Non-Interventional Study Protocol C1071024

COMPARATIVE EFFECTIVENESS OF ELRANATAMAB (PF-06863135) IN CLINICAL STUDY C1071003 VERSUS STANDARD OF CARE IN REAL-WORLD EXTERNAL CONTROL ARMS IN PATIENTS WITH TRIPLE-CLASS REFRACTORY MULTIPLE MYELOMA

Statistical Analysis Plan (SAP)

Version: 1.0

Author: PPD , MPH, PPD
PPD , MPH, PPD

Date: 02 November 2022

TABLE OF CONTENTS

1. LIST OF ABBREVIATIONS

Abbreviation	Term
Anti-CD38	Anti-CD38 monoclonal antibodies
AIPW	Augmented Inverse Probability Weights
BCMA	B-Cell Maturation Antigen
BM	Bone Marrow
CI	Confidence Interval
CR	Complete Response
DOR	Duration of Response
ECOG	Eastern Cooperative Oncology Group
EHR	Electronic Health Record
EMA	European Medicines Agency
FDA	Food and Drug Administration
FLC	Free Light Chain
GBS	Guillain-Barre Syndrome
GVHD	Graft Versus Host Disease
HBV	Hepatitis B Virus
HCV	Hepatitis C Virus
HIV	Human Immunodeficiency Virus
IMiD	Immunomodulatory Drug
IMWG	International Myeloma Working Group
IPTW	Inverse Probability of Treatment Weights
LOT	Line of Therapy
MM	Multiple Myeloma
OR	Objective Response
ORR	Objective Response Rate
PD	Progressive Disease
PI	Proteasome Inhibitor
POEMS	Polyneuropathy, Organomegaly, Endocrinopathy, Monoclonal Gammopathy, and Skin Changes
PR	Partial Response
PS	Propensity Scores
QT	Time from the beginning of the QRS Complex to the End of the T Wave
QTcF	Corrected QT (Fridericia Method)
RW	Real-World
RWD	Real-World Data
sCR	Stringent Complete Response
SMD	Standardized Mean Difference
SOC	Standard of Care
TCR	Triple-Class Refractory
TTR	Time to Treatment Response
ULN	Upper Limit of Normal

Abbreviation	Term
UPEP	Urine Protein Electrophoresis
US	United States
VGPR	Very Good Partial Response

2. AMENDMENTS FROM PREVIOUS VERSION(S)

Not applicable.

3. INTRODUCTION

In this document, any text taken directly from the non-interventional (NI) study protocol is *italicized*.

This Statistical Analysis Plan (SAP) describes the analyses and reporting for protocol C1071024, version 1.0, dated August 15, 2022.

This NI study aims to assess the comparative effectiveness of elranatamab (PF-06863135) versus standard of care (SOC) treatment in triple-class refractory multiple myeloma (TCR MM) patients using external control arms for the open-label, multicenter, non-randomized single-arm Phase 2 Study C1071003. To reduce the potential for bias, external control arms will be constructed from selected fit-for-purpose real-world data (RWD) sources (ie, reliable and relevant),[1] and appropriate comparative effectiveness methods and statistical techniques (eg, inverse probability of treatment weighting, IPTW) will be applied.

This SAP provides details on the research methods to meet the requirements of the United States (US) Food and Drug Administration (FDA) and European Medicines Agency (EMA) on the utility of external control arms derived from RWD in decision making.[2-4]

The SAP was prepared based on the review of the following study protocols:

- Comparative Effectiveness of elranatamab (PF-06863135) in Clinical Study C1071003 Versus Standard of Care (SOC) in Real-World (RW) External Control Arms in Patients with Triple-Class Refractory Multiple Myeloma (TCR MM), Protocol C1071024, Version 1.0, dated August 15, 2022.

- Study C1071003: MagnetisMM-3, an open-label, multicenter, non-randomized phase 2 study of elranatamab (PF-06863135) monotherapy in participants with MM who are refractory to at least one proteasome inhibitor, one immunomodulatory drug, and one anti-CD38 antibody. Protocol Amendment 9, 29 July 2022.

3.1. Study Design

This retrospective cohort study builds upon 2 previous studies, which identified 2 cohorts of RW TCR MM patients to use as external control arms for Study C1071003. The cohorts were identified from 2 US-based oncology electronic health record (EHR) databases, Flatiron Health, and COTA. In the current study, 2 external arms will be constructed from these 2 cohorts of RW TCR MM patients to maximize comparability to patients from Study C1071003 (see Section 4.1 for more detail).

MM patients eligible for selection into external control arms will be those patients who are refractory to at least 1 proteasome inhibitor (PI), 1 immunomodulatory drug (IMiD), and 1 anti-CD38 monoclonal antibody (anti-CD38) and have started at least 1 new treatment since documentation of TCR status. Refractory is defined as having disease progression, according to International Myeloma Working Group (IMWG) criteria or clinical assessment, while on therapy or within 60 days of the last dose in any line of therapy (LOT), regardless

of response. However, if a subsequent LOT is initiated and a progression is observed after at least 30 days of the LOT's start date, the patient will be considered refractory to the subsequent LOT even if the progression occurred within 60 days after the last dose of the preceding LOT.

In the RW setting, no single SOC currently exists for TCR MM patients, and combinations of treatments are frequently used in lieu of monotherapy (5). *In this study, the term "SOC" refers to all standard treatment options available for TCR MM patients.* See Appendix A for the list of available treatments. *Selection of TCR MM patients initiating a new treatment in the external control arm enables comparability of patients at a similar stage in disease progression following TCR documentation.*

*For the primary analyses, differences in baseline and key covariate characteristics including treatment history and disease-related characteristics at the index date between patients in Study C1071003 and each external control arm will be balanced using IPTW (*see Section 8.2.2 for more detail*).*

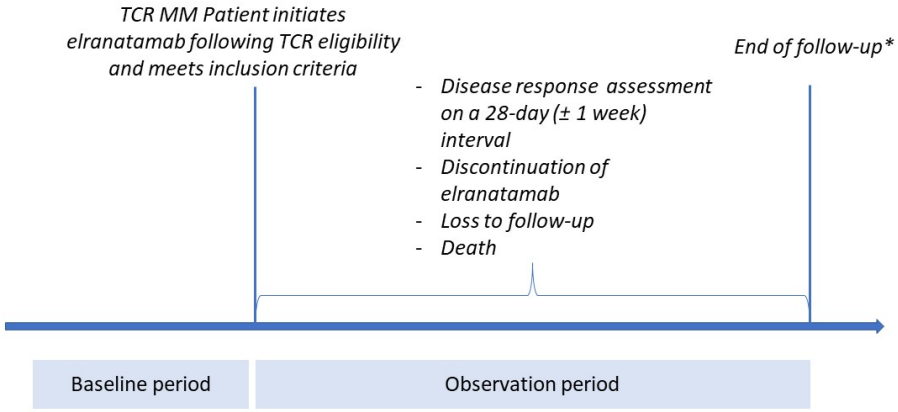
A series of sensitivity analyses will be conducted to assess the robustness of the estimates from the primary analysis using doubly robust estimation with augmented inverse probability weights (AIPW), to evaluate the effect of alternative inclusion/exclusion criteria, and, to evaluate any differences in the magnitude of treatment effect in a sub-group of patients without prior exposure to B-cell maturation antigen (BCMA)-directed therapy (details in Section 8.2.7, Section 5.3, and Section 8.3.5).

Finally, a quantitative bias analysis (nullification analysis) will be performed to evaluate the robustness of results in the presence of potential threats to internal validity (details in Section 8.2.8).

A schematic figure of Study C1071003 and external control arms is provided in Figure 1 *and,* Figure 2 *respectively.*

The date of initiation of the first regimen after TCR MM eligibility will be defined as the index date. The study period will be comprised of the baseline period (time preceding the index date) and the observational period (time following the index date). The observational period will span from the index date to the earliest of death, or latest available patient's record, whichever comes first. Clinical outcomes of interest will be objective response rate (ORR), time to treatment response (TTR), and duration of response (DOR).

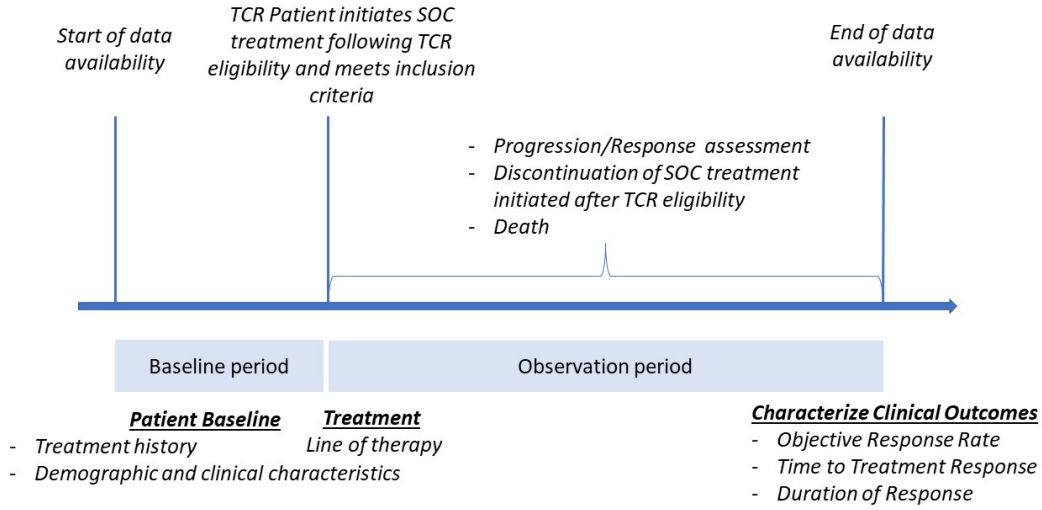
Figure 1. Baseline and Observation Periods in Study C1071003



** For this protocol, all participants will be followed for at least 9 months from the date of enrollment*

Abbreviations: MM=multiple myeloma; TCR=triple-class refractory.

Figure 2. Baseline and Observation Periods in External Control Arms.



Abbreviations: MM=multiple myeloma; TCR=triple-class refractory; SOC=standard of care

3.1.1. Study Population

Study C1071003 population

Study C1071003 is an open-label, multi-center, non-randomized Phase 2 study of elranatamab (PF-06863135) monotherapy.[6] To determine the effects of prior BCMA-directed therapy on the response to elranatamab monotherapy, Study C1071003 enrolled 2 independent and parallel cohorts, 1 with patients who are naïve to BCMA-directed therapies

(Cohort A; approximately 120 patients) and the other with patients previously exposed to BCMA-directed therapy (Cohort B; approximately 60 patients).

Populations of RW TCR MM patients

The cohorts of RW TCR MM patients for the external control arm were previously extracted from Flatiron Health and COTA databases that collect information on MM patients treated across the US. Flatiron and COTA have been selected according to data availability and as fit-for-purpose for fulfilling study objectives (US FDA, 2018 Framework). The dates for the study periods for each RW database have been selected to align as closely as possible to each other and to Study C1071003 which will include at least 9 months of follow-up.

3.1.2. Data Sources

Study C1071003

Study C1071003 is an open-label, multi-center, non-randomized Phase 2 study of PF-06863135 (elranatamab) monotherapy (6)*. To determine the effects of prior BCMA-directed therapy on the response to PF-06863135 monotherapy, Study C1071003 enrolled 2 independent and parallel cohorts: Cohort A with patients who are naïve to BCMA-directed therapies (approximately 120 patients) and Cohort B with patients previously exposed to BCMA-directed therapy (approximately 60 patients).*

Flatiron Health

The Flatiron Health database is a longitudinal, demographically, and geographically diverse database derived from EHR data. Flatiron includes data from over 280 community cancer centers and academic institutions (~800 sites of care) representing more than 2.4 million active US cancer patients available for analysis. The source population is the overall population reported in the EHR and includes patients managed in at least 1 of the US oncology centers taking part in the Flatiron Health network from 01 January 2011 onwards.

COTA

COTA maintains a multidisciplinary data curation approach. The COTA database is a longitudinal database derived from the EHR of healthcare provider sites including academic institutions, community centers, and hospital systems representing 500,000 patients from over 200 sites of care in the US. Data elements are standardized across sources and ontologies to create a single, structured dataset to cover the full longitudinal history of a patient's clinical care.

3.1.3. Combined Study Datasets

Combined study datasets will be created from the Study C1071003 arm plus the selected external control arm based on the data source and eligibility criteria.

A total of 6 combined study datasets will be created depending upon the Study C1071003 cohort (Cohort A or Cohorts A and B) selected, the RWD source, and inclusion/exclusion criteria applied to identify the external control arms, as indicated below.

To clarify the Data Analysis section of the study protocol (Section 9.7), the primary analyses will focus on the comparison of Cohort A from Study C1071003 to the two RWD sources. The comparison of both Cohorts A and B combined from Study C1071003 to the two RWD sources has been identified as an alternative analysis. As such, for the primary analyses:

1. Study C1071003 Cohort A arm plus external control arm selected using critical eligibility criteria from Flatiron Health

2. Study C1071003 Cohort A arm plus external control arm selected using critical eligibility criteria from COTA

For the sensitivity analyses:

3. Study C1071003 Cohort A arm plus external control arm selected using expanded eligibility criteria from Flatiron Health

4. Study C1071003 Cohort A arm plus external control arm selected using expanded eligibility criteria from COTA

For the alternative analyses using Study C1071003 Cohort A and Cohort B:

5. Study C1071003 Cohort A and Cohort B arms plus external control arm selected using critical eligibility criteria from Flatiron Health

6. Study C1071003 Cohort A and Cohort B arms plus external control arm selected using critical eligibility criteria from COTA

3.1.4. Treatment/Cohort Labels

The following treatment & cohort labels will be used:

- Study C1071003 Cohort A patients will be labelled as “Elranatamab”.
- Study C1071003 Cohort B patients will be labelled as “Elranatamab & prior BCMA”.
- Study C1071003 cohort of all patients will be labelled as “Elranatamab with or without prior BCMA exposure”.
- External control arm patients will be labelled as “Standard of Care” or “SOC”.
- External control arm patients without prior BCMA-will be labelled as “SOC & BCMA naïve”.
- External control arm patients with prior exposure to BCMA-directed therapy will be labelled as “SOC & prior BCMA”.

3.2. Study Objectives, Endpoints, and Estimand

Objectives	Endpoints
Primary	
To compare ORR among TCR MM patients treated with elranatamab in Study C1071003 with a comparable population of TCR MM patients receiving SOC therapy	ORR per IMWG
Secondary	
1. To compare TTR in TCR MM patients treated with elranatamab in Study C1071003 with a comparable population of TCR MM patients receiving SOC therapy	TTR per IMWG
2. To compare DOR in TCR MM patients treated with elranatamab in Study C1071003 with a comparable population of TCR MM patients receiving SOC therapy	DOR per IMWG

Primary estimand

Treatment effect of elranatamab compared to SOC on the ORR. The estimand has the following attributes:

- Population:

  TCR MM patients, as defined by the inclusion and exclusion criteria to reflect the targeted population of the treatment, who received at least 1 treatment dose (elranatamab or SOC).

- Variable:

  Objective response (OR), from the date of first dose until the first documentation of progressive disease (PD), death or start of new anticancer therapy, whichever occurs first.

- Intercurrent events:

  All data collected after the following events will be excluded: the first documentation of PD, death, the start of new anticancer therapy, or the end of available data in the patient's record.

- Population-level summary measure:

  The Average Treatment Effect (ATE) expressed as a relative risk (RR) or the ratio of the ORR in patients treated with elranatamab versus those treated with SOC, including the 2-sided 95% confidence interval (CI) and p-value.

4. HYPOTHESES AND DECISION RULES

4.1. Statistical Hypotheses

Primary endpoint

It will be tested whether the primary endpoint ORR is different between elratanamab and SOC patients.

H_0: $ORR_{Elratanamab} = ORR_{SOC}$ vs.

H_1: $ORR_{Elratanamab} \neq ORR_{SOC}$

Secondary endpoints

Secondary endpoints are time-to-event endpoints, specifically TTR and DOR.
Thus, the following pairs of hypotheses will be tested:

H_0: $TTR_{Elratanamab} = TTR_{SOC}$ vs.

H_1: $TTR_{Elratanamab} \neq TTR_{SOC}$

H_0: $DOR_{Elratanamab} = DOR_{SOC}$ vs.

H_1: $DOR_{Elratanamab} \neq DOR_{SOC}$

4.2. Statistical Decision Rules

The alpha level will be 0.05, 2-sided. No adjustments of the level of significance for multiple comparisons will be made because the participants from external control arm are not randomized, but actual RW patients.[7]

If the 2-sided p-value for a pair of hypotheses is ≤0.05, the test decision is that the treatment effects according to the endpoint are different and the null hypothesis will be rejected. If p-value is > 0.05, the null hypothesis will not be rejected.

5. ANALYSIS SETS/POPULATIONS

5.1. Full Analysis Set

Comparability of patients between Study C1071003 and the external control arms is 1 of the key considerations to minimize bias. Due to the missingness present within RWD sources, there is an inherent tradeoff between analyzable sample size and the degree of comparability of patients from Study C1071003 with the external control arm sample. In other words, the more closely the inclusion/exclusion criteria for the external control arms are aligned to Study C1071003, the smaller the available sample size.

From each population of RW TCR MM patients, we will establish 2 samples for external control arms to assess potential variation in study estimates based on different selection criteria.

1. **Critical eligibility criteria sample**: Defined based on a limited set of criteria that have shown to have the strongest influence on future outcomes.

2. **Expanded eligibility criteria sample**: Defined based on a more extensive set of selection criteria beyond the critical set.

The critical eligibility criteria sample will constitute the primary analysis set. Sensitivity analyses will be conducted using the expanded eligibility criteria to assess the potential influence of the selection process on the observed effects (ie, primary analyses will be repeated using this sample).

See Table 1 for information on the inclusion and exclusion criteria of Study C1071003 that is available and applied for RW patients (signified with a checkmark symbol where the critical and/or expanded selection criteria definitions are applicable).

Table 1. Inclusion and Exclusion Criteria for Study C1071003, Flatiron Health, and COTA.

Patient criteria per Study C1071003	Implementation in Flatiron Health		Implementation in COTA	
	Critical	Expanded	Critical	Expanded
Inclusion Criteria				
Male or female patients age ≥18 years	ü	ü	ü	ü
Willing and able to comply with all scheduled visits, treatment plans, laboratory tests, lifestyle considerations, and other study procedures	X	X	X	X
Prior diagnosis of MM as defined according to IMWG criteria (Rajkumar et al, 2014)	ü	ü	ü	ü
Measurable disease, based on IMWG criteria as defined by at least 1 of the following a) Serum M-protein ≥0.5 g/dL by SPEP b) Urinary M-protein excretion ≥200 mg/24 hours by UPEP c) Serum immunoglobulin FLC≥10 mg/dL (≥100 mg/L) and abnormal serum immunoglobulin kappa to lambda FLC ratio (<0.26 or >1.65)	ü	ü	ü	ü
Patients are TCR defined as being refractory to all 3 of the following: a) Refractory to at least 1 IMiD b) Refractory to at least 1 PI c) Refractory to at least 1 anti-CD38 antibody Relapsed/refractory to last anti-MM regimen	ü	ü	ü	ü
Initiated at least 1 anti-MM systemic therapy after becoming TCR eligible. The first systemic treatment initiation after becoming TCR eligible must not comprise any study or investigational agent. *	ü	ü	ü	ü
ECOG performance status ≤2	ü	ü	ü	ü
Adequate hepatic function characterized by all of the following: a) Total bilirubin ≤2 x ULN (≤3 x ULN if documented Gilbert's syndrome) b) aspartate aminotransferase (AST) ≤2.5 x ULN c) alanine aminotransferase (ALT) ≤2.5 x ULN	X	ü	X	ü
Adequate renal function, defined by an estimated creatinine clearance ≥30 mL/min (according to the Cockcroft Gault formula, by 24-hour urine collection for creatinine clearance, or according to the local institutional standard method)	X	ü	X	ü

Patient criteria per Study C1071003	Implementation in Flatiron Health		Implementation in COTA	
	Critical	Expanded	Critical	Expanded
Adequate BM function characterized by all of the following a) Absolute neutrophil count ≥1.0 × 109/L (use of granulocyte-colony stimulating factors is permitted if completed at least 7 days before planned start of dosing) b) Platelets ≥25 × 109/L (transfusion support is permitted if completed at least 7 days before the planned start of dosing) c) Hemoglobin ≥8 g/dL (transfusion support is permitted if completed at least 7 days before the planned start of dosing)	X	ü	X	ü
Left ventricular ejection fraction ≥40% as determined by a multigated acquisition scan or echocardiogram	X	X	X	X
Resolved acute effects of any prior therapy to baseline severity or Common Terminology Criteria for Adverse Events Grade ≤1	X	X	X	X
Exclusion Criteria				
Active plasma cell leukemia	ü	ü	ü	ü
Amyloidosis	ü	ü	ü	ü
Previous treatment with an anti-BCMA bispecific antibody.	X	X	X	X
Previous administration with an investigational drug within 30 days (or as determined by the local requirement) or 5 half-lives preceding the first dose of study intervention used in this study (whichever is longer).	ü	ü	ü	ü
Smoldering MM	ü	ü	ü	ü
Stem cell transplant within 12 weeks before enrolment or active GVHD.	ü	ü	ü	ü
Any other active malignancy within 3 years before enrolment, except for adequately treated basal cell or squamous cell skin cancer, or carcinoma *in situ*.	ü	ü	ü	ü
POEMS syndrome	X	X	X	X
Impaired cardiovascular function or clinically significant cardiovascular diseases, defined based on the history of any of the following conditions within 6 months before enrolment: a) Acute myocardial infarction or acute coronary syndromes (eg unstable angina, coronary artery bypass graft, coronary angioplasty or stenting, symptomatic pericardial effusion) b) Clinically significant cardiac arrhythmias (eg uncontrolled atrial fibrillation or uncontrolled paroxysmal supraventricular tachycardia) c) Thromboembolic or cerebrovascular events (eg transient ischemic attack, cerebrovascular accident, deep vein thrombosis, or pulmonary embolism) d) Prolonged QT syndrome (or triplicate average QTcF >470 msec).	X	ü	X	ü
Active HBV, hepatitis C virus HCV, SARS-CoV-2, known HIV , or any active, uncontrolled bacterial, fungal, or viral infection. Active infections must be resolved at least 14 days before enrolment	ü	ü	ü	ü
Other surgical (including major surgery within 14 days prior to enrolment), medical, or psychiatric conditions including recent (within the past year) or active suicidal ideation/behavior or laboratory abnormality that may increase the risk of study participation or, in the investigator's judgment, make the patient inappropriate for the study.	X	X	X	X
Ongoing Grade ≥2 peripheral sensory or motor neuropathy.	X	ü	X	X
History of any grade peripheral sensory or motor neuropathy with prior BCMA-directed therapy (Cohort B).	X	ü	X	ü
History of GBS or GBS variants, or history of any Grade ≥3 peripheral motor neuropathy	X	ü	X	X

Patient criteria per Study C1071003	Implementation in Flatiron Health		Implementation in COTA	
	Critical	Expanded	Critical	Expanded
Investigator site staff or Pfizer employees directly involved in the conduct of the study, site staff otherwise supervised by the investigator, and their respective family members.	X	X	X	X
Known or suspected hypersensitivity to the study intervention or any of its excipients.	X	X	X	X
Live attenuated vaccine must not be administered within 4 weeks of the first dose of the study intervention.	X	X	X	X

Note:
* This criterion is not among the inclusion/exclusion criteria of Study C10710003 and is applied only to RW patients.
Abbreviations: AST=aspartate aminotransferase; ALT=alanine aminotransferase; BCMA=B-cell maturation antigen; BM=bone marrow; ECOG=Eastern Cooperative Oncology Group; HCV=hepatitis C virus; FLC=free light chain; GBS= Guillain-Barre syndrome; GVHD=graft versus host disease; HBV=hepatitis B virus; HIV= human immunodeficiency virus; IMiD=immunomodulatory drug; IMWG=International Myeloma Working Group; MM=multiple myeloma; PI=proteasome inhibitor; POEMS=polyneuropathy, organomegaly, endocrinopathy, monoclonal gammopathy, and skin changes; QTcF=corrected QT (Fridericia method); SARS-CoV2=severe acute respiratory syndrome coronavirus 2; SPEP=serum protein electrophoresis; TCR=triple-class refractory; ULN=upper limit of normal; UPEP=urine protein electrophoresis.

The flowchart of patients will be updated upon the creation of external control arms using the critical and expanded eligibility criteria.

5.2. Safety Analysis Set

A safety analysis set will include patients from Study C1071003 Cohort A who received at least 1 dose of elranatamab and RW patients selected using critical eligibility criteria.

Overall, 2 safety analysis sets will be created: 1 with RW patients identified from Flatiron Health database, and the second with RW patients identified from the COTA database.

These sets will be the primary analysis population for evaluating effectiveness endpoints and participant characteristics.

5.3. Other Analysis Set

Sensitivity analysis sets

These sensitivity analysis sets will include patients from Study C1071003 Cohort A who received at least 1 dose of elranatamab and RW patients selected using expanded eligibility criteria.

As such, 2 sensitivity analysis sets will be created: 1 with RW patients identified from Flatiron Health database, and the second with RW patients identified from the COTA database.

5.4. Alternative Analysis Sets Using Study C1071003 Cohort A and B

These analysis sets will include an alternative grouping of all patients from Study C1071003 (ie, Cohort A and Cohort B rather than just Cohort A).

As such, 2 alternative analysis sets will be created: 1 with Study C1071003 patients from Cohort A and B, and RW patients identified from Flatiron Health database using critical eligibility criteria, and the second with Study C1071003 patients from Cohort A and B, and RW patients identified from the COTA database using critical eligibility criteria.

6. ENDPOINTS AND COVARIATES

6.1. Exposure Definition

In each analysis set, patients will be classified into 1 of 2 treatment groups according to the therapy received after TCR eligibility as those treated with elranatamab and those treated with a SOC regimen (any standard treatment option available for RW TCR MM patients, see Appendix A for the list of SOC treatments).

6.2. Effectiveness Endpoints

Definitions of the outcomes will be aligned, where possible, with Study C1071003 (Table 2). *In Study C1071003, the capture of ORR is defined according to IMWG criteria.*

Table 2. Definitions of Comparative Effectiveness Outcomes in Study C1071003 and Real-World Data Sources.

	Outcome	***Study C1071003***	***Flatiron Health***	***COTA***
1	*Objective response rate (ORR)*	*The proportion of patients with an OR based on blinded independent central review per IMWG criteria. OR is defined as having a best overall response of confirmed sCR, CR, VGPR, and PR per IMWG criteria, from the date of the first dose until confirmed PD, death or start of new anticancer therapy, whichever occurs first.*	*The proportion of patients who achieved at least VGPR or PR* *Based on IMWG criteria**	*The proportion of patients who achieved sCR, CR, VGPR, or PR* *Based on IMWG criteria**
2	*Time to response (TTR)*	*For patients with an OR per IMWG criteria, the time from the first dose to first documentation of OR that is subsequently confirmed.*	*For subjects with an OR (based on PR or VGPR), time from the initiation of the first line after TCR to the first documented OR (first date at which at least VGPR or PR was documented).*	*For subjects with an OR, time from the initiation of the first line after TCR to the first documented OR.*
3	*Duration of response (DOR)*	*For patients with an OR per IMWG criteria, the time from the first documentation of OR that is subsequently confirmed, until confirmed PD per IMWG criteria, or death due to any cause, whichever occurs first.*	*Among patients who achieved an OR (based on PR or VGPR), the time from the first documentation of OR until progression or death due to any cause, whichever occurs first.*	*Among patients who achieved an OR, the time from the first documentation of OR until progression or death due to any cause, whichever occurs first.*

Note:
* Details on the operational definition of IMWG criteria used in Flatiron and COTA to define response, and a side-by-side comparison with the criteria used in Study C1071003 are provided in the Protocol C1071024 (as Document 001 and 002).
Abbreviations: CR=complete response; IMWG=International Myeloma Working Group; PD=progressive disease; PR=partial response; RWD=real world database; sCR=stringent complete response; TCR=triple-class refractory; VGPR=very good partial response.

Objective response rate (ORR)

ORR is defined as the proportion of participants with an OR per IMWG criteria. OR will be assessed over the period from the index date until the first documentation of progression, death, or the start of new anticancer therapy.

For the analysis set consisting of patients from Study C1071003 and COTA, the OR will be defined as PR or better (sCR + CR + VGPR + PR).

For the analysis set consisting of patients from Study C1071003 and Flatiron Health, the OR will be defined as PR or better (at least VGPR + PR).

Time to response (TTR)

TTR will be estimated in the subgroup of patients with the OR per IMWG as the time from the index date to the first documentation of OR. No censoring will be performed.
TTR will be calculated as:
TTR (weeks) = [date of first objective response – index date + 1 day] / 7 days

Duration of response (DOR)

DOR will be estimated in the subgroup of patients with the OR per IMWG, as the time from the first documentation of OR, until confirmed PD per IMWG criteria, or death due to any cause, whichever occurs first.

DOR will be estimated as:
DOR (months) = [date of censoring event - first date of objective response + 1] /30.4375

The follow-up will be censored as follows:

- For patients who start a new line of therapy prior to an event (confirmed PD per IMWG criteria or death due to any cause), censoring will occur on the date the new anticancer therapy was initiated.

- For patients who do not have an event (confirmed PD per IMWG criteria or death due to any cause), censoring will occur on the data cutoff date.

6.3. Safety Endpoints

Safety is not evaluated in this study as it is unrelated to the study objectives.

6.4. Other Endpoints

Not applicable.

6.5. Covariates

Baseline characteristics, including demographics, comorbidity profile, disease characteristics, laboratory test/results, and treatment patterns will be measured to describe treatment groups at the index date and to perform further statistical adjustments to control for baseline confounding (eg, via IPTW). Adaptations of variable definitions or the use of proxy variables may be necessary if the preferred variables are unavailable in each data source.

Measurement of baseline characteristics will occur on or prior to the index date; if multiple measures are available prior to the index date, the most recent measurement will be selected for analysis.

Some variables, including adequate hepatic function (as defined in Table 3), adequate renal function (as defined in Table 3), adequate BM function (as defined in Table 3), impaired cardiovascular function (as defined in Table 3) will be used only to describe treatment groups at index date.

All variables will be reviewed for data completeness. Descriptive statistics will be provided (Section 8.2.1).

See Table 3 for the operational definition of variables.

6.5.1. Priority Covariates

Since small sample sizes may limit the number of covariates that can be accounted for in the statistical adjustment (eg, IPTW), priority will be given to addressing imbalances in the following covariates: age, sex, ISS stage, ECOG performance status, time since initial MM diagnosis, number of pre-index treatment lines, cytogenetic risk, and extramedullary disease (this variable is available only for the analysis set that include patients from Study C1071003 and COTA).

6.5.2. Additional Covariates

Additional covariates listed in Table 3 may be included in the propensity score model for estimation of IPTW if their inclusion substantially improves covariate balance and provided that sufficient data are available to avoid a large reduction in the sample size by entering them into the model.

Table 3. Operational Definitions of Variables

Variable	Operational definition	Role
Age (Years)	On the index date	Priority covariate
Sex	Male, female	Priority covariate
ISS stage	Stages I, II, III, missing Within 90 days before or on the index date, if feasible. ISS stage will be derived based on measurement of beta-2 microglobulin and serum albumin (See Section 6.5.3)	Priority covariate
ECOG performance status	ECOG=0, 1, 2 Within 90 days before or on the index date, if feasible	Priority covariate
Time since initial MM diagnosis (Days)	From the date of MM diagnosis to the index date	Priority covariate
High cytogenetic risk	Yes, No High risk if any of the following chromosomal abnormalities: t(4;14), t(14,16), del(17p) Before or on the index date, if feasible	Priority covariate
Number of pre-index treatment lines	N, % 1, 2, 3, 4, 5, … Between the date of MM diagnosis and index date	Priority covariate
Penta-refractory status	Yes, No. Penta-drug refractory (refractory to 2 IMIDs, 2 PIs and 1 anti-CD38) At time of TCR eligibility	Priority covariate
Extramedullary disease	Yes, No Yes: presence of any plasmacytoma (extramedullary and paramedullary) with a soft tissue component. Identified on or before the index date. Note, the variable is not reported in Flatiron Health	Priority covariate for analysis set that include patients from Study C1071003 and COTA. This variable is not available in Flatiron Health.
Race/Ethnicity	White, non-white	Additional covariate
Body mass index (Kg/m^2)	Identified from the most recent measure on or before the index date	Additional covariate
Treatment setting	Community-based, academic-based. Note, treatment settings are not reported in COTA At the index date	Additional covariate
Presence of bone lesions	Yes, No Identified on or before the index date.	Additional covariate
Duration of prior therapy (Days)	Between the date of MM diagnosis and index date	Additional covariate
Stem cell transplantation	Yes, No On or before the index date.	Additional covariate
Radiation therapy	N, % Yes, No Within 12 months before or on the index date, if feasible Note, the variable is not reported in Flatiron Health	Additional covariate
Charlson Comorbidity Index (CCI)	Within 12 months before or on the index date. CCI is reported for RW patients. For patients from Study C1071003, CCI will be derived from MedDRA classification[8]	Additional covariate
Impaired cardiovascular	Yes, No	Baseline characteristic to describe

Variable	Operational definition	Role
function	Within 12 months before or on the index date. See Section 6.5.4 for the definition of the impaired cardiovascular function for external control arms, not applicable for participants of Study C1071003	the treatment group, cannot be used as a covariate because this was an exclusion criterion in Study C1071003.
Basal cell or squamous cell skin cancer, or carcinoma *in situ*	Yes, No Before or on the index date.	Additional covariate
Adequate hepatic function	Normal, Impaired Within 90 days before or on the index date, if feasible. See Section 6.5.4 for the definition of adequate hepatic function for external control arms, not applicable for participants of Study C1071003	Baseline characteristic to describe the treatment group, cannot be used as a covariate because this was an inclusion criterion in Study C1071003.
Adequate renal function	Within 90 days before or on the index date, if feasible. See Section 6.5.4 for the definition of adequate renal function for external control arms, not applicable for participants of Study C1071003	Baseline characteristic to describe the treatment group, cannot be used as a covariate because this was an inclusion criterion in Study C1071003.
Adequate BM function	Yes, No Within 90 days before or on the index date, if feasible See Section 6.5.4 for the definition of adequate BM function for external control arms, not applicable for participants of Study C1071003	Baseline characteristic to describe the treatment group, cannot be used as a covariate because this was an inclusion criterion in Study C1071003.
Aspartate aminotransferase (microkat/L)	Within 90 days before or on the index date	Additional covariate and to define variable “adequate hepatic function”
Alanine aminotransferase (microkat/L)	Within 90 days before or on the index date	Additional covariate and to define variable “adequate hepatic function”
Creatinine clearance (mL/min)	Within 90 days before or on the index date	Additional covariate and to define variable “adequate renal function”
ANC ($\times 10^9$/L)	Within 90 days before or on the index date	Additional covariate and to define variable “adequate BM function”
Platelets ($\times 10^9$/L)	Within 90 days before or on the index date	Additional covariate and to define variable “adequate BM function”
Hemoglobin (g/L)	Within 90 days before or on the index date	Additional covariate and to define variable “adequate BM function”
Billirubin (mg/dL)	Within 90 days before or on the index date	Additional covariate to define variable “adequate hepatic function”
Calcium in serum or plasma (mmol/L)	Within 90 days before or on the index date	Additional covariate
Lactate dehydrogenase (U/L)	Within 90 days before or on the index date	Additional covariate
Beta-2 microglobulin (mg/L)	Within 90 days before or on the index date	To derive variable “ISS”
Serum albumin (g/dL)	Within 90 days before or on the index date	Additional covariate To derive variable “ISS”

Based on the distribution and proportion of missing values, continuous laboratory values may be categorized to limit the influence of outliers or to preserve sample size.

In efforts to preserve sample size while accounting for measured factors that were deemed a priori to be likely sources of covariate imbalance during modeling, essential covariates may be categorized if missingness exceed 10% for a given data source.

A graphical approach may be employed for each continuous covariate with substantial missingness to assess the relationship between values of the covariate and the values of the propensity score. The resulting plots will be used to identify data-driven cut points based on means, tertiles or quartiles of the distribution to better capture the functional form of the relationship. Additional analyses incorporating these continuous covariates without categorization will also be performed, and in cases where substantially improved balance is possible, the latter approach will be preferred.

6.5.3. Definition of the International Staging System (ISS)

ISS is a risk stratification algorithm that groups MM patients by their survival prognosis. ISS is defined based on the values of 2 lab tests: beta-2 microglobulin and serum albumin.[9]

Values for beta-2 microglobulin and serum albumin measured within 90 days before or on the index date will be used.

ISS stage will be defined as described in Table 4.

Table 4. ISS Definition

Stage	Lab test values
I	Serum beta-2 microglobulin <3.5 mg/L Serum albumin ≥ 3.5 g/dL
II	Not ISS stage I or III
III	Serum beta-2 microglobulin >5.5 mg/L

6.5.4. Definition of Other Covariates

Table 5 gives the definition for impaired cardiovascular function and adequate BM, renal and hepatic functions for patients from RW external control arms.

Table 5. Definition for Impaired Cardiovascular Function and Adequate BM, Renal and Hepatic Functions.

Covariate	Definition for external control arms
Impaired cardiovascular function	Impaired cardiovascular function or clinically significant cardiovascular diseases, defined as any of the following conditions within 12 months before the index date, inclusive. Flatiron Health a) Acute myocardial infarction or acute coronary syndromes b) Cardiac arrhythmias c) Thromboembolic or cerebrovascular events COTA Reported myocardial infarction, congestive heart failure, or cerebrovascular disease.
Adequate renal function	Yes/No Adequate renal function measured within 90 days before or on the index date, will be defined as an estimated creatinine clearance ≥30 mL/min (according to the Cockcroft-Gault formula, National Kidney Foundation https://www.kidney.org/professionals/KDOQI/gfr_calculatorCoc).
Adequate hepatic function	Yes/No Adequate hepatic function measured within 90 days before or on the index date. Characterized by the following: a) Total bilirubin ≤2 x upper limit of normal (ULN) (≤3 x ULN if documented Gilbert's syndrome), b) Aspartate aminotransferase (AST) ≤2.5 x ULN, and c) Alanine aminotransferase (ALT) ≤2.5 x ULN
Adequate BM function	Within 90 days before or on the index date, if feasible. As following: a) Absolute neutrophil count ≥1.0 × 10^9/L, b) Platelets ≥25 × 10^9/L, and c) Hemoglobin ≥8 g/dL

Abbreviations: ALT=alanine transaminase; AST=aspartate transaminase; BM=bone marrow; ULN=upper limit of normal.

7. HANDLING OF MISSING VALUES

Missingness in dates - partial handling of dates

Exact index dates will be required, without any missing components (eg, the day component), and no specific handling of the partial index date is needed.

Any missing day component for the date of death will be imputed as if it had occurred on the 15th of the corresponding month. For patients with a last record that is identified in the same month as the date of death, the date of the last record will be imputed as the date of death.

For the baseline characteristics listed in Section 6.5, a missing day component will be imputed as the last day of the corresponding month to decide whether the measurement lies within the baseline period for qualifying as the baseline value. A missing month component will not be imputed, but if it is clear from the year component that the covariate falls into the baseline period, the measurement will be taken into consideration. A missing year component will not be imputed.

The strategy of handling missing values for the baseline characteristics will be decided after investigation of the proportions of missing values in external control arms.

Missingness in measured laboratory values

Missing lab measurements will not be imputed. However, if a lab measurement is recorded, but the ULN is missing, it will be imputed using last observation carried forward (LOCF). If imputation by LOCF is not possible, the missing ULN will be imputed by sex-stratified mean value.

8. STATISTICAL METHODOLOGY AND STATISTICAL ANALYSES

8.1. Summary of the Analyses

All statistical analyses described in this section will be carried out separately for each combined study dataset. Table 6 provides a summary of analyses by external control arm sample.

Table 6. Summary of Statistical Analyses by Study Sample.

	Study datasets based on the critical eligibility criteria RW sample	***Study datasets based on the expanded eligibility criteria RW sample***
Primary analyses	*Descriptive statistics of elranatamab (Cohort A) and SOC-treated patients* *<u>Primary objective</u>* *Naïve comparison of ORR in elranatamab (Cohort A) versus SOC-treated patients* *IPTW comparison of ORR in elranatamab (Cohort A) versus SOC-treated patients* *<u>Secondary objectives</u>* *Naïve comparison of TTR, and DOR in elranatamab (Cohort A) versus SOC-treated patients* *IPTW comparison of TTR, and DOR in elranatamab (Cohort A) versus SOC-treated patients*	
Sensitivity analyses	*Doubly robust comparisons of ORR in elranatamab (Cohort A) versus SOC-treated patients* *Quantitative bias assessments when comparing ORR (using raw, IPTW, doubly robust estimates), and DOR (using raw and IPTW) in elranatamab (Cohort A) versus SOC-treated patients*	*Descriptive statistics of elranatamab (Cohort A) and SOC-treated patients* *Naïve comparison of ORR, TTR, and DOR in elranatamab (Cohort A) versus SOC-treated patients* *IPTW comparison of ORR, TTR, and DOR in elranatamab (Cohort A) versus SOC-treated patients* *Doubly robust comparisons of ORR in elranatamab (Cohort A) versus SOC-treated patients*
Alternative analyses	Descriptive statistics of all elranatamab (Cohort A and Cohort B) and SOC-treated patients Naïve, IPTW, and doubly robust comparisons of ORR in all elranatamab (Cohort A and Cohort B) versus SOC-treated patients Naïve and IPTW comparisons of TTR, and DOR in all elranatamab (Cohort A and Cohort B) versus SOC-treated patients	

In this study, measures will be taken to reduce potential bias. Additionally, an analysis firewall will be implemented throughout different stages and processes of this study such that the PS balancing steps are conducted independently of and before knowledge of the outcome, where possible.

8.2. Statistical Methods

8.2.1. Descriptive Statistics

Tabular summaries of baseline patient demographic and clinical characteristics for each analysis set and treatment arm will be provided. Summary statistics will include frequencies and percentages (categorical variables) and mean/median, minimum/maximum with standard deviation/interquartile range (continuous variables).

For reporting conventions, mean, median, and SD will be rounded to 1 decimal place. Percentages will be rounded to 1 decimal place.

Any p-value will be reported with 4 decimal places and values below 0.0001 will be reported as "p<0.0001".

8.2.2. Propensity Scores and Inverse Probability of Treatment Weighting

In analyses of RWD, an important consideration in the identification of potentially causal effects is the control for confounding. IPTW is a well-established method for causal inference in nonrandomized studies. Estimating the propensity score (PS) is a form of dimensionality reduction, in which several individual characteristics relevant to treatment assignment and the outcome, or those related only to outcome development are used to estimate the conditional probability that the patient is assigned to a given treatment. IPTW is a quasi-experimental approach used to create a pseudo-population in which the covariates are independent of the treatment assignment, thereby permitting an unbiased estimate of the average treatment effect (ATE).

The PS will be estimated using logistic regression models, where the dependent variable is a binary indicator of the treatment arm (elranatamab versus SOC). Covariates in the logistic regression will include the priority covariates described in Section 6.5.1 and, optional, some additional covariates. Diagnostics of the estimated PS will be applied by examining the distribution of the propensity scores in each treatment group, the area of PS overlap, or any outliers. The estimated PS will be used to generate IPTWs; the weights may be subsequently stabilized and/or truncated, after review of diagnostic plots and summary statistics. *A standardized mean difference (SMD) of ≥10% will be used to identify potential residual confounding and as an indicator of covariate imbalance between the treatment arms requiring further investigation.[10] Grossly higher SMDs (over 25%) may lead to refining the logistic model for deriving the weights.*[11,12].

Assumptions

Observational research relies on methods for causal inference when deriving unbiased estimates. Hernán and Robins (13) list 3 conditions for the valid use of causal inference methods such as IPTW: conditional exchangeability, positivity, and consistency.

1. The conditional exchangeability assumption allows an observational study to be conceptualized as a conditionally randomized trial, where the probability to receive the treatment is depending on the covariates, but not on unmeasured variables.

   Essentially, this assumption leads to the postulation of no unmeasured confounding. Potential effects of unmeasured confounding will be checked in a sensitivity analysis as described in Section 8.2.8.

2. The positivity assumption specifies that conditional on the covariates, every patient has a probability > 0 to receive either treatment.

   As this is a carefully designed study with inclusion and exclusion criteria derived from the experimental arm of the clinical trial, positivity is expected to be a reasonable assumption. However, positivity will be assessed by checking the distribution of PS by treatment group (elranatamab or SOC). If extreme PSs are observed, the reasons will be investigated (for example, resulting from some high values of a specific covariate) and if necessary, covariate value ranges might be harmonized to improve positivity (14). Any such action will be documented in the study report.

   The PS distributions of both treatment groups will be evaluated graphically using density plots.

3. The assumption of treatment consistency specifies that there is no ambiguity in defining a treatment. This assumption is also known as “treatment is well-defined”.

   If feasible, this assumption will be checked by comparisons against those individual SOC treatments, where the sample size is sufficiently high (at least 50 patients available for 1 specific SOC). These SOC treatments should have similar treatment effects to fulfill the assumptions of SOC treatment consistency.

The complete overlap of propensity score distributions does not constitute a necessary assumption for being able to estimate the ATE by weighting methods. The PS distributions of both treatment groups will be displayed graphically to visually inspect the range of overlap.

PS model building

SAS PROC LOGISTIC will be used to estimate PSs. The PSs will be estimated as the probability of initiating elranatamab versus SOC conditional on patients’ characteristics measured at baseline. The PS model will include all the priority covariates described in Section 6.5.1, which are sufficiently available (eg, missingness <10%). Given the adequate sample size, some of the additional covariates described in Section 6.5.2 may be included in the analyses.

The decision to keep each additional covariate in the PS model will be based on the resulting change in SMD following inclusion of this variable in the model, as well as the magnitude of its influence on the estimated propensity scores. If covariates are deemed to be particularly relevant given their role as inclusion or exclusion criteria for the trial, they may be

incorporated into the model even if their effect on SMDs is determined to be modest. In situations where covariates are highly correlated, or when categorical covariates include sparse strata, non-convergence or model instability may result.
As a mitigation strategy, covariates that result in non-convergence may be omitted from the model. All these measures will be described in the study report.

Estimation of IPTW

IPTWs will be estimated for patients initiating elranatamab as the inverse of the propensity score (IPTW=1/propensity score). For patients initiating SOC, the IPTW weights will be estimated as the inverse of 1 minus the estimated propensity score (IPTW=1/(1-propensity score)).

To reduce the influence of extreme values of the estimated IPTW on the standard error size, the weights will be stabilized by the inclusion of a numerator in the IPTWs, which will be an overall probability of being treated with elranatamab for patients from Study C1071003 or the overall probability of being treated with SOC for RW patients.[15] When treatment=elranatamab, the stabilized IPTW=P*IPTW, and when treatment=SOC, the stabilized IPTW=(1-P)*IPTW, where P is the probability of treatment with elranatamab without considering covariates.[16] The distribution of the estimated stabilized IPTW will be evaluated.

If extreme weights occur, truncation will be used as needed to address potential variance inflation. The threshold will be carefully selected in full consideration of the bias-variance tradeoff inherent in weight truncation.[17] The impact of truncation at different levels (eg, 99th percentile, 95th percentile) on the overall weight distribution will be explored.

Balance assessment

To assess the balance that is produced by applying a PS method, SMDs for the differences in the distribution of covariates across treatment groups will be assessed.

SAS macro STDDIFF.SAS will be used to estimate SMD.[18]

Standardized mean differences will be estimated for means (continuous variables) and prevalence (dichotomous variables) of each covariate and used to assess imbalances in population characteristics between elranatamab and SOC-treated patients. Counts, SMDs, and p-values for both the unweighted and weighted samples (via IPTW) will be reported. The SMD is preferred over p-values because of its robustness to sample size.[19]

A standardized mean difference (SMD) of ≥20% will be used as an indicator of covariate imbalance between the treatment arms requiring further investigation.[11,12,20]

8.2.3. Naïve Comparison of ORR

For each treatment group, number, and percentage of patients with overall response (composite) will be reported. Frequencies will also be reported for each response category: sCR, CR, VGPR and PR (when using COTA) or those with at least VGPR and PR (when using Flatiron Health).

ORR for each treatment group will be estimated as the number of patients with OR over the total number of patients in the RW TCR cohort. The 2-sided exact Clopper-Pearson (21) 95% CI will be estimated and reported for the ORR.

ORR will be compared between treatment groups using an unadjusted log-binomial model. The risk ratio will be reported together with corresponding 95% CIs. The analysis will be carried out using SAS PROC GENMOD.

8.2.4. Naïve Comparison of TTR and DOR

TTR will be measured only in patients with the OR and no censoring will be applied. TTR will be summarized by treatment group, using descriptive statistics. The median time, 25th and 75th percentiles will be reported. Sign test will be used to compare median TTR in patients treated with elranatamab versus those treated with SOC.[22] The analysis will be performed using SAS PROC UNIVARIATE.

DOR will be analyzed using Kaplan-Meier (KM) methods. Estimates will be plotted. The median time, 25th and 75th percentiles will be reported. The 95% CI for the median will be estimated using the Brookmeyer and Crowley method.[23] Click or tap here to enter text. and the 95% CI for the 25th percentile will be estimated via the Klein and Moeschberger method.[24]

DOR will be compared between treatment groups using hazard ratios (HR) estimated from unadjusted Cox proportional hazard models. The analysis will be carried out using SAS PROC PHREG.

The proportional hazards assumption will be checked, and in case of deviations, a restricted mean survival time model will be applied instead of the Cox proportional hazards regression model.

When the proportional hazards assumption is met, the log-cumulative hazard plot should show a constant HR over time. If moderate deviations of the proportional hazards' assumption are observed, the HR will be interpreted as a weighted average of the HR over the follow-up period and 95% CIs will be obtained via bootstrapping.[25] If severe violations are observed, a weighted restricted mean survival time model will be applied (using SAS PROC RMST).
Linearity of the relationship between the log-hazard and the treatment will be assessed by plotting the Martingale residuals.[26] Deviance residuals will also be plotted to assess the presence of influential observations (ie, check for outliers). The presence of any outliers will be noted and addressed, if appropriate.

8.2.5. Weighted Analysis for ORR

The dichotomous outcome of ORR will be compared between treatment groups using a weighted log-binomial model. Risk ratio estimates will be reported together with corresponding 95% CIs. The model convergence will be checked to ensure that the estimated probabilities are bounded between [0,1].

The analysis will be performed using SAS PROC GENMOD. Robust standard errors will be estimated using the REPEATED SUBJECT statement; the WEIGHT statement will be used to incorporate IPTW.[27-29]

8.2.6. Weighted Analyses for TTR and DOR

These analyses will be conducted only in patients with the OR.[30] TTR will be summarized in weighted sample by treatment group, using descriptive statistics. The median time, 25th and 75th percentiles will be reported. Sign test will be used to compare median TTR in patients treated with elranatamab versus those treated with SOC.[22] The analysis will be performed using SAS PROC UNIVARIATE with WEIGHT statement to implement IPTW.

DOR will be compared between treatment groups using hazard ratios (HR) estimated from weighted Cox proportional hazard models. The proportional hazards assumption will be checked, and in case of deviations, a restricted mean survival time model will be applied instead of the Cox proportional hazards regression model.

Kaplan-Meier figures will be used to visually examine weighted survival probabilities over the follow-up period.

The analysis will be carried out using SAS PROC PHREG with the WEIGHT statement to implement IPTW. The COVS (AGGREGATE) option and ID statements will be used to obtain model estimates with corresponding robust standard errors.[31]

8.2.7. Doubly Robust Estimation for ORR

Analyses employing a doubly robust estimator will require fitting 2 models:

(1) Model for treatment or exposure status

(2) Model for the outcome of interest

As with other causal inference methods, valid and unbiased estimates require assumptions of no unmeasured confounding (exchangeability), positivity (the experimental treatment assumption), no interference, and consistency.[32] Assuming these assumptions are upheld, if 1 of these 2 models is correctly specified, the other can be misspecified, and the resulting doubly robust estimates may remain consistent and unbiased.[33]

The augmented inverse probability weighting (AIPW) approach will be used to estimate a PS model for weighting, and subsequently augment the inverse probability weights using predicted values from the outcome model. The AIPW estimator incorporates an adjustment term that stabilizes the estimator when the propensity scores get close to zero or one (34). *The general procedure for estimation of PS, IPTW, and assessment of resulting improvements in covariate balance is provided in Section 8.2.2.*

Standard errors and confidence intervals for the doubly robust estimators will be obtained via a robust sandwich-type estimator of variance.[35] Diagnostics will be applied to assess whether the doubly robust model appears to be well-specified. Covariate density plots and SMD will be used to compare balance on covariates before and following AIPW.[34]

SAS PROC CAUSALTRT will be used to estimate the ORR with robust SE.

8.2.8. Nullification Analysis

Nullification analysis will be applied to assess the potential influence of unmeasured confounding on the observed associations.

Statistical software package to use for estimation of E-values

A sensitivity analysis will be conducted to assess the potential influence of unmeasured confounders on all estimands described in the primary analyses. The EValue package in R will be used to estimate E-values for ORR and DOR, which will quantify the minimum strength of association on the risk ratio scale that an unmeasured confounder must exhibit with both treatment status and outcome, given measured covariate values, to nullify an observed association between treatment and outcome.[36-38]

Calculation of E-values for the primary endpoint estimates and 95% CI

Table 7. Calculation of the E-value

Effect Measure	Calculation of Approximate E-value
HR (or risk ratio) for rare outcomes	$E-value = HR + \sqrt{HR \times (HR-1)}$
HR (or risk ratio) for common outcomes	When the outcome is common (>15% at the end of follow-up), an E-value may be obtained by applying the following approximation: $E-value = (1-0.5 \times \sqrt{HR})/(1-0.5 \times \sqrt{1/HR})$

Citation: Vanderweele & Ding, 2017.[38]

Rather than estimating a confidence interval for the E-value directly, the analyst may consider statistical uncertainty in the approximate E-value for a given measure of association by estimating a second E-value for the corresponding 95% CI. For CIs corresponding to ratio measures that contain the null (1.0), the E-value for the CI is also 1.0. If the CI does not contain the null, the analyst may compute the E-value for the 95% CI by determining which bound is closest to the null using the following formulas, where LL represents the lower limit of the 95% CI and UL the upper limit:

- If LL ≤ 1, then E-value = 1; if LL > 1, then E-value = $LL + \sqrt{(LL \times LL - 1)}$
- If UL ≥ 1, then E-value = 1; if UL < 1, then let UL* = 1/UL and E-value = $UL^* + \sqrt{(UL^* \times (UL^* - 1))}$

Interpretation of the E-Value

Given E-values obtained for the estimands specified in the primary analyses, the magnitude of the E-value corresponds to the minimum magnitude of residual confounding required to explain an estimated ORR or DOR. For these analyses, an E-value of 2.0, for example, will be interpreted to mean that the odds ratio or HR for the association between a residual

confounder and both treatment and outcome would need to be 2.0 or greater to explain the observed ORR or DOR.[38] For unmeasured confounders with a weaker association with treatment and outcome, the E-value provides support for the hypothesis that the observed association cannot be nullified by unmeasured confounding alone.

Selection of suspected unmeasured confounders

A short list of suspected unmeasured confounders will be identified; these are variables that were observed in the trial, but not available (or only with substantial missingness) in the RWD. The E-value will be computed and reflects the minimum strength of association the confounder would need to have with the exposure and outcome, conditional on the measured covariates, to fully explain away the observed treatment effect.[38]

8.3. Statistical Analyses

8.3.1. Safety Analyses

Safety data is not evaluated in this study as it is unrelated to the study objectives.

8.3.2. Analyses of ORR

The primary analysis will be performed using safety analysis set (analysis set with patients from Cohort A) as described in Section 8.2.3 and Section 8.2.5.
The sensitivity analysis employing a doubly robust estimator will be conducted using the safety analysis set as described in Section 8.2.7.

8.3.3. Analyses of TTR and DOR

The primary analysis will be performed using the safety analysis set (analysis set with patients from Cohort A) as described in Section 8.2.4 and Section 8.2.6.

8.3.4. Sensitivity Analysis Based on RW Sample Identified Using Expanded Eligibility Criteria

To evaluate robustness of primary analysis results, a sensitivity analysis will be conducted using the analysis sets that include patients from Study C1071003 Cohort A who received at least 1 dose of elranatamab and RW patients selected using expanded eligibility criteria. Analyses of ORR will be conducted as described in Section 8.2.3, Section 8.2.5, and Section 8.2.7. Analyses of DOR and TTR will be conducted as described in Section 8.2.4 and Section 8.2.6.

8.3.5. Alternative Analysis

This alternative analysis will be performed using all patients from Study 1071003 (ie, Cohort A and Cohort B) and RW patients using critical eligibility criteria.

Analyses of ORR will be conducted as described in Section 8.2.3, Section 8.2.5, and Section 8.2.7. Analyses of DOR and TTR will be conducted as described in Section 8.2.4 and Section 8.2.6.

8.3.6. Quantitative Bias Assessments

Nullification analysis will be conducted to assess the potential influence of unmeasured confounding. This sensitivity analysis will be applied for the primary analysis only.

This analysis will take place as outlined in Section 8.2.8.

8.4. Software

SAS 9.4 or higher or R v4.2.0 or higher (nullification analysis) will be used for the analyses.

9. REFERENCES

1. Framework for FDA's Real-World Evidence Program. 2018 [cited 2022 Jun 29]; Available from: www.fda.gov.

2. ICH E10 Choice of control group in clinical trials | European Medicines Agency [Internet]. 2001 [cited 2022 May 24]. Available from: https://www.ema.europa.eu/en/ich-e10-choice-control-group-clinical-trials.

3. Rare Diseases: Natural History Studies for Drug Development | FDA [Internet]. [cited 2022 May 24]. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/rare-diseases-natural-history-studies-drug-development.

4. Framework for FDA's Real-World Evidence Program. 2018 [cited 2022 May 24]; Available from: www.fda.gov.

5. Mikhael J. Treatment Options for Triple-class Refractory Multiple Myeloma. Clin Lymphoma Myeloma Leuk [Internet]. 2020 Jan 1 [cited 2022 May 24];20(1):1–7. Available from: https://pubmed.ncbi.nlm.nih.gov/31767529/.

6. Pfizer. MagnetisMM-3, an open-label, multicenter, non-randomized phase 2 study of elranatamab (pf-06863135) monotherapy in participants with MM who are refractory to at least one proteasome inhibitor (PI), one immunomodulatory drug, and one anti-CD38 antibody. Final Protocol Amendment 7, 11 November 2021. 2021.

7. No adjustments are needed for multiple comparisons - PubMed [Internet]. [cited 2022 Jul 17]. Available from: https://pubmed.ncbi.nlm.nih.gov/2081237/.

8. Putrik P, Ramiro S, Lie E, Michaud K, Kvamme MK, Keszei AP, et al. Deriving common comorbidity indices from the MedDRA classification and exploring their performance on key outcomes in patients with rheumatoid arthritis. Rheumatology (Oxford) [Internet]. 2018 Mar 1 [cited 2022 Sep 6];57(3):548–54. Available from: https://pubmed.ncbi.nlm.nih.gov/29272517/.

9. Palumbo A, Avet-Loiseau H, Oliva S, Lokhorst HM, Goldschmidt H, Rosinol L, et al. Revised International Staging System for Multiple Myeloma: A Report From International Myeloma Working Group. Journal of Clinical Oncology [Internet]. 2015 Sep 9 [cited 2022 Jun 11];33(26):2863. Available from: /pmc/articles/PMC4846284/.

10. Austin P. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. Stat Med [Internet]. 2009 [cited 2022 Jul 17];28:3083–107. Available from: www.interscience.wiley.com.

11. Rubin DB. Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation. Health Services and Outcomes Research Methodology 2001 2:3 [Internet]. 2001 [cited 2022 Oct 20];2(3):169–88. Available from: https://link.springer.com/article/10.1023/A:1020363010465.

12. Stuart EA. Matching methods for causal inference: A review and a look forward. Stat Sci [Internet]. 2010 Feb 2 [cited 2022 Oct 20];25(1):1. Available from: /pmc/articles/PMC2943670/.

13. Hernán MA, Robins JM (2020). Causal Inference: What If. Boca Raton: Chapman & Hall/CRC.

14. Petersen ML, Porter KE, Gruber S, Wang Y, van der Laan MJ. Diagnosing and responding to violations in the positivity assumption.

15. Austin PC. Variance estimation when using inverse probability of treatment weighting (IPTW) with survival analysis. Stat Med [Internet]. 2016 Dec 30 [cited 2022 Sep 27];35(30):5642–55. Available from: https://pubmed.ncbi.nlm.nih.gov/27549016/.

16. Xu S, Ross C, Raebel MA, Shetterly S, Blanchette C, Smith D. Use of stabilized inverse propensity scores as weights to directly estimate relative risk and its confidence intervals. Value Health [Internet]. 2010 [cited 2022 Sep 27];13(2):273–7. Available from: https://pubmed.ncbi.nlm.nih.gov/19912596/.

17. Crowson CS, Schenck LA, Green AB, Atkinson EJ, Therneau TM. The Basics of Propensity Scoring and Marginal Structural Models. 2013.

18. Yang D, Dalton JE. A unified approach to measuring the effect size between two groups using SAS ®. :335–2012.

19. Austin PC. Using the standardized difference to compare the prevalence of a binary variable between two groups in observational research. Commun Stat Simul Comput. 2009 Jun;38(6):1228–34.

20. Mccaffrey DF, Griffin BA, Almirall D, Slaughter ME, Ramchand R, Burgette LF. A tutorial on propensity score estimation for multiple treatments using generalized boosted models. Stat Med [Internet]. 2013 Aug 30 [cited 2022 Oct 31];32(19):3388–414. Available from: https://pubmed.ncbi.nlm.nih.gov/23508673/.

21. Clopper CJ, Pearson ES. The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial. Biometrika. 1934 Dec;26(4):404.

22. PROC UNIVARIATE: Tests for Location: Base SAS(R) 9.3 Procedures Guide: Statistical Procedures [Internet]. [cited 2022 Oct 20]. Available from: https://support.sas.com/documentation/cdl/en/procstat/63963/HTML/default/viewer.htm#procstat_univariate_sect029.htm.

23. R, Crowley J. A Confidence Interval for the Median Survival Time. Biometrics. 1982 Mar;38(1):29.

24. Klein JP, Moeschberger ML. Survival analysis: techniques for censored and truncated data. 2003 [cited 2022 May 1];536. Available from: https://books.google.com/books/about/Survival_Analysis.html?id=aO7xBwAAQBAJ.

25. Stensrud MJ, Hernán MA. Why Test for Proportional Hazards? JAMA [Internet]. 2020 Apr 14 [cited 2022 Jun 11];323(14):1401–2. Available from: https://pubmed.ncbi.nlm.nih.gov/32167523/.

26. Hosmer DW, Lemeshow S, May S. Applied Survival Analysis: Regression Modeling of Time to Event Data: Second Edition. Applied Survival Analysis: Regression Modeling of Time to Event Data: Second Edition [Internet]. 2011 Oct 17 [cited 2022 Jun 11];1–401. Available from: https://onlinelibrary.wiley.com/doi/book/10.1002/9780470258019.

27. Robins JM, Hernán MÁ, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology [Internet]. 2000 [cited 2022 Oct 16];11(5):550–60. Available from: https://pubmed.ncbi.nlm.nih.gov/10955408/.

28. Faries DE, Leon AC, Haro JMaria, Obenchain RL, SAS Institute. Analysis of observational health care data using SAS. 2010;436.

29. Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC. [Internet]. 2020 [cited 2022 Oct 17]. Available from: https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/.

30. Jagannath S, Lin Y, Goldschmid.t H, Reece D, Nooka A, Senin A, et al. KarMMa-RW: comparison of idecabtagene vicleucel with real-world outcomes in relapsed and refractory multiple myeloma. Blood Cancer J [Internet]. 2021 Jun 1 [cited 2022 Oct 20];11(6). Available from: https://pubmed.ncbi.nlm.nih.gov/34145225/.

31. Austin PC. Variance estimation when using inverse probability of treatment weighting (IPTW) with survival analysis. Stat Med [Internet]. 2016 Dec 12 [cited 2022 Jun 9];35(30):5642. Available from: /pmc/articles/PMC5157758/.

32. Funk MJ, Westreich D, Wiesen C, Stürmer T, Brookhart MA, Davidian M. Doubly Robust Estimation of Causal Effects. Am J Epidemiol. 2011 Apr 1;173(7):761–7.

33. Bang H, Robins JM. Doubly Robust Estimation in Missing Data and Causal Inference Models. Biometrics. 2005 Dec;61(4):962–73.

34. Glynn AN, Quinn KM. An Introduction to the Augmented Inverse Propensity Weighted Estimator. Political Analysis [Internet]. 2010 Dec 14 [cited 2022 Oct 31];18(1):36–56. Available from: https://www.cambridge.org/core/journals/political-analysis/article/abs/an-introduction-to-the-augmented-inverse-propensity-weighted-estimator/4B1B8301E46F4432C4DCC91FE20780DB.

35. Lin DY, Wei LJ. The Robust Inference for the Cox Proportional Hazards Model. J Am Stat Assoc. 1989 Dec;84(408):1074.

36. Smith LH, Vanderweele TJ. Bounding Bias Due to Selection. Epidemiology [Internet]. 2019 Jul 1 [cited 2022 Jun 11];30(4):509. Available from: /pmc/articles/PMC6553568/.

37. Mathur MB, VanderWeele TJ. Sensitivity Analysis for Unmeasured Confounding in Meta-Analyses. J Am Stat Assoc [Internet]. 2020 Jan 2 [cited 2022 Jun 11];115(529):163. Available from: /pmc/articles/PMC7518377/.

38. van der Weele TJ, Ding P. Sensitivity Analysis in Observational Research: Introducing the E-Value. Ann Intern Med [Internet]. 2017 Aug 15 [cited 2022 Jun 11];167(4):268–74. Available from: https://pubmed.ncbi.nlm.nih.gov/28693043/.

LIST OF TABLES

LIST OF FIGURES

Appendices

Appendix A. List of Treatments Available for MM

Treatment	Class
bendamustine	Alkylating agent
cisplatin	Alkylating agent
cyclophosphamide	Alkylating agent
melphalan	Alkylating agent
melphalan flufenamide	Alkylating agent
adriamycin	Anthracycline
idarubicin	Anthracycline
liposomal doxorubicin (Caelyx ®/Myocet ®)	Anthracycline
ADC	Anti-BCMA (ADC)
belantamab mafodotin (Blenrep ®)	Anti-BCMA (ADC)
WVT078	Anti-BCMA (bispecific)
BsAb	Anti-BCMA (bispecific)
CAR-T	Anti-BCMA (CAR-T)
idecabidecabtagene vicleucel (Abecma ®)	Anti-BCMA (CAR-T)
other anti-BCMA	Anti-BCMA (other)
venetoclax (Venclexta ® or Venclyxto ®)	BCL2 inhibitor
daratumumab (Darzalex ®)	CD38-directed mAb
dexamethasone	Corticosteroid
prednisone	Corticosteroid
panobinostat	HDAC
lenalidomide (Revlimid ®)	IMiD
pomalidomide (Pomalyst ® or Imnovid ®)	IMiD
thalidomide	IMiD
elotuzumab (Empliciti ®)	MAb
isatuximab (Sarclisa ®)	MAb
selinexor (Xpovio ®/Nexpovio ®)	Nuclear export inhibitor
bortezomib (Velcade ®)	PI
carfilzomib (Kyprolis ®)	PI
ixazomib (Ninlaro ®)	PI
etoposide	Podophyllotoxin Derivative
vincristine/leurocristine (Oncovin ®)	Vinca Alkaloid