Statistical Analysis Plan

Study ID: 207503

Official Title of Study: DREAMM 7: A Multicenter, Open-Label, Randomized Phase III Study to Evaluate the Efficacy and Safety of the Combination of Belantamab Mafodotin, Bortezomib, and Dexamethasone (B-Vd) Compared With the Combination of Daratumumab, Bortezomib and Dexamethasone (D-Vd) in Participants With Relapsed/Refractory Multiple Myeloma

NCT ID: NCT04246047

Date of Document: 17-November-2023

Information Type:	Statistical Analysis Plan (SAP)

TITLE PAGE

Protocol Title:	Reporting and Analysis Plan for**:** DREAMM 7: A Multicenter, Open-Label, Randomized Phase III Study to Evaluate the Efficacy and Safety of the Combination of Belantamab Mafodotin, Bortezomib, and Dexamethasone (B-Vd) Compared with the Combination of Daratumumab, Bortezomib and Dexamethasone (D-Vd) in Participants with Relapsed/Refractory Multiple Myeloma
Study Number:	207503
Compound Number:	GSK2857916
Abbreviated Title:	Phase III study of GSK2857916, bortezomib, and dexamethasone versus daratumumab, bortezomib, and dexamethasone in participants with relapsed/refractory multiple myeloma (DREAMM 7)
Acronym:	DREAMM-7

Sponsor Name: GSK Research & Development Limited

Regulatory Agency Identifier Number(s)

Registry	**ID**
Clinicaltrials.gov	NCT04246047
EudraCT	2018-003993-29
IND	119333

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

TABLE OF CONTENTS

LIST OF TABLES

VERSION HISTORY

SAP Version	Approval Date	Protocol Version (Date) on which SAP is Based	Change	Rationale
SAP Original Version	06 Mar 2023	TMF-15039862 (19 Dec 2022)	Not Applicable	Original version
SAP Amendment 1	27 Sep 2023	GSK Document Number TMF-15691281 (20 Sep 2023)	1. Addition of interim analyses for PFS and OS; 2. Order of key secondary endpoints changed; 3. Multiplicity adjustment strategy detailed; 4. Increased targeted number of PFS events for primary PFS analysis; 5. Administrative updates to add clarification and/or remove discrepancies	Requirement for increased OS data maturity at the time of Primary PFS analysis and acknowledgment of the importance of OS endpoint.
SAP Amendment 2	17 Nov 2023	GSK Document Number TMF-15691281 (20 Sep 2023)	1. Section 2.1: Added clarifying footnotes to Figure 1. 2. Section 4: • Added detail for source of lab summaries. • Added clarifying footnotes to Tables 1, 3, and 4. • Added detail on source of randomization/strata. • Added display details for concordance between IRC and Inv-Assessed PD. • Added detail for possible additional sensitivity analyses.	Updates related to late dry run

SAP Version	Approval Date	Protocol Version (Date) on which SAP is Based	Change	Rationale
			• Added details for Best Overall Response for participants within measurable disease at baseline. • Added clarifying details for PFS2 analysis and Table 5. • Refined details for MMRM models for PRO endpoints. • Added detail on visit-slotting and OSDI table. • Added detail on calculated dose delay summary. • Added detail on death and SAE summaries. • Added detail on KVA grade summaries. • Added wording for summaries of hepatobiliary laboratory events. • Added wording for summaries of neutropenia events. • Added subgroup analyses for prior Len and refractory to Len. • Added benefit-risk forest plot. • Added details for planned analyses to Table 13.	

SAP Version	Approval Date	Protocol Version (Date) on which SAP is Based	Change	Rationale
			• Added information for re-randomized participants. • Added references.	

1. INTRODUCTION

The purpose of this SAP (Amendment 2) is to describe the planned analyses to be included in the CSR for Study 207503. 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[1]
Primary	
The primary objective of this study is to compare the efficacy of belantamab mafodotin in combination with bortezomib and dexamethasone (bor/dex) with that of daratumumab in combination with bor/dex in participants with RRMM	• Progression-Free Survival (PFS), defined as the time from the date of randomization until the earliest date of documented disease progression or death due to any cause
Key Secondary	
To compare the efficacy of belantamab mafodotin in combination with bor/dex with that of daratumumab in combination with bor/dex in participants with RRMM	• Overall Survival (OS), defined as the time from the date of randomization until 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 • Minimal Residual Disease (MRD) negativity rate, defined as the percentage of participants who are MRD negative by next-generation sequencing (NGS)
Secondary	
To further assess the efficacy of belantamab mafodotin in combination with bor/dex with that of daratumumab in combination with bor/dex in terms of other efficacy outcomes in participants with RRMM	• Complete Response Rate (CRR), defined as the percentage of participants with a confirmed complete response (CR) or better (i.e., CR, stringent complete response (sCR)) • Overall Response Rate (ORR), defined as the percentage of participants with a confirmed partial response (PR) or better (i.e., PR, VGPR, CR, sCR) • Clinical Benefit Rate (CBR), defined as the percentage of participants with a confirmed minimal response (MR) or better per International Myeloma Working Group (IMWG) • Time to Response (TTR), defined as the time between the date of randomization and the first documented evidence of

Objectives	Endpoints[1]
	response (PR or better) among participants who achieve confirmed PR or better • Time to Progression (TTP), defined as the time from the date of randomization until the earliest date of documented PD or death due to PD • PFS2, defined as time from randomization to disease progression 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 belantamab mafodotin when administered in combination with bor/dex	• Incidence of adverse events (AEs) and changes in laboratory parameters • Ocular findings on ophthalmic exam
To further describe the exposure to belantamab mafodotin when administered in combination with bor/dex	• Plasma concentrations of belantamab mafodotin, and cys-mcMMAF
To assess anti-drug antibodies (ADAs) against belantamab mafodotin	• Incidence and titers of ADAs against belantamab mafodotin
To evaluate the safety and tolerability of belantamab mafodotin based on self-reported symptomatic adverse effects when administered in combination with bor/dex	• Maximum post-baseline PRO-CTCAE score for each item attribute
To evaluate and compare changes in symptoms and health-related quality of life (HRQOL)	• Change from baseline in HRQOL as measured by EORTC QLQ-C30 and EORTC IL52 (disease symptoms domain from the EORTC QLQ-MY20)
Exploratory	
To further assess the efficacy of belantamab mafodotin in combination with bor/dex with that of daratumumab in combination with bor/dex in terms of additional efficacy outcomes in participants with RRMM	• 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 • VGPR rate, defined as the percentage of participants with a confirmed Very Good Partial Response (VGPR) or better (i.e., VGPR, CR, sCR) • Sustained MRD negativity rate: defined as the percentage of participants with MRD negativity confirmed by NGS minimum of one year apart, per IMWG criteria

Objectives	Endpoints[1]
To further evaluate the safety and tolerability of belantamab mafodotin when administered in combination with bor/dex	• Changes in safety assessments, including vital signs
To evaluate self-reported ocular symptomatic adverse effects of belantamab mafodotin in combination with bor/dex	• Changes from baseline in symptoms and related impacts as measured by OSDI
To further evaluate and compare changes in health-related quality of life (HRQOL) and symptoms	• Change from baseline in EQ-5D-3L • Change from baseline in PGIS and change in PGIC over time
To further evaluate the impact of side effects on QOL	• Change from baseline in FACT-GP5
To assess imaging plus MRD-negativity rate	• Imaging plus MRD-negativity rate, defined as the percentage of participants who are MRD negative by NGS and who have no evidence of disease on PET-CT
To evaluate and compare healthcare resource utilization (HCRU)	• Number of office/outpatient/hospital clinic visits by specialty • Number of emergency room/urgent care facility visits • Number and duration of in-patient hospitalizations (total nights, including duration by wards [intensive care unit vs. general ward]) • Use of supportive care medication
To further describe the pharmacokinetic of belantamab mafodotin when administered in combination with bor/dex	• Derived pharmacokinetic parameter values of belantamab mafodotin, and cys-mcMMAF, as data permit
To explore the exposure-response relationship between belantamab mafodotin exposure and clinical endpoints in participants treated with belantamab mafodotin in combination with bor/dex	• Belantamab mafodotin exposure (e.g., concentration, Cmax, or AUC) vs. efficacy and safety endpoints (e.g., PFS, ORR, CRR, corneal events)
Explore the relationship between clinical response and biologic characteristics including, but not limited to, BCMA expression on tumor cells and sBCMA concentrations	• Assess various biomarkers at baseline and on-treatment, by tumor and blood-based analysis of DNA, RNA, and protein including but not limited to evaluating baseline BCMA expression and/or immune status in tumor tissue and in the tumor microenvironment and/or serum soluble BCMA levels, and their relationship to clinical response

1. All categories of disease response (sCR, CR, VGPR, PR, SD, PD) used in the calculation of study endpoints will be determined by an IRC using IMWG 2016 criteria.

ADA = anti-drug antibodies; AE = adverse event; BCMA = B-cell maturation antigen; bor/dex = bortezomib/dexamethasone; CBR = Clinical Benefit Rate; CRR = complete response rate; DNA = deoxyribonucleic acid; DoR = duration of response; EORTC QLQ C30 = European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire 30 item Core module; EORTC IL52 = European Organisation for Research and Treatment of Cancer Item Library 52; HCRU = health care resource utilization; HRQoL = health-related quality of life; IMWG = International Myeloma Working Group; MRD = minimal residual disease; NGS = Next-generation

Objectives	Endpoints[1]
sequencing; ORR = overall response rate; OS = overall survival; PD = progressive disease; PFS = progression-free survival; PR = partial response; PRO-CTCAE = Patient Reported Outcomes version of the Common Terminology Criteria for Adverse Events; QOL = quality of life; RRMM = relapsed/refractory multiple myeloma; TTR = time to response; TTBR = time to best response; TTP = time to progression; VGPR = very good partial response.	

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-Vd compared to D-Vd 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-Vd vs D-Vd
	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-Vd vs D-Vd
	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-Vd vs D-Vd
	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-Vd vs D-Vd
	Supplementary 4 (S4)	PFS	ITT	• Disease assessments between scheduled visits: treatment policy • New anti-myeloma therapy: while on treatment	Hazard ratio for B-Vd vs D-Vd

Objective	Estimand Category	Estimand			Population Level Summary Measure
		Variable/ Endpoint	Analysis Set	Intercurrent Event Strategy	
				• Extended loss to follow-up: while on treatment • Treatment discontinuation: composite • 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-Vd vs D-Vd
Key Secondary Objectives: Superiority of B-Vd compared to D-Vd in OS, DoR, and 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-Vd vs D-Vd
	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-Vd vs D-Vd
	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

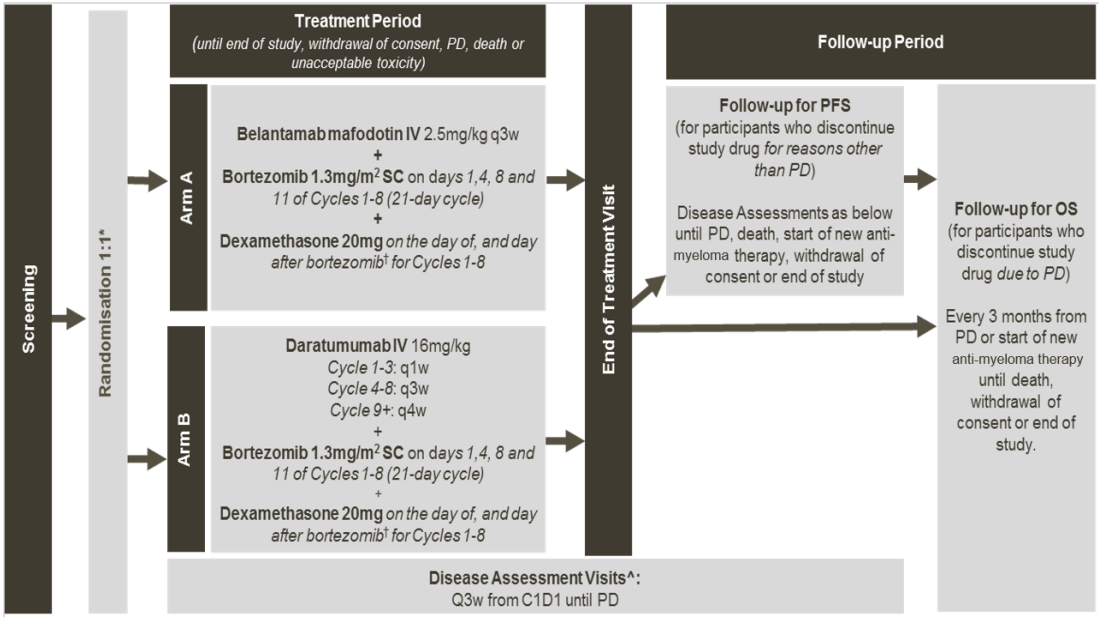
Objective	Estimand Category	Estimand			
		Variable/ Endpoint	Analysis Set	Intercurrent Event Strategy	Population Level Summary Measure
Secondary/Exploratory Objectives (Efficacy): To demonstrate the superiority of B-Vd vs D-Vd in CRR/ ORR/ CBR/ TTR/ TTP/ PFS2/ TTBR/ VGPR+ in participants with relapsed/refractory multiple myeloma (RRMM)[1]	Primary	CRR	ITT	• New anti-myeloma treatment: while on treatment • Treatment discontinuation: treatment policy	>=CR percentage by treatment arm
		ORR	ITT	• New anti-myeloma treatment: while on treatment • Treatment discontinuation: treatment policy	>=PR percentage by treatment arm
		CBR	ITT	• New anti-myeloma treatment: while on treatment • Treatment discontinuation: treatment policy	>=MR percentage 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-Vd vs D-Vd
		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
		VGPR+	ITT	• New anti-myeloma treatment: while on treatment • Treatment discontinuation: treatment policy	>=VGPR percentage by treatment arm
		TTBR	Participants with a confirmed PR	• New anti-myeloma treatment: while on treatment • Treatment discontinuation: treatment policy	Descriptive summary of

Objective	Estimand Category	Estimand			
		Variable/ Endpoint	Analysis Set	Intercurrent Event Strategy	Population Level Summary Measure
			or better in the ITT		median TTBR by treatment arm

[1]. Eligible participants must have a confirmed diagnosis of MM, been previously treated with at least 1 prior line of therapy and must have documented disease progression during, or following, the most recent line of therapy; see inclusion/ exclusion criteria for details.

1.2. Study Design

Overview of Study Design and Key Features



* Stratification: Prior lines of treatment (1 vs 2/3 vs ≥4), R-ISS (I vs II/III), Prior bortezomib (yes vs no).

† Reduce starting dose of dexamethasone to 10 mg for participants older than 75 years of age, who have a body-mass index of less than 18.5, who had previous unacceptable side effects associated with glucocorticoid therapy, or who are unable to tolerate the starting dose.

PD = Progressive Disease; C1D1 = Cycle 1 Day 1; IV = intravenous; SC = subcutaneous; Q1W = once weekly; Q3W = once every 3 weeks; Q4W = once every 4 weeks; PFS = Progression-free Survival; OS = Overall Survival.

Design Features	**Overall Design:** This is a multicenter Phase III, randomized, open-label study evaluating the efficacy and safety of the combination of belantamab mafodotin and bor/dex versus with the combination of daratumumab and bor/dex in participants with RRMM. **Disclosure Statement:** This is a randomized, parallel group treatment study with two arms and no blinding. **Number of Participants:** Approximately 600 participants will be screened to achieve about 478 participants randomized in a 1:1 ratio between the 2 study arms. There will be a [redacted]% global enrollment cap on North East Asia Countries. In these countries, the respective regulatory authorities require a sufficient number of their country populations 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, prior treatment with bortezomib and revised international staging system (R-ISS, see Appendix 13 of Protocol Amendment 6) at screening, and centrally randomized in a 1:1 ratio to either arm. No more than [redacted]% of participants with 2 or more prior lines of treatment will be enrolled. No cross-over will be allowed.

<table>
<tr><th colspan="2">Overview of Study Design and Key Features</th></tr>
<tr><td>Study intervention</td><td>Treatment Arm A: Belantamab mafodotin 2.5 mg/kg (IV) Q3W to progression. Cycles 1 through 8: bortezomib 1.3 mg/m2 (SC) on Days 1, 4, 8, and 11 of every 21-day cycle; and dexamethasone 20 mg (IV or PO) on the day of and the day after bortezomib treatment.

Treatment Arm B: Daratumumab 16 mg/kg (IV) weekly for Cycles 1 through 3 (Weeks 1-9; 21-day cycles, total of 9 doses), on Day 1 of Cycles 4 through 8 (Weeks 10 – 24; 21-day cycles, total of 5 doses), and then every 4 weeks from Cycle 9 (Week 25) onwards until progression (28-day cycles). For Cycles 1 through 8: bortezomib 1.3 mg/m2 (SC) on Days 1, 4, 8, and 11 of every 21-day cycle; and dexamethasone 20 mg (IV or PO, but IV prior to first daratumumab dose) on the day of and the day after bortezomib treatment.

Treatment will continue in both arms until progressive disease, death, unacceptable toxicity, withdrawal of consent, or end of study, whichever occurs first. Dose delays or reductions may be required following potential drug-associated toxicities.</td></tr>
<tr><td>Study intervention Assignment</td><td>Randomization list will be done centrally using a randomization schedule generated by the GSK Clinical Statistics Department in RandALL NG, which will assign participants in a 1:1 ratio to Treatment Arm A and Treatment Arm B. Separate randomization lists will be generated for any extension cohorts required.</td></tr>
<tr><td>Planned Analysis</td><td>
<table>
<tr><th>Analyses / Timing</th><th>Endpoints for analyses</th><th>Data to be used</th></tr>
<tr><td>Safety review by IDMC/ Reviewed periodically starting from when CCI [redacted] and then every [redacted] or as requested by the IDMC thereafter. Ad hoc meetings may be convened at the discretion of the IDMC or if requested by the sponsor.</td><td>Key safety (AEs, SAEs, AESIs, deaths, ocular, exposure, dose modifications, laboratory parameters), descriptive efficacy summaries (e.g. best response categories, and counts of PFS/OS events upon request) and study population summaries.</td><td>All data available at the time of the data cut</td></tr>
<tr><td>Interim Analysis 1 (IA1)
CCI [redacted]</td><td>Minimally, key safety, study population and PFS. CCI [redacted]

CCI [redacted]</td><td>All data available at the time of the data cut</td></tr>
</table>
</td></tr>
</table>

Overview of Study Design and Key Features			
	Primary PFS analysis / Interim Analysis 2 (IA2). CCI [redacted]	All endpoints. CCI [redacted] CCI [redacted]	All data available at the time of the data cut
	Interim Analysis 3 (IA3) CCI [redacted]	CCI [redacted]	All data available at the time of the data cut
	Final analysis CCI [redacted]		All data available
Abbreviations: IA=interim analysis; PFS=progression-free survival; OS=overall survival.			

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 2 treatment groups.

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

where, θ is the PFS HR (belantamab mafodotin/bor/dex vs. daratumumab/bor/dex arm).

Key Secondary Endpoints

a) Overall Survival (OS)

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

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

where, θ_1 is the OS HR (belantamab mafodotin/bor/dex vs. daratumumab/bor/dex arm).

b) Duration of Response

The following statistical hypotheses will be 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 the belantamab mafodotin/bor/dex arm and μ_0 is the RMDOR for the daratumumab/bor/dex arm.

c) MRD Negativity

The following key secondary hypotheses may be tested:

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

where,

P_0 = proportion of participants with MRD negativity Arm B (daratumumab/bor/dex)

P_1 = proportion of participants with MRD negativity Arm A (belantamab mafodotin/bor/dex).

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 the three key secondary endpoints OS, DoR, 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. For the second family, a weighted Bonferroni procedure will be applied across OS and DoR. Alpha will be split between the endpoints, with a larger proportion assigned to OS initially. Testing of MRD will be conditional on the successful rejection of the null hypothesis for OS, aligned with a step-down (or hierarchical) testing procedure [Bretz, 2009; Li, 2017]. The multiple testing strategy 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 $i = 1, 2, 3, 4$ denote the index indicating PFS, OS, DoR and MRD negativity, respectively.

PFS Testing

PFS will be tested across 2 planned analyses: an analysis for efficacy (IA1) and the primary PFS analysis/IA2. The Lan DeMets approach, which 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 IA1 and the primary PFS analysis/IA2, since these analyses provide the opportunity to make a claim of efficacy. All boundaries (see Section 4.7.2) will be adjusted based on the actual number of PFS events observed at the time of analysis.

Testing of key secondary endpoints: OS, DoR, MRD Negativity

Testing of OS and DoR will be conditional on rejection of H_1(PFS). Alpha will be split such that $^4/_5$ of alpha (i.e. 2%) will be initially allocated to testing H_2(OS) and $^1/_5$ of alpha (i.e. 0.5%) will be allocated to testing H_3(DoR).

H_3(DoR) will only be tested using data available at IA1. Note that if H_1(PFS) fails to be rejected at IA1 but is later rejected at Primary PFS/IA2, then the full 0.5% alpha propagated to test H_3(DoR) can be used to test DoR IA1 data at the time of Primary PFS/IA2. If H_3(DoR) is rejected (at IA1 or Primary PFS/IA2), the 0.5% alpha allocated to DoR will be propagated so that H_2(OS) will be tested at the 2.5% level.

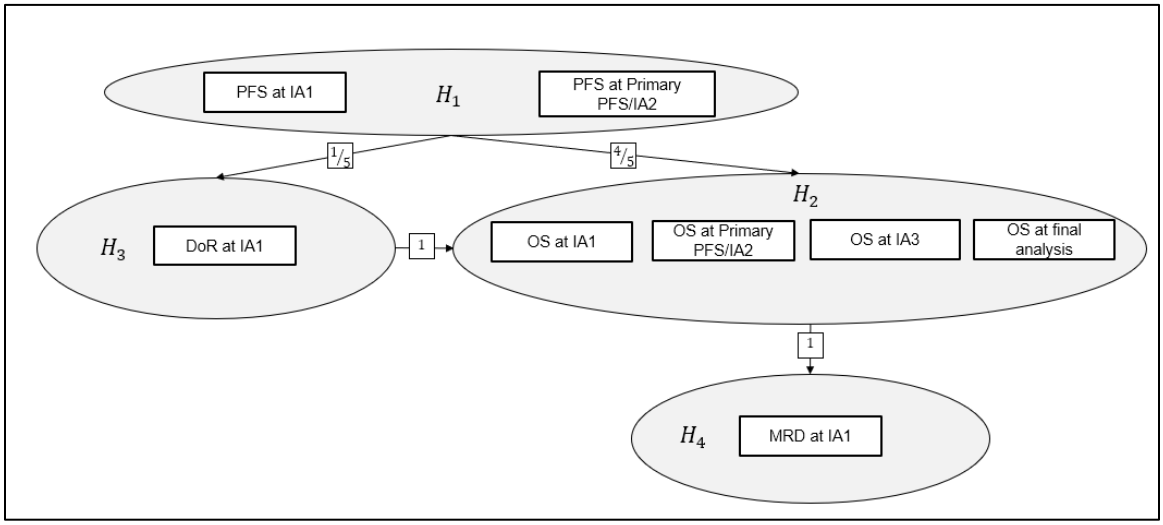
OS will be tested across 4 planned analyses: IA1, primary PFS analysis/IA2, IA3, and at the OS final analysis. The Lan DeMets approach that approximates the O'Brien and Fleming spending function will be used [Lan, 1983]. The efficacy boundaries will be adjusted based on the actual number of OS events observed at the time of analysis and the alpha allocated.

H_4(MRD) testing will be conditional on rejection of H_2(OS). Regardless of the timing of rejection of H_2(OS):

1) H_4(MRD) will only be tested using data available at IA1, similar to DoR.
2) The full alpha allocated to OS (2% or 2.5% conditional on successful rejection of H_3(DoR)) will be propagated.

The remaining secondary efficacy endpoints will be analyzed without alpha adjustment.

Figure 1 Multiple Testing Strategy



Abbreviations: DoR=Duration of Response; IA=Interim Analysis; MRD=MRD Negativity Rate; PFS=Progression-Free Survival; OS=Overall Survival. H_i denotes the one-sided null hypothesis for the primary and key secondary endpoints, where i=1,2,3,4 denotes the index indicating PFS, OS, DoR 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. H_1will 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. The efficacy boundaries will be adjusted based on the observed number of events at the time of analysis.

3. ANALYSIS SETS

Population	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 and some study populations summaries.	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 take at least 1 dose of study treatment. Participants will be analyzed according to the treatment they actually received.	Safety Population
Intent-to-Treat (ITT)	ITT Population will consist of all randomized participants whether or not randomized treatment was administered.	Efficacy

Population	Definition / Criteria	Analyses Evaluated
	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.	
Modified ITT (mITT)	Participants who met all criteria below will be included: • Have received at least 1 prior line of therapy • With measurable disease at baseline[1] • Randomized and received at least one dose of planned study treatment (belantamab mafodotin or daratumumab) o Patient randomized to the belantamab mafodotin arm that received daratumumab will be excluded and vice versa o Patient randomized but never treated will be excluded	Efficacy (sensitivity analysis of primary endpoint and key secondary endpoint)
Pharmacokinetic	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.	PK

[1] Measurable disease at baseline is defined as, at baseline, a patient has at least one of the following measurements: Serum M-protein ≥0.5 g/dL (≥5 g/L) or Urine M-protein ≥200 mg/24 h or 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).

For safety reviews, analysis populations will be labelled as "dummy" populations to account for the masking of treatment groups.

4. STATISTICAL ANALYSES

4.1. General Considerations

4.1.1. General Methodology

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

The stratified log-rank test and stratified Cox proportional hazards models will include the randomization stratification factors as "strata". Unless otherwise specified, the stratification factors entered for randomization will be used in the primary analysis. If there is any mis-stratification, a supplementary analysis will be performed using the stratification data based on the clinical database.

Confidence intervals will use 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 in preference to 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. MRD assessment will be based on central lab values.

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.

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, only central lab values will be used.

For ECG analyses, if the latest, non-missing pre-dose values is from triplicate, the participant level baseline is defined as the mean of triplicate baseline assessments.

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 from randomization until the earliest date of PD, determined by an IRC, according to IMWG criteria [Kumar, 2016], or death due to any cause. The analyses of PFS will be based on the ITT Analysis Set, unless otherwise specified.

Determination of dates of PFS events and dates for censoring are described in Table 1.

4.2.2. Planned Analyses of PFS

- Primary 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 1 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 1 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
1	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
2	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
3	Progression documented at scheduled visits and Progression documented without extended loss-to-follow-up time [4]	Date of assessment of progression	Event

#	Situation	Date of Event (Progression/Death) or Censoring	Outcome Event (Progression/Death) Or Censored
4	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*
5	With post-baseline assessment but no progression (or death)	Date of last 'adequate' assessment of response [2]	Censored
6	No adequate post-baseline assessment before start of new anti-myeloma therapy (prior to documented disease progression or death) [3].	Randomization	Censored
		(S2) Date of starting new anti-myeloma therapy	*(S2) Event*
7	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*
8	Death before first scheduled assessment (or death at Baseline or without any adequate assessments)	Date of death	Event
9	Death between adequate assessment visits	Date of death	Event
10	Death without extended loss-to-follow-up time [4]	Date of death	Event
11	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 3 weeks, a window of 49 days (6 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 49 days, PFS will be censored at the last adequate disease	Censored

#	Situation	Date of Event (Progression/Death) or Censoring	Outcome Event (Progression/Death) Or Censored
		assessment prior to PD/death.	
		(S3) Date of death or progression	*(S3) Event*
12	*(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: 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 response is sCR, CR, VGPR, PR, MR, or SD.
[3]. If PD or death and new anti-myeloma therapy occur on the same day assume the progression or death was documented first (e.g., 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 = 6 weeks + 7-day window = 49-day window; Without extended loss-to-follow-up time is defined as: <= 49 days; after an extended loss-to-follow-up time is defined as: >49 days. More details about extended loss-to-follow-up time are provided in Section 7.2.9.
[5]. Treatment discontinuation of any component due to physician decision → unconfirmed disease progression.

Refer to Table 3 for information regarding the derivation of confirmed response.

Interim PFS Analysis (IA1)

An interim PFS analysis will be conducted when approximately CCI PFS events (CCI % information fraction) are observed. Minimal safety and efficacy outputs will be produced in order for the 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 IA1 using the primary estimand (see Section 4.7.2 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. IA2 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.

Primary PFS Analysis (IA2)

If PFS at IA1 is not statistically significant, the primary PFS analysis will be conducted after observing approximately CCI PFS events in the randomized participants

contributing to the analysis. Assuming successful PFS, OS will be tested at the appropriate alpha level (see Section 4.7.2 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 DoR and MRD negativity will be based on IA1 data.

4.2.3. 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. Kaplan-Meier estimates for the median PFS and the first and third quartiles will be presented, along with 95% CIs. CIs for quartiles will be estimated using Brookmeyer-Crowley method [Brookmeyer, 1982].

The treatment difference in PFS will be compared by the one-sided stratified log-rank test. The stratified log-rank test (stratified by 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 ITT Analysis Set.

Hazard ratio (HR) and its corresponding 95% CI will be estimated from Cox proportional hazard model stratified by randomization factors with treatment arm as the sole explanatory variable. The 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 using the Randomization and Trial Supply Management (RTSM) system (i.e., RAMOS) will be used in the primary analysis. If there is any mis-stratification, supplementary analyses will be performed using the stratification data based on the clinical database (eCRF).

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 difference in PFS will be tested by the stratified log-rank test. • A stratified Cox proportional hazard model 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 IA1, analyses may be repeated at the primary PFS analysis/IA2, 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 cut-off 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 may 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 2.

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 response 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 2 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.

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

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:

2) 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.

4.2.4.5. Sensitivity Analysis of PFS Primary Estimand Considering Local Efficacy Labs

A sensitivity analysis may be performed to allow the use of baseline efficacy assessments from local labs, using investigator-assessed responses.

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

Depending on the number of COVID-19 related deaths, 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.

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 endpoint(s)

- **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 of last contact. The last contact date will be determined by the maximum collection/assessment date from among selected data domains within the clinical database; details will be provided in a separate Output and Programming Specification (OPS) document. When calculating overall survival, all deaths following subsequent anti-cancer therapy will be included. This is the primary estimand of OS, and there is no supplementary estimand of OS.

  Note: attempts to obtain survival status of routine visits 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.

  The last known alive date will be determined by the latest collection/assessment date from among selected data domains within the clinical database; details will be provided in a separate Output and Programming Specification (OPS) document. When calculating overall survival, all deaths following subsequent anti-cancer

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.

- **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 will be provided. Additionally, the ratio of the RMDORs (Arm A/Arm B) and associated 95% CI will be calculated.

- **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 of key secondary endpoints DoR and MRD negativity will be based on data available at the time of IA1. 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. Regardless of timing of PFS statistical significance, formal testing (if applicable) of DoR and MRD negativity will be based on IA1 data.

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)
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* test stratified by the three randomization factors (number of prior lines of therapy (1, 2/3, 4+), prior bortezomib use (yes, no), and R-ISS stage at screening (I, II/III)) at the one-sided 0.025 alpha level. The p-values presented will be 2-sided (5%), and such significance only declared if MRD negativity rate is in favor of GSK2857916 2.5 mg/kg + Bor/Dex (which is equivalent to one-sided 2.5%). A supportive one-sided p-value will be calculated also from fisher's exact test. • Note: MRD interpretation will be based the one-sided CMH p-value. 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 4.2.4.1for RMST method.

- *Analysis of OS based on Stratification Data from the Clinical Database:* An additional sensitivity analysis may 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 repeated at the time of PFS IA1 (conditional upon PFS statistical significant at IA1) as well as primary PFS analysis. DoR will additionally be analysed as follows:

- *RMDOR analysis*: Using investigator-assessed response according to IMWG (based on ITT)

- *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). In addition, pending on maturity of data, the survival probability at 6, 12, and 18 months with 95% CI may be estimated using Kaplan-Meier method.

As an exploratory analysis, a conventional DoR analysis may 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., Kaplan-Meier estimates, Cox proportional hazards model stratified by randomization factors). P-values will not be produced.

MRD Negativity

MRD additional analyses, as described in Section 4.3.1.2, will also be repeated as follows at the time of PFS IA1 (conditional upon PFS statistical significance at IA1) 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 response*

- Using the stratified Cochran Mantel Haenszel test (only if PFS and DoR are statistically significant), based on eCRF stratification (if needed).

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** may be repeated for the responder supplementary estimand 1 at the time of PFS IA1 (conditional upon PFS statistical significant at IA1) 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 endpoint(s)

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 IA1 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.

4.3.2.1. 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 3 for details). Additionally, if participants do not have measurable disease at baseline, they can only be assessed for at least a complete response or progressive disease, per IMWG [Kumar, 2016]. Therefore, in these cases BOR can only be assigned as sCR, CR, PD, or NE. Participants without measurable disease at baseline but 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.

Table 3 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 adequate disease assessment	SD
12	PD (due to reasons other than imaging, i.e., plasmacytoma or bone lesion) [4]	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 49 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 49 days (6 weeks + 7-day window) of PD at first time point before further adequate disease assessment OR	NE

#	Response at any given visit	Response at Subsequent Disease Assessment[1]	Confirmed Response at the given visit
		No subsequent disease assessment: participant **discontinued study** before further adequate disease assessment	
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	NE
15	SD	Any OR No subsequent disease assessment	SD
16	PD due to Imaging (plasmacytoma or bone lesion)[3]	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 adequate (not 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 will only apply if there is no previous confirmed response. Otherwise, *confirmed response at the given visit* will be the last confirmed response category. For 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.

[3] Additionally, per IMWG, in patients without measurable SPEP(IFX)/UPEP(IFX)/FLC levels, a 25% increase in bone marrow plasma-cell percentage irrespective of baseline status (absolute increase must be ≥10%) will also be assessed as Progressive Disease.
[4] This can be a single PD response assessment (unconfirmed), or any response (other than PD) from Investigator-Assessed response, and study treatment discontinuation due to Physician decision → Unconfirmed progressive disease, as per the eCRF.

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

- **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.

- **Clinical Benefit Rate (CBR)** is defined as the percentage of participants with a confirmed minimal response (MR) or better.

- **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 4.

Table 4 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 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 or death due to PD)[4]	Randomization	Censored
With adequate post-baseline assessment and new anti-myeloma treatment started (prior to documented disease progression or death due to PD)[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

Situation	Date of Event (Progression/Death due to PD) or Censored	Outcome Event (Progression/Death due to PD) Or Censored
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 49 days (6 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 49 days, TTP will be censored at the last adequate disease assessment prior to PD/death.	Censored

[1]Adequate baseline assessment is defined as at baseline, a patient has been tested for Serum M-protein, Urine M-protein and Serum FLC assay, and 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) were met.
[2]Extended loss-to-follow-up time = 6 weeks + 7-day window = 49 day window; without extended loss-to-follow-up time is defined as: <= 49 days; after an extended loss-to-follow-up time is defined as: >49 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 response is sCR, CR, VGPR, PR, MR, or SD
[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.

- **Progression-free survival on subsequent line of therapy (PFS2**) is defined as time from randomization (in months) to disease progression after initiation of 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 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 5. For the PFS2 analysis progression (after anti-myeloma therapy) will be based on investigator-assessed response per IMWG. Additionally, if

a patient has discontinued the study, and death details have been retrieved after the end of study date, the end of study date will be used as last contact date.

Table 5 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 line of new anti-myeloma therapy ended (if 1st new 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 line of new 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	min(Last date known alive, end of study date)

[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.

- **CBR**: summaries of CBR (MR or better) by treatment arms will be provided in the same way as ORR.

- **TTR**: TTR 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).

- **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]. TTP analysis will also be conducted using Cox proportional hazards model stratified by applicable randomization factors. 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. 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 may 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 IA1 or required for IDMC review of the benefit:risk.

With the exception of PFS2, all secondary efficacy endpoint analyses will be repeated for the primary estimand but instead using the investigator-assessed response (PFS2 primary analysis is using investigator-assessed response only).

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

4.3.3. Pharmacokinetic Analyses

All pharmacokinetic analyses will be based on the Pharmacokinetic population, unless otherwise specified.

Plasma concentrations of belantamab mafodotin (ADC), total mAb, and cys-mcMMAF will be summarized using descriptive statistics, graphically presented (where appropriate) and listed.

Linear and semi-logarithmic individual concentration-time profiles and mean and/or median profiles (when appropriate) may be plotted for belantamab mafodotin (ADC), total mAb, and cys-mcMMAF.

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, EORTC QLQ-IL52 (disease symptom domain of 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.

The analysis of EORTC QLQ-C30 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 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 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 in Table 6.

Table 6 PRO-CTCAE Levels and Related Code Values

	Levels and related code values				
Response scale	0	1	2	3	4
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. A listing of the PRO-CTCAE score will be provided for each attribute (frequency, severity, interference, presence).

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 7.2.8.1 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, physical functioning, role 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, physical functioning, role functioning, and global health status/QOL) 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 across all scores: 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. Additionally, models will be run only when there are a minimum of 10 participants within each arm.

4.3.5.3. European Organization for Research and Treatment of Cancer Item Library 52 (disease symptoms domain from the EORTC Quality of Life Questionnaire 20-item Multiple Myeloma module) (EORTC QLQ-MY20)

The EORTC 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, 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. Details of deriving domain scores can be found in Section 7.2.8.2.

From the above EORTC QLQ-MY20 domain scores, summaries will be provided for only the disease symptoms domain (EORTC IL52). For the disease symptoms domain (EORTC IL52) the following outputs will be provided:

- The descriptive summary of the actual value and change from baseline by visit
- Summary of the number (%) of patients with improvement in score ≥ 5 and ≥ 10 points by visit.

EORTC IL52 will also be analyzed similarly to EORTC QLQ-C30.

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 each domain will also be provided. 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 response.

Longitudinal changes from baseline by treatment group for EORTC QLQ-IL52 domain score will be explored using a restricted maximum likelihood-based mixed model for repeated measures (MMRM), using the same approach described in Section for EORTC QLQ-C30 analysis.

4.3.5.4. Compliance of PRO-CTCAE, EORTC QLQ-C30, and EORTC QLQ-IL52

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

- Number of patients 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:
    - PRO-CTCAE: with at least one non-missing item score
    - EORTC QLQ-C30: 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 patients with an evaluable baseline form and at least one evaluable post-baseline form, divided by the number of patients expected to complete the baseline form.

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

4.4. Exploratory Endpoint(s) Analyses

Exploratory efficacy endpoints will be analysed at the primary PFS analysis only, unless PFS demonstrates statistical significance at IA1 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 Efficacy Analyses

- **TTBR**, 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. 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).

- **Very good partial response rate (VGPR+),** 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. Summaries of VPPR+ (i.e., VGPR or better including sCR, CR, VGPR) by treatment arms will be provided in the same way as ORR.

Sensitivity analyses

All sensitivity/supportive analyses will be performed at the time of primary PFS analysis. Sensitivity analyses of TTBR and VGPR+ will be repeated for the primary estimand but instead using the investigator-assessed response.

4.4.2. 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 7 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 7 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-tau)	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
Ctau, 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, may use the most current population pharmacokinetic model. 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.

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

4.4.3. 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.

4.4.4. Exploratory Patient Reported Outcome Analyses

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

4.4.4.1. Ocular Safety Disease Index (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 of ≥ minimal clinically important difference (MCID) from baseline score will be summarized at selected time points for (Total score, Ocular Symptoms subscale, Vision-related Function subscale). The number and percentage will be provided for total score and each sub scale (higher score indicates worsening). MCIDs for total score and each sub-scale are listed in Table 8 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 8 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 shift table, showing a summary of worst post-baseline driving and reading levels, will be created. OSDI will also be included in the compliance display summarized in Section 4.3.5.4.

4.4.4.2. 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 his/her 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).

EQ-5D-3L data will be collected and the analysis will be described in a separate analysis plan.

4.4.4.3. 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].

PGIS and PGIC data will be collected and the analysis will be described in a separate analysis plan.

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

FACT-GP5 is a single item from the FACT-GP5, 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 by visit and study intervention. Time points include but are not limited to worst-case post-baseline, end of treatment and last follow-up visit.

4.4.5. Medical Resource Utilization and Health Economics

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

4.4.6. Exploratory MRD Negativity Endpoints

These analyses will be based on the ITT Analysis Set.

4.4.6.1. 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 [insufficient information to determine MRD status]) 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 difference between the treatment arms will be provided along with the corresponding 95% exact CIs.

4.4.6.2. 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. P-values will not be provided.

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.

4.4.7. 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

Extent of exposure to belantamab mafodotin, daratumumab, bortezomib, and dexamethasone will be summarized.

The start date of the overall study treatment is defined as the first dose date of belantamab mafodotin, daratumumab, bortezomib, or dexamethasone, whichever is earlier (i.e. the first study drug start date).

The overall duration of exposure to study treatment (defined in Section 7.2.1) will be calculated and summarized in cycles, using mean, median, standard deviation, minimum, and maximum.

Descriptive statistics of dose intensity will be summarized over all cycles for belantamab mafodotin, bortezomib, and dexamethasone, and within Cycles 1-3, 4-8, and 9+, for daratumumab.

The dose intensity calculation is described below:

- Dose intensity (units/3 weeks) = cumulative actual dose divided by duration of exposure in 3 weeks (duration of exposure in days / 21); will be used for belantamab mafodotin (all cycles), and daratumumab, bortezomib, dexamethasone (cycles 1-8).

- Dose intensity (units/4 weeks) = cumulative actual dose divided by duration of exposure in 3 weeks (duration of exposure in days / 28); will be used for daratumumab (cycles 9+).

- Duration of exposure in days used for the dose intensity calculation is defined as: end date of the cycle – first dose date + 1 day.
    - The end date of the cycle is defined as the cycle start date + 20 days for belantamab mafodotin (all cycles), and daratumumab, bortezomib, dexamethasone (cycles 1-8).
    - The end date of the cycle is defined as the cycle start date + 27 days for daratumumab (cycles 9+).
    - The end date of the last cycle will be calculated as the earliest of: the calculated end date of the last cycle, treatment discontinuation date, or the death date, if the participant discontinues study or dies before the expected end of the last cycle.

Note: Dose intensity units will depend on treatment being summarized (belantamab mafodotin, daratumumab, bortezomib, dexamethasone). Specifics on treatment units can be found in study protocol (Version: GSK Document Number TMF-15691281). Dose intensity for dexamethasone will be split by participants with a first dose of 20mg versus those with a first dose of 10mg. 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)$^{0.5378}$

If baseline height or weight are missing, we will use the closest height/weight date to baseline.

Relative dose intensity (RDI) will also be summarized for belantamab mafodotin, bortezomib, and dexamethasone separately. For daratumumab, RDI will be calculated within for cycles 1-4, 4-8, and 9+. Relative dose intensity is calculated as a percent and is defined as 100*(mean overall dose intensity divided by planned dose intensity). Planned dose intensity for each treatment is calculated as:

- Belantamab Mafodotin = 2.5 mg/kg
- Daratumumab = Cycles 1-3: 48 mg/kg, Cycles 4-8 and 9+: 16 mg/kg
- Bortezomib = 5.2 mg/m^2
- Dexamethasone = 160 mg for participants with a first dose of 20mg, and 80mg for participants with a first dose of 10mg

Summaries of Dose Modifications:

The summaries of dose modifications will be provided. All the dose reductions, infusion interruptions, and dose delays will be summarized or listed.

Dose reductions will be summarised by number of reductions and reasons for reductions. Dose delays will be summarised by number of delays, reasons for the delays, delay duration (days), and reason for dose delay. The number and percentage of the delays for intervals of 1-21, 22-42 and >42, will be computed. For bortezomib and dexamethasone the delay intervals will be defined as 1-8, 9-15, 16-22, and >22 days.

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 dose) was applied.

Duration of delays is defined as period from the expected start date of dose to subsequent actual dosing date following dose delay. Calculation: (actual start date of current dose - expected start date of dose). Expected start date of dose = actual start date of previous dose + 21. For participants on Daratumumab in cycle 9+, expected start date of dose = actual start date of previous dose + 28. When there are multiple doses within a cycle, the expected start date = the expected off-treatment from the previous dose. Duration of

delay will be missing if dosing did not resume prior to data cut off (i.e., if the delay was ongoing or participant subsequently discontinued treatment).

An additional summary of calculated dose delays will be created for belantamab mafodotin and daratumumab, by deriving 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 (all cycles) and daratumumab (cycles 1-8): xx days = 24 days (Q3W + 3 days)

    b) daratumumab (cycles 9+): xx days = 31 days (Q4W + 3 days)

2. Count an additional delay from a participant's last dose of belantamab mafodotin/daratumumab to "end of study". For "end of study", consider the following:

    a) Date of death

    b) Date of decision to discontinue treatment

    c) Treatment discontinuation date

    d) Start date of new anti-myeloma therapy

    e) Last contact date

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

Duration of Follow-Up will be summarized and is defined as the time from randomization to last contact or death.

4.5.2. Adverse Events

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.

An overview summary of AEs, including counts and percentages of participants with any AE, AEs related to any 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 delays, SAEs, SAEs related to study treatment, fatal SAEs, and fatal SAEs related to study treatment will be produced.

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 SOC and PT.

The relationship between MedDRA SOC, PT, and Verbatim Text will be displayed.

Adverse events will be coded using the standard Medical Dictionary for Regulatory Affairs (MedDRA dictionary) and graded by the investigator according to the NCI-CTCAE, (version 5) 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 Preferred term (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 two ways: 1) in descending order by PT only and 2) in descending order by SOC and PT. In the SOC row, the number of participants with multiple events under the same SOC will be counted once.

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 in two ways: 1) by maximum grade sorted by PT in descending order and 2) in descending order 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.

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

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

Details of the planned displays are provided in the OPS.

4.5.2.1. Adverse Events of Special Interest

Adverse events of special interest (AESI) for belantamab mafodotin are corneal events (corneal adverse events), thrombocytopenia and infusion-related reactions. A comprehensive list of MedDRA terms based on clinical review will be used to identify

each type of event. Preferred terms for thrombocytopenia will be identified based on specific queries in the eCRF. Other AESI will be identified based on list of terms of interests which will be produced in Integrated Coding Dictionary System by Clinical Dictionary Development & Management and provided to Statistics and Programming.

Corneal events associated with belantamab mafodotin will be graded according to the KVA scale. Other treatment-related ocular AEs are to be reported based on NCI-CTCAE v5.0 criteria for eye disorders. 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 Section 6.6 of the protocol (Study Protocol: GSK Document Number TMF-15691281). Dose modifications for belantamab mafodotin corneal events will be based on grading of corneal events according to the guidelines of Keratopathy Visual Acuity (KVA) Scale (Study Protocol: GSK Document Number TMF-15691281).

Summaries of the number and percentage of participants with these events will be provided for each type of events separately by preferred term and maximum grade. The time to onset and duration of first occurrence for each type of events 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 <24 hours, (1-21, 22-42, 43-63, >63 days) will be reported. The number and percentage of participants who have duration of first occurrence (1-21, 22-42, >42 days) will be reported. 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 (One, Two, Three or more), maximum grade, maximum grade for events related to study intervention, worst outcome 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.

For thrombocytopenia, number, and percentage of participants with grade 3 or 4 platelet count decreased (based on lab data) and concomitant grade 2 or above bleeding event will

be summarized. A bleeding event will be considered as concomitant only if the start date is within ± 3 days of the lab event.

For infusion related reactions, events would only be considered IRR if the event 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.

Details of the planned displays are provided in the OPS.

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 treatment (>30 days or ≤30 days) and primary cause of death (disease under study, SAE possibly related to study treatment, or other). 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.

SAEs are included in the listing of all adverse events. Separate supportive listings with participant-level details will be generated for:

- Fatal SAEs
- Non-Fatal SAEs

4.5.2.3. Adverse Events Leading to Discontinuation and Dose Modification

The following categories of AEs will be summarized separately by PT and separate supportive listings will be generated with participant level details for those participants:

- AEs Leading to Permanent Discontinuation of Study Treatment
- AEs Leading to Dose Interruptions or Delays
- AEs Leadings to Dose Reductions

4.5.2.4. Ocular Findings from Ophthalmic Exam

Ocular Exam and Visual Acuity

As outlined in study protocol (Version: GSK Document Number TMF-15691281), 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 Arm A. The ocular findings from ophthalmic exams will be summarized descriptively:

- 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; a possible worsened vision is defined as a change from baseline >=0.12 to <0.3; a definite worsened vision is defined as a change from baseline >=0.3 logMAR score.

- Number (%) of Subjects with a Decline in Best Corrected Visual acuity (BCVA) to LP or NLP due to Corneal Exam Findings Anytime Post-Baseline

- Summary of Best Corrected Visual Acuity Test (BCVA) Scores (logMAR score) due to Corneal Event (KVA Scale)

- Summary of Worst Post-Baseline Shift in Best Corrected Visual Acuity (BCVA) Snellen Equivalent

Calculated CTCAE: Changes in Best Corrected Visual Acuity (BCVA) will be 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, for CTCAE grades 2+. Worst post-baseline BCVA change will be calculated using CTCAE grades.

Corneal Exam

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

- Microcyst-like deposits (No to Yes)
- Subepithelial haze (No to Yes)
- Stromal opacity (No to Yes)
- Corneal epithelial defect (No to Yes)
- Superficial punctate keratopathy severity (No to yes)

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-Vd). 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 or 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).

Unless otherwise specified, for the following analyses, KVA scale events will be summarized by treatment arm, 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.

- **Summary of Characteristics of Keratopathy Visual Acuity (KVA) Scale**
  Participants with any event, number of events, events characteristics (serious, requiring hospitalization), number of occurrences, outcome, 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 by excluding all participants who were consented to the original protocol.

- **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 of post-treatment exposure, time to resolution post-treatment exposure, outcome of last event, time to last ocular exam date since last dose, time to resolution 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 Investigator Reported Keratopathy Visual Acuity (KVA) Scale Events Grade 2 or Above Time to Resolution**
  A summary of number of occurrences of grade 2 or above Investigator reported KVA scale events, number of resolved occurrences of grade 2 or above Investigator reported KVA scale events, and duration of occurrences of grade 2 or above Investigator reported KVA scale events, will be produced.

- **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 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/Daratumumab

only. The end date used will be the last recorded infusion irrespective of treatment discontinuation. 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

An additional summary of reported grade for first and second corneal event of Grade 2 or above (GSK/KVA [Investigator-Reported]) will also be produced.

Dose Modifications

Additionally, a dose modification display will be created combining AE and KVA source data. Dose modifications (reduction, interruption / delay) 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.

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. Details of the planned displays are provided in the OPS. 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. These summaries will display the number and percentage of participants with a maximum post-baseline grade increasing from their baseline grade. Increase in grade from baseline will be summarized as “Increase to Grade X” for grades 1, 2, 3, and 4. Additionally, “Increases to Grades X to Y” for grades 1 to 4,

2 to 4, and 3 to 4. Missing baseline grade will be assumed as grade 0. 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, 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 chemical chemistry.

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

A supporting listing of laboratory data for participants with abnormalities of potential clinical concern will be provided. A separate listing of laboratory data with character values will also 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. The medical conditions data for participants with liver stopping events will be listed. The substance use data for participants with liver stopping events will be listed.

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

4.5.3.2. Vital Signs

Values of vital signs (temperature, systolic and diastolic blood pressure, heart rate) as well as the change from baseline will be summarized by scheduled visit using mean, median, standard deviation, minimum and maximum.

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'. Temperature (C) will be categorized into 'Low' (<35), 'Normal' (36-37), and 'High' (>=38) groups. 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 may be provided using the collected values based on Fridericia's formula.

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). A supporting listing will also be provided.

4.5.3.6. Other Risks

Although not a protocol defined AESI, neutropenia has also been identified as an event warranting further investigation based on emerging data. The summary of event characteristics display produced for AESIs will be repeated for 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.

The severity of neutropenia will be graded utilizing NCI-CTCAE v5.0 criteria. A comprehensive list of MedDRA preferred terms for neutropenia based on clinical review will be used to identify each type of event. Neutropenia will be identified based on list of terms of interests which will be produced in Integrated Coding Dictionary System by Clinical Dictionary Development & Management and provided to Statistics and Programming.

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.

4.6. 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 randomized/RTSM strata).

The following subgroup analyses (see Table 9) 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 9 Subgroup Analyses

Subgroup	Categories
Prior Lines of Therapy	1 vs. 2/3 vs. ≥4, and 1 vs. >1
Prior Bortezomib	No, Yes
Prior Lenalidomide	No, Yes
Refractory to Lenalidomide	No, Yes
R-ISS Stage	I vs. II/III
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)
Time to Relapse with 1 Prior Line Of Therapy	Relapse <18 months vs >=18 months
Cytogenetic Risk[1]	High Risk, Standard Risk, Missing or Not Evaluable
Extramedullary Disease	No, Yes

[1] High risk is defined as at least one high-risk abnormality – T(4;14), T(14;16), or 17p13del. Standard risk is defined as negative results for all three high-risk abnormalities - T(4;14), T(14;16), or 17p13del. All other cases will be considered as missing or not evaluable.

For the "Time to Relapse with 1 Prior Line Of Therapy" subgroup analysis, *relapse* is defined as the time from the start date of first line of therapy to date of progressive disease (PD) on that line of therapy. If PD date is not available, we will use date of randomization onto this study. Relapse will be calculated in days and converted to months to define <18 months and ≥18 months categories to be used for the subgroup analysis. This subgroup analysis will be performed for participants with only 1 prior line of therapy.

All subgroup analyses will be performed at the time of primary PFS analysis, unless PFS demonstrates statistical significance at the IA1. 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. Benefit-Risk Forest Plot

A forest plot showing specified primary and key secondary efficacy endpoints, and specified safety endpoints may be generated. The plot will contain proportion of benefit/risk within each treatment arm, as well as hazard ratios (efficacy) or odds ratios (safety) with 95% confidence intervals. Additional details will be described within the OPS.

4.7. Planned Analyses

4.7.1. Periodic IDMC Safety Reviews and CPMS Early Access

Safety data will be reviewed periodically starting from when CCI [redacted], and then every CCI [redacted] or as requested by the IDMC thereafter.

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 several time points (e.g., prior to interim and primary PFS analyses) 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.

4.7.2. Interim Analyses

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

Analysis	Purpose	Timing
Interim Analysis 1 (IA1)	Efficacy	CCI [redacted]
Primary PFS Analysis / Interim Analysis 2 (IA2)	Efficacy. This will also be the planned primary analysis of PFS.	
Interim Analysis 3 (IA3)	Efficacy	
Final OS Analysis (Final OS)	Final OS analysis	

Table 10 presents the PFS efficacy stopping boundaries, while Table 11 and Table 12 presents the OS efficacy stopping boundaries according to 2% or 2.5% alpha, respectively. 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.

Table 10 Stopping Boundaries for Interim Analyses for PFS Efficacy (based on CCI Targeted PFS Events)

Information Fraction	~N of Events	Cum. α Spent	Boundary (p-value)	Boundary (HR)	Boundaries Crossing Probabilities (Incremental)	
					Under H0	Under H1
CCI [redacted]						

Table 11 Stopping Boundaries for Interim Analyses for OS Efficacy (based on 2% Alpha Allocation)

Information Fraction	~N of Events	Cum. α Spent	Boundary (p-value)	Boundary (HR)	Boundaries Crossing Probabilities (Incremental)	
					Under H0	Under H1
CCI [redacted]						

Table 12 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
CCI [redacted]						

4.7.3. Sequence of Interim and Other Planned Analyses

All planned analyses are listed in Table 13.

CCI [redacted]. The analysis requirements and expected timing will be detailed in the country-specific SAP, if applicable.

CCI [redacted]. The details of these analyses including the associated alpha adjustment, if any, will be described in an updated SAP.

Table 13 Details of Planned Analyses

Analyses	Main Purpose	Timing	Endpoints included	Alpha adjustment for Primary and Key Secondary Endpoints
Safety review by IDMC	Safety review	Periodically starting from when CCI [redacted], and then every CCI [redacted] 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 1 (IA1)	Early Efficacy PFS	CCI [redacted]	Minimally, key safety, study population and PFS. CCI [redacted] CCI [redacted]	CCI [redacted]

Analyses	Main Purpose	Timing	Endpoints included	Alpha adjustment for Primary and Key Secondary Endpoints
Primary PFS analysis / Interim Analysis 2 (IA2)	Primary PFS analysis	CCI [redacted]	All endpoints. CCI [redacted] *An endpoint will not be re-tested once statistically significant. For these endpoints, updates (without formal hypothesis testing) will be provided.*	CCI [redacted]
Interim Analysis 3 based on OS (IA3)	Early Efficacy OS	CCI [redacted]		

Analyses	Main Purpose	Timing	Endpoints included	Alpha adjustment for Primary and Key Secondary Endpoints
				CCI [redacted]
Final Analysis	Final OS analysis	CCI [redacted]		

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

There were no changes or deviations to the originally planned statistical analysis specified in Protocol 207503 (DREAMM 7) [GSK Document Number TMF-15691281 , Dated: 2023-SEP-20].

5. SAMPLE SIZE DETERMINATION

Primary Endpoint PFS

Based on data from the CASTOR study, the median PFS in Treatment Arm B is expected to be approximately 16.7 months [Spencer, 2018]. It is expected that treatment with belantamab mafodotin in combination with bor/dex will lead to a CCI [redacted]

The primary PFS analysis will be conducted after observing approximately CCI PFS events. With CCI , the study has a power of CCI [redacted]
This calculation assumes participants are randomized to the two treatment arms in a 1:1 randomization ratio. Assuming that a total of 478 participants will be randomized in a 1:1 ratio to Arm A or Arm B and a uniform enrollment rate of CCI participants per month, enrollment will continue for approximately 16 months. It is estimated that the targeted CCI PFS events will be observed approximately CCI months from the time when the first participant is randomized under H1, assuming an annual dropout rate of [redacted]%. These calculations were conducted using East 6.4.

There will be a CCI % global enrollment cap on North East Asia Countries. 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 enrolment 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, which is based on approximately CCI events. 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

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 daratumumab-bortezomib-dexamethasone (DVd) arm is expected to be around 49 months [Spencer, 2018; Sonneveld, 2023 (CASTOR); Meletios, 2023 (POLLUX); Stewart, 2017 (ASPIRE)]. It is hypothesized that treatment with belantamab mafodotin will result in a CCI% reduction in the hazard rate for OS, i.e., an expected CCI (which corresponds to an increase in median OS to CCI months under the exponential model assumption). In order to ensure 80% power to test the null hypothesis: OS HR = 1, versus the specific alternative hypothesis: OS HR = CCI, a total of CCI deaths need to be observed (~CCI% power). This calculation assumes analysis by a one-sided log-rank test at the overall 2.5% level of significance, participants randomized to the two treatment arms in a 1:1 allocation ratio, and a group sequential design with a Lan DeMets (O'Brien-Fleming) alpha spending function [Lan, 1983] using CCI 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. If OS is tested at the 2% level, under the same assumptions as stated above the study will provide approximately CCI% power to demonstrate superiority of OS for B-Vd vs. D-Vd.

These calculations were made using the software package East 6.5.

Key Secondary Endpoint: Duration of Response

Duration of Response (DoR), as one of the key secondary endpoints, will be formally statistically tested, provided that the primary endpoint PFS is statistically significant.

Comparison of restricted mean DOR (RMDOR, see Section 4.3.1.2 for further details) between the two treatment arms will be based on a one-sided Z test at the overall 0.5% level of significance. Adjustments will be made as per multiplicity strategy in Section 2.1

Key Secondary Endpoint: MRD Negativity

MRD Negativity, 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 IA1. Based on available data from literature, the proportion of participants with MRD Negativity as assessed by NGS with a 10^{-5} sensitivity, in the Dara/bor/dex arm is expected to be around CCI% [Spencer, 2018]. It is hypothesized that treatment with belantamab mafodotin will result in a CCI% absolute increase in MRD negativity to CCI%. 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., 478 participants), the power to detect a difference in the MRD negativity between the 2 treatment arms is approximately CCI%. 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 CCI% power to detect a difference in MRD negativity between the two treatment arms.

These calculations were made using the software package PASS 2019,v19.0.1.

6. RE-RANDOMIZED PARTICIPANTS

There are two participants that were randomized, not treated, re-screened, and re-randomized within a short timeframe. Given these participants were randomized multiple times, with multiple sets of baseline data, the recommended approach to preserve the balance in prognostic factors achieved by randomization is to retain both randomizations (Yelland , 2015).

All analyses using the ITT analysis population will retain both randomizations for the two participants (counting as 4 participants). An additional analysis may be performed for the primary estimand using only the data from the initial randomized participants.

7. SUPPORTING DOCUMENTATION

7.1. 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-cancer therapy, surgical/medical procedures, duration of follow-up and exposure will be based on GSK Core Data Standards.

7.1.1. Participant Disposition

A summary of the number of participants in each of the analysis set described will be provided. 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 summary of screening status and reasons for screen failure will also be produced for the All Screened Analysis Set.

A summary 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 may also be produced.

7.1.2. Demographic and Baseline Characteristics

The demographic characteristics (e.g., age, race, ethnicity, 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 sex and ethnicity. Age categories will be reported to meet differing regulatory and study-specific requirements.

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 and listed.

Disease characteristics at screening including but will not be limited to: International Staging System (ISS) at screening, relapsed or refractory disease, 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, genetics, and high-risk cytogenetics will be summarized and listed. Stratification factors may also be included, based on RTSM and eCRF/vendor data.

Medical conditions collected at screening will be listed and 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.

Prior anti-cancer therapy for multiple myeloma participants will also be summarized by type of therapy, and drug class. A summary of multiple myeloma participants' refractory to prior anti-cancer therapy by drug class will be provided.

Anti-cancer radiotherapy will be listed.

7.1.3. Protocol Deviations

Important protocol deviations will be summarized.

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 freezing the database to ensure all important deviations (where possible without knowing the study intervention details) are captured and categorised in the protocol deviations dataset.

- This dataset will be the basis for the summaries of important protocol deviations.

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

7.1.4. 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. Multi-ingredient products will be summarized by their separate ingredients rather than as a combination of ingredients. 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 and listed separately.

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. Supporting listings will also be provided.

7.1.5. Prior and Subsequent Anti-Myeloma Therapies

Prior anti-multiple myeloma (anti-MM) therapy will be coded using GSK Drug coding dictionary, then summarized by ingredient. A summary of prior lines of therapy may also be produced. Prior anti-cancer therapy for multiple myeloma participants will also be summarized by type of therapy, and drug class. A summary of multiple myeloma participants' refractory to prior anti-cancer therapy by drug class will be provided.

Prior and follow-up/subsequent anti-myeloma therapy will be coded using GSK Drug coding dictionary, then summarized by type of therapy and drug class. "Drug class" is identified by clinical in an external file.

A listing of prior and subsequent anti-myeloma therapy will show the relationship between ATC Level 1, ingredient, and verbatim text. 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.

A separate summary of participants' refractory to prior anti-myeloma therapy by drug class will be provided. Anti-myeloma radiotherapy will only be listed.

Anti-cancer radiotherapy will be listed.

7.1.6. 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.

7.1.7. 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 and listings 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.

7.2. Appendix 2 Data Derivations Rule

7.2.1. Extent of Exposure Calculations

Extent of Exposure
• Number of days of exposure to study drug will be calculated based on the formula: GSK2857916, Bor/Dex, and Dara Cycles 1-8: **Duration of Exposure = Last date of the study drug – First date of the study drug + 1** Dara Cycles 9+: **Duration of Exposure = Last date of the study drug – First 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. 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).

- Participants who were randomized but did not report a treatment start date will be categorised as having zero days of exposure.

- The cumulative dose will be based on the formula:

Cumulative Dose (units) = Sum of Dose at Each Cycle

For GSK2857916:

- Dose intensity will be calculated based on the formula:

Dose intensity (mg/kg/3 week) = Cumulative Dose/((duration of exposure)/21)

For Bor/Dex (up to and including Cycle 8):

- Dose intensity will be calculated based on the formula:

Dose intensity (units/3 week) = Cumulative Dose/((duration of exposure)/21)

    - Units for Bor: mg/m2
    - Units for Dex: mg

For Dara:

- Dose intensity will be calculated based on the formulas:

For cycles 1-3:

Dose intensity (mg/kg/3 week) = Cumulative Dose//((duration of exposure)/21)

For cycles 4-8:

Dose intensity (mg/kg/3 week) = Cumulative Dose//((duration of exposure)/21)

For cycles 9+:

Dose intensity (mg/kg/4 week) = Cumulative Dose//((duration of exposure)/28)

7.2.2. Criteria for Potential Clinical Importance

See OPS.

7.2.3. Study Period

See OPS.

7.2.4. Study Day and Reference Dates

See OPS.

7.2.5. 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.

7.2.6. Multiple measurements at One Analysis Time Point

See OPS.

7.2.7. Handling of Partial Dates

See OPS.

7.2.8. Patient Reported Outcome Analyses

7.2.8.1. 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). The below image shows the details.

Table 1: Scoring the QLQ-C30 version 3.0

	Scale	Number of items	Item range*	Version 3.0 Item numbers	Function scales
Global health status / QoL					
Global health status/QoL (revised)†	QL2	2	6	29, 30	
Functional scales					
Physical functioning (revised)†	PF2	5	3	1 to 5	F
Role functioning (revised)†	RF2	2	3	6, 7	F
Emotional functioning	EF	4	3	21 to 24	F
Cognitive functioning	CF	2	3	20, 25	F
Social functioning	SF	2	3	26, 27	F
Symptom scales / items					
Fatigue	FA	3	3	10, 12, 18	
Nausea and vomiting	NV	2	3	14, 15	
Pain	PA	2	3	9, 19	
Dyspnoea	DY	1	3	8	
Insomnia	SL	1	3	11	
Appetite loss	AP	1	3	13	
Constipation	CO	1	3	16	
Diarrhoea	DI	1	3	17	
Financial difficulties	FI	1	3	28	

* *Item range* is the difference between the possible maximum and the minimum response to individual items; most items take values from 1 to 4, giving *range* = 3.

† (revised) scales are those that have been changed since version 1.0, and their short names are indicated in this manual by a suffix "2" – for example, PF2.

Scores for each scale and single-item measure are averaged and transformed linearly to a score ranging from 0–100 (see image below for details). 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].

Technical Summary

In practical terms, if items $I_1, I_2, \ldots I_n$ are included in a scale, the procedure is as follows:

Raw score
Calculate the raw score
$RawScore = RS = (I_1 + I_2 + ... + I_n)/n$

Linear transformation
Apply the linear transformation to 0-100 to obtain the score S,

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

Symptom scales / items: $S = \{(RS-1)/range\} \times 100$

Global health status / QoL: $S = \{(RS-1)/range\} \times 100$

Range is the difference between the maximum possible value of *RS* and the minimum possible value. The QLQ-C30 has been designed so that all items in any scale take the same range of values. Therefore, the range of *RS* equals the range of the item values. Most items are scored 1 to 4, giving *range* = 3. The exceptions are the items contributing to the global health status / QoL, which are 7-point questions with *range* = 6, and the initial yes/no items on the earlier versions of the QLQ-C30 which have *range* = 1.

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, The EORTC QLQ-C30 Scoring Manual (3rd Edition) 2001].

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].

7.2.8.2. EORTC QLQ-MY20

The EORTC 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, 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.

Table 1. Scoring the QLQ-MY20

	Scale	Number of items (n)	Item range*	QLQ-MY20 item numbers ($I_1, I_2, \ldots, I_n$)
Symptom scales				
Disease Symptoms	DS	6	3	31 - 36
Side Effects of Treatment [a]	SE	10	3	37 - 46
Functional scales / items				
Body Image	BI	1	3	47
Future Perspective	FP	3	3	48 - 50

* "Item range" is the difference between the possible maximum and the minimum response to individual items. All items are scored 1 to 4, giving range = 3.

[a] Item 42 must only be scored if applicable to the patient.

From the above EORTC QLQ-MY20 domain scores, summaries will be provided for only the disease symptoms domain (EORTC IL52). For the disease symptoms domain (EORTC IL52) the following outputs will be provided:

- The descriptive summary of the actual value and change from baseline by visit
- Summary of the number (%) of patients with improvement in score ≥ 5 and ≥ 10 points by visit.

EORTC IL52 will also be analyzed similarly to EORTC QLQ-C30.

As with the QLQ-C30, QLQ-MY20 domain scores are also averaged and transformed linearly to a score ranging from 0–100 (see below for details).

1) Raw score

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

$Raw\ Score = RS = \{(I_1 + I_2 + \ldots + I_n)/n\}$

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:

Symptom scales: $S=\{(RS-1)/range\} \times 100$

Functional scales: $S= \{1-(RS-1)/range\} \times 100$

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.

7.2.9. 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 3 weeks, a window of 49 days (6 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 49 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 49 days, then PFS will be censored at the randomization date.

7.2.10. Derivation of KVA grade (Sponsor-Assessed KVA Grade)

The following hybrid approach (programming algorithm and manual review) will be used to derive KVA grade at each ocular exam visit:

1. At eye level (i.e., separate for each eye), derive KVA grade by programming algorithm:

- At eye level (i.e., separately for each eye), derive Corneal exam grade and Visual acuity grade then combine for KVA grade by programming algorithm:
- At eye level, identify "Not gradable by programming algorithm" visits for each eye:

- Corneal exam grade is 'Not gradable by programming algorithm' for an eye at following visits:
  a. All visits, if at baseline examination any of the following conditions are met:
     Corneal epithelial exam is reported as "Abnormal" OR not reported
  b. Any visit after a Cataract surgery is reported

- Visual acuity grade is 'Not gradable by programming algorithm' for an eye at following visits:
  a. All visits, if at baseline examination any of the following conditions are met:
     Best corrected visual acuity is 20/200 or worse OR not reported.
  b. Any visit after a Cataract surgery is reported

- KVA grade is 'Not gradable by programming algorithm' for an eye at following visits:

  Any visit where Corneal exam grade or Visual acuity grade is 'Not gradable by programming algorithm'.
  At eye level, derive Corneal exam grade and Visual acuity grade for visits that are NOT "Not gradable by programming algorithm" based on the algorithm below:

KVA grade	Grade 1	Grade 2	Grade 3	Grade 4

Corneal examination finding(s) at visit*	Mild superficial punctate keratopathy and no 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

- When there are multiple findings on corneal examination at a visit, the corneal exam grade for the eye will be determined by the worst-case
- 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'.
- Location of sub-epithelial haze was not collected prior to protocol amendment 1. Presence of sub-epithelial haze reported without corresponding location prior to the amendment would be considered sufficient information for corneal exam grade 2 [*applies to D7 only*].
-

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

Change in BCVA lines were calculated using the following table:

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
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

- Reset Corneal exam grade to missing if there is any missing information that may potentially indicate a higher grade (e.g. presence or absence of corneal erosion or ulcer). Note: if a higher grade is already determined, missing data which would only indicate a lower grade would not result in resetting the grade to missing.
- Reset Overall KVA grade to "Not gradable by programming algorithm" or missing at a given visit at eye level based on corneal exam grade and visual acuity grade based on the following conditions being met:

a. If corneal exam grade or visual acuity grade is "Not gradable by programming algorithm" then Overall KVA is "Not gradable by programming algorithm", OR b. If corneal exam grade or visual acuity grade is missing then Overall KVA is missing, OR c. 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, KVA grade for all "Not gradable by programming algorithm" visits will be manually reviewed and graded based on Medical/Safety review
3. At eye level, determine overall KVA grade based on hybrid approach at a visit:
• Assign the higher grade of corneal exam grade and visual acuity grade as overall grade. If grade from one component is missing, assign the grade from the non-missing component grade as the overall grade.

7.2.11. List of Abbreviations

Abbreviation	Description
ADA	Anti-drug antibody
ADC	Antibody drug conjugate
AE	Adverse event
ALT	Alanine transaminase
AUC	Area under the curve
BCMA	B cell maturation antigen
BNP	B-type natriuretic peptide
BP	Blood pressure
CBR	Clinical benefit rate
CI	Confidence interval
CL	Clearance
Cmax	Maximum observed concentration
CR	Complete response
CTCAE	Common Terminology Criteria for Adverse Events
CT	Computed tomography
Ctrough	Concentration at trough
DNA	Deoxyribonucleic acid
DOR	Duration of response
ECG	Electrocardiogram
ECOG (PS)	Eastern Cooperative Oncology Group (Performance Status)
eCRF	Electronic case report form
eGFR	Estimated glomerular filtration rate
EQ-5D-3L	EuroQol Group EQ 5D 3 Level version
EOI	End of infusion
EORTC IL52	European Organisation for Research and Treatment of Cancer - Disease Symptoms domain of EORTC-QLQ-MY20
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
EOT	End of treatment
FLC	Free light chain
HR	Hazard ratio
HRQoL	Health-related quality of life
ICF	Informed consent form
IMWG	International Myeloma Working Group
IRC	Independent Review Committee
ITT	Intent-To-Treat
kg	Kilogram
KVA	Keratopathy Visual Acuity

Abbreviation	Description
λz	Terminal phase elimination rate constant
L	Liter
LVEF	Left ventricular ejection fraction
mAb	Monoclonal antibody
MedDRA	Medical Dictionary for Regulatory Activities
mg	Milligram
µg	Microgram
min	Minute
mm	Millimeter
MM	Multiple myeloma
MMAF	Microtubular inhibitor monomethyl auristatin-F
MMRM	Mixed Model Repeated Measures
MR	Minimal response
MRD	Minimal residual disease
NCI	National Cancer Institute
NCI-CTCAE	National Cancer Institute- Common Toxicity Criteria for Adverse Events
NGS	Next generation sequencing
ORR	Overall response rate
OPS	Output and Programming Specification
OS	Overall survival
PD	Progressive disease
PET	Positron emission tomography
PFS	Progression-free survival
PFS2	Progression-free Survival on Subsequent Line of Therapy
PK	Pharmacokinetic(s)
PopPK	Population PK
PRO-CTCAE	Patient-Reported Outcome Version of the Common Toxicity Criteria for Adverse Events
PDMP	Protocol Deviation Management Plan
PR	Partial response
Q3W	Every three weeks
QTc	Corrected QT interval
QTcF	Frederica's QT Interval Corrected for Heart Rate
RMDOR	Restricted mean duration of response
RRMM	Relapsed/refractory multiple myeloma
SAE	Serious adverse event
SAP	Statistical Analysis Plan
SAS	Statistical Analysis System
sBCMA	Soluble B cell maturation antigen
sCR	Stringent complete response
SD	Stable disease
$t_{1/2}$	Terminal phase half-life
tlast	Last time point where the concentration is above the limit of quantification

Abbreviation	Description
tmax	Time to Cmax
TTR	Time to response
TTBR	Time to best response
ULN	Upper limit of normal
V	Volume of distribution
VGPR	Very good partial response

7.2.12. Trademarks

Trademarks of the GSK Group of Companies
NONE

Trademarks not owned by the GSK Group of Companies
NONMEM
SAS
WinNonlin

8. 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 Aug 1;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.

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.

GSK Document Number TMF-15039862 Study ID207503. DREAMM 7: A Multicenter, Open-Label, Randomized Phase III Study to Evaluate the Efficacy and Safety of the Combination of Belantamab Mafodotin, Bortezomib, and Dexamethasone (B-Vd) Compared with the Combination of Daratumumab, Bortezomib and Dexamethasone (D-Vd) in Participants with Relapsed/Refractory Multiple Myeloma. Report Date 19-DEC-2022.

GSK Document Number TMF-15691281 Study ID207503. DREAMM 7: A Multicenter, Open-Label, Randomized Phase III Study to Evaluate the Efficacy and Safety of the

Combination of Belantamab Mafodotin, Bortezomib, and Dexamethasone (B-Vd) Compared with the Combination of Daratumumab, Bortezomib and Dexamethasone (D-Vd) in Participants with Relapsed/Refractory Multiple Myeloma. Report Date 20-SEP-2023.

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.

Keith Stewart, David Siegel, Heinz Ludwig, et al. Overall Survival (OS) of Patients with Relapsed/Refractory Multiple Myeloma (RRMM) Treated with Carfilzomib, Lenalidomide, and Dexamethasone (KRd) Versus Lenalidomide and Dexamethasone (Rd): Final Analysis from the Randomized Phase 3 Aspire Trial. Blood 2017; 130 (Supplement 1): 743.

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;70:659-663.

Li X, Wulfsohn MS, Koch GG. Considerations on testing secondary endpoints in group sequential design. Statistics in Biopharmaceutical Research. 2017 Oct 2;9(4):333-7.

Meletios A. Dimopoulos, Albert Oriol, Hareth Nahi, et al. Overall Survival With Daratumumab, Lenalidomide, and Dexamethasone in Previously Treated Multiple Myeloma (POLLUX): A Randomized, Open-Label, Phase III Trial. Journal of Clinical Oncology 2023;41:8: 1590-1599.

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.

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.

Sonneveld P, Chanan-Khan A, Weisel K,et. al. Overall Survival With Daratumumab, Bortezomib, and Dexamethasone in Previously Treated Multiple Myeloma (CASTOR): A Randomized, Open-Label, Phase III Trial. J Clin Oncol. 2023 Mar 10;41(8):1600-1609.

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: 2079-87.

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.

Yelland LN, Sullivan TR, Voysey M, Lee KJ, Cook JA, Forbes AB. Applying the intention-to-treat principle in practice: Guidance on handling randomisation errors. Clin Trials. 2015 Aug;12(4):418-23.