejection of the hypothesis and regardless of the timing of rejection, the full alpha allocated to testing the hypothesis can be propagated.
[2] Alpha will be adjusted to account for multiple testing of an endpoint across timepoints using the Lan DeMets approach that approximates the O'Brien and Fleming spending function [Lan, 1983]. The efficacy boundaries will be adjusted based on the observed number of events at the time of analysis.

All data available at the time of data cut will be used and all analyses will be performed once the analysis specific criteria have been met and following the steps indicated below:

- All required database cleaning activities have been completed and database lock has been declared by Data Management.
- All criteria for unblinding the randomization codes have been met.
- Randomization codes have been distributed.

4.8. Changes to Protocol Defined Analyses

Clarification added: Sustained MRD negativity rate definition clarified handling of indeterminate results. Sustained MRD negativity rated is defined as the percentage of participants who achieve MRD negative status assessed by NGS at 10^{-5} threshold at least twice, a minimum of 12 months apart and with no MRD positive (or indeterminate) result in between, during the time of confirmed CR or better response per IRC-assessment according to IMWG.

Terminology of supplementary vs sensitivity has been revised for some planned analyses, but the technical details have not changed.

Revised multiplicity strategy. Equivalent to assigning 0 weight to DoR in the weighted Bonferroni procedure defined within the protocol, a hierarchical procedure will be used such that OS is tested at 2.5% level conditional on successful rejection of the null hypothesis associated with PFS only, and MRD Negativity is tested at 2.5% level conditional on successful rejection of the null hypothesis associated with OS. DoR will remain key secondary, but will not be formally tested (descriptive only).

Clarified that an adequate assessment is defined as an assessment where the *confirmed* response is sCR, CR, PR, VGPR, MR or SD.

5. SAMPLE SIZE DETERMINATION

Primary Endpoint PFS

Based on data from the OPTIMISMM study [Richardson, 2019], the median PFS in the PVd arm is expected to be around 12 months. It is expected that treatment with B-Pd will lead to a 40% reduction in the risk of progression or death, i.e., an expected PFS HR of 0.6, which corresponds to an increase in median PFS from 12 months to 20 months under the exponential assumption.

To ensure >90% power to test the null hypothesis: PFS HR = 1, versus the specific alternative hypothesis: PFS HR = 0.6, a total of approximately 173 PFS events are needed. The calculation assumes a comparison of PFS by log-rank test at overall 1-sided alpha level of 2.5% with 1:1 randomization ratio, and two interim analyses: an interim analysis for harm using gamma spending function with parameter of -3 when observing ~25% PFS events and an early efficacy analysis using Lan De Mets O'Brien Fleming alpha spending function [Lan, 1983]. The calculation further assumes approximately 302 participants to be randomized in a 1:1 ratio to receive B-Pd or PVd, with a uniform enrollment rate of 11.2 participants per month and enrollment period of approximately 27 months. It is estimated that the targeted 173 PFS events will be observed approximately 35 months from the time when the first participant is randomized under H_1, assuming an annual dropout rate of 5%. These calculations were conducted using the software package EAST v6.5.

If the number of participants required by local regulatory agencies are not recruited within the planned recruitment target, enrollment may continue in separate cohorts until the country enrollment requirements, as required by local regulatory bodies, have been reached. Additional participants that are enrolled in separate cohorts will not be included in the analysis portion of the study planned for the marketing application. However, these additional participants will be included in country-specific supplemental analyses, requested by the applicable regulatory authorities concerned, as detailed in the country-specific SAP.

Key Secondary Endpoint Overall Survival

Overall Survival (OS), as one of the key secondary endpoints, will be formally statistically tested, provided that the primary endpoint PFS is statistically significant. Using available data from literature, the median OS in the PVd arm is expected to be around 44 months (OS HR = 0.67) [Siegel, 2018; Dimopoulos, 2020; San-Miguel, 2016; Richardson, 2021; Sonneveld, 2022]. Additionally, this calculation assumes the same number of participants are enrolled in this study to provide sufficient power for the primary PFS endpoint (i.e., approximately 302 participants), analysis by a one-sided log-rank test at overall 1-sided alpha level of 2.5% with 1:1 randomization ratio, no dropouts and accrual as defined for PFS sample size assumptions, above. The calculation further allows for 3 interim analyses at 50%, 60%, and 75% information fraction using Lan DeMets O'Brien Fleming alpha spending boundaries [Lan,1983]. This corresponds approximately to 83% power at the end of study when approximately 217 OS events will be observed. The information fraction may shift dependent on the actual timing of

analyses and the observed OS events at that time and the boundaries will be adjusted accordingly.

This calculation was conducted using EAST v6.5.

Key Secondary Endpoint Minimal Residual Disease Negativity Rate

MRD Negativity Rate, as one of the key secondary endpoints, will be formally statistically tested, provided that the primary endpoint PFS and OS is statistically significant and will be based on the data available at IA2. Based on available data from literature, the proportion of participants with MRD Negativity in the PVd arm is expected to be around 12% [Spencer, 2018]. It is hypothesized that treatment with belantamab mafodotin will result in a 15% absolute increase in MRD negativity to 27%. Based on the same number of participants that are planned to be enrolled in this study to provide sufficient power for the primary endpoint (i.e., approximately 302 participants), the power to detect a difference in the MRD negativity between the two treatment arms is approximately 88%. This calculation assumes analysis by a 1-sided Fisher's exact test at the overall 2.5% level of significance, participants randomized to the 2 treatment arms in a 1:1 allocation ratio. Assuming MRD negativity is tested at the 2% level of significance, the study will provide approximately 86% power to detect a difference in MRD negativity between the two treatment arms. These calculations were made using EAST v6.5.

Key Secondary Endpoint Duration of Response

Duration of Response (DoR), will not be formally statistically tested.

See Section 4.3.1.2 for further details on the comparison of restricted mean DoR (RMDOR).

5.1. Sample Size Re-estimation

Note: the interim analysis for harm (IA1) was performed prior to the change in targeted number of events from 139 PFS events to 173 PFS events and prior to changing the PFS events from derived to IRC assessed events. Details of the intended sample size re-estimation at the time of IA1 are provided below.

A sample size re-estimation (SSR) may be considered at the time of the interim analysis for harm (IA1) to ensure adequate power to demonstrate the treatment effect. Details will also be described in the IDMC charter. Adequate firewalls will be maintained to ensure the integrity of the study.

The current number of events for the primary PFS analysis is based on 85% power. The key idea is to evaluate conditional power (CP) at the interim look at 25% information fraction. If the CP is either too low or too high, we do not alter the number of events for the final analysis. However, if conditional power falls in a range that we deem promising, then the number of events may be increased, participant to a pre-determined upper limit [Mehta, 2011]. The pre-defined upper limit for this study is set to be the number of events

for the Primary PFS analysis based on 90% power, an increase from approximately 139 events to approximately 163 events.

We propose the following guidelines for sample size re-estimation, depending on the zone into which CP falls at the interim look.

SSR Decision Rule based on CP:

Maintain zone: SSR won't be done if interim result is too disappointing to avoid a large study without clinical significance of the result, e.g., CP in [0, CP_{min})

Maintain zone: SSR won't be done if interim result is good enough to ensure a positive study without SSR, e.g., CP in (CP_{max}, 1]

Promising zone: SSR may be done if interim result is promising but does not have sufficient power for final success, e.g., CP in [CP_{min}, CP_{max}]

The number of events for current study based on 85% may be increased to number of events based on 90% power.

Table 18 shows the criteria and decision rules for sample size re-estimation.

Table 18 Response Criteria and Decision Rules for Sample Size Re-Estimation

Criteria	Outcome	Decision
If CP ≤ 0.40 or CP ≥ 0.80	Maintain	No increase in number of events
If 0.40 < CP < 0.80	Promising Zone	Increase number of events to 163 events [i.e., number of events based on 90% power

The Statistical Data Analysis Center (SDAC) will provide the IDMC with the following operating characteristics for the SSR evaluation:

- Estimates of CP, HR, standard error, number of events, information fraction, number of participants.

6. SUPPORTING DOCUMENTATION

6.1. Appendix 1 Study Population Analyses

Unless otherwise specified, the study population analyses will be based on the ITT Analysis Set. Study population analyses including analyses of participant's disposition, protocol deviations, demographic and baseline characteristics, prior and concomitant medications, disease characteristics, prior and subsequent anti-myeloma therapy, surgical/medical procedures, duration of follow-up and exposure will be based on GSK Core Data Standards.

6.1.1. Participant Disposition

A summary of the number of participants in each of the analysis set described will be provided. A listing of participants excluded from any population will also be provided and a separate listing of planned and actual treatment for each participant will also be produced. In addition, the number of participants enrolled by country and site will be summarized by treatment arm using the "Enrolled" population. A summary of participant status and reason for study withdrawal will be provided. This display will show the number and percentage of participants who withdrew from the study, including primary reasons for study withdrawal. Reasons for study withdrawal will be presented in the order they are displayed in the eCRF. A listing will also be provided. A summary and listing of screening status and reasons for screen failure will also be produced for the All Screened Analysis Set.

Summaries of study treatment status will be provided. This display will show the number and percentage of participants who are ongoing or discontinued study treatment and a summary of the primary reasons for discontinuation of study treatment. Reasons for study treatment discontinuation will be presented in the order they are displayed in the eCRF. A listing of study treatment discontinuation will be generated. The listing will include last dose date, and reasons for study treatment discontinuation.

It is anticipated that patient accrual will be spread thinly across sites, so data from all participating sites will be integrated and site-effect will not be considered in the statistical analyses. Summaries of data by site are unlikely to be informative.

Summaries of study status and treatment status by relationship to the COVID-19 pandemic will be included. A summary of visits impacted by the COVID-19 pandemic will be produced. Plots of enrolment over time by prior line of therapy may also be produced.

6.1.2. Demographic and Baseline Characteristics

The demographic characteristics (e.g., age, race, ethnicity, region, sex, baseline height, baseline body weight and baseline BMI) will be summarized. Age, height, weight and BMI will be summarized using the mean, standard deviation, minimum, median, and maximum. The count and percentage will be computed for age groups of interest (e.g.

≤18, 19-64, ≥65, 65-<75 and ≥75 years), race, region, sex and ethnicity. Age categories will be reported to meet differing regulatory and study-specific requirements. Key demographic characteristics will be listed.

Race and racial combinations may be summarized and listed.

Disease history and characteristics (e.g., time since initial diagnosis in years, stage at initial diagnosis, date of initial diagnosis) at initial diagnosis and screening will be summarized.

Disease characteristics at screening including but will not be limited to: International Staging System (ISS) at screening, extramedullary disease, extramedullary disease location, lytic bone lesions, myeloma immunoglobulin, myeloma light chain, type of multiple myeloma, lines of therapy completed prior to screening (categories and summary statistics), prior stem cell transplant, time to relapse after initiation of 1L treatment (categories and summary statistics), time to relapse after initiation of latest treatment (categories and summary statistics), high-risk cytogenetics (individual cytogenetics and cytogenetic profile categories, including double hit multiple myeloma), other cytogenetic abnormalities, baseline renal impairment status per eGFR (categories and summary statistics), best response to most recent prior anti-MM therapy will be summarized..

Medical conditions collected at screening will be summarized according to past and current and by cancer-related and non-cancer related categories.

Substance use, including smoking history and alcohol use will be summarized.

If there are stratification errors, then a summary of stratification factors will be produced comparing IVRS data vs eCRF data. Prior lines of therapy (1 vs 2/3 vs ≥4), prior bortezomib use (yes vs no), ISS status (I vs II/III) and prior anti-CD38 treatment (yes vs no) will be included. Participants will be categorized as follows:

- Consistent stratification factors (prior lines of therapy, prior bortezomib use and ISS status for those randomized prior to PA1, and prior lines of therapy, prior bortezomib use, prior anti-CD38 treatment for those randomized after PA1)
- Consistent prior lines of therapy and prior bortezomib use
- Discrepant stratification factor (excluding missing IVRS data):
    - Prior lines of therapy
    - Prior bortezomib use
    - ISS status
    - Prior anti-CD38 treatment
- Number of discrepant stratification factors:
    - 0
    - 1

- 2
- 3

A supportive listing of randomized and actual stratification will also be produced.

6.1.3. Protocol Deviations

Protocol deviations will be tracked by the study team throughout the conduct of the study. These protocol deviations will be reviewed to identify those considered as important as follows:

- Data will be reviewed prior to unblinding and locking the database to ensure all deviations leading to analysis population exclusions are captured and categorised in the protocol deviations ADaM dataset (note these exclusions are not captured in the SDTM dataset).
- This dataset will be the basis for the summaries of important protocol deviations.

Important protocol deviations will be listed and summarized as well as deviations that result in exclusion from analysis sets, if any.

In addition to the overall summary of important protocol deviations, separate summaries will be produced for important protocol deviations related to COVID-19, and important protocol deviations not related to COVID-19 respectively.

6.1.4. Prior and Concomitant Medications

Concomitant medications will be coded using both the GSK Drug and WHO Drug dictionaries. However, the summary will be based on GSK Drug dictionary only. The summary of concomitant medications will show the number and percentage of participants taking concomitant medications by Ingredient. Anatomical Therapeutic Chemical (ATC) classification Level 1 (Body System) information will be included in the dataset created but will not appear on the listing or summary.

Concomitant medications will be summarized by base ingredient. Each participant is counted once within each ingredient. For example, if a participant takes Amoxycillin on two separate occasions, the participant is counted only once under the ingredient “Amoxycillin”.

Note: In order to be considered a concomitant medication, the concomitant medication must have been taken at some point during the on-treatment study phase.

Prophylactic medication for infusion-related reactions will be summarized by drug class and drug name.

Blood products or blood supportive care products with onset date within the on-treatment window will be included in the summary tables. The frequency and percentage of participants using blood products and blood supportive care products after the start of study medication will be provided.

6.1.5. Prior and Subsequent Anti-Myeloma Therapies

Prior and follow-up/ subsequent anti-multiple myeloma (anti-MM) therapy will be coded using GSK Drug coding dictionary, then summarized by ingredient. A summary of multiple myeloma participants' refractory to prior anti-myeloma therapy by drug class will be provided.

Prior and follow-up/ subsequent anti-myeloma therapy will also be summarized by type of therapy and drug class. "Drug class" is identified by clinical in an external file. Sub-classes of interest will also be summarized.

A summary of the best response to the most recent prior anti-MM therapy will be provided. A summary of the number of prior anti-MM therapy regimens will also be produced.

6.1.6. Study Intervention Compliance

Summaries of study treatment exposure and dose modifications (e.g., number of dose reductions, number of dose delays) will further characterize compliance. These analyses are defined in Section 4.5.1.

6.1.7. Additional Analyses Due to the COVID-19 Pandemic

A participant is defined as having a suspected, probable, or confirmed COVID-19 infection during the study if the answer is "Confirmed", "Probable" or "Suspected" to the case diagnosis question from the COVID-19 coronavirus infection assessment eCRF.

Summaries of the numbers of participants with a suspected, probable, or confirmed COVID-19 infection, and of COVID-19 test results will be based on GSK Core Data Standards and will be summarized using the Safety Analysis Set. A Standardised MedDRA Query (SMQ) will be used to identify all COVID-19 AEs.

The incidence of AEs and SAEs (Fatal and Non-Fatal) of COVID-19, COVID-19 AEs leading to study drug discontinuation, and COVID-19 AEs leading to study withdrawal, will be obtained from standard AE and SAE summaries.

6.2. Appendix 2 Data Derivations Rule

6.2.1. Criteria for Potential Clinical Importance

See OPS.

6.2.2. Study Period

See OPS.

6.2.3. Study Day and Reference Dates

See OPS.

6.2.4. Definitions of Assessment Windows for Analyses

For data summaries by visit, scheduled visits with nominal visit description as well as the worst-case post baseline will be displayed. Unscheduled visits will not be displayed or slotted into a visit window but will be included in the derivation of worst-case post baseline assessment, with the exception of PRO analyses where unscheduled visits will be slotted (See OPS document). All un-scheduled visits will be displayed in the listing.

6.2.5. Multiple Measurements at One Analysis Time Point

See OPS.

6.2.6. Handling of Partial Dates

See OPS.

6.2.7. Patient Reported Outcome Analyses

6.2.8. EORTC QLQ-C30

The EORTC QLQ-C30 is a 30-item questionnaire containing both single- and multi-item measures [Aaronson, 1993]. These include five functional scales (Physical, Role, Cognitive, Emotional, and Social Functioning), three symptom scales (Fatigue, Pain, and Nausea/Vomiting), a Global Health Status/QoL scale, and six single items (Constipation, Diarrhea, Insomnia, Dyspnea, Appetite Loss, and Financial Difficulties).

Scores for each scale and single-item measure are averaged and transformed linearly to a score ranging from 0–100. This is performed similarly to the steps for EORTC QLQ-MY20/EORTC IL52 (see below). A high score for functional scales and for Global Health Status/QoL represent better functioning ability or HRQoL, whereas a high score for symptom scales and single items represents significant symptomatology [Proskorovsky, 2014].

Handling of missing items:

Single-item measures: if the item is missing, the score S will be set to missing.

Scales requiring multiple items: if at least half of the items from the scale are available, the score S will be calculated based on available items. If more than half of the items from the scale are missing, the score S will be set to missing [Fayers, 2021].

Minimal Important Difference (MID):

In a sample of patients who received chemotherapy for either breast cancer or small-cell lung cancer (n=246, n=80 respectively), the mean change in EORTC QLQ-C30 score between baseline and follow-up was about 5 to 10 points on a 0-100 scale for patients who indicated "a little" change on the Subjective Significance Questionnaire (SSQ), either for better or for worse [Osoba, 1998].

6.2.8.1. EORTC QLQ-MY20/EORTC IL52

For EORTC QLQ-MY20/EORTC IL52, the raw scores from the following multi-item scales and single-item measures raw scores will be calculated by averaging the items that contribute to the scale or single item: disease symptom scales (includes bone aches or pain, back pain, hip pain, arm or shoulder pain, chest pain, and pain increasing with activity), side effects of treatments (including drowsiness, thirst, feeling ill, dry mouth, hair loss, upset by hair loss, tingling hands or feet, restlessness/agitation, acid indigestion/heartburn, and burning or sore eyes), future perspective (includes worry about death and health in the future, and thinking about illness), and body image scale is a single-item scale that addresses physical attractiveness.

1) Raw Score

For each multi-item scale, calculate the average of the corresponding items.

$$\text{Raw Score = RS} = \left\{\frac{(I_1+I_2+\cdots+I_n)}{n}\right\}$$

For the single-item measure, the score of the concerning item corresponds to the raw score.

2) Linear Transformation

To obtain the Score S, standardize the raw score to a 0 – 100 range following the appropriate transformation:

$$\text{Functional scales:} \quad \text{S} = \left\{1 - \frac{(RS-1)}{range}\right\} x\ 100$$

$$\text{Symptom scales:} \quad \text{S} = \left\{\frac{(RS-1)}{range}\right\} x\ 100$$

A high score for Disease Symptoms and Side Effects of Treatment represents a high level of symptomatology or problems, whereas a high score for Future Perspective and Body Image represents better outcomes [Proskorovsky, 2014].

Missing items can be handled similarly to EORTC QLQ-C30.

6.2.9. Extended Loss to Follow-up or Extended Time Without an Adequate Assessment

For participants, if two or more scheduled disease assessments are missed and are then followed by an assessment of PD or death, PFS will be censored at the last adequate assessment prior to PD or death. When the scheduled disease assessment is every 4 weeks, a window of 63 days (8 weeks + 7-day window) will be used to determine whether there was an extended time without adequate assessment. That is, if the time difference between PD/death and last adequate assessment is more than 63 days, then PFS will be censored at the last adequate assessment prior to PD/death. In case there is no adequate assessment between PD/death and randomization date, and the time difference between PD/death and randomization date is more than 63 days, then PFS will be censored at the randomization date.

6.2.9.1. Derivation of Sponsor-Assessed KVA Grade

The following hybrid approach (medical/safety review, assisted by programming algorithm) will be used to assign sponsor-assessed KVA grade at each ocular exam visit:

1. At eye level (i.e., separate for each eye), derive corneal exam grade and visual acuity grade and then combine for KVA grade by programming algorithm:
    a. At eye level, identify 'Not gradable by programming algorithm' visits for each eye:
        i. Corneal exam grade is 'Not gradable by programming algorithm' for an eye at following visits:
            1. All visits if at baseline examination corneal epithelial exam is reported as "Abnormal" OR not reported
            2. Any visit after a cataract surgery is reported
        ii. Visual acuity grade is 'Not gradable by programming algorithm' for an eye at the following visits:
            1. All visits, if at baseline examination best corrected visual acuity is 20/200 or worse OR not reported
            2. Any visit after a cataract surgery is reported
        iii. KVA grade is 'Not gradable by programming algorithm' for an eye at any visit where corneal exam grade or visual acuity grade is 'Not gradable by programming algorithm' for the respective eye.

b. At eye level, derive Corneal exam grade, Visual acuity grade and overall KVA grade for visits that are NOT "Not gradable by programming algorithm" based on the algorithm below:

   i. Corneal exam grade is assigned per Table 19 below.

      1. When there are multiple findings on corneal examination at a visit, the corneal exam findings grade for the eye will be determined by the worst-case

      2. If insufficient information is reported to preclude a higher grade, e.g. no information regarding presence or absence of corneal erosion or ulcer or stromal opacity is missing expected corresponding location, then corneal exam grade is 'Missing'

   ii. Visual acuity grade is assigned per Table 20 and Table 21, below.

   iii. At eye level, determine overall KVA at a visit:

      1. Assign the higher grade of corneal exam grade and visual acuity grade as overall grade. If grade from one component is missing, assign the overall KVA grade as missing.

      2. If corneal exam grade is 0 and visual acuity grade is grade 2+ then Overall KVA is "Not gradable by programming algorithm"

2. At eye level, Corneal exam grade, Visual acuity grade and Overall KVA grade for all "Not gradable by programming algorithm" visits will be manually reviewed and graded based on Medical/Safety review. If required, programmatically determined grade (including missing values) may be revised based on Medical/Safety review.

3. Once a grade is assigned for each eye (right/left) for each component (Overall KVA, CE and VA), the worst eye will be calculated – if a missing value is present, the non-missing value will be used.

Table 19 Corneal Exam Grade

KVA grade	Grade 1	Grade 2	Grade 3	Grade 4
Corneal examination finding(s) at visit*	Mild superficial punctate keratopathy and no superficial punctate keratopathy at baseline	Moderate superficial punctate keratopathy OR any of (patchy microcyst-like deposits, peripheral sub-epithelial haze, new peripheral stromal opacity).	Severe superficial punctate keratopathy OR any of (diffuse microcyst-like deposits, central sub-epithelial haze, new central stromal opacity).	Corneal erosion or ulcer

Table 20 Visual Acuity Grade

KVA grade	Grade 1	Grade 2	Grade 3	Grade 4
Change in BCVA from Baseline	Decline from baseline of 1 line on Snellen Visual Acuity	Decline from baseline of 2 or 3 lines (and Snellen Visual Acuity not worse than 20/200)	Decline from baseline by more than 3 lines (and Snellen Visual Acuity not worse than 20/200)	Snellen Visual Acuity worse than 20/200

Table 21 Change in BCVA lines

Baseline Best Corrected Visual Acuity (BCVA)	Grade 1 (*1 line decrease from Baseline BCVA*)	Grade 2 (*2-3 lines decrease from Baseline BCVA*)	Grade 3 (*>3 lines decrease from Baseline BCVA but not worse than 20/200*)	Grade 4 (*BCVA worse than 20/200*)
20/10	20/12.5	20/15 to 20/16 or 20/20	20/25 to 20/200	Worse than 20/200
20/12.5	20/15 to 20/16	20/20 to 20/25	20/30 to 20/200	Worse than 20/200
20/15 to 20/16	20/20	20/25 to 20/30 or 20/32	20/40 to 20/200	Worse than 20/200
20/20	20/25	20/30 to 20/40	20/50 to 20/200	Worse than 20/200
20/25	20/30 or 20/32	20/40 to 20/50	20/60 to 20/200	Worse than 20/200
20/30 to 20/32	20/40	20/50 to 20/60 or 20/63	20/70 to 20/200	Worse than 20/200
20/40	20/50	20/60 to 20/70 or 20/80	20/100 to 20/200	Worse than 20/200
20/50	20/60 or 20/63	20/70 to 20/100	20/125 to 20/200	Worse than 20/200
20/60 to 20/63	20/70 or 20/80	20/100 to 20/125	20/150 to 20/200	Worse than 20/200
20/70 to 20/80	20/100	20/125 to 20/150 or 20/160	20/200	Worse than 20/200
20/100	20/125	20/150 to 20/160 or 20/200	N/A	Worse than 20/200

Baseline Best Corrected Visual Acuity (BCVA)	**Grade 1** **(*1 line decrease from Baseline BCVA*)**	**Grade 2** **(*2-3 lines decrease from Baseline BCVA*)**	**Grade 3** **(*>3 lines decrease from Baseline BCVA but not worse than 20/200*)**	**Grade 4** **(*BCVA worse than 20/200*)**
20/125	20/150 or 20/160	20/200	N/A	Worse than 20/200
20/150 to 20/160	20/200	N/A	N/A	Worse than 20/200
Worse than 20/160	N/A	N/A	N/A	Any further reduction from baseline is considered Grade 4

6.2.10. Trademarks

Trademarks of the GlaxoSmithKline Group of Companies
NONE

Trademarks not owned by the GlaxoSmithKline Group of Companies
EAST
NONMEM
SAS
WinNonlin

6.2.11. List of Abbreviations

Abbreviation	Description
ADA	Anti-Drug Antibodies
ADaM	Analysis Data Model
AE	Adverse Event
AESI	Adverse Events of Special Interest
AIC	Akaike's Information Criteria
Anti-MM	Anti-Multiple Myeloma
A&R	Analysis and Reporting
BOR	Best Overall Response
Bor/Dex	Bortezomib/Dexamethasone
CDISC	Clinical Data Interchange Standards Consortium
CI	Confidence Interval
CMH	Cochran Mantel Haenszel
CP	Conditional Power
CPMS	Clinical Pharmacology Modelling & Simulation
CR	Complete Response
CRR	Complete Response Rate
CS	Clinical Statistics
CSR	Clinical Study Report
CTCAE	Common Terminology Criteria for Adverse Events
CTR	Clinical Trial Register
CV_b / CV_w	Coefficient of Variation (Between) / Coefficient of Variation (Within)
DBF	Database Freeze
DBR	Database Release
DOB	Date of Birth
DoR	Duration of Response
DP	Decimal Places
ECG	Electrocardiogram
eCRF	Electronic Case Record Form
EMA	European Medicines Agency
FDA	Food and Drug Administration
FDAAA	Food and Drug Administration Amendments Act of 2007
GSK	GlaxoSmithKline
HR	Hazard Ratio
IA	Interim Analysis
ICH	International Conference on Harmonization
IDMC	Independent Data Monitoring Committee
IDSL	Integrated Data Standards Library
IMMS	International Modules Management System
IMWG	International Myeloma Working Group
IP	Investigational Product
ISS	International Staging System
ITT	Intent-To-Treat
KVA	Keratopathy Visual Acuity

Abbreviation	Description
LVEF	Left ventricular ejection fraction
MedDRA	Medical Dictionary for Regulatory Activities
mITT	Modified Intent-To-Treat
MMRM	Mixed Model Repeated Measures
MRD	Minimal Residual Disease
NE	Not Evaluable
NGS	Next Generation Sequencing
OPS	Output and Programming Specification document
ORR	Overall Response Rate
OS	Overall Survival
PCI	Potential Clinical Importance
PD	Pharmacodynamic
PDMP	Protocol Deviation Management Plan
PFS	Progression-Free Survival
PK	Pharmacokinetic
PP	Per Protocol
PopPK	Population PK
QC	Quality Control
QRS	Q, R, and S waves in ECG
QT	Q and T waves in ECG
QTc	Corrected Q and T waves in ECG
QTcF	Frederica's QT Interval Corrected for Heart Rate
QTcB	Bazett's QT Interval Corrected for Heart Rate
RMDOR	Restricted Mean Duration of Response
RMST	Restricted Mean Survival Time
SAC	Statistical Analysis Complete
SAE	Serious adverse event
SAP	Statistical Analysis Plan
sCR	Stringent Complete Response
SDAC	Statistical Data Analysis Center
SDSP	Study Data Standardization Plan
SDTM	Study Data Tabulation Model
SOP	Standard Operation Procedure
SSR	Sample Size Re-estimation
TA	Therapeutic Area
TFL	Tables, Figures & Listings
TTR	Time to Response
TTP	Time to Progression

7. REFERENCES

Aaronson NK, Ahmedzai S, Bergman B, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst. 1993 Mar 3;85(5):365-376.

Amit O, Mannino F, Stone AM, Bushnell W, Denne J, Helterbrand J, Burger HU. Blinded independent central review of progression in cancer clinical trials: results from a meta-analysis. European Journal of Cancer. 2011;47(12):1772-8

Basch E, Reeve BB, Mitchell SA, Clauser SB, Minasian LM, Dueck AC, et al. Development of the National Cancer Institute patient-reported outcomes version of the common terminology criteria for adverse events (PRO-CTCAE). J Natl Cancer Inst. 2014; 106:1-11.

Bretz F, Maurer W, Brannath W, et al. A graphical approach to sequentially rejective multiple test procedures. Statistics in medicine. 2009;28(4):586-604.

Brookmeyer R, Crowley J. A confidence interval for the median survival time. Biometrics, 1982; 38:29-41.

Cocks K, Cohen D, Wisløff F, Sezer O, Lee S, Hippe E, et al. An international field study of the reliability and validity of a disease-specific questionnaire module (the QLQ-MY20) in assessing the quality of life of patients with multiple myeloma. European Journal of Cancer. 2007;43(11):1670-8.

Dimopoulos MA, Lonial S, White D, Moreau P, Weisel K, San-Miguel J, Shpilberg O, Grosicki S, Špička I, Walter-Croneck A, Magen H, Mateos MV, Belch A, Reece D, Beksac M, Spencer A, Oakervee H, Orlowski RZ, Taniwaki M, Röllig C, Einsele H, Matsumoto M, Wu KL, Anderson KC, Jou YM, Ganetsky A, Singhal AK, Richardson PG. Elotuzumab, lenalidomide, and dexamethasone in RRMM: final overall survival results from the phase 3 randomized ELOQUENT-2 study. Blood Cancer J; 2020; 10: 91

Dougherty BE, Nichols JJ, Nichols KK. Rasch analysis of the Ocular Surface Disease Index (OSDI). Invest Ophthalmol Vis Sci. 2011 ;52(12): 8630-8635.

Eliason L, Loubert A, Gorsh B, Meunier J, Kleinman D, Sapra S, Regnault A. Methodology for Evaluating Meaningful Change Thresholds in Patient-Reported Outcome (PRO) Measures for Corneal Events in Patients with Relapsed/Refractory Multiple Myeloma (RRMM) Receiving Single-Agent Belantamab Mafodotin (BELAMAF). International Society for Pharmacoeconomics and Outcomes Research. European Congress. Nov. 16-19, 2020.

Fayers PM, Aaronson NK, Bjordal K, Groenvold M, Curran D, Bottomley A on behalf of the EORTC Quality of Life Group. EORTC QLQ-C30 Scoring Manual (3rd edition). Brussels: EORTC, 2001.

Guy W (ed). ECDEU Assessment Manual for Psychopharmacology. Rockville, MD: US Department of Health, Education, and Welfare Public Health Service Alcohol, Drug Abuse, and Mental Health Administration, 1976

Haycock GB, Schwartz GJ, Wisotsky DH. Geometric method for measuring body surface area: a height-weight formula validated in infants, children, and adults. J Pediatr. 1978 Jul;93(1):62-6. doi: 10.1016/s0022-3476(78)80601-5. PMID: 650346.

Huang B, Tian L. Utilizing restricted mean duration of response for efficacy evaluation of cancer treatments. Pharm Stat. 2022 Sep;21(5):865-878. doi: 10.1002/pst.2198. Epub 2022 Feb 21. PMID: 35191170.

Kenward MG, Roger JH. An improved approximation to the precision of fixed effects from restricted maximum likelihood. Computational Statistics & Data Analysis. 2009 May 15;53(7):2583-95.

Klein J, Moeschberger M. Survival analysis: techniques for censored and truncated data. Springer Science and Business Media. 2005

Kumar S, Paiva B, Anderson KC, et al. International Myeloma Working Group consensus criteria for response and minimal residual disease assessment in multiple myeloma. Lancet Oncol. 2016;17(8):e328-346.

Lan KKG, DeMets DL. Discrete sequential boundaries for clinical trials. Biometrika. 1983;70:649-653.

Li Y, Xu Y, Liu L, et al. Population pharmacokinetics of pomalidomide. J Clin Pharmacol. 2015;55(5):563-572

Li X, Wulfsohn MS, Koch GG. Considerations on Testing Secondary Endpoints in Group Sequential Design. Statistics in Biopharmaceutical Research. 2017;9(4):333- 337.

Mehta CR, Pocock SJ. Adaptive increase in sample size when interim results are promising: a practical guide with examples. Stat Med. 2011 Dec 10;30(28):3267-84. doi: 10.1002/sim.4102. Epub 2010 Nov 30. PMID: 22105690.

National Cancer Institute (NCI) CTCAE Version 5. Available at: https://ctep.cancer.gov/protocolDevelopment/electronic_applications/docs/CTCAE_v5_Quick_Reference_5x7.pdf.

NCCN (National Comprehensive Cancer Network) Multiple myeloma guidelines. 2018. https://www.nccn.org/professionals/physician_gls/pdf/myeloma.pdf (Accessed 3 Sept 2018).

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

Proskorovsky I, Lewis P, Williams CD, et al. Mapping EORTC QLQ-C30 and QLQ-MY20 to EQ-5D in patients with multiple myeloma. Health Qual Life Outcomes. 2014 Mar 11; 12:35.

Rajkumar SV, Dimopoulos, M.A., Palumbo, A., Blade, J., Merlini, G., Mateos, M.V., et al. International Myeloma Working Group updated criteria for the diagnosis of multiple myeloma. Lancet. 2014; 15: e538-48.

Richardson PG, Oriol A, Beksac M, et al. Pomalidomide, bortezomib, and dexamethasone for patients with relapsed or refractory multiple myeloma previously treated with lenalidomide (OPTIMISMM): a randomised, open-label, phase 3 trial. Lancet Oncol. 2019 May 13. pii: S1470-2045(19)30152-4. doi: 10.1016/S1470-2045(19)30152-4. [Epub ahead of print]

Richardson PG, Kumar SK, Masszi T, Grzasko N, Bahlis NJ, Hansson M, Pour L, Sandhu I, Ganly P, Baker BW, Jackson SR, Stoppa AM, Gimsing P, Garderet L, Touzeau C, Buadi FK, Laubach JP, Cavo M, Darif M, Labotka R, Berg D, Moreau P; Final Overall Survival Analysis Of The Tourmaline-Mm1 Phase IIi Trial Of Ixazomib, Lenalidomide, And Dexamethasone In Patients With Relapsed Or Refractory Multiple Myeloma. J Clin Oncol; 2021;39:2430-2442

San-Miguel JF, Hungria VT, Yoon SS, Beksac M, Dimopoulos MA, Elghandour A, Jedrzejczak WW, Günther A, Nakorn TN, Siritanaratkul N, Schlossman RL, Hou J, Moreau P, Lonial S, Lee JH, Einsele H, Sopala M, Bengoudifa BR, Binlich F, Richardson PG. Overall survival of patients with relapsed multiple myeloma treated with panobinostat or placebo plus bortezomib and dexamethasone (the PANORAMA 1 trial): a randomised, placebo-controlled, phase 3 trial. Lancet Haematol. 2016;3:e506-e515.

Schiffman RM, Christianson MD, Jacobsen G, Hirsch JD, Reis BL. Reliability and validity of the Ocular Surface Disease Index. Arch Ophthalmol. 2000;118:615-21.

Siegel DS, Dimopoulos MA, Ludwig H, Facon T, Goldschmidt H, Jakubowiak A, San-Miguel J, Obreja M, Blaedel J, Stewart AK; Improvement in Overall Survival With Carfilzomib, Lenalidomide, and Dexamethasone in Patients With Relapsed or Refractory Multiple Myeloma. J Clin Oncol. 2018;36:728-734

Sonneveld P, Chanan-Khan A, Weisel K, Nooka AK, Masszi T, Beksac M, Spicka I, Hungria V, Munder M, Mateos MV, Mark TM, Levin MD, Ahmadi T, Qin X, Garvin Mayo W, Gai X, Carey J, Carson R, Spencer A. Overall Survival With Daratumumab, Bortezomib, and Dexamethasone in Previously Treated Multiple Myeloma (CASTOR): A Randomized, Open-Label, Phase III Trial. J Clin Oncol. 2023;41(8)1600-09. Published online 2022 Nov 22. doi: 10.1200/JCO.21.02734

Spencer A, Lentzsch S, Weisel K, et al. Daratumumab plus bortezomib and dexamethasone versus bortezomib and dexamethasone in relapsed or refractory multiple myeloma: updated analysis of CASTOR. Haematologica. 2018;103(12):2079–2087. doi:10.3324/haematol.2018.194118

Uno H, Tian L, Claggett B, Wei LJ. A versatile test for equality of two survival functions based on weighted differences of Kaplan–Meier curves. Statistics in medicine. 2015; 34(28):3680-95.

Zhao L, Claggett B, Tian L, et al. On the restricted mean survival time curve in survival analysis. Biometrics. 2016; 72:215-221.