Study Protocol P2 C1-001

25/09/2023

Version 3.2

TABLE OF CONTENTS

DOCUMENT HISTORY

Version	Date	Description
V1.0	05/05/2023	**First submission to EMA**
V2.0	08/06/2023	**Second version following comments from EMA**
V3.0	28/06/2023	**Third version following comments from EMA**
V3.1	08/08/2023	**Third version – EUPAS register number added**
V3.2	25/09/2023	**Third version – ACI Varha excluded from the study protocol**

Study Title	DARWIN EU® - Multiple myeloma: patient characterisation, treatments and survival in the period 2012-2022
Protocol version identifier	V3.2
Date of last version of protocol	25/09/2023
EU PAS register number	EUPAS105033
Active substance	N/A
Medicinal product	N/A
Research question and objectives	The overall objective of this study is to characterise patients with multiple myeloma diagnosed in the period 2012-2022. The specific objectives of this study are: 1. To describe demographic and clinical characteristics of patients with multiple myeloma at the time of diagnosis. 2. To describe multiple myeloma treatments (including combinations and regimen types, e.g. induction, maintenance, etc.). 3. To describe multiple myeloma treatment sequences. 4. To estimate the overall survival of incident multiple myeloma patients during the study period (2012-2022). All results will be reported by country/database, overall and stratified by age and sex when possible.
Country(-ies) of study	Estonia, France, Germany, Spain, The Netherlands
Author	Talita Duarte-Salles (tduarte@darwin-eu.org); Daniel Prieto-Alhambra (d.prietoalhambra@darwin-eu.org)

LIST OF ABBREVIATIONS

Acronyms/terms	Description
ADHD	Attention Deficit Hyperactivity Disorder
ASCT	Autologous Stem Cell Transplant
CAR	Chimeric Antigen Receptors
CDM	Common Data Model
CDWBordeaux	Clinical Data Warehouse of Bordeaux University Hospital
COPD	Chronic obstructive pulmonary disease
DA	Disease Analyzer
DARWIN EU®	Data Analysis and Real World Interrogation Network
DOI	Declaration Of Interests
EHR	Electronic Health Records
EMA	European Medicines Agency
EBB	Estonian Biobank
EGCUT	Estonian Genome Center at the University of Tartu
GERD	Gastro-esophageal reflux disease
GP	General Practitioner
IKNL	Netherlands Cancer Registry
IMASIS	Institut Municipal Assistència Sanitària Information System
IMiDs	Immunomodulatory agents
MGUS	Monoclonal Gammapathy of Unknown Significance
OMOP	Observational Medical Outcomes Partnership
PCT	Primary Care Teams
PI	Proteasome Inhibitors
PSMar	Parc Salut Mar
SIDIAP	Sistema d'Informació per al Desenvolupament de la Investigació en Atenció Primària
SINE	Selective inhibitors of nuclear export
SNOMED	Systematized Nomenclature of Medicine

1. TITLE

DARWIN EU® - Multiple myeloma: patient characterisation, treatments and survival in the period 2012-2022

2. RESPONSIBLE PARTIES – STUDY TEAM

Study team Role	Names	Organisation
Study Project Manager/Principal Investigator	Talita Duarte-Salles Daniel Prieto-Alhambra	Erasmus MC Erasmus MC/University of Oxford
Epidemiologist	Talita Duarte-Salles Katia Verhamme	Erasmus MC Erasmus MC
Clinical Domain Expert	Daniel Prieto-Alhambra	Erasmus MC/University of Oxford
Data Analysts/programmers	Maarten van Kessel Ross Williams Edward Burn	Erasmus MC Erasmus MC University of Oxford
Data Partner*	**Names**	**Organisation – Database**
Local Study Coordinator/Data Analyst	James Brash Hanne van Ballegooijen Núria Mercadé Miguel-Angel Mayer Angela Leis Juan Manuel Ramirez Raivo Kolde Romain Griffier Peter Prinsen	IQVIA - DA Germany IQVIA - DA Germany IDIAPJGol - SIDIAP PSMAR - IMASIS PSMAR - IMASIS PSMAR - IMASIS University Tartu - Estonian Biobank University of Bordeaux - CDWBordeaux Netherlands Cancer Registry - IKNL

*Data partners' role is only to execute code at their data source, review and approve their results. These people do not have an investigator role.

Data analysts/programmers do not have an investigator role and thus declaration of interests (DOI) for these people is not needed.

3. ABSTRACT (STAND ALONE SUMMARY OF THE STUDY PROTOCOL)

Title

DARWIN EU® - Multiple myeloma: patient characterisation, treatments and survival in the period 2012-2022.

Rationale and Background

Multiple myeloma is a rare type of blood cancer with an estimated overall crude and age-standardized incidence rates of 6.8 and 2.9 per 100,000 persons in 2020 in Europe, respectively (IARC – Cancer Today). Survival rates have improved due to the better management of the disease and the development in recent years of new medicines such as immunomodulatory agents, proteasome inhibitors and monoclonal

antibodies. However, there is still unmet need for new medicines for patients who do not respond to existing therapies.
The rarity of multiple myeloma makes it challenging to have a clear picture across Europe of the characteristics of these patients at the time of diagnosis, the different therapies they receive in subsequent lines and their overall survival. The goal of this study is to inform these aspects, which are important from a regulatory point of view to provide context and help understand the added value of the newest medicines under development or recently approved.

Research question and Objectives

The overall objective of this study is to characterise patients with multiple myeloma diagnosed in the period 2012-2022.
The specific objectives of this study are:

1. To describe demographic and clinical characteristics of patients with multiple myeloma at the time of diagnosis.
2. To describe multiple myeloma treatments (including combinations and regimen types, e.g. induction, maintenance, etc.).
3. To describe multiple myeloma treatment sequences.
4. To estimate the overall survival of incident multiple myeloma patients during the study period (2012-2022).

All results will be reported by country/database, overall and stratified by age and sex when possible.

Research Methods

Study design

Population-based cohort study.

Population

The study population will include all individuals identified in the database between 01/01/2012 and 31/12/2022. Participants with a diagnosis of cancer (any, excluding non-melanoma skin cancer) any time prior to the diagnosis of multiple myeloma will be excluded.

Additional eligibility criteria will be applied for each study objective: cohort 1) at least 365 days of prior history available before date of multiple myeloma diagnosis (=index date) will be applied for large-scale characterisation (objective 1), cohort 2) a minimum follow-up time of 30 days will be applied to capture cancer treatment (objective 2) and treatment sequences (objective 3), and, cohort 3) a minimum of 1 year of potential follow-up time for the survival analyses (objective 4).

Variables

Two main outcomes of interest will be studied: treatment/s initiated at index date, 1 to 30, 1 to 90 and/or 1 to 365 days post index date, and death. For the former, a pre-specified list of multiple myeloma treatments will be generated (Objectives 2-3). Overall survival in patients with incident multiple myeloma will also be calculated based on the registered date of death.

All co-morbidities and co-medications will be used for large-scale patient characterisation, identified as concept/code and descendants. A list of pre-specified co-morbidities and co-medications will also be described.

Data sources

1. IQVIA Disease Analyzer Germany (IQVIA DA Germany), Germany
2. Sistema d'Informació per al Desenvolupament de la Investigació en Atenció Primària (SIDIAP), Spain
3. Institut Municipal Assistencia Sanitaria Information System (IMASIS), Spain
4. Estonian Biobank, Estonia
5. Clinical Data Warehouse of Bordeaux University Hospital (CDWBordeaux), France
6. Netherlands Cancer Registry (IKNL), The Netherlands

Sample size

No sample size has been calculated as this is a descriptive Disease Epidemiology Study where we are interested in the characteristics of all incident multiple myeloma patients. Based on a preliminary feasibility assessment the expected number of multiple myeloma records in the included databases for this study will be approximately 88470.

Data analyses

Large-scale patient-level characterisation will be conducted (objective 1). Age and sex at time of multiple myeloma diagnosis will be described for each of the generated study cohorts. Medical history will be assessed for anytime –and up to 365 days before index date, for 365 to 31 days before index date, for 30 to 1 day before index date, and at index date. Medication use history will be reported for 365 to 31 days before index date, for 30 to 1 day before index date, and at index date. We will also report medication use for 1 to 30, 1 to 90, and 1 to 365 days post index date.

The number and % of patients receiving each of a pre-specified list of multiple myeloma treatments (see Appendix 1) and treatment combinations (objective 2) will be described at index date, 1 to 30, 1 to 90, and 1 to 365 days post index date. When available, treatment regimen types will also be described. Additionally, sunburst plots and Sankey diagrams will be used to describe treatment patterns and sequences over time (objective 3).

Overall survival (objective 4) will be calculated using data on time at risk of death from any cause and the Kaplan-Meier method. Results will be reported as plots of the estimated survival curves as well as the estimated probability of survival at years 1, 3, and 5. This analysis will be conducted only for databases that collect systematically data on mortality.

For all analyses n and % will be reported. A minimum cell count of 5 will be used when reporting results, with any smaller counts reported as "<5". All analyses will be reported by country/database, overall and stratified by age and sex when possible (minimum cell count reached). Additionally, in order to capture treatments availability and survival changes over time, sunburst plots, Sankey diagrams and KM curves will be further stratified by study periods (2012-2017 and 2018-2022).

4. AMENDMENTS AND UPDATES

Number	Date	Section of study protocol	Amendment or update	Reason
Version 2.0	08/06/2023	All	Update	Update following EMA's assessment
Version 3.0	28/06/2023	8.4, 8.7, Appendix II	Update	Update following EMA's assessment
Version 3.1	08/08/2023	Document history	Update	EUPAS register number added
Version 3.2	25/09/2023	3, 7, 8	Update	ACI Varha excluded from the protocol

5. MILESTONES

STUDY SPECIFIC DELIVERABLE	TIMELINE
Draft Study Protocol	5th May 2023
Final Study Protocol	1st June 2023
Creation of Analytical code	June 2023
Execution of Analytical Code on the data	July/August 2023
Interim Study Report (if applicable)	Not applicable
Draft Study Report	September 2023
Final Study Report	To be confirmed

6. RATIONALE AND BACKGROUND

Multiple myeloma is a hematologic malignancy that is characterized by the abnormal proliferation of plasma cells in the bone marrow with substantial morbidity and mortality. The most common presenting signs and symptoms of multiple myeloma are anemia, bone pain, renal impairment, fatigue, hypercalcemia, infection and weight loss(1). Globally, it accounts for approximately 1% of all cancer cases and 13% of hematologic malignancies. The incidence of multiple myeloma varies geographically, it is more common and rising in developed countries with the highest incidences in Asia, Europe, and North America. In Europe, the crude and age-standardized incidence rates of multiple myeloma were 6.8 and 2.9 per 100,000 persons in 2020, respectively, with a higher incidence observed in men than women and a median age at diagnosis of 72 years(2,3).

Survival rates of patients with multiple myeloma have improved in the past decades due to the better management of the disease and the development of new medicines such as immunomodulatory agents (IMiDs), proteasome inhibitors (PIs) and monoclonal antibodies(3). More recently, the approval of chimeric antigen receptors (CAR) T-cell therapies represents a promising immunotherapeutic approach in the treatment of advanced relapsed multiple myeloma(4).

To date, the standard treatment for multiple myeloma is largely dependent on patient health status, underlying comorbidities, tumour stage and age at time of diagnosis. For those younger than 70 to 75 years and no comorbidities, the preferred treatment comprises a triplet novel agent regimen, typically including an IMiD and PI in combination with glucocorticoids, followed by autologous stem cell transplant (ASCT) and maintenance therapy with low-dose IMiD or PI. Induction with novel agents and low-dose maintenance therapy is recommended for those patients unable to undergo ASCT(3,5). However, there is still unmet need for new medicines for patients who do not respond to existing therapies.

The rarity of multiple myeloma makes it challenging to have a clear picture across Europe of the characteristics of these patients at the time of diagnosis, the different therapies they receive in subsequent

lines and their overall survival. The goal of this study is to inform these aspects, which are important from a regulatory point of view to provide context and help understand how new medicines may add value for patients.

7. RESEARCH QUESTION AND OBJECTIVES

The overall objective of this study is to characterise patients with multiple myeloma diagnosed in the period 2012-2022.

The specific objectives of this study are:

1. To describe demographic and clinical characteristics of patients with multiple myeloma at the time of diagnosis.
2. To describe multiple myeloma treatments (including combinations and regimen types, e.g. induction, maintenance, etc.).
3. To describe multiple myeloma treatment sequences.
4. To estimate the overall survival of incident multiple myeloma patients during the study period (2012-2022).

All results will be reported by country/database, overall and stratified by age and sex when possible.

Table 1: Primary and secondary research questions and objective

Objective:	To describe demographic and clinical characteristics, treatments, treatment sequences, and overall survival of patients with incident multiple myeloma, stratified by age, sex, and country/database.
Hypothesis:	N/A
Population *(mention key inclusion-exclusion criteria):*	All individuals with a first diagnosis of multiple myeloma identified in the database between 01/01/2012 and 31/12/2022. Participants with a diagnosis of cancer (any, excluding non-melanoma skin cancer) any time prior to the diagnosis of multiple myeloma will be excluded. Additional eligibility criteria will be applied for each study objective: cohort 1) at least 365 days of prior history available before date of cancer diagnosis (=index date) will be applied for lar-scale characterisation (objective 1), cohort 2) a minimum follow-up time of 30 days will be applied to capture cancer treatment (objective 2) and treatment sequences (objective 3), and cohort 3) a minimum of 1 year of potential follow-up time for the survival analyses (objective 4).
Exposure:	N/A
Comparator:	N/A
Outcome:	Two main outcomes of interest will be studied: treatment/s initiated within 30, 90 and/or 365 days after diagnosis, and death. For the former, a pre-specified list of multiple myeloma treatments will be generated (objectives

	2-3). Overall, 1-, 3-, and 5-year survival in patients with multiple myeloma will also be calculated based on the registered date of death (objective 4).
Time *(when follow up begins and ends):*	Follow-up will start from date of first multiple myeloma diagnosis (index date) until the earliest of the following: 1) loss to follow-up, 2) end of data availability, or 3) date of death.
Setting:	Inpatient and outpatient setting from 6 databases in 5 European countries.
Main measure of effect:	Proportions and probability of survival

8. RESEARCH METHODS

8.1 Study type and Study Design

This will be a **patient-level characterisation** study classified as "off-the-shelf" (C1) and as described in the DARWIN EU® Complete Catalogue of Standard Data Analyses. A retrospective cohort study of all incident multiple myeloma cases will be conducted.

Table 2. Description of Potential Study Types and Related Study Designs

STUDY TYPE	STUDY DESIGN	STUDY CLASSIFICATION
Patient-level characterisation	Cohort analysis.	Off-the-shelf (C1)

8.2 Study Setting and Data Sources

This study will be conducted using routinely collected health data from 6 databases in 5 European countries. All databases were previously mapped to the OMOP CDM.

Data sources:

1. IQVIA Disease Analyzer Germany (IQVIA DA Germany), Germany
2. Sistema d'Informació per al Desenvolupament de la Investigació en Atenció Primària (SIDIAP), Spain
3. Institut Municipal Assistencia Sanitaria Information System (IMASIS), Spain
4. Estonian Biobank, Estonia
5. Clinical Data Warehouse of Bordeaux University Hospital (CDWBordeaux), France
6. Netherlands Cancer Registry (IKNL), The Netherlands

We selected 6 out of the 10 databases onboarded in DARWIN EU® in 2022. The selection of databases for this study was performed based on data reliability and relevance for the proposed research question. The selected databases fulfil the criteria required for a patient-level characterisation study allowing for large-scale characterisation, while covering different settings and regions of Europe.

Complete hospital-based cancer treatment data (needed for objectives 2 and 3) will be available in all databases except SIDIAP and IQVIA DA Germany. In turn, any potential outpatient therapies (e.g. pain management for bone complications) will be captured in these primary care datasets. Additionally, with the exception of IQVIA DA Germany, all databases have information on date of death (needed for objective 4).

The included databases also contain high quality data on multiple myeloma diagnosis. IKNL is a national oncological hospital registry in the Netherlands which includes data from all cancer patients in the country since 1989 (https://iknl.nl/en/ncr). The IMASIS database is linked to the hospital registry of tumours which is one of the oldest hospital-based cancer registries in Spain (created in 1978).(6) The EBB database is also regularly linked to the national cancer registry.(7) In SIDIAP, the diagnosis of multiple myeloma has been previously validated and a high sensitivity was reported (80%)(8). The diagnosis of multiple myeloma in French hospitals has also been validated with a reported sensitivity from 70-90%.(9)

https://genomics.ut.ee/en/content/estonian-biobankDetailed information on the selected data sources and their ability to answer the study research questions are described in **Table 3**.

Table 3. Description of the selected Data Sources

Country	Name of Database	Justification for Inclusion	Health Care setting	Type of Data	Number of active subjects	Data lock for the last update	Ability to answer study objectives
DE	IQVIA DA Germany	Covers primary care setting with information on cancer diagnoses.	Primary care and outpatient specialist care	EHR	8.5 million	31/12/2022	1
ES	SIDIAP	Covers primary care setting, data on cancer diagnoses previously validated, data on date of death available.	Primary care with hospital linkage	EHR	5.8 million	31/12/2022	1, 4
ES	IMASIS	Covers secondary care setting, database has information on cancer treatment in the in- and outpatient settings, mortality and other outcomes for in-house patients.	Secondary care (in and outpatients)	EHR	0.6 million	31/12/2022	1 to 4
ET	EBB	Contains information on 200,000 participants with not only genetic information but also health insurance claims, digital prescriptions, discharge information and causes of death through linkage with the national death register. Data is linked to cancer registry.	Biobank	Claims	0.2 million	31/12/2021	1 to 4
FR	CWDBordeaux	Covers secondary care setting, database has information on cancer treatment, mortality and other outcomes for in-house patients.	Secondary care (in and outpatients)	EHR	1.9 million	10/04/2023	1 to 4
NL	IKNL	Cancer registry data with high quality information on cancer diagnoses, mortality, and cancer treatment.	Cancer registry	Registry	3.5 million	31/03/2023 with incident cancer patients included up to 01/01/2022	2 to 4

DE = Germany, EBB = Estonian Biobank, ES = Spain, ET = Estonia, FR = France, NL = The Netherlands, SIDIAP = Sistema d'Informació per al Desenvolupament de la Investigació en Atenció Primària, IMASIS = Institut Municipal Assistencia Sanitaria Information System, DA = Disease Analyzer, CWDBordeaux = Clinical Data Warehouse of Bordeaux University Hospital, IKNL = Netherlands Cancer Registry.

IQVIA Disease Analyser (DA) Germany, Germany

DA Germany is collected from extracts of patient management software used by GPs and specialists practicing in ambulatory care settings(6). Data coverage includes more than 34M distinct person records out of at total population of 80M (42.5%) in the country and collected from 2,734 providers. Patient visiting more than one provider are not cross identified for data protection reasons and therefore recorded as separate in the system. Dates of service include from 1992 through present. Observation time is defined by the first and last consultation dates. Germany has no mandatory GP system and patient have free choice of specialist. As a result, data are collected from visits to 28.8% General, 13.4% Orthopaedic Surgery, 11.8% Otolaryngology, 11.2% Dermatology, 7.7% Obstetrics/Gynaecology, 6.2% various Neurology and Psychiatry 7.0% Paediatric, 4.6% Urology, 3.7% Cardiology, 3.5% Gastroenterology, 1.5% Pulmonary and 0.7% Rheumatology practices. Drugs are recorded as prescriptions of marketed products. No registration or approval is required for drug utilisation studies.

Information System for Research in Primary Care (SIDIAP), Spain

SIDIAP is collected from EHR records of patients receiving primary care delivered through Primary Care Teams, consisting of GPs, nurses and non-clinical staff(7). The Catalan Health Institute manages 328 out of 370 such Primary Care Teams with a coverage of 5.8M patients, out of 7.8M people in the Catalan population (74%). The database started to collect data in 2006. The mean follow-up is 15 years. The observation period for a patient can be the start of the database (2006), or when a person is assigned to a Catalan Health Institute primary care centre. Date of exit can be when a person is transferred-out to a primary care centre that does not pertain to the Catalan Health Institute, or date of death, or date of end of follow-up in the database. Drug information is available from prescriptions and from dispensing records in pharmacies. Drugs not prescribed in the GP setting might be underreported; and disease diagnoses made at specialist care settings are not included. Studies using SIDIAP data require previous approval by both a Scientific and an Ethics Committee.

Institut Municipal Assistència Sanitària Information System (IMASIS), Spain

The Institut Municipal Assistència Sanitària Information System (IMASIS) is the Electronic Health Record (EHR) system of Parc de Salut Mar Barcelona (PSMar) which is a complete healthcare services organisation. Currently, this information system includes and shares the clinical information of two general hospitals (Hospital del Mar and Hospital de l'Esperança), one mental health care centre (Centre Dr. Emili Mira) and one social-healthcare centre (Centre Fòrum) including emergency room settings, which are offering specific and different services in the Barcelona city area (Spain). At present, IMASIS includes clinical information more than 1 million patients with at least one diagnosis and who have used the services of this healthcare system since 1990 and from different settings such as admissions, outpatients, emergency room and major ambulatory surgery. The diagnoses are coded using The International Classification of Diseases ICD-9-CM and ICD-10-CM. The average follow-up period per patient in years is 6.37 (SD±6.82). IMASIS-2 is the anonymized relational database of IMASIS which is used for mapping to OMOP including additional sources of information such as the Tumours Registry.

Estonian Biobank (EBB), Estonia

The Estonian Biobank (EBB) is a population-based biobank of the Estonian Genome Center at the University of Tartu (EGCUT)(8). Its cohort size is currently close to 200,000 participants ("gene donors" ≥ 18 years of age), which closely reflects the age, sex and geographical distribution of the Estonian population. Estonians represent 83%, Russians 14%, and other nationalities 3% of all participants. Genomic GWAS analyses have

been performed on all gene donors. The database also covers health insurance claims, digital prescriptions, discharge reports, information about incident cancer cases, and causes of death from national sources for each donor.

Clinical Data Warehouse of Bordeaux University Hospital (CDWBordeaux), France

The clinical data warehouse of the Bordeaux University Hospital comprises electronic health records on more than 2 million patients with data collection starting in 2005. The hospital complex is made up of three main sites and comprises a total of 3,041 beds (2021 figures). The database currently holds information about the person (demographics), visits (inpatient and outpatient), conditions and procedures (billing codes), drugs (outpatient prescriptions and inpatient orders and administrations), measurements (laboratory tests and vital signs) and dates of death (in or out-hospital death)(9).

Netherlands Cancer Registry (IKNL), The Netherlands

The IKNL compiles clinical data of all individuals newly diagnosed with cancer in the Netherlands. Cancer registration clerks register newly diagnosed cancer patients since 1989 on a national basis, with 3 million patients included. Over the past 30 years, this registry has provided clinicians and researchers with a wealth of clinical data (e.g., patient and tumour characteristics, primary treatment, outcome) on cancer patients of all ages. See https://iknl.nl/en for more information.

8.3 Study Period

The study period will be from 01/01/2012 to end of available data in each of the data sources (see **Table 3** for more details).

8.4 Follow-up

Study participants will be followed up from their date of first multiple myeloma diagnosis (index date) until the earliest of the following: 1) loss to follow-up, 2) end of data availability, or 3) date of death. In the survival analysis, patients in 1) will be censored at the time of loss to follow-up. Patients in 2) will be administratively censored at end of data availability.

Table 4: Operational Definition of Time 0 (index date) and other primary time anchors

Study population name(s)	Time Anchor Description	Number of entries	Type of entry	Washout window	Care Setting[1]	Code Type[2]	Diagnosis position	Incident with respect to…	Measurement characteristics /validation	Source of algorithm
All patients with incident multiple myeloma eligible for the study	Date of incident multiple myeloma diagnosis	Single entry	Incident	Anytime prior to multiple myeloma diagnosis	IP, OP, OT	SNOMED	Any	Any cancer diagnosis except non-melanoma skin cancer	N/A	N/A

[1] IP = inpatient, OP = outpatient, OT = other, n/a = not applicable

8.5 Study Population with inclusion and exclusion criteria

The study population will include all individuals with a first diagnosis of multiple myeloma identified in the database between 01/01/2012 and 31/12/2022. Participants with a diagnosis of cancer (any, excluding non-melanoma skin cancer) any time prior to the diagnosis of multiple myeloma will be excluded (see https://atlas-demo.ohdsi.org/#/conceptset/1866029/included for a complete list of codes).

Multiple myeloma is a malignant uncontrolled proliferation of plasma cells derived from one single clone. It can affect many organs, typically bones and calcium metabolism, kidneys, immune system, blood, and more rarely neurologic. Diagnosis is typically obtained after bone marrow biopsy, where plasma cells, monoclonal kappa or lambda light chains will be present. The most important differential diagnosis is monoclonal gammapathy of unknown significance (MGUS) or 'smoldering multiple myeloma'. For this study, cases will be identified based on a record indicating a diagnosis or observation of multiple myeloma. Conditions in the OMOP CDM use the Systematized Nomenclature of Medicine (SNOMED) as the standard vocabulary for diagnosis codes. A clinical description along with a preliminary code list is provided in **Appendix 1**. This list was previously developed to define multiple myeloma for the DARWIN EU® study EUPAS50800.

Additional eligibility criteria will be applied for each study objective:

Cohort 1) for large-scale characterisation (objective 1), at least 365 days of prior history available before date of multiple myeloma diagnosis (=index date) will be required. This is needed to ensure a minimum prior observation time to exclude prevalent cases and to identify individuals' characteristics (e.g., comorbidities and drug history);

Cohort 2) for cancer treatments (objective 2) and treatment sequences (objective 3), a minimum follow-up time of 30 days will be applied to allow time to capture treatments, and finally;

Cohort 3) for survival analysis (objective 4), a minimum of 1 year of potential follow-up time will be required.

Table 5. Operational Definitions of Inclusion Criteria

Criterion	Details	Order of application	Assessment window	Care Settings[1]	Code Type	Diagnosis position	Applied to study populations:	Measurement characteristics /validation	Source for algorithm
Prior database history of 1 year (objective 1)	Study participants will be required to have a year of prior history observed before contributing observation time	After	1 year	IP, OP, OT	N/A	N/A	All study participants with incident multiple myeloma	N/A	N/A
Minimum potential follow-up (objectives 2 and 3)		After	30 days	IP, OP, OT	N/A	N/A	All study participants with incident multiple myeloma	N/A	N/A
Minimum potential follow-up time (objective 4)	Only participants with a diagnosis of multiple myeloma (index date) occurring one year prior to end of data availability in the database will be included	After	1 year	IP, OP, OT	N/A	N/A	All study participants with incident multiple myeloma	N/A	N/A

[1] IP = inpatient, OP = outpatient, OT = other, n/a = not applicable

Table 6. Operational Definitions of Exclusion Criteria

Criterion	Details	Order of application	Assessment window	Care Settings[1]	Code Type	Diagnosis position	Applied to study populations:	Measurement characteristics/validation	Source for algorithm
History of cancer diagnosis	Participants with a diagnosis of cancer (any, excluding non-melanoma skin cancer) any time prior to the diagnosis of multiple myeloma or prior to the start of the study period	After	Anytime prior to multiple myeloma diagnosis	IP, OP, OT	SNOMED	Any	All study participants with incident multiple myeloma	N/A	N/A

1 IP = inpatient, OP = outpatient, OT = other, n/a = not applicable

8.6 Variables

8.6.1. Exposure/s

N/A

8.6.2. Outcome/s

Two main outcomes of interest will be studied: treatment/s initiated within 30, 90 and/or 365 days after diagnosis, and death. For the former, a pre-specified list of multiple myeloma treatments will be generated (Objectives 2-3).

Multiple Myeloma treatments will include chemotherapies (melphalan, bendamustine, doxirubicin, cisplatin, cyclophosphamide, etoposide, vincristine), IMiDs (thalidomide, lenalidomide, pomalidomide), PI (bortezomib, carfilzomib, ixazomib, venetoclax), monoclonal antibodies (daratumumab, isatuximab, *denosumab, elotuzumab*), nuclear export inhibitor (selinexor), CAR T-cell (lisocabtagene maraleucel, idecabtagene vicleucel, brexucabtagene autoleucel, axicabtagene ciloleucel, tisagenlecleucel), glucocorticoids (dexamethasone, prednisone), bisphosphonates (zoledronate, pamidronate, clodronate, ibandronate, etidronate), and others (panobinostat)(3,10). Please check **Appendix 1 Table 2** for a preliminary list of codes to identify these treatments.

Overall survival in patients with incident multiple myeloma will also be calculated based on the registered date of death for the latter. Individuals will contribute with survival time as per the follow-up described in section 3.4.

Table 7. Operational Definitions of Outcome

Outcome name	Details	Primary outcome?	Type of outcome	Washout window	Care Settings[1]	Code Type	Diagnosis Position	Applied to study populations:	Measurement characteristics/ validation	Source of algorithm
Multiple myeloma treatments	Preliminary code lists provided in Appendix 1 Table 2	Yes	Counts	N/A	IP and OP care	RxNorm	N/A	All study participants with incident multiple myeloma	N/A	N/A
Overall survival		Yes	Time	N/A	IP and OP care	Date of death	N/A	All study participants with incident multiple myeloma	N/A	N/A

[1] IP = inpatient, OP = outpatient, n/a = not applicable

8.6.3. Other covariates, including confounders, effect modifiers and other variables

Age at multiple myeloma diagnosis will be described. The following age grouping will be used: 0-17; 18-44; 45-59; 60-69; 70 and over. The sex (male/ female) of study participants will also be identified.

All co-morbidities and co-medications recorded prior and at index date will be used for large-scale patient characterisation, identified as concept/code and descendants. Additionally, a list of pre-specified co-morbidities and co-medications will be described. These will include:

- Medical History: Asthma, COPD, Chronic Liver disease, Crohn's Disease, Diabetes mellitus, Gastro-esophageal reflux disease (GERD), GI-Bleeding, Human Immunodeficiency Virus (HIV), Hyperlipidemia, Hypertension, Obesity, Osteoarthritis, Pneumonia, Psoriasis, Renal impairment, Ulcerative Collitis, Urinary Tract infection, Viral Hepatitis, Visual system disorder [General] -- Schizophrenia, Dementia, Parkinson, Depressive disorder, Anxiety, Attention Deficit Hyperactivity Disorder (ADHS) [Neurology] --- Any cancer except non-melanoma skin cancer (for quality assessment purposes, this should be 0 in our study population before index date).
- Medication use: Agents acting on the renin-angiotensin system, Antibacterials for systemic use, Antidepressants, Antiepileptics, Antiinflammatory and antirheumatic products, Antineoplastic agents, Antipsoriatics, Antithrombotic agents, Beta blocking agents, Calcium channel blockers, Diuretics, Drugs for acid related disorders, Drugs for obstructive airway diseases, Drugs used in diabetes, Immunosuppressants, Lipid modifying agents, Opioids, Psycholeptics, Psychostimulants, agents used for ADHD and nootropics [General] -- contraceptives [contraceptives].

Table 8. Operational Definitions of Covariates

Characteristic	Details	Type of variable	Assessment window	Care Settings[1]	Code Type	Diagnosis Position	Applied to study populations:	Measurement characteristics /validation	Source for algorithm
Co-morbidities	Large-scale patient-level characterisation with regard to underlying comorbidities	Counts	At index date (ID), for 30 to 1 day before ID, for 365 to 31 days before ID, at any time and up to 366 days before ID	OP, IP, OT	SNOMED	N/A	N/A	N/A	N/A
Concomitant medication	Large-scale patient-level characterisation with regard to use of concomitant drugs	Counts	At index date (ID), for 30 to 1 day before ID, for 365 to 31 days before ID, 1 to 30 post ID, 1 to 90 post ID, and 1 to 365 days post ID	OP, IP, OT	RxNorm	N/A	N/A	N/A	N/A

[1] IP = inpatient, OP = outpatient, OT = other, n/a = not applicable

8.7 Study size

No sample size has been calculated as this is a descriptive Disease Epidemiology Study where we are interested in the characteristics of all incident multiple myeloma patients. Based on a preliminary feasibility assessment the expected number of multiple myeloma records in the included databases for this study will be approximately 88470. Please note that this number is based on the overall number of multiple myeloma registers in each database with no filter by study period or inclusion and exclusion criteria.

8.8 Analysis

Table 9. Description of Study Types and Type of analysis

STUDY TYPE	STUDY CLASSIFICATION	TYPE OF ANALYSIS
Disease epidemiology - Patient-level characterisation	Off-the-shelf (C1)	- large-scale characterisation - patient-level characteristics - Progression to a pre-specified outcome - Standard care description

8.8.1 Federated Network Analyses

Analyses will be conducted separately for each database. Before study initiation, test runs of the analytics are performed on a subset of the data sources or on a simulated set of patients and quality control checks are performed. Once all the tests are passed, the final package is released in the version-controlled Study Repository for execution against all the participating data sources.

The data partners locally execute the analytics against the OMOP-CDM in R Studio and review and approve the by default aggregated results before returning them to the Coordination Centre. Sometimes multiple execution iterations are performed, and additional fine tuning of the code base is needed. A service desk will be available during the study execution for support.

The study results of all data sources are checked after which they are made available to the team in the Digital Research Environment (DRE) and the Study Dissemination Phase can start. All results are locked and timestamped for reproducibility and transparency.

8.8.2 Patient privacy protection

Cell suppression will be applied as required by databases to protect people's privacy. Cell counts < 5 will be masked.

8.8.3 Statistical model specification and assumptions of the analytical approach considered

R-packages

We will use the R packages "CohortDiagnostics" for the patient-level characterization of demographics and clinical characteristics, "TreatmentPatterns" for the patient-level characterisation of treatments, and "CohortSurvival" for the estimation of overall survival.

Patient-level characterisation

Large-scale patient-level characterisation will be conducted (objective 1). Age and sex at time of MM diagnosis will be described for each of the generated study cohorts. The index date will be the date of the multiple myeloma diagnosis for each patient. Medical history will be assessed for anytime –and up to 365 days before index date, for 365 to 31 days before index date, for 30 to 1 day before index date, and at index date. Medication use history will be reported for 365 to 31 days before index date, for 30 to 1 day before index date, and at index date. We will also report medication use for 1 to 30, 1 to 90, and 1 to 365 days post index date. These time windows were defined based on the options currently available in the standard analytical tools that will be used in this project. Co-variates to be presented in a summary baseline characteristics table will be pre-defined as described in section 8.6.3.

The number and % of patients receiving each of a pre-specified list of multiple myeloma treatments (objective 2, as listed in section 8.6.2) will be described at index date, 1 to 30, 1 to 90, and 1 to 365 days post index date. When available, treatment regimen types will also be described. Additionally, sunburst plots and Sankey diagrams will be used to describe treatment patterns over time (objective 3). Sankey diagrams will be censured at end of treatment or end of follow-up as described in section 8.

Overall survival (objective 4) will be calculated using data on time at risk of death from any cause and the Kaplan-Meier (KM) method. Results will be reported as plots of the estimated survival curves as well as the estimated probability of survival at years 1, 3, and 5. Individuals who are lost to follow-up will be censored at the time of loss of follow-up. The KM approach implicitly assumes censoring occurs at random. This analysis will be conducted only for databases that collect systematically data on mortality (all except IQVIA DA Germany).

For all analyses n and % will be reported. A minimum cell count of 5 will be used when reporting results, with any smaller counts reported as "<5". All analyses will be reported by country/database, overall and stratified by age and sex when possible (minimum cell count reached). Additionally, in order to capture treatments availability and survival changes over time, sunburst plots, Sankey diagrams and 1-year survival will be further stratified by study periods (2012-2017 and 2018-2022).

8.9 Evidence synthesis

Results from analyses described in section 8.8 will be presented separately for each database and no meta-analysis of results will be conducted.

9. DATA MANAGEMENT

9.1 Data management

All databases are mapped to the OMOP common data model. This enables the use of standardised analytics and tools across the network since the structure of the data and the terminology system is harmonised. The OMOP CDM is developed and maintained by the Observational Health Data Sciences and Informatics (OHDSI) initiative and is described in detail on the wiki page of the CDM: https://ohdsi.github.io/CommonDataModel and in The Book of OHDSI: http://book.ohdsi.org

The analytic code for this study will be written in R. Each data partner will execute the study code against their database containing patient-level data and will then return the results set which will only contain aggregated data. The results from each of the contributing data sites will then be combined in tables and figures for the study report.

9.2 Data storage and protection

For this study, participants from various EU member states will process personal data from individuals which is collected in national/regional electronic health record databases. Due to the sensitive nature of this personal medical data, it is important to be fully aware of ethical and regulatory aspects and to strive to take all reasonable measures to ensure compliance with ethical and regulatory issues on privacy.

All databases used in this study are already used for pharmaco-epidemiological research and have a well-developed mechanism to ensure that European and local regulations dealing with ethical use of the data and adequate privacy control are adhered to. In agreement with these regulations, rather than combining person level data and performing only a central analysis, local analyses will be run, which generate non-identifiable aggregate summary results.

The output files are stored in the DARWIN Digital Research Environment. These output files do not contain any data that allow identification of subjects included in the study. The DRE implements further security measures in order to ensure a high level of stored data protection to comply with the local implementation of the General Data Protection Regulation (GDPR) (EU) 679/20161 in the various member states.

10. QUALITY CONTROL

General database quality control

A number of open-source quality control mechanisms for the OMOP CDM have been developed (see Chapter 15 of The Book of OHDSI http://book.ohdsi.org/DataQuality.html). In particular, it is expected that data partners will have run the OHDSI Data Quality Dashboard tool (https://github.com/OHDSI/DataQualityDashboard). This tool provides numerous checks relating to the conformance, completeness and plausibility of the mapped data. Conformance focuses on checks that describe the compliance of the representation of data against internal or external formatting, relational, or computational definitions, completeness in the sense of data quality is solely focused on quantifying missingness, or the absence of data, while plausibility seeks to determine the believability or truthfulness of data values. Each of these categories has one or more subcategories and are evaluated in two contexts: validation and verification. Validation relates to how well data align with external benchmarks with expectations derived from known true standards, while verification relates to how well data conform to local knowledge, metadata descriptions, and system assumptions.

Study specific quality control

When defining multiple myeloma, a systematic search of possible codes for inclusion was previously identified using CodelistGenerator R package (https://github.com/darwin-eu/CodelistGenerator). This software allows the user to define a search strategy and using this will then query the vocabulary tables of the OMOP CDM so as to find potentially relevant codes. The codes returned were then reviewed by two clinical epidemiologist to consider their relevance. This work was done previously as part of study the

DARWIN EU® study EUPAS50800. In addition, the CohortDiagnostics R package (https://github.com/OHDSI/CohortDiagnostics) will be run to assess the use of different codes across the databases contributing to the study and identify any codes potentially omitted in error. This will allow for a consideration of the validity of the study cohort of patients with multiple myeloma in each of the databases, and inform decisions around whether multiple definitions are required.

The study code will be based on three R packages currently being developed to (1) characterise demographic and clinical characteristics, (2) characterise treatment patterns, and (3) estimate overall survival using the OMOP CDM. These packages will include numerous automated unit tests to ensure the validity of the codes, alongside software peer review and user testing. The R package will be made publicly available via GitHub.

11. LIMITATIONS OF THE RESEARCH METHODS

The study will be informed by routinely collected health care data and so data quality issues must be considered. In particular, the identification of multiple myeloma patients and the recording of the co-morbidities may vary across databases and while relatively few false positives would be expected, false negatives may be more likely especially for databases without patient-level linkage to secondary care data. Nevertheless, the diagnosis of multiple myeloma has been previously validated in the SIDIAP database(8), and databases from or linked to cancer registries (IKNL, IMASIS, and EBB) are known to contain high quality data on cancer diagnoses.

Since date of death is not available in IQVIA DA Germany, overall survival will not be estimated in this database. In addition, the recording of events used for patient characterisation may vary across databases; and in databases with information on cancer treatment, the recording of treatment use may be incomplete. This may occur particularly for new available treatments such as CAR T-cell drugs which might not have been mapped to the OMOP CDM.

12. MANAGEMENT AND REPORTING OF ADVERSE EVENTS/ADVERSE REACTIONS

In agreement with the guideline on good pharmacovigilance practice (EMA/873138/2011), there will be no requirement for expedited reporting of adverse drug reactions as only secondary data will be used in this study.

13. GOVERNANCE BOARD ASPECTS

All data sources require approval from their respective IRB boards, with the exception of IQVIA DA Germany which will not require any further specific approvals to undertake this study.

14. PLANS FOR DISSEMINATING AND COMMUNICATING STUDY RESULTS

14.0 Study Report

A PDF report including an executive summary, and the specified tables and/or figures will be submitted to EMA by the DARWIN EU® CC upon completion of the study.

An interactive dashboard incorporating all the results (tables and figures) will be provided alongside the pdf report. The full set of underlying aggregated data used in the dashboard will also be made available if requested.

15. OTHER ASPECTS

N/A

16. REFERENCES

1. Kyle RA, Gertz MA, Witzig TE, Lust JA, Lacy MQ, Dispenzieri A, et al. Review of 1027 patients with newly diagnosed multiple myeloma. Mayo Clin Proc. 2003 Jan;78(1):21–33.

2. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021 May;71(3):209–49.

3. Dimopoulos MA, Moreau P, Terpos E, Mateos MV, Zweegman S, Cook G, et al. Multiple myeloma: EHA-ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up†. Ann Oncol. 2021 Mar;32(3):309–22.

4. Gagelmann N, Sureda A, Montoto S, Murray J, Bolaños N, Kenyon M, et al. Access to and affordability of CAR T-cell therapy in multiple myeloma: an EBMT position paper. Lancet Haematol. 2022 Oct;9(10):e786–95.

5. Rajkumar SV. Multiple myeloma: 2016 update on diagnosis, risk-stratification, and management. Am J Hematol. 2016 Jul;91(7):719–34.

6. Planas Domingo J, Gallén Castillo M, Malats Riera N, Porta Serra M, Guasch Jordán I, Cardona Hernández T, et al. [Tumors registered in Hospital del Mar (Barcelona). Descriptive analysis from 1978 to 1986]. Rev Clin Esp. 1988 Sep;183(4):175–9.

7. Leitsalu L, Haller T, Esko T, Tammesoo ML, Alavere H, Snieder H, et al. Cohort Profile: Estonian Biobank of the Estonian Genome Center, University of Tartu. Int J Epidemiol. 2015 Aug;44(4):1137–47.

8. Recalde M, Manzano-Salgado CB, Díaz Y, Puente D, Garcia-Gil MDM, Marcos-Gragera R, et al. Validation Of Cancer Diagnoses In Electronic Health Records: Results From The Information System For Research In Primary Care (SIDIAP) In Northeast Spain. Clin Epidemiol. 2019 Dec 3;11:1015–24.

9. Palmaro A, Gauthier M, Conte C, Grosclaude P, Despas F, Lapeyre-Mestre M. Identifying multiple myeloma patients using data from the French health insurance databases: Validation using a cancer registry. Medicine (Baltimore). 2017 Mar;96(12):e6189.

10. Rathmann W, Bongaerts B, Carius HJ, Kruppert S, Kostev K. Basic characteristics and representativeness of the German Disease Analyzer database. Int J Clin Pharmacol Ther. 2018 Oct;56(10):459–66.

11. Recalde M, Rodríguez C, Burn E, Far M, García D, Carrere-Molina J, et al. Data Resource Profile: The Information System for Research in Primary Care (SIDIAP). Int J Epidemiol. 2022 Dec 13;51(6):e324–36.

12. Brat GA, Weber GM, Gehlenborg N, Avillach P, Palmer NP, Chiovato L, et al. International electronic health record-derived COVID-19 clinical course profiles: the 4CE consortium. Npj Digit Med. 2020 Aug 19;3(1):109.

13. Mhaskar R, Kumar A, Miladinovic B, Djulbegovic B. Bisphosphonates in multiple myeloma: an updated network meta-analysis. Cochrane Haematological Malignancies Group, editor. Cochrane Database Syst Rev [Internet]. 2017 Dec 18 [cited 2023 Apr 10];2017(12). Available from: http://doi.wiley.com/10.1002/14651858.CD003188.pub4

17. ANNEXES

Appendix I: Definition of Multiple Myeloma Diagnosis and Treatments

Clinical description

Overview

Multiple myeloma (MM) is a malignant uncontrolled proliferation of plasma cells derived from one single clone. MM can affect many organs, typically bones and calcium metabolism (e.g. fractures or bone pain, hypercalcemia), kidneys (renal failure/ckd/aki), immune system (e.g. infection), blood (anemia, clotting), and more rarely neurologic.

The etiology of MM is unknown, and it is one of the most common blood cancers, with an estimated crude and age-standardized incidence rates of 6.8 and 2.9 per 100,000 persons in 2020 in Europe, respectively, with a higher incidence observed in men than women and a median age at diagnosis of 72 years(2,3).

Presentation

Bone pain, fractures, and hypercalcemia are the most common manifestations of MM. Radiology would show (on top of fracture/pathologic fracture where present) lytic/osteolytic lesions in plain x-ray or ct scan. Bone lysis results in hypercalcemia, and related complications.

Other typical features that lead to diagnosis include renal failure (ckd or aki) or bacterial infections like pneumonia or pyelonephritis. Recurrent infections are common.

Bloods and urine labs are typically used to investigate/diagnose MM, and will show hypogammaglobulinemia, bence-jones proteins, hypercalcemia, and other alterations of proteinogram.

Assessment

Assessment will typically include imaging e.g. x-ray, ct scan, and sometimes bone gammagraphy or similar techniques. Also, bloods and urine are included often during investigation/diagnosis, typically including full blood counts, biochemistry, immunology, proteinogram, serum electrophoresis, β2Microglobulin, LDH and serum albumin, and bence jones proteins

Diagnosis is typically obtained after bone marrow biopsy, where plasma cells monoclonal kappa or lambda light chains will be present.

The most important differential diagnosis is MGUS (monoclonal gammapathy of unknown significance) or 'smoldering multimple myeloma'. These are far more common than MM, and sometimes preclude or transition to MM proper. It is estimated that about 1% of patients with MGUS transition every year to MM. Patients with MGUS would not have a bone marrow biopsy, but they will/should have repeat bloods over time to monitor/rule out MM.

Diagnostic Criteria

Monoclonal Gammopathy of Undetermined Significance (MGUS)

- Serum monoclonal protein (non-IgM type) <30 g/L
- Clonal bone marrow plasma cells <10%

- Absence of myeloma-defining events or amyloidosis that can be attributed to the plasma cell proliferative disorder

Smoldering Multiple Myeloma (Asymptomatic Myeloma)
Both criteria must be met:
-Serum monoclonal protein (IgG or IgA) ≥30 g/L or urinary monoclonal protein ≥500 mg per 24 h and/or clonal bone marrow plasma cells 10–60%
-Absence of myeloma-defining events or amyloidosis

Symptomatic Multiple Myeloma Clonal bone marrow plasma cells or biopsy-proven bony or extramedullary plasmacytomaa and any one or more of the following myeloma-defining events:

Evidence of one or more indicators of end-organ damage that can be attributed to the underlying plasma cell proliferative disorder, specifically:

- Hypercalcemia: serum calcium >0.25 mmol/L (>1 mg/dL) higher than the upper limit of normal or >2.75 mmol/L (>11 mg/dL)
- Renal insufficiency: creatinine clearance <40 mL/minb or serum creatinine >177 µmol/L (>2 mg/dL)
- Anemia: hemoglobin value of >20 g/L below the lower limit of normal, or a hemoglobin value <100 g/L
- Bone lesions: one or more osteolytic lesions on skeletal radiography, CT, or PET-CTc
- Any one or more of the following biomarkers of malignancy: i.Clonal bone marrow plasma cell percentagea ≥60%; ii.Involved: uninvolved serum free light chain ratiod ≥100; and/or iii.>1 focal lesion on MRI studies

Plan Common treatments for MM include:

- Immunomodulatory drugs (IMiD) like Thalidomide, Lenalidomide, Pomalidomide
- Proteasome inhibitors (PI) like Bortezomib, Carfilzomib, Ixazomib
- Antibodies/monoclonal antibodies like Daratumumab, Elotuzumab, Isatuximab, Belantamab mafodotin
- Selective inhibitors of nuclear export (SINE) like Selinexor
- Histone deacetylase inhibitors like Panobinostat
- Alkylating agents like Melphalan, Cyclophosphamide, Bendamustine, Melflufen
- Cellular therapy like Idecabtagene vicleucel
- Glucocorticoids like Dexamethasone or Prednisone

Prognosis
Medial survival ranges from 30 to 60 months, with some advanced diagnoses surviving for no more than 6 months.

Strengtheners
Any of the treatments mentioned above, use of imaging and specific bloods also mentioned above.

Table 1: Code list for multiple myeloma.

Id	Name
4224628	Amyloid light chain amyloidosis due to multiple myeloma
4258135	Asymptomatic multiple myeloma
4043447	Bone marrow: myeloma cells
4094548	Extramedullary plasmacytoma
37209514	Hypogammaglobulinemia due to multiple myeloma
4111355	IgA myeloma
4112310	IgD myeloma
4111356	IgG myeloma
4259972	Indolent multiple myeloma
4188299	Kappa light chain myeloma
4197600	Lambda light chain myeloma
4082464	Light chain myeloma
37016161	Light chain nephropathy due to multiple myeloma
437233	Multiple myeloma
4210177	Multiple myeloma
436059	Multiple myeloma in remission
4214660	Multiple solitary plasmacytomas
4019477	Myeloma-associated amyloidosis
4137433	Myeloma kidney
4043713	Neuropathy due to multiple myeloma
4079684	Non-secretory myeloma
42538151	Osteoporosis co-occurrent and due to multiple myeloma
4137510	Osteosclerotic myeloma
133154	Plasma cell leukemia
4028859	Plasma cell leukemia
760936	Plasma cell leukemia in relapse
133158	Plasma cell leukemia in remission
4190641	Plasma cell myeloma - category
4190642	Plasma cell myeloma/plasmacytoma
4163558	Plasma cell myeloma/plasmacytoma
4216139	Plasmacytoma
4024874	Plasmacytoma
4300702	Primary cutaneous plasmacytoma
764229	Relapse multiple myeloma

4184985	Smoldering myeloma
4145040	Solitary osseous myeloma

Multiple Myeloma Treatments

Table 2: Preliminary Code list for Multiple Myeloma treatments.

Class	Treatment	ATC code	ConceptID
Chemotherapies	Melphalan	L01AA03	21601392
	Bendamustine	L01AA09	21601397
	Doxorubicin	L01DB01	21603732
	Cisplatin	L01XA01	21603748
	Cyclophosphamide	L01AA01	21601390
	Etoposide	L01CB01	21603718
	Vincristine	L01CA02	21601448
IMiDs	Thalidomide	L04AX02	21603928
	Lenalidomide	L04AX04	21603930
	Pomalidomide	L04AX06	43534824
PI	Bortezomib	L01XX32	21603804
	Carfilzomib	L01XG02	947958
	Ixazomib	L01XG03	947997
	Venetoclax	L01XX52	1123676
Monoclonal antibodies	Daratumumab	L01XC24	1123898
	Isatuximab	L01XC38	947956
	Denosumab	M05BX04	21604169
	Elotuzumab	L01XC23	1123776
Nuclear export inhibitor	Selinexor	L01XX66	715834
CAR T-cell	Lisocabtagene maraleucel	L01XX88	739860
	Idecabtagene vicleucel	L01XL07	36026872
	Brexucabtagene autoleucel	L01XL06	37002393
	Axicabtagene ciloleucel	L01XX70	947849
	Tisagenlecleucel	L01XX71	947930
Glucocorticoids	Dexamethasone	H02AB02	21602730
	Prednisone	H02AB07	21602735
Bisphosphonates	Zoledronic acid	M05BA08	21604156
	Pamidronic acid	M05BA03	21604151
	Clodronic acid	M05BA02	21604150
	Ibandronic acid	M05BA06	21604154
	Etidronate	M05BA01	21604149
Others	Panobinostat	L01XH03	947750

Appendix II: ENCePP checklist for study protocols

Study title:
DARWIN EU® - Multiple myeloma: patient characterisation, treatments and survival in the period 2012-2022

EU PAS Register® number: N/A
Study reference number (if applicable): N/A

Section 1: Milestones	Yes	No	N/A	Section Number
1.1 Does the protocol specify timelines for				
1.1.1 Start of data collection[1]	☐	☒	☐	5
1.1.2 End of data collection[2]	☒	☐	☐	
1.1.3 Progress report(s)	☐	☐	☒	
1.1.4 Interim report(s)	☐	☐	☒	
1.1.5 Registration in the EU PAS Register®	☐	☒	☐	
1.1.6 Final report of study results.	☒	☐	☐	

Comments:

Section 2: Research question	Yes	No	N/A	Section Number
2.1 Does the formulation of the research question and objectives clearly explain:	☐	☐	☒	6, 7
2.1.1 Why the study is conducted? (e.g. to address an important public health concern, a risk identified in the risk management plan, an emerging safety issue)	☒	☐	☐	
2.1.2 The objective(s) of the study?	☒	☐	☐	
2.1.3 The target population? (i.e. population or subgroup to whom the study results are intended to be generalised)	☒	☐	☐	
2.1.4 Which hypothesis(-es) is (are) to be tested?	☐	☐	☒	
2.1.5 If applicable, that there is no *a priori* hypothesis?	☐	☐	☒	

Comments:

[1] Date from which information on the first study is first recorded in the study dataset or, in the case of secondary use of data, the date from which data extraction starts.
[2] Date from which the analytical dataset is completely available.

Section 3: Study design		**Yes**	**No**	**N/A**	**Section Number**
3.1	Is the study design described? (e.g., cohort, case-control, cross-sectional, other design)	☒	☐	☐	8.1
3.2	Does the protocol specify whether the study is based on primary, secondary or combined data collection?	☒	☐	☐	8.2
3.3	Does the protocol specify measures of occurrence? (e.g., rate, risk, prevalence)	☒	☐	☐	8.8
3.4	Does the protocol specify measure(s) of association? (e.g., risk, odds ratio, excess risk, rate ratio, hazard ratio, risk/rate difference, number needed to harm (NNH))	☐	☐	☒	
3.5	Does the protocol describe the approach for the collection and reporting of adverse events/adverse reactions? (e.g. adverse events that will not be collected in case of primary data collection)	☐	☐	☒	

Comments:

This is a descriptive study and no measure of association or collection or reporting of adverse events/reactions will be reported.

Section 4: Source and study populations		**Yes**	**No**	**N/A**	**Section Number**
4.1	Is the source population described?	☒	☐	☐	8.2/8.5
4.2	Is the planned study population defined in terms of:				8.5
	4.2.1 Study time period	☒	☐	☐	
	4.2.2 Age and sex	☒	☐	☐	
	4.2.3 Country of origin	☒	☐	☐	
	4.2.4 Disease/indication	☒	☐	☐	
	4.2.5 Duration of follow-up	☒	☐	☐	
4.3	Does the protocol define how the study population will be sampled from the source population? (e.g., event or inclusion/exclusion criteria)	☒	☐	☐	8.5

Comments:

Section 5: Exposure definition and measurement		**Yes**	**No**	**N/A**	**Section Number**
5.1	Does the protocol describe how the study exposure is defined and measured? (e.g. operational details for defining and categorising exposure, measurement of dose and duration of drug exposure)	☐	☐	☒	

Section 5: Exposure definition and measurement	**Yes**	**No**	**N/A**	**Section Number**
5.2 Does the protocol address the validity of the exposure measurement? (e.g., precision, accuracy, use of validation sub-study)	☐	☐	☒	
5.3 Is exposure categorised according to time windows?	☐	☐	☒	
5.4 Is intensity of exposure addressed? (e.g., dose, duration)	☐	☐	☒	
5.5 Is exposure categorised based on biological mechanism of action and taking into account the pharmacokinetics and pharmacodynamics of the drug?	☐	☐	☒	
5.6 Is (are) (an) appropriate comparator(s) identified?	☐	☐	☒	

Comments:

No exposures are described. The use of medicines is described as outcomes in this protocol.

Section 6: Outcome definition and measurement	**Yes**	**No**	**N/A**	**Section Number**
6.1 Does the protocol specify the primary and secondary (if applicable) outcome(s) to be investigated?	☒	☐	☐	8.6
6.2 Does the protocol describe how the outcomes are defined and measured?	☒	☐	☐	8.6
6.3 Does the protocol address the validity of outcome measurement? (e.g. precision, accuracy, sensitivity, specificity, positive predictive value, use of validation sub-study)	☒	☐	☐	8.6
6.4 Does the protocol describe specific outcomes relevant for Health Technology Assessment? (e.g. HRQoL, QALYs, DALYS, health care services utilisation, burden of disease or treatment, compliance, disease management)	☐	☐	☒	

Comments:

Section 7: Bias	**Yes**	**No**	**N/A**	**Section Number**
7.1 Does the protocol address ways to measure confounding? (e.g., confounding by indication)	☐	☐	☒	
7.2 Does the protocol address selection bias? (e.g. healthy user/adherer bias)	☐	☐	☒	

Section 7: Bias	**Yes**	**No**	**N/A**	**Section Number**
7.3 Does the protocol address information bias? (e.g. misclassification of exposure and outcomes, time-related bias)	☒	☐	☐	

Comments:

This is a descriptive study and no associations will be assessed.

Section 8: Effect measure modification	**Yes**	**No**	**N/A**	**Section Number**
8.1 Does the protocol address effect modifiers? (e.g., collection of data on known effect modifiers, sub-group analyses, anticipated direction of effect)	☐	☐	☒	

Comments:

This is a descriptive study and no associations will be assessed.

Section 9: Data sources	**Yes**	**No**	**N/A**	**Section Number**
9.1 Does the protocol describe the data source(s) used in the study for the ascertainment of:				
9.1.1 Exposure? (e.g., pharmacy dispensing, general practice prescribing, claims data, self-report, face-to-face interview)	☐	☐	☒	
9.1.2 Outcomes? (e.g., clinical records, laboratory markers or values, claims data, self-report, patient interview including scales and questionnaires, vital statistics)	☒	☐	☐	8.6
9.1.3 Covariates and other characteristics?	☒	☐	☐	8.6
9.2 Does the protocol describe the information available from the data source(s) on:				
9.2.1 Exposure? (e.g. date of dispensing, drug quantity, dose, number of days of supply prescription, daily dosage, prescriber)	☐	☐	☒	
9.2.2 Outcomes? (e.g. date of occurrence, multiple event, severity measures related to event)	☒	☐	☐	8.6
9.2.3 Covariates and other characteristics? (e.g., age, sex, clinical and drug use history, co-morbidity, co-medications, lifestyle)	☒	☐	☐	8.6
9.3 Is a coding system described for:				
9.3.1 Exposure? (e.g. WHO Drug Dictionary, Anatomical Therapeutic Chemical (ATC) Classification System)	☐	☐	☒	
9.3.2 Outcomes? (e.g., International Classification of Diseases (ICD), Medical Dictionary for Regulatory Activities (MedDRA))	☒	☐	☐	8.6
9.3.3 Covariates and other characteristics?	☒	☐	☐	8.6

Section 9: Data sources	**Yes**	**No**	**N/A**	**Section Number**
9.4 Is a linkage method between data sources described? (e.g. based on a unique identifier or other)	☐	☐	☒	

Comments:

Section 10: Analysis plan	**Yes**	**No**	**N/A**	**Section Number**
10.1 Are the statistical methods and the reason for their choice described?	☒	☐	☐	8.8
10.2 Is study size and/or statistical precision estimated?	☐	☐	☒	8.7
10.3 Are descriptive analyses included?	☒	☐	☐	8.8
10.4 Are stratified analyses included?	☒	☐	☐	8.8
10.5 Does the plan describe methods for analytic control of confounding?	☐	☐	☒	
10.6 Does the plan describe methods for analytic control of outcome misclassification?	☐	☐	☒	
10.7 Does the plan describe methods for handling missing data?	☐	☒	☐	
10.8 Are relevant sensitivity analyses described?	☒	☐	☐	8.8

Comments:

Section 11: Data management and quality control	**Yes**	**No**	**N/A**	**Section Number**
11.1 Does the protocol provide information on data storage? (e.g., software and IT environment, database maintenance and anti-fraud protection, archiving)	☒	☐	☐	9.2
11.2 Are methods of quality assurance described?	☒	☐	☐	10
11.3 Is there a system in place for independent review of study results?	☐	☐	☒	

Comments:

Section 12: Limitations	**Yes**	**No**	**N/A**	**Section Number**
12.1 Does the protocol discuss the impact on the study results of:				

Section 12: Limitations	Yes	No	N/A	Section Number
12.1.1 Selection bias?	☒	☐	☐	11
12.1.2 Information bias?	☒	☐	☐	
12.1.3 Residual/unmeasured confounding? (e.g., anticipated direction and magnitude of such biases, validation sub-study, use of validation and external data, analytical methods).	☐	☐	☒	
12.2 Does the protocol discuss study feasibility? (e.g. study size, anticipated exposure uptake, duration of follow-up in a cohort study, patient recruitment, precision of the estimates)	☒	☐	☐	8.2

Comments:

Section 13: Ethical/data protection issues	Yes	No	N/A	Section Number
13.1 Have requirements of Ethics Committee/ Institutional Review Board been described?	☒	☐	☐	13
13.2 Has any outcome of an ethical review procedure been addressed?	☐	☐	☒	
13.3 Have data protection requirements been described?	☒	☐	☐	9.2

Comments:

Section 14: Amendments and deviations	Yes	No	N/A	Section Number
14.1 Does the protocol include a section to document amendments and deviations?	☒	☐	☐	4

Comments:

Section 15: Plans for communication of study results	Yes	No	N/A	Section Number
15.1 Are plans described for communicating study results (e.g., to regulatory authorities)?	☒	☐	☐	14
15.2 Are plans described for disseminating study results externally, including publication?	☐	☒	☐	14

Comments:

Name of the main author of the protocol: Talita Duarte-Salles

Date: 25/09/2023

Signature: T. Duarte-Salles