NON-INTERVENTIONAL STATISTICAL ANALYSIS PLAN FOR SECONDARY DATA COLLECTION STUDY

VERSION HISTORY

Version	Effective Date	Change Type (New, Revise, Admin)	Summary of Revisions
1.0		New	

Non-Interventional Study Protocol C1071031

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

Authors: PPD, PhD, STATLOG Inc.

PPD, MPH, STATLOG Inc.

Date: 15 May 2023

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Abbreviation	Term
ADC	Antibody-Drug Conjugate
AE	Adverse Event
AIPW	Augmented Inverse Probability Weights
ALT	Alanine Aminotransferase
Anti-CD38	Anti-CD38 monoclonal antibodies
AST	Aspartate Aminotransferase
ATE	Average Treatment Effect
ATT	Average Treatment Effect Amongst Treated
BCMA	B-Cell Maturation Antigen
BM	Bone Marrow
CAR	Chimeric Antigen Receptor
CI	Confidence Interval
DOR	Duration of Response
ECOG	Eastern Cooperative Oncology Group
EHR	Electronic Health Record
EMA	European Medicines Agency
EORTC QLQ-C30	European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Version 3.0
EORTC QLQ-CIPN20	European Organization for Research and Treatment of Cancer Quality of Life Questionnaire for Chemotherapy-Induced Peripheral Neuropathy
EORTC QLQ-MY20	European Organization for Research and Treatment of Cancer Multiple Myeloma Questionnaire
EQ-5D	European Quality of Life Five Dimension
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
HCP	Healthcare Provider
HIV	Human Immunodeficiency Virus
HRQOL	Health-Related Quality of Life
IMiD	Immunomodulatory Drug
IMWG	International Myeloma Working Group
IPTW	Inverse Probability of Treatment Weights
ISS	International Staging System
LOCF	Last Observation Carried Forward
LOT	Line of Therapy
LVEF	Left Ventricular Ejection Fraction

MCID	Minimally Clinically Important Difference
MM	Multiple Myeloma
MMRM	Mixed Effects Model With Repeat Measures
M-protein	Monoclonal Immunoglobulin Protein
MUGA	Multigated Acquisition Scan
OS	Overall Survival
PD	Progressive Disease
PGIC	Patient Global Impression of Change
PI	Proteasome Inhibitor
PFS	Progression-Free Survival
POEMS	Polyneuropathy, Organomegaly, Endocrinopathy, Monoclonal Gammopathy, and Skin Changes
PRO	Patient-Reported Outcomes
PS	Propensity Scores
QoL	Quality of Life
QT	Time from the beginning of the QRS Complex to the End of the T Wave
QTcF	Corrected QT (Fridericia Method)
REML	Restricted Maximum Likelihood
R-ISS	Revised International Staging System
RRMM	Relapsed/Refractory Multiple Myeloma
RW	Real-World
RWD	Real-World Data
SAP	Statistical Analysis Plan
SARS-Cov2	Severe Acute Respiratory Syndrome- Coronavirus 2
SLAMF7	Signaling Lymphocytic Activation Molecule Family Member 7
SMD	Standardized Mean Difference
SOC	Standard of Care
SPEP	Serum Protein Electrophoresis
TCR	Triple-Class Refractory
TOI	Time-to-clinically meaningful Improvement
TOW	Time-to-clinically meaningful Worsening
ULN	Upper Limit of Normal
UPEP	Urine Protein Electrophoresis
US	United States
VAS	Visual Analog Score

1. AMENDMENTS FROM PREVIOUS VERSION(S)

Not applicable

2. 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 C1071031, version 1.0, dated May 1, 2023.

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 C1071031, Version 1.0, dated May 1, 2023.
- 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 1 proteasome inhibitor, 1 immunomodulatory drug, and 1 anti-CD38 antibody. Protocol Amendment 9, 29 July 2022.

2.1. Study design

This retrospective cohort study will use patient-level data from single-arm clinical Study C1071003 and external control arms identified previously from RWD sources. To maximize comparability, the eligibility criteria for the participants from Study C1071003 have been applied to patients from the RWD sources (see Section 4.1 for more detail).

MM patients eligible for selection into external control arms are those patients who are refractory to at least 1 PI, 1 IMiD, and 1 anti-CD38 and have started at least 1 new treatment since the documentation of TCR status. According to International Myeloma Working Group (IMWG) criteria or clinical assessment, refractory is defined as having disease progression while on therapy or within 60 days of the last dose in any line of therapy (LOT) regardless of response. In Flatiron Health and COTA, if a subsequent LOT was initiated and a progression was observed after at least 30 days of the LOT start, the patient is considered refractory to the subsequent LOT even if the progression occurred within 60 days after the last dose of the preceding LOT. Details on the operational definitions of IMWG criteria for progression used in Flatiron Health, COTA, C1071013, and C1071014 and a side-by-side comparison with the criteria used in Study C1071003 are provided in Study Protocol C1071031.

In the RW setting, no single SOC currently exists for TCR MM patients, and combinations of treatments are frequently used instead 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.*

The date of initiation of the first regimen after TCR MM eligibility will be defined as the index date. For the comparative analysis of progression-free survival (PFS) and overall survival (OS), only patients with an index date occurring between 16 November 2015, and 30 June 2022 will be selected (the first anti-CD38 therapy was approved by the FDA on 16 November 2015) from COTA and Flatiron Health. CCI [redacted] *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 the latest available patient record, whichever comes first. Clinical outcomes of interest will be PFS and OS. For the exploratory analysis, PROs measured using self-administered questionnaires will be assessed.*

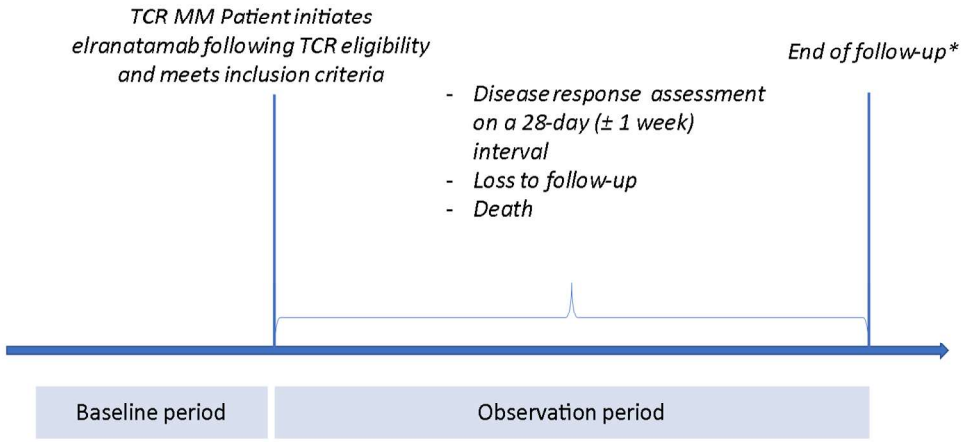
*In the main analysis comparing PFS and OS, differences in baseline characteristics (demographic characteristics, treatment history and disease-related characteristics) between patients in Study C1071003 and each external control arm will be balanced using IPTW (*see Section 7.2.5 for more detail*).*

Sensitivity analyses of PFS and OS will be conducted to evaluate the effect of alternative methods to address confounding and selection bias, including the doubly robust method, and using alternative inclusion/exclusion criteria. 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 7.2.6, Section 4.3, and Section 7.2.7).

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

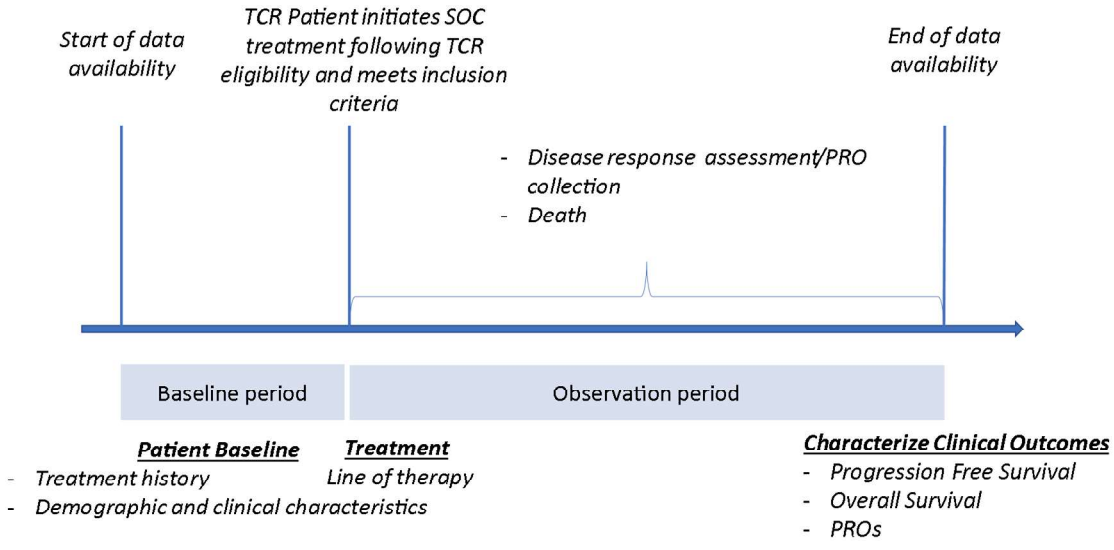
Figure 1. Baseline and observation periods in Study C1071003



- *For this protocol, all participants will be followed for about 15 months from the date of enrolment*

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

Figure 2. Baseline and observation periods in external control arms.



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

2.1.1. 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; 123 patients) and the other with patients previously exposed to BCMA-directed therapy (Cohort B; 64 patients). Because few patients were exposed to BCMA-directed therapy in our RW sources, the focus of the comparisons will be on Cohort A in Study C1071003, though additional analyses will include both Cohorts A and B.*

Populations of RW TCR MM patients

The cohorts of RW TCR MM patients for the external control arm will be identified from EHR databases Flatiron Health and COTA and prospective observational studies C1071013 and C1071014. These databases have been selected according to data availability and as fit-for-purpose for fulfilling study objectives. The eligible period for the index dates of patients from each RW database has been selected to align as closely as possible to each other and Study C1071003.

2.1.2. Data sources

Study C1071003

Study C1071003 is an open-label, multi-center, non-randomized Phase 2 study. (6). *The study aims to evaluate whether single-agent PF-06863135 (elranatamab) can provide clinical benefit in participants with RRMM who are refractory to at least 1 PI, 1 IMiD and 1 Anti-CD38. Elranatamab is a heterodimeric humanized full-length bispecific IgG2 kappa monoclonal antibody against BCMA and CD3* (6).

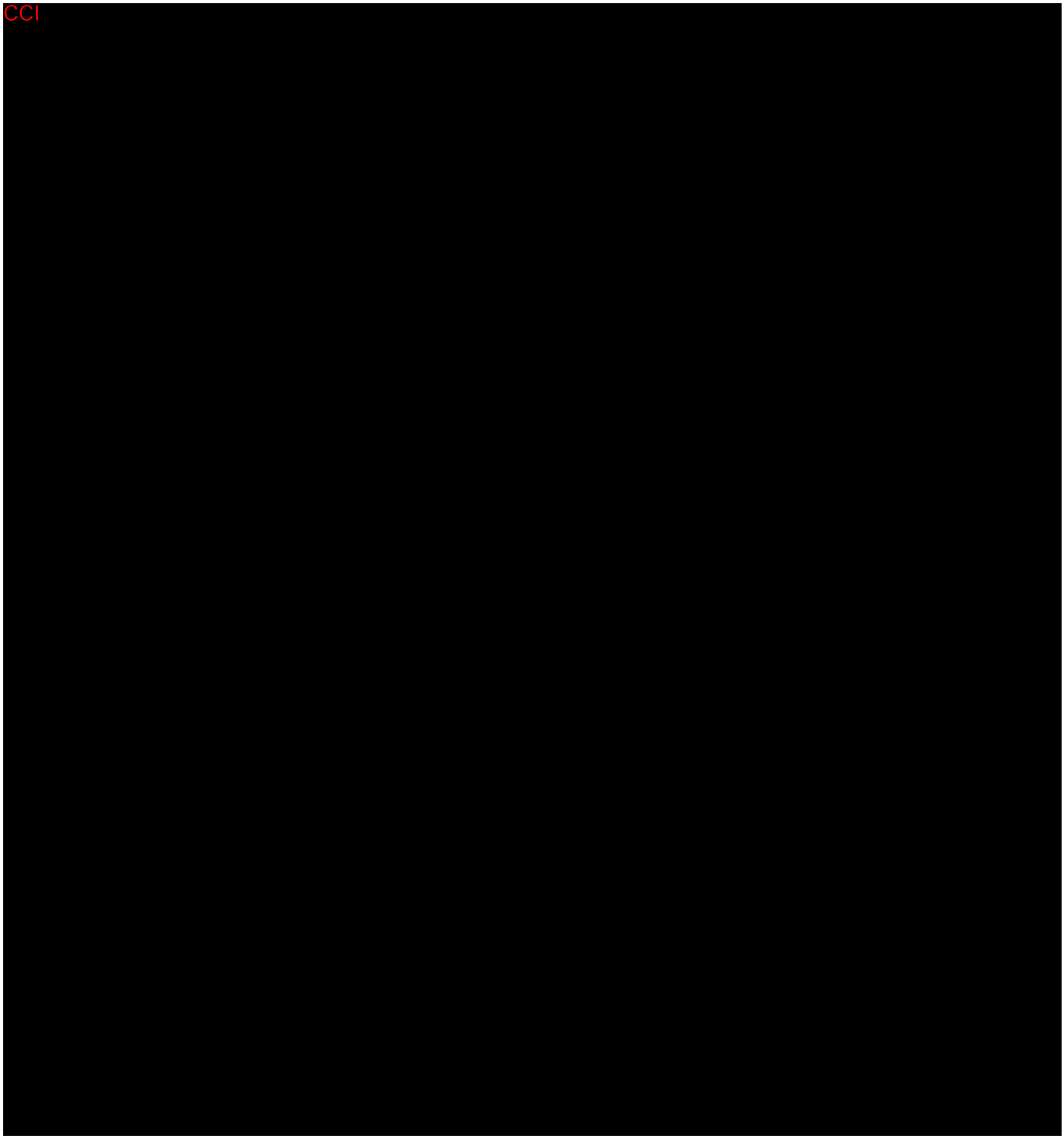
Flatiron Health

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.

CCI



2.1.3. Treatment/Cohort labels

The following treatment & cohort labels will be used:

- Study C1071003 Cohort A patients will be labeled as “Elranatamab”.
- Study C1071003 cohort of all patients (Cohort A and Cohort B) will be labeled as “Elranatamab with or without prior BCMA exposure”.
- External control arm patients will be labeled as “Standard of Care” or “SOC”.

2.2. Study objectives

Objectives	Endpoints
Primary	
1. To compare PFS among TCR MM patients treated with elranatamab in Study C1071003 with a comparable cohort of TCR MM patients receiving SOC therapy from the COTA database	PFS
2. To compare PFS among TCR MM patients treated with elranatamab in Study C1071003 with a comparable cohort of TCR MM patients receiving SOC therapy from the Flatiron Health database	PFS
Secondary	
1. To compare OS in TCR MM patients treated with elranatamab in Study C1071003 with a comparable cohort of TCR MM patients receiving SOC therapy from the COTA database	OS
2. To compare OS in TCR MM patients treated with elranatamab in Study C1071003 with a comparable cohort of TCR MM patients receiving SOC therapy from the Flatiron Health database	OS
Exploratory	
To compare the change of PROs in TCR MM patients treated with elranatamab in Study C1071003 with a comparable cohort of TCR MM patients receiving SOC therapy from the prospective observational studies C1071013 and C1071014	QLQ-C30 domains: Global Health Score Physical functioning Role functioning Emotional functioning Cognitive functioning

	Social functioning Fatigue Pain QLQ-MY20 domains: Disease symptoms Side effects EQ-5D index score EQ-5D VAS PGIC score

Primary estimand

Treatment effect of elranatamab compared to SOC on PFS. 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:

  Progression free survival (PFS), time from the date of the first dose until confirmed progressive disease (PD) per IMWG criteria or death due to any cause, whichever occurs first.

- Intercurrent events:

  Initiation of a new anticancer therapy for participants of Study C1071003 or initiation of a new LOT for RW patients.

- Population-level summary measure:

  The average treatment effect (ATE) expressed as hazard ratio (HR) or the ratio of the probability of PD or death in participants of Study C1071003 versus RW patients treated with SOC, including the 2-sided 95% confidence interval (CI) and p-value.

3. HYPOTHESES AND DECISION RULES

3.1. Statistical hypotheses

Primary endpoint

It will be tested whether the primary endpoint PFS is different between elranatamab and SOC patients.

H_0: $PFS_{Elranatamab} = PFS_{SOC}$ vs.

H_1: $PFS_{Elranatamab} \neq PFS_{SOC}$

Secondary endpoint

The secondary endpoint is OS, for which the following pairs of hypotheses will be tested:

H_0: $OS_{Elranatamab} = OS_{SOC}$ vs.

H_1: $OS_{Elranatamab} \neq OS_{SOC}$

Exploratory endpoints (if sample size permits)

The exploratory endpoints are PROs.

For each PRO, the following pairs of hypotheses will be tested:

H_0: $PRO_{Elranatamab} = PRO_{SOC}$ vs. H_1: $PRO_{Elranatamab} \neq PRO_{SOC}$

3.2. 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 the 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 the p-value is > 0.05, the null hypothesis will not be rejected.

4. ANALYSIS SETS/POPULATIONS

4.1. 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.

For the comparative analysis of PFS and OS, from populations of RW TCR MM patients extracted from COTA and Flatiron Health databases, we will establish 2 samples of external control arms to assess potential variations 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 be used to define the main 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, main analyses will be repeated using this sample).

CCI

[redacted]

See Table 1 for information on the Study C1071003 inclusion and exclusion criteria which will be applied for RW patients depending on data availability (a checkmark symbol indicates where the critical and/or expanded selection criteria definitions are applicable).

Table 1. Inclusion and exclusion criteria for study C1071003 and RW data sources.

Patient criteria per Study C1071003	*Implementation in Flatiron Health*		*Implementation in COTA*		*Implementation in C1071013 and C1071014*
	Critical eligibility criteria	*Expanded eligibility criteria*	*Critical eligibility criteria*	*Expanded eligibility criteria*	*Critical eligibility criteria*
Inclusion Criteria					
Male or female patients aged≥18 years	✓	✓	✓	✓	✓
Willing and able to comply with all scheduled visits, treatment plans, laboratory tests, lifestyle considerations, and other study procedures	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*
Prior diagnosis of MM as defined according to IMWG criteria (Rajkumar et al, 2014)	✓	✓	✓	✓	✓
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 Urine Protein Electrophoresis (UPEP)* *c) Serum immunoglobulin Free Light Chain (FLC)≥10 mg/dL (≥100 mg/L) and abnormal serum immunoglobulin kappa to lambda FLC ratio (<0.26 or >1.65)*	✓	✓	✓	✓	✓
Patients are TCR, which is 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*	✓	✓	✓	✓	✓
*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.**	✓	✓	✓	✓	✓
ECOG performance status ≤2	✓	✓	✓	✓	✓
Adequate hepatic function characterized by all of 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* *c) alanine aminotransferase (ALT) ≤2.5 x ULN*	*X*	✓	*X*	✓	*X*
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*	✓	*X*	✓	*X*
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*	✓	*X*	✓	*X*
Left ventricular ejection fraction ≥40% as determined by a multigated acquisition scan (MUGA) or echocardiogram	*X*	*X*	*X*	*X*	*X*
Resolved acute effects of any prior therapy to baseline severity or Common Terminology Criteria for AE Grade ≤1	*X*	*X*	*X*	*X*	*X*

Exclusion Criteria					
Active plasma cell leukemia	✓	✓	✓	✓	✓
Amyloidosis	✓	✓	✓	✓	✓
Previous treatment with an anti-BCMA bispecific antibody.	*X*	*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).	✓	✓	✓	✓	✓
Smoldering MM	✓	✓	✓	✓	✓
Stem cell transplant within 12 weeks before enrolment or active Graft Versus Host Disease (GVHD).	✓	✓	✓	✓	✓
Any other active malignancy within 3 years before enrolment, except for adequately treated basal cell or squamous cell skin cancer, or carcinoma in situ.	✓	✓	✓	✓	✓
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 Time from the beginning of the QRS Complex to the End of the T Wave (QT) syndrome (or triplicate average Corrected QT (Fridericia Method) (QTcF) >470 msec).*	*X*	✓	*X*	✓	*X*
Active Hepatitis B Virus (HBV), Hepatitis C Virus (HCV), Severe Acute Respiratory Syndrome- Coronavirus 2 (SARS-CoV2), known Human Immunodeficiency Virus (HIV), or any active, uncontrolled bacterial, fungal, or viral infection. Active infections must be resolved at least 14 days before enrolment.	✓	✓	✓	✓	✓
Other surgical (including major surgery within 14 days before 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*	*X*
Ongoing Grade ≥2 peripheral sensory or motor neuropathy.	*X*	✓	*X*	*X*	*X*
History of any grade peripheral sensory or motor neuropathy with prior BCMA-directed therapy (Cohort B).	*X*	✓	*X*	✓	*X*
History of Guillain-Barre Syndrome (GBS) or GBS variants, or history of any Grade ≥3 peripheral motor neuropathy	*X*	✓	*X*	*X*	*X*
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*	*X*
Known or suspected hypersensitivity to the study intervention or any of its excipients.	*X*	*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*	*X*

***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.*

Note:
[a]This criterion is not among the inclusion/exclusion criteria of Study C10710003 and is applied only to RW patients.
***Legend**: ✓: this criterion can be applied using this RW dataset;*
X: this criterion cannot be applied/assessed due to a lack of information in this RW dataset.

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

4.2. Safety analysis sets

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 comprise the main analysis population for the comparative analysis of PFS and OS.

4.3. Other analysis sets

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

Additional analysis sets using Study C1071003 Cohorts A and B

These analysis sets will include an alternative grouping of all patients from Study C1071003 (ie, Cohort A and Cohort B pooled together rather than just Cohort A). 2 alternative analysis sets will be created: 1 with Study C1071003 patients from Cohorts A and B, and RW patients identified from the Flatiron Health database using critical eligibility criteria, and the second with Study C1071003 patients from Cohorts A and B, and RW patients identified from the COTA database using critical eligibility criteria.

4.4. PRO analysis sets

The PRO analysis sets will be created from participants from Study C1071003 Cohort A and Cohort B who received at least 1 dose of elranatamab and the combined RW patients selected from prospective observational studies C1071013 and C1071014 using critical eligibility criteria. Only patients who completed a baseline PRO will be included.

For the primary PRO analysis, the analysis set will include participants from Study C1071003 Cohort A and Cohort B and patients from observational studies C1071013 and C1071014 CCI [redacted]

Two additional analysis sets will be created for descriptive analyses:

- Participants from Study C1071003 Cohort A and Cohort B and all patients from observational studies C1071013 and C1071014 CCI [redacted]
- Participants from Study C1071003 Cohort A and patients from observational studies C1071013 and C1071014 CCI [redacted]

4.5. Subgroups

Subgroup analyses will be performed to only include treatments frequently available for patient population in non-US countries. The list of treatments for sub-group analyses is provided in Appendix B.

5. ENDPOINTS AND COVARIATES

5.1. 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).

5.2. Effectiveness endpoints

Definitions of the outcomes will be aligned, where possible, with Study C1071003 (Table 2).

Table 2. Definitions of comparative effectiveness outcomes in Study C1071003 and RW data sources.

	Outcome	*Study C1071003*	*Flatiron Health*	*COTA*
1	*Progression-free survival (PFS)*	*Time from the date of the first dose until confirmed PD per IMWG criteria or death due to any cause, whichever occurs first.*	*Time from initiation of the first line after TCR to either the date of progression* or death due to any cause, whichever occurs first.*	*Time from initiation of the first line after TCR to either the date of progression* or death due to any cause, whichever occurs first.*
2	*Overall survival (OS)*	*Time from the date of the first dose until death due to any cause.*	*Time from initiation of the first line after TCR until the date of death due to any cause.*	*Time from initiation of the first line after TCR until the date of death due to any cause.*

Note:
* COTA uses a third-party obituary data source to capture mortality data. COTA defines a progression as an increase of ≥25% from the lowest response value in any one or more of the following criteria: SPEP with an absolute increase > 0.5 g/dL; 24-hour UPEP with an absolute increase > 200 mg/24 h; in patients without measurable serum and urine M-protein, the absolute increase of >10 mg/dL in the difference between involved and uninvolved FLC levels; or an absolute bone marrow plasma cell percentage > 10%.
Flatiron Health's mortality variable is created through an amalgamation of structured data elements, unstructured documents, and linking to external mortality sources and the Social Security Death Index (SSDI). Flatiron Health defines a progression as an increase of ≥25% from baseline/nadir value in any one or more of the following: an absolute increase in serum M-protein by SPEP by ≥0.5 g/dL; serum M-protein ≥1 g/dL if the lowest M component was ≥5 g/dL; an absolute increase in urine M-protein by UPEP by ≥200 mg/24 h; in patients without measurable serum and urine M-protein levels, an absolute increase in the difference between involved and uninvolved FLC levels of >10 mg/dL.
Abbreviations: IMWG=International Myeloma Working Group; PD=progressive disease; TCR=triple-class refractory.

Progression Free Survival (PFS)

PFS will be estimated among all patients in the analysis as follows:

PFS (months) = [date of event or censoring – index date +1] / 30.4375.

See Table 3 for information on PFS censoring rules in Study C1071003 that is available and applied for RW patients.

Table 3. PFS censoring rules in Study C1071003 and RW data Sources.

Study C1071003	Implementation in Flatiron Health	Implementation in COTA
Participants who do not have an event (confirmed PD per IMWG criteria or death due to any cause).	Patients without an event will be censored on the earliest date between the latest available record for the patient or the data cut-off date.	Patients without an event will be censored on the earliest date between the latest available record for the patient or the data cut-off date.
Participants who start a new anticancer therapy prior to an	Patients who start a new line before an event will be censored	Patients who start a new line before an event will be censored

event will be censored on the date of the last adequate disease assessment before the new anticancer therapy.	on the day before the start date of the next line of therapy.	on the day before the start date of the next line of therapy.
Participants with an event after a gap of 2 or more missing disease assessments will be censored on the date of the last adequate disease assessment before the gap.	Cannot be implemented.	Cannot be implemented.
Participants who do not have an adequate post-baseline disease assessment will be censored on the date of first dose of study intervention unless death occurs on or before the time of the second planned disease assessment (ie, ≤70 days after the date of first dose), in which case the death will be considered an event.	Cannot be implemented.	Cannot be implemented.

Overall Survival (OS)

OS will be estimated among all patients in the analysis as follows:

OS (months) = [date of death or censoring – index date +1] / 30.4375

See Table 4 for information on OS censoring rules in Study C1071003 that is available and applied for RW patients.

Table 4. OS censoring rules in Study C1071003 and RW data sources.

Study C1071003	Implementation in Flatiron Health	Implementation in COTA
Participants who do not have an event, who withdraw consent, or who are lost to follow-up.	Patients without an event will be censored on the earliest date between the latest available record for the	Patients without an event will be censored on the earliest date between the latest available record for the

	patient or the data cut-off date.	patient or the data cut-off date.

5.3. Safety endpoints

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

5.4. Other endpoints

CCI

[redacted]

CCI

[redacted]

CCI

[redacted]

5.5. Covariates

Baseline covariates will be selected to compare patients from elranatamab and external control arms, and to perform further statistical adjustments to control for baseline confounding (eg, via IPTW).

Baseline covariates to be captured will include those related to patient demographic characteristics, disease characteristics, comorbidity profile, laboratory measurements, and MM treatment patterns, including LOT. Each variable will be taken on or before the index date; if before, the most recent measurement will be used. All variables listed below will be used to describe the study cohorts and to adjust for baseline confounding. Since small sample sizes limit the number of covariates that can be reasonably accounted for via statistical adjustment (eg, IPTW), a systematic literature review was conducted to identify the variables most strongly and consistently correlated with outcomes in RWD studies conducted among relapsed/refractory multiple myeloma (RRMM) patients. CCI

[redacted]

Additional confounders identified in the Phase 1 analysis will also be included in the Phase 2 analysis to optimize the balance in clinical characteristics between participants from Study C1071003 and each RWD source. These confounders have been selected based on their clinical importance and relevance to the disease prognosis/severety and disease complications (level of serum albumin, serum calcium level, presence of bone lesions, extramedullary disease [EMD], hemoglobin, and serum creatinine), liver dysfunction (levels of bilirubin, aspartate aminotransferase [AST], and alanine aminotransferase [ALT]), and burden of comorbid

conditions (Charlson comorbidity index [CCI]). Of note, the presence of EMD is only available for the analyses using combined datasets of Study C1071003 and COTA patients.

See Table 7 for the operational definition of the covariates.

Table 7. Operational definitions of covariates

Variable	Operational definition
Age, years	On the index date
Sex	Male, female
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 5.5.1)
ECOG performance status	ECOG=0, 1, 2 Within 90 days before or on the index date, if feasible
Time since initial MM diagnosis, days	From the date of MM diagnosis to the index date
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
Number of pre-index treatment lines	N, % 1, 2, 3, 4, 5, … Between the date of MM diagnosis and index date
Penta-refractory status	Yes, No. Penta-drug refractory (refractory to 2 IMIDs, 2 PIs and 1 anti-CD38) At time of TCR eligibility
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
Presence of bone lesions	Yes, No Identified in patients without extramedullary disease on or before the index date.
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 (13)
Aspartate aminotransferase (microkat/L)	Within 90 days before or on the index date
Alanine aminotransferase (microkat/L)	Within 90 days before or on the index date
Creatinine clearance (mL/min)	Within 90 days before or on the index date
Hemoglobin (g/L)	Within 90 days before or on the index date
Billirubin (mg/dL)	Within 90 days before or on the index date
Calcium in serum or plasma (mmol/L)	Within 90 days before or on the index date
Serum albumin (g/dL)	Within 90 days before or on the index date

Multiple imputation will be used to address the presence of missing values for covariates in RWD. Multiple imputation will be implemented for covariates with up to 30 % of missing values (14). If the missingness excess 30%, the covariate will be omitted from the analysis, or it will be categorized using a separate category for patients with missing values.

Based on their distribution, continuous laboratory values may be categorized or transformed to limit the influence of outliers.

5.5.1. 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 (15).

Values for beta-2 microglobulin and serum albumin measured within 90 days before or on the index date will be used. If patients do not have beta-2 microglobulin or serum albumin tests available during this window, their ISS stage will be considered missing.

ISS stage will be defined as described in Table 8.

Table 8. 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. 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), therefore no specific handling of the partial index date will be needed.

Date of death

Incomplete dates of death (ie, missing day) will be imputed in the RW data to match as closely as possible the rules of Study C1071003 as described in Table 9.

Table 9. Handling of incomplete or missing date of death in Study C1071003 and RW data sources.

Study C1071003	Implementation in Flatiron Health	Implementation in COTA
Missing or partial death dates will be imputed based on the last contact date: • If the date of death is missing, it will be imputed as the day after the date of the last contact. • If the day or both day and month are missing, the date of death will be imputed to the maximum of the full	Flatiron Health provides only the month and year of death. Therefore the date of death will be imputed as the middle of the month (ie, the 15th), unless the patient's last record date falls within the month of death, in this case, the date of death will be imputed as the date of the last record.	COTA provides the precise date of death for most patients. When the date of death is missing, COTA by default imputes the date of the death as the middle of the month when the month is known, and as the middle of the year when only the year is known.

(non-imputed) day after the date of the last contact and the following: ✓ Missing day: 1st day of the month and year of death ✓ Missing day and month: January 1st of the year of death.		For patients whose month of death is known, the date of death will be imputed as the last record date if it falls within 15 days of the date of death provided by COTA (otherwise the default date of death will be kept). For patients whose year of death only is known, the date of death will be imputed as the last record date if it falls within 182 days of the date of death provided by COTA (otherwise the default date of death will be kept).

For the baseline characteristics listed in Section 5.5, an adjustment for missing day information will be performed by imputing the last day of the corresponding month. If this imputed date lies within the baseline period, the measurement will be qualified as the baseline value.

Missingness in measured baseline characteristics

Investigations into the proportion of missing values for each baseline covariate (Section 5.5) will be conducted. If the proportion of missing values is ≤ 30% for a given covariate, multiple imputation will be performed (Section 7.2.2) (14).

When the ULN of lab measurements is missing, the ULN will be imputed using the last observation carried forward (LOCF) method. If imputation by LOCF is not possible, the missing ULN will be imputed by sex-stratified mean value.

7. STATISTICAL METHODOLOGY AND STATISTICAL ANALYSES

7.1. Summary of the analyses

All statistical analyses described in this section will be carried out separately for each analytical set. Table 10 provides a summary of the analyses stratified by external control arm sample.

Table 10. Summary of statistical analyses by study sample.

	Critical eligibility criteria samples selected from COTA or Flatiron Health	***Expanded eligibility criteria samples selected from***	***Critical eligibility criteria sample selected from C1071013 and C1071014***

		COTA or Flatiron Health	
Main analyses	• *IPTW comparison of PFS in elranatamab (Cohort A) versus SOC-treated patients selected from COTA* • *IPTW comparison of PFS in elranatamab (Cohort A) versus SOC-treated patients selected from Flatiron Health* • *IPTW comparison of OS in elranatamab (Cohort A) versus SOC-treated patients selected from COTA* • *IPTW comparison of OS in elranatamab (Cohort A) versus SOC-treated patients selected from Flatiron Health*		
Sensitivity analyses	• *Doubly robust comparisons of PFS in elranatamab (Cohort A) versus SOC-treated patients selected from COTA* • *Doubly robust comparisons of PFS in elranatamab (Cohort A) versus SOC-treated patients selected from Flatiron Health* • *Doubly robust comparisons of OS in elranatamab (Cohort A) versus SOC-treated patients selected from COTA* • *Doubly robust comparisons of OS in elranatamab (Cohort A) versus SOC-treated patients selected from Flatiron Health* • *Quantitative bias assessments (using IPTW estimates) when comparing PFS in elranatamab (Cohort A) versus SOC-treated patients selected from COTA* • *Quantitative bias assessments (using IPTW estimates) when comparing PFS in elranatamab (Cohort A) versus SOC-treated patients selected from Flatiron Health* • *Quantitative bias assessments (using IPTW estimates) when comparing OS in elranatamab (Cohort A) versus SOC-treated patients selected from COTA* • *Quantitative bias assessments (using IPTW estimates) when comparing OS in elranatamab (Cohort A) versus SOC-treated patients selected from Flatiron Health*	• IPTW comparison of PFS in elranatamab (Cohort A) versus SOC-treated patients selected from COTA • IPTW comparison of PFS in elranatamab (Cohort A) versus SOC-treated patients selected from Flatiron Health • IPTW comparison of OS in elranatamab (Cohort A) versus SOC-treated patients selected from COTA • IPTW comparison of OS in elranatamab (Cohort A) versus SOC-treated patients	

		selected from Flatiron Health	
Additional analyses	• *Comparisons of PFS in PS matched elranatamab (Cohort A) and SOC-treated patients selected from COTA* • *Comparisons of PFS in PS matched elranatamab (Cohort A) and SOC-treated patients selected from Flatiron Health* • *Comparisons of OS in PS matched elranatamab (Cohort A) and SOC-treated patients selected from COTA* • *Comparisons of OS in PS matched elranatamab (Cohort A) and SOC-treated patients selected from Flatiron Health* • *IPTW comparison of PFS in elranatamab (Cohort A+Cohort B) versus SOC-treated patients selected from COTA* • *IPTW comparison of PFS in elranatamab (Cohort A+Cohort B) versus SOC-treated patients selected from Flatiron Health* • *IPTW comparison of OS in elranatamab (Cohort A+Cohort B) versus SOC-treated patients selected from COTA* • *IPTW comparison of OS in elranatamab (Cohort A+Cohort B) versus SOC-treated patients selected from Flatiron Health* • *Subgroup analyses to compare PFS and OS in participants of Study C1071003 and patients from Flatiron Health and COTA*		• *Number and percentage of participants who completed QLQ-C30 domains, QLQ-MY20 domains, EQ-5D index, EQ-5D-VAS, and PGIC at each follow-up visit* • *Descriptive statistics of QLQ-C30 domains, QLQ-MY20 domains, EQ-5D index, EQ-5D-VAS, and PGIC in elranatamab and SOC-treated patients, including mean and absolute change in PRO scores (final assessment score minus baseline score)* *Based on the results of the descriptive analysis of completeness, and the timing of assessments in the two treatment arms, the following analyses may be conducted:* • *Comparison of change from baseline on the QLQ-C30 domains, QLQ-MY20 domains, EQ-5D index, EQ-5D-VAS, and PGIC using naïve and adjusted mixed-effect models*

7.2. Statistical methods

7.2.1. Descriptive statistics

Tabular summaries of baseline patient demographic and clinical characteristics for each treatment arm and analysis set 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".

7.2.2. Multiple imputation

Multiple imputation will be performed to address missing values only in baseline covariates described in Section 5.5 under the assumption of missingness at random (MAR). MAR assumes that there might be systematic differences between the missing and observed values of prognostic characteristics, but these can be entirely explained by other observed prognostic characteristics, the treatment variable, and the observed overall survival. SAS PROC MI will be used to perform multiple imputation by chained equations (MICE) for missing values using the fully conditional specification (FCS) method (16). This approach imputes values for every variable conditional on all other variables (by specifying an imputation model for each variable). The technique is iterative and proceeds via Gibbs sampling if the initial joint distribution defined by the specified conditional distributions exists (17). The variables to be included in the imputation model will be baseline characteristics that will be used to estimate the PS, measured covariates related to missingness or those correlated with the covariate of interest, and an outcome variable (18–23). The method assumes the normal distribution of continuous variables in the imputation model and linear relationships between variables in the model. In case of violation of these assumptions, predictive mean matching and/or variable transformations will be used (19). The number of imputed datasets will be specified according to Bodner's conservative, simulation-based criteria (24).

Diagnostics to assess and evaluate the resulting imputation models obtained via MICE will be conducted to determine whether convergence has been achieved (25). Visual assessment will be performed using plots to examine the observed and imputed data and compare their distributions. Convergence will be evaluated by plotting the mean and variance of each imputation run across iterations, to confirm that there are no apparent trends. If convergence has not been achieved, the number of iterations will be increased until means and variances stabilize.

For each imputed dataset, the comparative analysis of PFS and OS will be conducted as described in Sections 7.2.3-7.2.8. For each outcome, the parameter estimates from imputed datasets will be combined to generate the final estimate using Rubin's rules that is based on assumption that variables estimated for each imputed dataset is normally distributed (26). If the statistics estimated from each imputed dataset are not normally distributed, the logarithm transformation will be used as recommended by Van Buuren (27) and the final estimate will be generated.

SAS PROC MIANALYZE will be run to combine the results of the analyses with each imputed dataset to generate the final estimate using Rubin's rules (26).

7.2.3. Propensity scores and inverse probability of treatment weighting

In analyses of RWD, an important consideration in the identification of potentially causal effects is the ability to effectively 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 an approach often used in non-randomized studies to create a pseudo-population in which the covariates are independent of the treatment assignment, thereby permitting an unbiased estimate of the ATE, provided that fundamental assumptions of causal inference are satisfied.

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 5.5 ***Error! Reference source not found.*** *and, optional, some additional covariates. Diagnostics of the estimated PS will be applied by examining the distribution of the PS in each treatment group to identify the degree of overlap and region of common support* (28)*. The estimated PS will be used to generate IPTWs, which may be stabilized to reduce variance and the effect of very large weights. A standardized mean difference (SMD) of ≥25% will be used as an indicator of covariate imbalance between the treatment arms requiring further investigation and may lead to refining the logistic model for deriving the weights*(29,30)*.*

Assumptions

Observational research relies on methods for causal inference when deriving unbiased estimates. Hernán and Robins (31) 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 7.2.7.

(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 (28). 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 PS. The PS will be estimated as the probability of initiating elranatamab versus SOC conditionally on patients' characteristics measured at baseline. The PS model will include all covariates described in Section 5.5.

Estimation of IPTW

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

To reduce the potential variance inflation induced by extreme weights, the weights will be stabilized by the inclusion of the overall probability of being treated with elranatamab for patients from Study C1071003 in the numerator, or by the overall probability of being treated with SOC for RW patients (32). 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 adjustment for covariates (33). The distribution of the estimated stabilized IPTW will be evaluated visually and by review of descriptive statistics.

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 (34). 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.

The SAS macro STDDIFF.SAS will be used to estimate SMD (35).

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 and SMDs 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 (36).

A standardized mean difference (SMD) of ≥20% will be used as an indicator of covariate imbalance between the treatment arms requiring further investigation (29,30,37).

7.2.4. Naïve comparison of PFS and OS

For each outcome, the total follow-up time for each treatment group will be reported. SAS PROC GENMOD will be used to estimate PFS and OS event rates (the events of PD and death, or death only, respectively) with corresponding two-sided 95% CIs using Poisson distribution (38).

SAS PROC LIFETEST will be used to produce the Kaplan-Meier estimates for PFS and OS (product-limit estimates) and to graphically display the survival functions together with a summary of associated statistics including the median PFS and OS time with 2-sided 95% CIs. The 95% CI for the median will be estimated using the Brookmeyer and Crowley method (39) and the 95% CI for the 25th percentile will be estimated via the Klein and Moeschberger method (40).

PFS and OS 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 model assumption (e.g. proportional hazard) will be checked by plotting log(-log(time-to-event)) versus log(time) and tested using the Schoenfeld residuals test. In the case the proportional hazards assumption is not met, a restricted mean survival time model will be applied instead of the Cox proportional hazards regression model (41).

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 (42). If severe violations are observed, a weighted restricted mean survival time model will be applied (using SAS PROC LIFETEST with RMST option).

Linearity of the relationship between the log-hazard and the treatment will be assessed by plotting the Martingale residuals (43). 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.

All estimates will be obtained for each imputed dataset. SAS PROC MIANALYZE will be run to combine the results and generate final estimates using Rubin's rules (26).

7.2.5. Weighted analyses for PFS and OS

SAS PROC LIFETEST with the WEIGHT statement will be used to produce the Kaplan-Meier estimates for PFS and OS and to graphically display the survival functions together with a summary of associated statistics including the median PFS and OS time with 2-sided 95% CIs (44). The 95% CI for the median will be estimated using the Brookmeyer and Crowley method (39) and the 95% CI for the 25th percentile will be estimated via the Klein and Moeschberger method (40).The pooled Kaplan-Meier estimate from all imputed datasets will be obtained applying Rubin's rules (26) and Greenwood's formula for the SE of the pooled estimates (45). Because median OS likely will not be reached in Study C1071003, for the main analysis (Section 7.3.2), the IPT weighted survival probabilities will be described using KM estimator and compared between two treatment groups using the log-rank test for 12 and 15 months of follow-up (46,47).

PFS and OS will be compared between treatment groups using hazard ratios (HR) estimated via weighted Cox proportional hazard models. The model assumption (eg, proportional hazard) will be checked by plotting log(-log(time-to-event)) versus log(time) and tested using the Schoenfeld residuals test. If the proportional hazards assumption is met, 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 (48). All estimates will be obtained for each imputed dataset. SAS PROC MIANALYZE will be run to combine the results and generate final estimates using Rubin’s rules (26).

In the case the proportional hazards assumption is not met, an IPT weighted restricted mean survival time model will be applied instead of the Cox proportional hazards regression model in R using the approach described by Conner et al (49). P-values for null hypothesis testing for pooled, IPT-weighted RMST estimates will be obtained using F-tests as described by Li et al. and recommended by van Buuren (50).

7.2.6. Doubly robust analyses of PFS and OS

A conditional average treatment effect (CATE) and its associated hazard ratio will be estimated using a semi-parametric approach described by Yadlowsky et al. (51,52)*. This estimator will provide a doubly robust comparison for PFS and OS between treatment arms. Standard errors and 95% confidence intervals are obtained for this estimator using the non-parametric bootstrap.*

Analyses employing a doubly robust estimator require fitting 2 models: a model for treatment or exposure status, and a second 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 (53)*. Assuming these assumptions are upheld, if at least 1 of these 2 models is*

correctly specified, resulting in a doubly robust estimate that may remain consistent and unbiased (54).

The R package precmed will be used to estimate the average treatment effect (ATE) via the atefitsurv function, which implements the semiparametric estimator for the hazard ratio developed by Yadlowsky et al.(51,52). Standard errors and confidence intervals for the doubly robust estimators will be obtained via non-parametric bootstrap.

All estimates will be obtained for each imputed dataset. SAS PROC MIANALYZE will be run to combine the results and generate final estimates using Rubin's rules (26).

7.2.7. Quantitative bias assessment

Quantitative bias assessment will be done using nullification analysis to assess the potential influence of unmeasured confounding on the HR estimates from the main analysis. A shortlist of suspected unmeasured confounders will be identified; these are variables that were observed in Study C1071003 but not available (or only with substantial missingness) in the RWD sources. 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 (55).

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 main analyses. The EValue package in R will be used to estimate E-values for PFS and OS, 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 (56–58).

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

Table 11. Calculation of the E-value

Effect Measure	Calculation of Approximate E-value
HR for rare outcomes	$E-value = HR + \sqrt{HR \times (HR-1)}$
HR 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 (58).

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 main analyses, the magnitude of the E-value corresponds to the minimum magnitude of residual confounding required to explain an estimated PFS or OS. For these analyses, an E-value of 2.0, for example, will be interpreted to mean that the 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 PFS or OS (58). 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 (58).

7.2.8. Matching of participants of Study C1071003 to RW patients

In order to estimate the ATT effect, 2 matched cohorts will be created from the participants of Study C1071003 and patients selected from COTA using critical eligibility criteria, and from the participants of Study C1071003 and patients selected from Flatiron Health.

Each participant of Study C1071003 will be matched to 1 RW patient without replacement (once a RW patient is selected to match a given participant of Study C1071003, this RW patient will no longer be available as a potential match for other participants of Study C1071003). Matching will be done on 0.2 standard deviations of the logit of the PS (14) estimated as described in Section 7.2.3. Previous studies showed that this caliper allowed elimination of about 99% of the bias due to measured confounders (59). Patients with no match will be excluded from the analysis.

Balance assessment will be obtained by estimating SMDs for the baseline covariates in the PS-matched cohorts using the SAS macro STDDIFF.SAS (35).

A standardized mean difference (SMD) of ≥20% will be used as an indicator of covariate imbalance between the matched cohorts requiring further investigation or the use of regression adjustment to address the imbalance (29,30,37).

PFS and OS will be compared using hazard ratios (HR) estimated from unadjusted Cox proportional hazard models. The analysis will be carried out using SAS PROC PHREG. The robust standard error will be estimated (60). HR estimates will be obtained for each imputed dataset. SAS PROC MIANALYZE will be run to combine the results and generate final HR estimates using Rubin's rules (26). In the case the proportional hazards assumption is not met, a restricted mean survival time model will be applied instead of the Cox proportional hazards regression model using SAS PROC LIFETEST with RMST option.

7.2.9. Analysis of PROs

Completion status

For each treatment group, at each time point, the number and percentage of participants who completed at each visit the QLQ-C30 domains, QLQ-MY20 domains, EQ-5D index, EQ-5D-VAS, and PGIC will be summarized. The timing of PRO completion by treatment group will be described.

Descriptive statistics of PROs in elranatamab and SOC-treated patients

For each domain of interest of QLQ-C30, QLQ-MY20, EQ-5D index, EQ-VAS, SAS PROC UNIVARIATE will be used to estimate mean (SD), median (IQR), minimum, maximum at each timepoint. This will be done based on observed values as well as changes in scores from baseline values. SAS PPROC FREQ will be used to summarize the number and percentage of patients at each level of PGIC scores.

- Participants from Study C1071003 Cohort A and Cohort B and patients from observational studies C1071013 and C1071014 CCI [redacted]
- Participants from Study C1071003 Cohort A and Cohort B and all patients from observational studies C1071013 and C1071014 CCI [redacted]
- Participants from Study C1071003 Cohort A and patients from observational studies C1071013 and C1071014 CCI [redacted]

If the sample size permits, the comparative analyses of PRO described below will be conducted. These analyses will be conducted only in the set of patients consisting of participants from Study C1071003 Cohort A and Cohort B and patients from observational studies C1071013 and C1071014 CCI [redacted]

Comparison of change from baseline in elranatamab and SOC-treated patients

Among patients for whom a baseline and at least 2 follow-up measures for the same PRO outcome are available, a repeated measures mixed-effects model will be fit to the data to examine the effects of time (visit) among patients treated with elranatamab versus SOC. Mixed-effects model with repeated measures (MMRM) will be carried out for the domains of QLQ-C30 and QLQ-MY20, EQ-5D index and EQ-VAS in order to examine the effects of time (visit) by cohort and overall. CCI [redacted]

Linear mixed model parameters will be estimated using restricted maximum likelihood (REML) when possible, as maximum likelihood may produce biased estimates of variance in linear mixed models. The unstructured covariance structure will be used to define covariance between random effects (using option "Type=UN" as a part of the RANDOM statement in PROC MIXED). For the degrees-of-freedom calculations the Kenward and Roger algorithm should be used (using option ddfm = kr") as a part of the MODEL statement in PROC MIXED).

If the model using PROC MIXED with the unstructured covariance structure cannot converge, the following hierarchical strategy will be used to choose the covariance structure.

- Unstructured
- Spatial Power if Unstructured does not converge
- Toeplitz if Spatial Power does not converge
- Auto Regressive (1) if Toeplitz does not converge
- Compound Structure if Auto Regressive (1) does not converge

If the outcome variable is not normally distributed (Gaussian), MMRM can be carried out using PROC GLIMMIX. The selection of distribution and link function might be data-driven. PROC UNIVARIATE will be used to create QQplots and test statistics for assessing normality. For example, if the distribution of continuous outcome is extremely skewed, gamma distribution with a log link could be chosen for modelling.

7.3. Statistical analyses

7.3.1. Safety analyses

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

7.3.2. Main analyses of PFS and OS

The main analysis will be performed using the safety analysis set (Section 4.2) as described in Section 7.2.4 and Section 7.2.5.

7.3.3. Sensitivity analyses – doubly robust comparison of PFS and OS in elranatamab (Cohort A) and SOC-treated patients

To evaluate robustness of the results from the main analysis, a sensitivity analysis will be conducted using the safety analysis set (Section 4.2) and doubly robust analysis as described in Section 7.2.6.

7.3.4. Sensitivity analyses – comparison of PFS and OS in elranatamab (Cohort A) and SOC-treated patients identified using expanded eligibility criteria

To evaluate robustness of the results from the main analysis, a sensitivity analysis will be conducted using the sensitivity analysis sets (Section 4.3) . IPTWs will be estimated as described in Section 7.2.3. Analyses of PFS and OS will be conducted as described in Section 7.2.4 and Section 7.2.5.

7.3.5. Sensitivity analysis - quantitative bias assessments

Nullification analysis will be conducted to assess the potential influence of unmeasured confounding. This sensitivity analysis will be applied for the results obtained in the main analysis using the safety analysis set (Section 4.2).

This analysis will take place as outlined in Section 7.2.7.

7.3.6. Additional analyses

Comparison of PFS and OS in elranatamab (Cohort A + Cohort B) versus SOC-treated patients

This analysis will be performed using the additional analysis set (Section 4.3). IPTWs will be estimated as described in Section 7.2.3. Analyses of PFS and OS will be conducted as described in Section 7.2.4 and Section 7.2.5.

Analyses of PROs in elranatamab versus SOC-treated patients

PROs (see Section 5.4) will be analyzed and compared as described in Section 7.2.9 using PRO analysis sets (Section 4.4).

Comparisons of PFS and OS in PS-matched elranatamab (Cohort A) and SOC-treated patients

To evaluate robustness of the results of the main analysis, an additional analysis will be conducted in PS-matched elranatamab and SOC-treated patients selected from safety analysis set (Section 4.2). Matching will be done as described in Section 7.2.8. Analysis will be done as described in Section 7.2.4 and Section 7.2.5.

Subgroup analysis

Because the treatment patterns described in COTA and Flatiron may not generalize to all non-US countries, subgroup analyses will be performed to only include treatments frequently available for patient populations in non-US countries. The list of treatments for sub-group analyses is provided in Appendix B.

Subgroup analyses will be conducted using the safety analysis set (Section 4.2) to compare PFS and OS.

If the sample size is sufficient, PFS and OS will be compared between participants of Study C1071003 and patients from Flatiron Health and COTA using naïve comparison and IPT weighted analysis as described in Section 7.2.4 and Section 7.2.5. For each treatment subgroups, the IPTWs will be estimated as described in Section 7.2.3.

In case of insufficient sample size, only naive comparison of PFS and OS will be provided as described in Section 7.2.4.

7.4. Software

SAS 9.4 or higher and R v4.2.0 or higher will be used for the analyses.

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. EORTC Quality of Life Website - EORTC - Quality of Life : EORTC – Quality of Life [Internet]. [cited 2023 Mar 16]. Available from: https://qol.eortc.org/
9. EORTC - Quality of Life. EORTC QLQ-C30 Scoring Manual [Internet]. [cited 2023 May 1]. Available from: https://qol.eortc.org/manuals/
10. 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. Eur J Cancer. 2007 Jul 1;43(11):1670–8.
11. Index value set calculators – EQ-5D [Internet]. [cited 2023 May 1]. Available from: https://euroqol.org/support/analysis-tools/index-value-set-calculators/

12. Dolan P. Modeling valuations for EuroQol health states. Med Care [Internet]. 1997 [cited 2023 Mar 6];35(11):1095–108. Available from: https://pubmed.ncbi.nlm.nih.gov/9366889/

13. 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/

14. Jagannath S, Lin Y, Goldschmidt 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/

15. 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/

16. Buuren S van, Groothuis-Oudshoorn K. Multivariate imputation by chained equations in R. J Stat Softw. 2011;45(3):1–67.

17. van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. Stat Methods Med Res. 2007 Jun 2;16(3):219–42.

18. White IR, Royston P. Imputing missing covariate values for the Cox model. Stat Med [Internet]. 2009 Jul 10 [cited 2023 Feb 20];28(15):1982–98. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/sim.3618

19. Austin PC, White IR, Lee DS, van Buuren S. Missing Data in Clinical Research: A Tutorial on Multiple Imputation. Can J Cardiol [Internet]. 2021 Sep 1 [cited 2023 Feb 20];37(9):1322–31. Available from: https://pubmed.ncbi.nlm.nih.gov/33276049/

20. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. Stat Med [Internet]. 2011 Feb 20 [cited 2023 Feb 20];30(4):377–99. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/sim.4067

21. Moons KGM, Donders RART, Stijnen T, Harrell FE. Using the outcome for imputation of missing predictor values was preferred. J Clin Epidemiol [Internet]. 2006 Oct [cited 2023 Mar 16];59(10):1092–101. Available from: https://pubmed.ncbi.nlm.nih.gov/16980150/

22. Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ [Internet]. 2009 Jun 29 [cited 2023 Mar 16];338(7713):157–60. Available from: https://www.bmj.com/content/338/bmj.b2393

23. Bartlett JW, Frost C, Carpenter JR. Multiple imputation models should incorporate the outcome in the model of interest. Brain [Internet]. 2011 Nov 1 [cited 2023 Mar 16];134(11):e189–e189. Available from: https://academic.oup.com/brain/article/134/11/e189/308833

24. Bodner TE. What Improves with Increased Missing Data Imputations? https://doi.org/101080/10705510802339072 [Internet]. 2008 Oct [cited 2023 Mar 16];15(4):651–75. Available from: https://www.tandfonline.com/doi/abs/10.1080/10705510802339072

25. van Buuren S. Flexible imputation of missing data. Boca Raton, Florida: Chapman & Hall/CRC; 2012.

26. PROC MIANALYZE: Overview :: SAS/STAT(R) 9.3 User's Guide [Internet]. [cited 2023 Feb 20]. Available from: https://support.sas.com/documentation/cdl/en/statug/63962/HTML/default/viewer.htm#statug_mianalyze_sect001.htm

27. van Buuren S. Flexible Imputation of Missing Data. Flexible Imputation of Missing Data [Internet]. 2012 Jan 1 [cited 2023 Mar 30];1–299. Available from: https://www.taylorfrancis.com/books/mono/10.1201/b11826/flexible-imputation-missing-data-stef-van-buuren

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

29. 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

30. 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/

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

32. 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/

33. 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/

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

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

36. 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.

37. 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/

38. Poisson Regression :: SAS/STAT(R) 13.1 User's Guide [Internet]. [cited 2023 Apr 27]. Available from: https://support.sas.com/documentation/cdl/en/statug/66859/HTML/default/viewer.htm#statug_genmod_gettingstarted01.htm

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

40. 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

41. Andersen PK, Pohar Perme M. Pseudo-observations in survival analysis. http://dx.doi.org/101177/0962280209105020 [Internet]. 2009 Aug 4 [cited 2023 Mar 16];19(1):71–99. Available from: https://journals.sagepub.com/doi/10.1177/0962280209105020?url_ver=Z39.88-2003&rfr_id=ori%3Arid%3Acrossref.org&rfr_dat=cr_pub++0pubmed

42. 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/

43. 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

44. SAS Help Center: The LIFETEST Procedure WEIGHT Statement [Internet]. [cited 2023 Feb 22]. Available from: https://documentation.sas.com/doc/en/pgmsascdc/9.4_3.4/statug/statug_lifetest_syntax08.htm

45. Morisot A, Bessaoud F, Landais P, Rébillard X, Trétarre B, Daurès JP. Prostate cancer: net survival and cause-specific survival rates after multiple imputation. BMC Med Res Methodol [Internet]. 2015 Jul 28 [cited 2023 May 14];15(1). Available from: https://pubmed.ncbi.nlm.nih.gov/26216355/

46. SAS Help Center: WEIGHT Statement [Internet]. [cited 2023 Apr 17]. Available from: https://documentation.sas.com/doc/en/pgmsascdc/9.4_3.4/statug/statug_lifetest_syntax08.htm

47. Xie J, Liu C. Adjusted Kaplan-Meier estimator and log-rank test with inverse probability of treatment weighting for survival data. Stat Med [Internet]. 2005 Oct 30 [cited 2023 Apr 17];24(20):3089–110. Available from: https://pubmed.ncbi.nlm.nih.gov/16189810/

48. 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/

49. Conner SC, Sullivan LM, Benjamin EJ, LaValley MP, Galea S, Trinquart L. Adjusted restricted mean survival times in observational studies. Stat Med [Internet]. 2019 Sep 10 [cited 2023 May 14];38(20):3832–60. Available from: https://pubmed.ncbi.nlm.nih.gov/31119770/

50. Li KH, Raghunathan TE, Rubin DB. Large-Sample Significance Levels from Multiply Imputed Data Using Moment-Based Statistics and an F Reference Distribution. J Am Stat Assoc. 1991 Dec;86(416):1065.

51. Yadlowsky S, Pellegrini F, Lionetto F, Braune S, Tian L. Estimation and Validation of Ratio-based Conditional Average Treatment Effects Using Observational Data. J Am Stat Assoc [Internet]. 2021 [cited 2023 Apr 4];116(533):1–18. Available from: https://pubmed.ncbi.nlm.nih.gov/33767517/

52. CRAN - Package precmed [Internet]. [cited 2023 Apr 4]. Available from: https://cran.r-project.org/web/packages/precmed/index.html

53. 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.

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

55. van der Weele TJ, Ding P. Sensitivity Analysis in Observational Research: Introducing the E-Value. Ann Intern Med [Internet]. 2017 Aug 15 [cited 2022 May 25];167(4):268–74. Available from: https://pubmed.ncbi.nlm.nih.gov/28693043/

56. 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/

57. 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/

58. 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/

59. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate Behav Res [Internet]. 2011 May [cited 2023 Mar 12];46(3):399–424. Available from: https://pubmed.ncbi.nlm.nih.gov/21818162/

60. Austin PC. The performance of different propensity score methods for estimating marginal hazard ratios. Stat Med [Internet]. 2013 Jul 20 [cited 2023 Mar 14];32(16):2837–49. Available from: https://pubmed.ncbi.nlm.nih.gov/23239115/

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 ®)	Anti-CD38-directed MAb
isatuximab (Sarclisa ®)	Anti-CD38-directed MAb
dexamethasone	Corticosteroid
prednisone	Corticosteroid
panobinostat	HDAC
lenalidomide (Revlimid ®)	IMiD
pomalidomide (Pomalyst ® or Imnovid ®)	IMiD
thalidomide	IMiD
elotuzumab (Empliciti ®)	MAb
selinexor (Xpovio ®/Nexpovio ®)	Nuclear export inhibitor
bortezomib (Velcade ®)	PI

carfilzomib (Kyprolis ®)	PI
ixazomib (Ninlaro ®)	PI
etoposide	Podophyllotoxin Derivative
vincristine/leurocristine (Oncovin ®)	Vinca Alkaloid

Appendix B. List of subgroups and comprising treatments

Subgroup	Treatments to be included in Subgroup Analysis
Subgroup 1	All regimens except those which include the following therapies: -Selinexor -Belantamab -CAR-T
Subgroup 2	Only including the following regimens: -Carfilzomib+Dexamethasone -Panabinostat+Bortezomib+Dexamethasone -Carfilzomib+Lenalidomide+Dexamethasone -Carfilzomib+Cyclophosphamide+Dexamethasone -Pomalidomide+Dexamethasone -Pomalidomide+Cyclophosphamide+Dexamethasone
Subgroup 3	All regimens except those which include the following therapies: -Selinexor -Belantamab -Panobinostat -Bendamustine
Subgroup 4	All regimens except those which include the following therapies: -Panobinostat -Venetoclax -Selinexor -Hyaluronidase -Abiraterone -Trastuzumab -Dabrafenib -Autologous stem cell transplant
Subgroup 5	Only including the following regimens: -Dexamethasone/Prednisolone+Bortezomib -Cyclophosphamide+Dexamethasone/Prednisolone+Bortezomib -Bendamustine -Pomalidomide -Dexamethasone/Prednisolone+Pomalidomide

	-Dexamethasone+Pomalidomide+Daratumumab -Dexamethasone/Prednisolone+Pomalidomide+Isatuximab -Cyclophosphamide+Dexamethasone/Prednisolone+Pomalidomide -Dexamethasone/Prednisolone+Carfilzomib -Dexamethasone/Prednisolone+Carfilzomib+Daratumumab -Cyclophosphamide+Dexamethasone+Carfilzomib -Dexamethasone+Pomalidomide+Carfilzomib -Dexamethasone/Prednisolone+Panobinostat -Dexamethasone/Prednisolone+Bortezomib+Panobinostat -Daratumumab -Dexamethasone/Prednisolone+Daratumumab -Dexamethasone/Prednisolone+Lenalidomide+Daratumumab -Dexamethasone/Prednisolone+Bortezomib+Daratumumab -Daratumumab+Carfilzomib+Lenalidomide+Dexamethasone/ Prednisolone -Daratumumab+Bortezomib+Cyclophosphamide+Dexamethasone/ Prednisolone -Dexamethasone/Prednisolone+Elotuzumab -Dexamethasone/Prednisolone+Ixazomib -Isatuximab -Dexamethasone/Prednisolone+Carfilzomib+Isatuximab -Isatuximab+Dexamethasone+Pomalidomide+Carfilzomib -Any other regimen containing Isatuximab -Belantamab -Selinexor -Selinexor+Dexamethasone/Prednisolone -Venetoclax+Bortezomib+Dexamethasone/Prednisolone -Cyclophosphamide

	-Pegylated liposomal doxorubicin -Melphalan+Dexamethasone/Prednisolone -Adriamycine+Cyclophosphamide+Dexamethasone/Prednisolone -Cyclophosphamide+Adriaymicine/Pegylated liposomal doxorubicin+Dexamethasone -Vincristine+Adriamycine+Cyclophosphamide+Dexamethasone