Study Protocol

P5-C1-002 and P5-C2-003

DARWIN EU® - Representativeness of the clinical trial population compared with the real-world population with resistant hypertension

17/07/2026

Version 4.0

Authors: Ellen Gerritsen, Dina Vojinovic

Public

CONTENTS

Study title	DARWIN EU® - Representativeness of the clinical trial population compared with the real-world population with resistant hypertension
Protocol version	V4.0
Date	17/07/2026
EUPAS number	EUPAS1000001065
Active Substance	**Drug class** — **ATC code** Diuretics — C03 Beta blocking agents — C07 Calcium channel blockers — C08 ACE inhibitors — C09A + C09B Angiotensin II receptor blockers — C09C + C09D
Medicinal Product	Not applicable
Research question and objectives	Research question: To what extent is the clinical trial population representative of the real-world population with resistant hypertension? Study objectives: 1. To determine the extent to which the *resistant hypertension trial population reflects the real-world resistant hypertension population* by comparison of demographic (age, sex), clinical, and treatment characteristics observed in the clinical trial with those derived from real-world data sources. 2. To characterise patients with *resistant hypertension*, including but not limited to age, sex, BMI/obesity status, chronic kidney disease stage, diabetes, atrial fibrillation, cardiovascular disease (aldosteronism), current treatment of care, and other relevant comorbidities.
Countries of study	P5-C1-002: Belgium, Denmark, the Netherlands, the United States P5-C2-003: Finland, Spain
Authors	Ellen Gerritsen, e.gerritsen@darwin-eu.org Dina Vojinovic, d.vojinovic@darwin-eu.org

LIST OF ABBREVIATIONS

Acronyms/terms	Description
ACE inhibitors	Angiotensin-converting enzyme inhibitors
ARB	Angiotensin receptor blockers
ATC	Anatomical Therapeutic Chemical
BMI	Body mass index
CC	Coordination centre
CDM	Common Data Model
DARWIN EU®	Data Analysis and Real-World Interrogation Network
DK-DHR	Danish Data Health Registries
DKMA	Danish Medicines Agency
DQD	Data Quality Dashboard
DRE	Digital Research Environment
DTZ	Data Transfer Zone
ECG	Electrocardiogram
EHR	Electronic Health Records
EMA	European Medicines Agency
ENCePP	European Network of Centres for Pharmacoepidemiology and Pharmacovigilance
EU	European Union
EUPAS	EU Post-Authorisation Studies Register
GDPR	General Data Protection Regulation
GP	General Practitioner
HbAC1	Haemoglobin A1c
FinOMOP-HUS	Hospital District of Helsinki and Uusimaa
ICD	International Classification of Diseases
IDIAPJGol	Institut d'Investigació en Atenció Primària de Salut Jordi Gol
IMASIS	Institut Municipal Assistència Sanitària Information System
IPCI	Integrated Primary Care Information
IQR	Interquartile range
IQVIA US - AmbEMR	IQVIA US - Ambulatory EMR
IRB	Institutional Review Board
LOINC	Logical Observation Identifiers Names and Codes
NSAIDs	Nonsteroidal anti-inflammatory drugs
OHDSI	Observational Health Data Sciences and Informatics
OMOP	Observational Medical Outcomes Partnership
PSMar	Consorci Mar Parc de Salut Barcelona
RAS	Renin-Angiotensin System inhibitors

Acronyms/terms	Description
RxNorm	Medical prescription normalised
SIDIAP	The Information System for Research in Primary Care
SNOMED	Systemised Nomenclature of Medicine
UZA	Universitair Ziekenhuis Antwerpen Structured Data
WHO	World Health Organisation

1. TITLE

DARWIN EU® - Representativeness of the clinical trial population compared with the real-world population with resistant hypertension

2. DESCRIPTION OF THE STUDY TEAM

Study team role	Names	Organisation
Principal Investigators	Ellen Gerritsen Dina Vojinovic	IQVIA
Data Scientists	Gargi Jadhav Akram Mendez	IQVIA
Study Manager	Natasha Yefimenko	Erasmus MC
Data source	**Names**	**Data Partner Organisation***
P5-C1-002		
Universitair Ziekenhuis Antwerpen Structured Data (UZA)	Annelies Verbiest Zarah Van Schoor	Universitair Ziekenhuis Antwerpen (UZA)
Danish Data Health Registries (DK-DHR)	Elvira Bräuner Susanne Bruun	Danish Medicines Agency (DKMA)
Integrated Primary Care Information (IPCI)	Katia Verhamme Marcel de Wilde Guido van Leeuwen Mees Mosseveld	Erasmus Medical Center
IQVIA US - Ambulatory EMR (IQVIA US - AmbEMR)	James Brash	IQVIA
P5-C2-003		
FinOMOP-HUS	Marianna Niemi Eric Fey Tiina Wahlfors	FinOMOP
Institut Municipal Assistència Sanitària Information System (IMASIS)	Juan Manuel Ramírez-Anguita Angela Leis Miguel-Angel Mayer	Consorci Mar Parc de Salut Barcelona (PSMar)
The Information System for Research in Primary Care (SIDIAP)	Laura Granés González Agustina Giuliodori Picco Talita Duarte Salles Elena Roel Herranz	Institut d'Investigació en Atenció Primària de Salut Jordi Gol (IDIAPJGol)

*Data partners do not have an investigator role. Data partners execute code at their data source, review, and approve their results.

3. ABSTRACT

Title

DARWIN EU® - Representativeness of the clinical trial population compared with the real-world population with resistant hypertension

Rationale and background

Resistant hypertension is defined by elevated systolic and diastolic blood pressure values despite concomitant use of at least three antihypertensive drugs of different classes, usually including ACE-inhibitors/angiotensin receptor blockers, calcium channel antagonists, and diuretics, and is associated with an increased risk of myocardial infarction, stroke, end-stage renal disease, and death. Baxdrostat, an aldosterone synthase inhibitor, is under evaluation for marketing authorisation for the treatment of resistant hypertension. A recent clinical trial examined its efficacy and safety in individuals with uncontrolled or resistant hypertension. This study aims to assess the representativeness of the Baxdrostat trial population with resistant hypertension by estimating the demographic, clinical, and treatment characteristics of the real-world population with resistant hypertension.

Research question and objectives

Research question

To what extent is the clinical trial population representative of the real-world population with resistant hypertension?

Objectives

The specific objectives of this study are:

1. To determine the extent to which the *resistant hypertension trial population reflects the real-world resistant hypertension population* by comparison of demographic (age, sex), clinical, and treatment characteristics observed in the clinical trial with those derived from real-world data sources.
2. To characterise patients with *resistant hypertension*, including but not limited to age, sex, BMI/obesity status, chronic kidney disease stage, diabetes, atrial fibrillation, cardiovascular disease (aldosteronism), current treatment of care, and other relevant comorbidities.

Methods

Study design

- Characterisation: A patient-level cohort analysis will be conducted to characterise demographic and clinical profile, and treatment of care of individuals with resistant hypertension, defined based on elevated blood pressure measurement values and concomitant use of antihypertensive medications (*objectives 1–2*).

The index date will be defined as the date of a blood pressure measurement value of ≥140/90 mmHg, recorded at least four weeks after initiation of the third concomitant prespecified antihypertensive drug class. Individuals will be followed up until the earliest of i) loss to follow-up, ii) death, iii) end of observation period (the latest available data), or iv) the study end date 31/12/2025, whichever occurs first.

Population

The study population will include males and females aged ≥18 years who meet the definition of resistant hypertension, within the study period from 01/01/2015 to 31/12/2025 (or the end of available data). Resistant hypertension will be defined by presence of:

- A recorded systolic blood pressure measurement value of ≥140 mmHg and/or a diastolic blood pressure measurement value of ≥90 mmHg, and
- Concomitant prescription or dispensing of at least three antihypertensive drug classes including:
    - a diuretic, and
    - an angiotensin-converting enzyme (ACE) inhibitor or angiotensin receptor blocker (ARB), and
    - either a calcium channel antagonist (*primary definition*) or a beta-blocker (*alternative definition*).

Exclusion criteria:

- Not applicable.

Variables

Exposure:

Concomitant prescription or dispensing records of at least three prespecified antihypertensive drug classes, including:

- a diuretic, and an ACE inhibitor or ARB, and a calcium channel antagonist (*primary definition*), or
- a diuretic, and an ACE inhibitor or ARB, and a beta-blocker (*alternative definition*)

Relevant covariates and other variables:

Demographics, clinical characteristics, concomitant conditions, and treatment characteristics.

Data sources

1. Belgium: Universitair Ziekenhuis Antwerpen Structured Data (UZA)
2. Denmark: Danish Data Health Registries (DK-DHR)
3. The Netherlands: Integrated Primary Care Information (IPCI)
4. The United States: IQVIA US - Ambulatory EMR (IQVIA US - AmbEMR)
5. Finland: Hospital District of Helsinki and Uusimaa (FinOMOP-HUS)
6. Spain: Institut Municipal Assistència Sanitària Information System (IMASIS)
7. Spain: The Information System for Research in Primary Care (SIDIAP)

Study size

No sample size has been calculated, as this is an exploratory study which will not test a specific hypothesis.

Statistical analysis

In the *main analysis*, patient-level characteristics will be assessed at least four weeks following initiation of the third prespecified antihypertensive drug class. As part of a *sensitivity analysis*, patient-level characteristics will also be assessed following a longer duration of concomitant treatment exposure of at least eight weeks and at least twelve weeks following initiation of the third prespecified antihypertensive drug class. A gap of 90 days between successive prescription or dispensing records will be permitted when defining continuous exposure. Additionally, resistant hypertension will be characterised using alternative gaps of 60 days and 180 days between successive prescriptions or dispensing records, with an eligible blood pressure measurement value recorded at least 4 weeks following concomitant use.

Characteristics of individuals with resistant hypertension will be described by means of large-scale characterisation and prespecified patient-level characteristics. Categorical variables will be summarised as counts and proportions, while continuous variables will be reported as minimum, q25, median, q75, and maximum. Demographic characteristics, including age and sex, will be assessed at the index date. Clinical characteristics (including measurements and female contraceptive status), concomitant conditions, and treatment characteristics will be assessed during prespecified time windows. These analyses will be conducted using the *CohortCharacteristics* R package based on Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) mapped data.

4. AMENDMENTS AND UPDATES

None.

5. MILESTONES

Study milestones and deliverables	Planned dates*
Final Study Protocol	To be confirmed by EMA
Creation of Analytical code	June/July 2026
Execution of Analytical Code on the data	July 2026
Draft Study Report	17 August 2026
Final Study Report	To be confirmed by EMA

*Planned dates are dependent on obtaining approvals from the internal review boards of the data sources and deliverables review rounds.

6. RATIONALE AND BACKGROUND

Resistant hypertension is defined as blood pressure values remaining ≥140 mmHg (systolic) and/or ≥90 mmHg (diastolic) despite treatment with appropriate lifestyle measures and maximum or maximally tolerated doses of antihypertensive drugs belonging to different drug classes, including a diuretic, a renin-angiotensin system (RAS) blocker, and a calcium channel blocker. Patients with resistant hypertension may have an increased risk of myocardial infarction, stroke, end-stage renal disease, and death.[1]

Baxdrostat, an aldosterone synthase inhibitor, is under evaluation for marketing authorisation for the treatment of resistant hypertension. A recent clinical trial examined its efficacy and safety in individuals with uncontrolled or resistant hypertension. However, clinical trial populations often represent a selected subset of patients and may not fully reflect the characteristics of patients observed in routine clinical practice.

This study aims to assess the representativeness of the Baxdrostat trial population with resistant hypertension by describing the demographic, clinical, and treatment characteristics of individuals with resistant hypertension in the real-world data.[2]

This study aims to characterise individuals with resistant hypertension in seven different data sources.

7. RESEARCH QUESTION AND OBJECTIVES

Research question

To what extent is the clinical trial population representative of the real-world population with resistant hypertension?

Objectives

The specific objectives of this study are:

1. To determine the extent to which the *resistant hypertension trial population reflects the real-world resistant hypertension population* by comparison of demographic (age, sex), clinical, and treatment characteristics observed in the clinical trial with those derived from real-world data sources.
2. To characterise patients with *resistant hypertension*, including but not limited to age, sex, BMI/obesity status, chronic kidney disease stage, diabetes, atrial fibrillation, cardiovascular disease (aldosteronism), current treatment of care, and other relevant comorbidities.

8. RESEARCH METHODS

8.1. Study design

A cohort study will be conducted using routinely collected health data from 7 data sources across 6 countries, including 5 European Union (EU) member states and the United States. The study will comprise:

- A patient-level characterisation study, which will be conducted to address *objectives 1–2*, assessing demographics, clinical characteristics (measurements, female contraceptive status), concomitant conditions, and treatment characteristics (**Figures 1–2**).

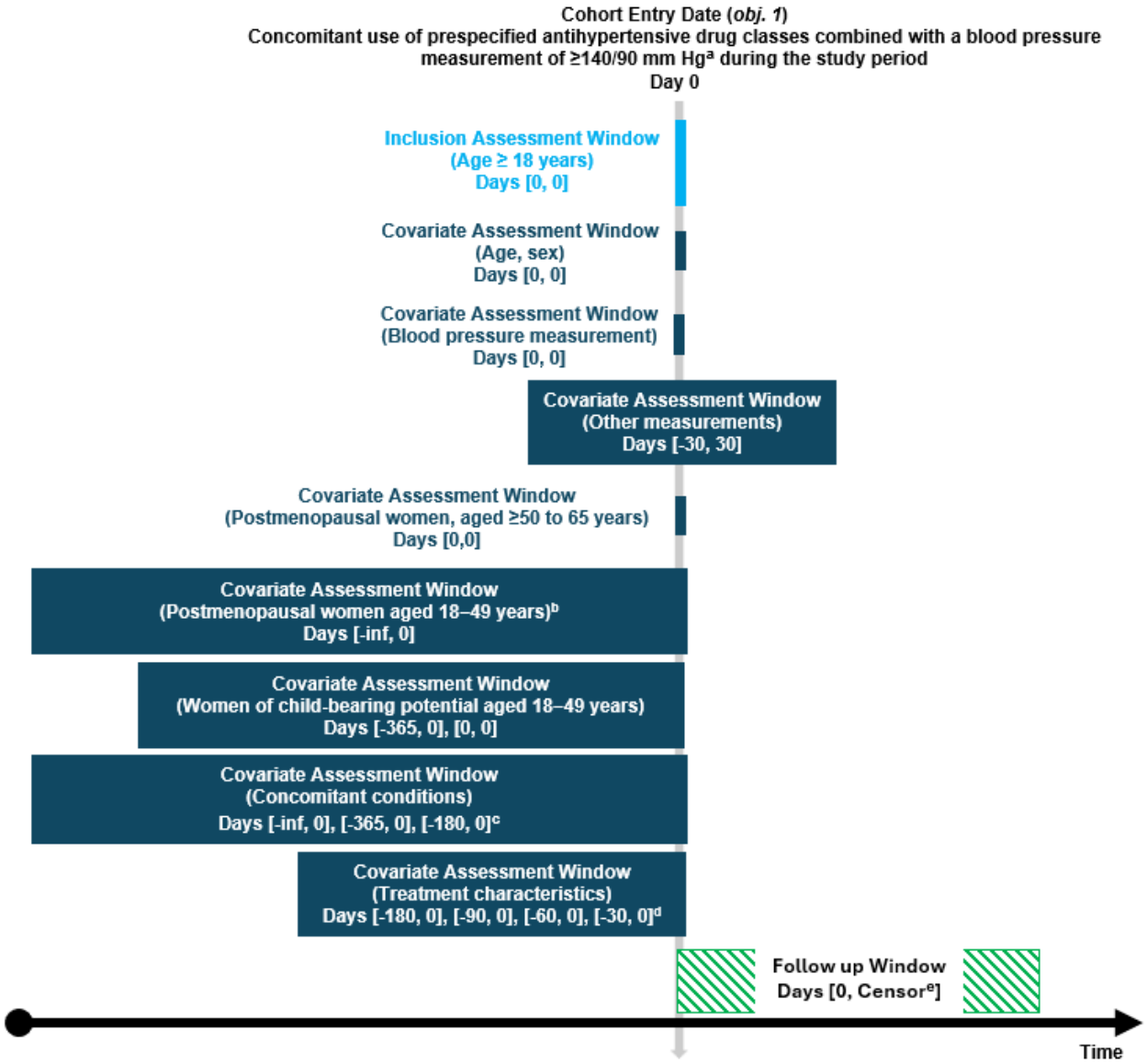

Figure 1. Graphical depiction of the study design *objective 1*.

a. Measured at least four weeks (*main analysis*) or eight or twelve weeks (*sensitivity analysis*) following initiation of the third concomitant antihypertensive drug class during the study period.
b. With a prior record of conditions or procedures indicating premature or early menopause.
c. Time window depends on type of condition, i.e. acute, chronic, beta-blocker indications, detailed information regarding the time window per type of condition is provided in **Section 8.6.3.**
d. Time window depends on type of drug, detailed information regarding the time window per drug is provided in **Section 8.6.3.**
e. Earliest of: loss to follow-up, death, end of observation period (the latest available data), or end of the study period.

In addition, potential prespecified indications for beta-blocker prescribing will be assessed 180 days prior to 7 days following the beta-blocker prescription as part of concomitant antihypertensive drug use (*alternative definition*), considering that there can be a delay in recording the indication of a prescription.

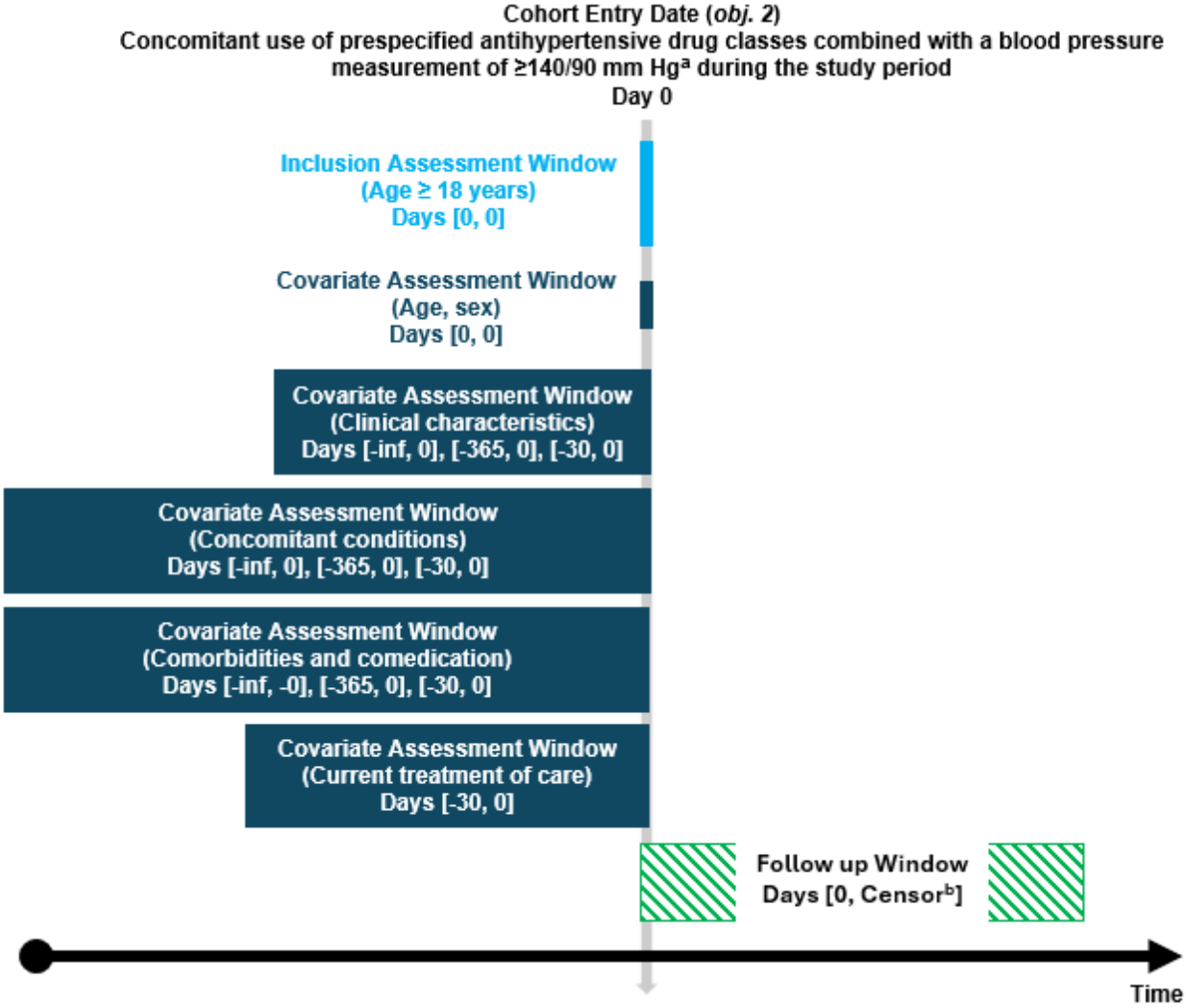

Figure 2. Graphical depiction of the study design *objective 2*.

a. Measured at least four weeks (*main analysis*) or eight or twelve weeks (*sensitivity analysis*) following initiation of the third concomitant antihypertensive drug class during the study period.
b. Earliest of: loss to follow-up, death, end of observation period (the latest available data), or end of the study period.

8.2. Follow-up

For patient-level characterisation, follow-up will start on the index date, defined as the date of the blood pressure measurement value of ≥140 mmHg (systolic) and/or ≥90 mmHg (diastolic) recorded at least four weeks following initiation of the third concomitant antihypertensive drug class during the study period, as defined in **Section 8.6.1**.

Follow-up will end on the earliest of: loss to follow-up, death, end of observation period (the latest available data), or the study end date 31/12/2025.

8.3. Study population with inclusion and exclusion criteria

For characterisation *(objectives 1–2)*, the study population will include adults with resistant hypertension identified within routinely collected healthcare data:

Inclusion criteria

- Individuals with a blood pressure measurement value of ≥140 mmHg (systolic) and/or ≥90 mmHg (diastolic)* observed during the study period (01/01/2015 to 31/12/2025 (or the latest available date)).
- Drug prescriptions or dispensing records indicating concomitant use of at least three prespecified antihypertensive drug classes, including a diuretic, an angiotensin-converting enzyme (ACE) inhibitor or angiotensin receptor blocker (ARB), and a calcium channel blocker (*primary definition*) or beta-blocker (*alternative definition*) for at least four weeks following initiation of the third drug class and at least four weeks prior to the blood pressure measurement value of ≥140/90 mmHg, as defined in **Section 8.6.1.**
- Individuals aged ≥18 years.

* = Recorded blood pressure measurement values will be assessed for plausibility using predefined thresholds. Recorded measurements with systolic ≥300 mmHg or diastolic blood pressure measurement values ≥200 mmHg will be excluded, as they are likely measurement or recording errors. Cut-offs are defined a priori based on physiological plausibility.

Exclusion criteria

- Not applicable.

8.4. Study setting and data sources

This study will be conducted using routinely collected data from 7 primary care, hospital, and registry data sources in the DARWIN EU® Network of data partners from 6 countries, including 5 EU member states and the United States (**Table 1, ANNEX I**). All data were *a priori* mapped to the OMOP CDM.

Data sources

P5-C1-002:

1. Belgium: Universitair Ziekenhuis Antwerpen Structured Data (UZA)
2. Denmark: Danish Data Health Registries (DK-DHR)
3. The Netherlands: Integrated Primary Care Information (IPCI)
4. The United States: IQVIA US - Ambulatory EMR (IQVIA US - AmbEMR)

P5-C2-003:

1. Finland: Hospital District of Helsinki and Uusimaa (FinOMOP-HUS)
2. Spain: Institut Municipal Assistència Sanitària Information System (IMASIS)
3. Spain: The Information System for Research in Primary Care (SIDIAP)

Table 1. Data sources.

Country	Name of Data source	Health Care setting	Type of Data	Number of active individuals*	Calendar period covered by each data source
Belgium	UZA	Hospital care	EHR, registries	20,100	2021 – 09/2025
Denmark	DK-DHR	Registry	EHR, registries	5,984,000	1995 – 12/2024
The Netherlands	IPCI	Outpatient General Practitioner Care	EHR	1,323,300	2006 – 06/2025
The United States	IQVIA US - AmbEMR	Outpatient General Practitioner Care	EHR	13,414,900	2006 – 05/2025
Finland	FinOMOP-HUS	Hospital care	EHR	810,110	2004 – 12/2025
Spain	IMASIS	Hospital care	EHR	252,900	1990 – 06/2025
Spain	SIDIAP	Outpatient General Practitioner Care	EHR	6,068,000	2006 – 12/2024

*= Number of active individuals is defined as the number of individuals present in the last six months of the respective data source. UZA = Universitair Ziekenhuis Antwerpen Structured Data; EHR = Electronic Health Records; DK-DHR = Danish Data Health Registries; FinOMOP-HUS = Hospital District of Helsinki and Uusimaa; IMASIS = Institut Municipal Assistència Sanitària Information System; SIDIAP = The Information System for Research in Primary Care; IPCI = Integrated Primary Care Information; IQVIA US - AmbEMR = IQVIA US - Ambulatory EMR.

Data sources selection

These data sources fulfil the criteria required in terms of data quality (including EMA Data Quality Framework requirements), completeness, timeliness, and representativeness for a characterisation study while covering different regions of Europe (**ANNEX II**).

8.5. Study period

The study period will be from 01/01/2015 to 31/12/2025 (or the latest available date) for each contributing data source.

It should be noted that for several data sources, the availability of the accurate data deviated from the start or end date of the study period. Detailed information about the study period per data partner can be found in **Section 8.4**.

8.6. Variables

8.6.1. Exposure

The primary exposure is defined as concomitant prescription or dispensing of at least three prespecified antihypertensive drug classes, including:

- a diuretic, and an ACE inhibitor or ARB, and a calcium channel antagonist (*primary definition*), or
- a diuretic, and an ACE inhibitor or ARB, and a beta-blocker (*alternative definition*)

Concomitant use is defined as an overlap in exposure to all three antihypertensive drug classes for at least 4 weeks. The concomitant treatment period is considered to start on the date of initiation of the third prespecified antihypertensive drug class. Individuals will be considered continuously exposed to a specific drug class if the subsequent prescription or dispensing record is within 90 days following the start date of the prior prescription of the same drug class. If there is no subsequent record observed during that gap, the individual will be considered exposed to that drug for 30 days following the start date of the last prescription. To assess the robustness of the continuous exposure definition, additional analyses will be performed using alternative gaps of 60 days and 180 days between successive prescriptions or dispensing records of the same drug class to characterise individuals with resistant hypertension with an eligible blood pressure measurement value recorded at least 4 weeks following concomitant use of at least three prespecified antihypertensive drug classes.

The preliminary concept sets used for the identification of exposures are described in **ANNEX IV**. These codes will be refined during the study execution, following the DARWIN EU® phenotyping standard processes, which involve the review of code lists by clinical experts and the review of phenotypes after their execution in the participating data sources. Exposures will be defined following input from the European Medicines Agency (EMA).

8.6.2. Outcome

The outcomes will be as follows:

- Characterisation of individuals with resistant hypertension (*objectives 1–2*)

The preliminary concept sets used for the identification of outcomes are described in **ANNEX IV**. These codes will be refined during the study execution, following the DARWIN EU® phenotyping standard processes, which involve the review of code lists and the review of phenotypes after their execution in the participating data sources. Phenotypes will be defined following input from the EMA.

8.6.3. Covariates, including confounders, effect modifiers, and other variables

The variables for characterisation of the real-world resistant hypertension population (*objective 1*) will be as follows:

- Demographics, defined at the index date:
    - Sex: Female/male
    - Age
- Data availability and observation time (proportion of individuals with one year of available observation):
    - At least 365 days of continuous observation prior to the index date
- Clinical characteristics:
    - Blood pressure measurement value, assessment window at the index date [0, 0]

- Other measurements, assessment window 30 days prior to and 30 days following the index date [-30, +30], including:
  - Glomerular filtration rate
  - Serum potassium
  - Serum sodium
  - Cortisol
  - Electrocardiogram (ECG), including QT corrected interval, if available
  - Heart rate
  - Haemoglobin A1c (HbAC1)

  For measurement variables other than blood pressure, analyses will be restricted to the availability of measurement records among individuals with resistant hypertension rather than the assessment of measurement values. For each variable, the proportion of individuals with at least one recorded measurement within the assessment window will be reported. This reflects expected variability in data completeness across data sources and ensures a consistent and feasible approach for real-world data within the scope of the study. The absence of a recorded measurement within the assessment window will be interpreted as lack of data availability rather than absence of the underlying clinical characteristic.
- Female reproductive status:
  - Postmenopausal females, defined as i) females aged ≥50 to 65 years at the index date or ii) females aged 18–49 years at the index date with a prior record (any time before index date) of conditions or procedures indicating premature or early menopause (primary ovarian insufficiency/failure, premature menopause, early menopause, bilateral oophorectomy, or hysterectomy with ovarian removal).
  - Females of child-bearing potential, defined as females aged 18–49 years at the index date, who do not meet criteria for postmenopausal status [0, 0]. Additionally, use of hormonal contraception will be assessed within 365 days prior to the index date [-365, 0] among these females of child-bearing potential.

- Concomitant conditions will be identified based on recorded diagnoses and procedures in the data sources and grouped according to clinical relevance.
  - Secondary causes of hypertension, assessment window 365 days prior to the index date [-365, 0] and any time prior to the index date [-inf, 0]:
    - Renal artery stenosis
    - Hyperthyroidism
    - Hypothyroidism
    - Pheochromocytoma
    - Cushing's syndrome
    - Aortic coarctation
  - Cardiovascular and metabolic comorbidities, assessment window 365 days prior to the index date [-365, 0] and any time prior to the index date [-inf, 0]:

        - Stroke
        - Acute coronary syndrome
        - Hypertensive encephalopathy
        - Heart failure
        - Cardiac hypertrophy
        - Atrial fibrillation
        - Diabetes mellitus
        - Hepatic impairment
    - Cardiac and structural conditions, assessment window 365 days prior to the index date [-365, 0] and any time prior to the index date [-inf, 0]:
        - Left ventricular outflow obstruction, such as obstructive hypertrophic cardiomyopathy and/or aortic valvular disease
        - Left bundle branch block and any cardiac arrhythmia
        - Long QT syndrome
    - Recent clinical conditions and interventions, assessment window 180 days prior to the index date [-180, 0]:
        - Percutaneous coronary intervention
        - Coronary artery bypass grafting
        - Heart failure
        - COVID-19 infection
    - Pregnancy and breastfeeding status, where available, assessment window 180 days prior to index date [-180, 0].
    - Adrenal insufficiency, assessment window 180 days prior to index date [-180, 0].
- Potential indications for beta-blocker prescribing (*alternative definition*), 180 days prior to 7 days following the beta-blocker prescription as part of concomitant antihypertensive drug use [-180, +7]:
        - Migraine
        - Heart failure
        - Coronary artery disease
        - Unknown
        - None
- Treatment characteristics:
    - Use of medication associated with QT prolongation, assessment window 30 days prior to index date [-30, 0].
    - Treatment with an ARB and ACE inhibitor, both taken simultaneously, assessment window 30 days prior to the index date [-30, 0].
    - Treatment with antiarrhythmic medications or potassium-sparing diuretics, assessment window 30 days prior to index date [-30, 0].

- Treatment with potassium binders, assessment window 60 days prior to index date [-60, 0].
- Treatment with strong inducers of cytochrome P450 (CYP) 3A, nonsteroidal anti-inflammatory drugs (NSAIDs), mineralocorticoid receptor antagonists, and/or use of systemic steroids, assessment window 90 days prior to index date [-90, 0].
- Treatment with cytotoxic therapy, assessment window 180 days prior to index date [-180, 0].

Please note that the abovementioned variables have been refined to ensure feasibility in the current study context. Hence, these variables may slightly differ from the inclusion and exclusion criteria used in the Baxdrostat trial (**Table S5**).

The variables for broader characterisation of the (real-world) resistant hypertension population (*objective 2*) will be as follows:

- Demographics including age and sex, defined at index date, will be consistent with those as part of *objective 1.*
- Clinical characteristics, assessment window 30 days prior to the index date [-30, 0], one year prior to the index date [-365, 0], and any time prior the index date [-inf, 0]:
  - Body mass index
  - Body height
  - Body weight

  For these clinical characteristics, analyses will be restricted to the availability of measurement records rather than the assessment of measurement values. For each variable, the proportion of individuals with at least one recorded measurement within the assessment window will be reported. This reflects expected variability in data completeness across data sources and ensures a consistent and feasible approach for real-world data within the scope of the study.
- Prespecified concomitant conditions will be identified based on recorded diagnoses and procedure codes in the data sources, assessment window 30 days prior to the index date [-30, 0], one year prior to the index date [-365, 0], and any time prior the index date [-inf, 0]:
  - Obesity
  - Chronic kidney disease and, if feasible, stage
  - Cardiovascular disease (aldosteronism)
  - Coronary artery disease/myocardial infarction
  - Arrhythmias
  - Migraine
- Large scale characterisation: The top 10 most frequent diagnostic codes (proxy for comorbidities) or prescribed medications (proxy for comedication) will be identified through large-scale characterisation. These will be assessed 30 days prior to the index date [-30, 0], one year prior to the index date [-365, 0], and any time prior the index date [-inf, 0].
- Current treatment of care: Prespecified antihypertensive medication drugs at ingredient level, assessment window 30 days prior to the index date [-30, 0].

The preliminary concept sets used for the identification of covariates are described in **ANNEX IV**. These codes will be refined during the study execution, following the DARWIN EU® phenotyping standard

processes, which involve the review of code lists and the review of phenotypes after their execution in the participating data sources. Phenotypes will be defined following input from the EMA.

8.7. Study size

No sample size has been calculated, as this is a patient-level characterisation study with a descriptive design, which will not test a specific hypothesis. Thus, the sample size is driven by the availability of data for individuals meeting the study definition of resistant hypertension. Based on a preliminary feasibility assessment, the number of recorded blood pressure measurement values in the data sources included in this study range from 3,900 (IQVIA US - AmbEMR) to 74,180,400 (SIDIAP).

8.8. Analysis

8.8.1. Federated network analyses

All analyses will be conducted separately for each data source and will be carried out in a federated manner, allowing analyses to be run locally without sharing individuals' data.

Before sharing the study package, test runs of the analytics will be performed on a subset of the data sources, and quality control checks will be performed. After all the tests are passed (see **ANNEX III**), the final package will be released in a version-controlled study repository for execution against all the participating data sources.

8.8.2. Data privacy protection

The data partners will locally execute the analytics against the OMOP CDM in R Studio and review and approve the default aggregated results. They will then be made available to the Principal Investigators and study team in secure online repository of Data Transfer Zone (DTZ). All results will be locked and timestamped for reproducibility and transparency. The study results of all data sources will be checked, after which they are made available to the team, and the Study Dissemination Phase can start. All analyses will be conducted separately for each data source, and will be carried out in a federated manner, allowing analyses to be run locally without sharing patient-level data. Cell counts <5 will be suppressed when reporting results to comply with the data source's privacy protection regulations.

8.8.3. Statistical model specification and assumptions of the analytical approach considered

Objectives 1–2

The patient-level characterisation of the resistant hypertension population (*objectives 1–2*) will be estimated based on OMOP CDM mapped data using the *CohortCharacteristics* R package (https://darwin-eu.github.io/CohortCharacteristics/), developed by DARWIN EU®.

The resistant hypertension population will be characterised in terms of demographics, clinical characteristics, concomitant conditions, comorbidities, and treatment characteristics. Demographics, clinical characteristics, concomitant conditions, comorbidities, and treatment characteristics are reported as counts and proportions for categorical variables and as medians with interquartile ranges (IQR) for continuous variables. For indications, the percentage of unknowns will also be reported.

Sensitivity analysis

Description of sensitivity analyses are presented by means of **Table 2**.

Table 2. Sensitivity analyses – rationale, strengths, and limitations.

Sensitivity analysis	What is being varied? How?	Why? (What do you expect to learn?)	Strengths of the sensitivity analysis compared to the primary	Limitations of the sensitivity analysis compared to the primary
Required concomitant use of beta-blocker (alternative definition)	Instead of a calcium channel antagonist being one of the required antihypertensive drugs for concomitant use (primary definition), concomitant use including a beta-blocker is required.	To align with the treatment combination of antihypertensive drugs used in the Baxdrostat trial.	With the inclusion of this additional treatment combination, we may capture additional individuals with resistant hypertension.	Beta-blockers can also be used for other indications than hypertension, i.e. migraine, heart failure, and coronary artery disease. Therefore, we will also assess the occurrence of these indications in the sensitivity analysis.
Longer concomitant exposure windows	The required minimum duration of concomitant exposure is extended to at least 8 weeks and at least 12 weeks following initiation of the third prespecified antihypertensive drug class, instead of 4 weeks in the main analysis.	To evaluate whether longer duration of concomitant treatment exposure influences the identification of blood pressure measurement values and other variables.	Longer exposure duration may better reflect sustained treatment use and allow sufficient time for treatment effects to stabilise.	A longer duration of concomitant treatment exposure may decrease the number of included individuals and may preferentially select individuals with more persistent treatment patterns.
Gap between successive prescriptions or dispensing records	Instead of permitting a gap of up to 90 days between successive prescription or dispensing records when defining continuous exposure, analyses will also be performed using alternative permissible gaps of 60 days and 180 days to characterise individuals with resistant hypertension with an eligible blood pressure measurement value recorded at least 4 weeks following concomitant use of at least three prespecified antihypertensive drug classes.	To assess the robustness of the findings to alternative permissible gaps between successive prescriptions or dispensing records.	Evaluating shorter and longer permissible gaps between successive prescriptions or dispensing records provides insight into whether the findings are robust.	A shorter permissible gap may exclude individuals with irregular prescription refills, whereas a longer permissible gap may include individuals who have discontinued treatment.

8.8.4. Output

Output will include the following:

A PDF report, including an executive summary and the following tables.

- Table 1. Mock table shell: Study attrition of individuals with recorded resistant hypertension, presented by data source (*objective 1–2*).

	Data source 1	Data source 2	Data source 3/4/5/6/7
Qualifying initial records (Individuals with a blood pressure measurement value of ≥140/90 mmHg)			
Cohort start date between 1/1/2015 and 31/12/2025			
Concomitant drug use of three prespecified antihypertensive drug classes initiated at least four weeks prior to the qualifying blood pressure measurement			
Age ≥18 years			
No observed cohort exit (censoring because of recorded death*) before index date			

* = If death is recorded prior to index date, the individual will be excluded

- Table 2. Mock table shell: Demographic characteristics of resistant hypertension population during the study period.

Variable name	Data source name						
	UZA	DK-DHR	FinOMOP-HUS	IMASIS	SIDIAP	IPCI	IQVIA US - AmbEMR
Number of subjects							
Age (years), median [q25–q75]							
Age (years), range (min–max)							
Sex, n (%)							
• Female							
• Male							
• None							

UZA = Universitair Ziekenhuis Antwerpen Structured Data, DK-DHR = Danish Data Health Registries, FinOMOP-HUS = Hospital District of Helsinki and Uusimaa, IMASIS = Institut Municipal d'Assistència Sanitària Information System, SIDIAP = Information System for Research in Primary Care, IPCI = Integrated Primary Care Information, IQVIA US - AmbEMR = IQVIA United States Ambulatory Electronic Medical Records; N = sample size.

- Table 3. Mock table shell: Prespecified measurements among the resistant hypertension population during the study period (*objective 1*).

Variable name	Data source name						
	UZA	DK-DHR	FinOMOP-HUS	IMASIS	SIDIAP	IPCI	IQVIA US - AmbEMR
Blood pressure measurement median [q25–q75]							
Blood pressure measurement, range (min–max)							
Glomerular filtration rate, n (%)							

Variable name	Data source name						
	UZA	DK-DHR	FinOMOP-HUS	IMASIS	SIDIAP	IPCI	IQVIA US - AmbEMR
Serum potassium, n (%)							
Serum sodium, n (%)							
Cortisol, n (%)							
Electrocardiogram							
Heart rate							
Hemoglobin AC1							

* = Please note that for measurement variables other than blood pressure, analyses will be restricted to the availability of measurement records rather than the assessment of measurement values. UZA = Universitair Ziekenhuis Antwerpen Structured Data, DK-DHR = Danish Data Health Registries, FinOMOP-HUS = Hospital District of Helsinki and Uusimaa, IMASIS = Institut Municipal d'Assistència Sanitària Information System, SIDIAP = Information System for Research in Primary Care, IPCI = Integrated Primary Care Information, IQVIA US - AmbEMR = IQVIA United States Ambulatory Electronic Medical Records; N = sample size.

- Table 4. Mock table shell: Female reproductive status among the resistant hypertension population during the study period (*objective 1*).

Variable name	Data source name						
	UZA	DK-DHR	FinOMOP-HUS	IMASIS	SIDIAP	IPCI	IQVIA US - AmbEMR
Postmenopausal, n (%)							
Females of child-bearing potential, n (%)							
Females of child-bearing potential on hormonal contraception, n (%)							

UZA = Universitair Ziekenhuis Antwerpen Structured Data, DK-DHR = Danish Data Health Registries, FinOMOP-HUS = Hospital District of Helsinki and Uusimaa, IMASIS = Institut Municipal d'Assistència Sanitària Information System, SIDIAP = Information System for Research in Primary Care, IPCI = Integrated Primary Care Information, IQVIA US - AmbEMR = IQVIA United States Ambulatory Electronic Medical Records; N = sample size.

- Table 5. Mock table shell: Prespecified concomitant conditions recorded among the resistant hypertension population in prespecified time window during the study period (*objective 1*).

Variable name	Data source name						
	UZA	DK-DHR	FinOMOP-HUS	IMASIS	SIDIAP	IPCI	IQVIA US - AmbEMR
Prespecified condition 1, n (%)							
Prespecified condition 2, n (%)							
Prespecified condition 3, n (%)							
Prespecified condition 4/5/etc., n (%)							

UZA = Universitair Ziekenhuis Antwerpen Structured Data, DK-DHR = Danish Data Health Registries, FinOMOP-HUS = Hospital District of Helsinki and Uusimaa, IMASIS = Institut Municipal d'Assistència Sanitària Information System, SIDIAP = Information System for Research in Primary Care, IPCI = Integrated Primary Care Information, IQVIA US - AmbEMR = IQVIA United States Ambulatory Electronic Medical Records; N = sample size.

- Table 6. Mock table shell: Treatment characteristics recorded among the resistant hypertension population in prespecified time window during the study period (*objective 1*).

Variable name	Data source name						
	UZA	DK-DHR	FinOMOP-HUS	IMASIS	SIDIAP	IPCI	IQVIA US - AmbEMR
Treatment 1, n (%)							
Treatment 2, n (%)							
Treatment 3, n (%)							
Treatment 4/5/etc., n (%)							

UZA = Universitair Ziekenhuis Antwerpen Structured Data, DK-DHR = Danish Data Health Registries, FinOMOP-HUS = Hospital District of Helsinki and Uusimaa, IMASIS = Institut Municipal d'Assistència Sanitària Information System, SIDIAP = Information System for Research in Primary Care, IPCI = Integrated Primary Care Information, IQVIA US - AmbEMR = IQVIA United States Ambulatory Electronic Medical Records; N = sample size.

- Table 7. Mock table shell: Prespecified measurements among the resistant hypertension population during the study period (*objective 2*).

Variable name	Data source name						
	UZA	DK-DHR	FinOMOP-HUS	IMASIS	SIDIAP	IPCI	IQVIA US - AmbEMR
Body mass index, n (%)							
Body height, n (%)							
Body weight, n (%)							

UZA = Universitair Ziekenhuis Antwerpen Structured Data, DK-DHR = Danish Data Health Registries, FinOMOP-HUS = Hospital District of Helsinki and Uusimaa, IMASIS = Institut Municipal d'Assistència Sanitària Information System, SIDIAP = Information System for Research in Primary Care, IPCI = Integrated Primary Care Information, IQVIA US - AmbEMR = IQVIA United States Ambulatory Electronic Medical Records; N = sample size.

- Table 8. Mock table shell: Prespecified concomitant conditions recorded among the resistant hypertension population in prespecified time window during the study period (*objective 2*).

Variable name	Data source name						
	UZA	DK-DHR	FinOMOP-HUS	IMASIS	SIDIAP	IPCI	IQVIA US - AmbEMR
Prespecified condition 1, n (%)							
Prespecified condition 2, n (%)							
Prespecified condition 3, n (%)							
Prespecified condition 4/5/etc., n (%)							

UZA = Universitair Ziekenhuis Antwerpen Structured Data, DK-DHR = Danish Data Health Registries, FinOMOP-HUS = Hospital District of Helsinki and Uusimaa, IMASIS = Institut Municipal d'Assistència Sanitària Information System, SIDIAP = Information System for Research in Primary Care, IPCI = Integrated Primary Care Information, IQVIA US - AmbEMR = IQVIA United States Ambulatory Electronic Medical Records; N = sample size.

- Table 9. Mock table shell: Top 10 conditions recorded among the resistant hypertension population in prespecified time window during the study period (*objective 2*).

Variable name	Data source name						
	UZA	DK-DHR	FinOMOP-HUS	IMASIS	SIDIAP	IPCI	IQVIA US - AmbEMR
Condition 1, n (%)							
Condition 2, n (%)							
Condition 3, n (%)							
Condition 4, n (%)							
Condition 5, n (%)							
Condition 6, n (%)							
Condition 7, n (%)							
Condition 8, n (%)							
Condition 9, n (%)							
Condition 10, n (%)							

UZA = Universitair Ziekenhuis Antwerpen Structured Data, DK-DHR = Danish Data Health Registries, FinOMOP-HUS = Hospital District of Helsinki and Uusimaa, IMASIS = Institut Municipal d'Assistència Sanitària Information System, SIDIAP = Information System for Research in Primary Care, IPCI = Integrated Primary Care Information, IQVIA US - AmbEMR = IQVIA United States Ambulatory Electronic Medical Records; N = sample size.

- Table 10. Mock table shell: Top 10 medication recorded among the resistant hypertension population in prespecified time window during the study period (*objective 2*).

Variable name	Data source name						
	UZA	DK-DHR	FinOMOP-HUS	IMASIS	SIDIAP	IPCI	IQVIA US - AmbEMR
Medication 1, n (%)							
Medication 2, n (%)							
Medication 3, n (%)							
Medication 4, n (%)							
Medication 5, n (%)							
Medication 6, n (%)							
Medication 7, n (%)							
Medication 8, n (%)							
Medication 9, n (%)							
Medication 10, n (%)							

UZA = Universitair Ziekenhuis Antwerpen Structured Data, DK-DHR = Danish Data Health Registries, FinOMOP-HUS = Hospital District of Helsinki and Uusimaa, IMASIS = Institut Municipal d'Assistència Sanitària Information System, SIDIAP = Information System for Research in Primary Care, IPCI = Integrated Primary Care Information, IQVIA US - AmbEMR = IQVIA United States Ambulatory Electronic Medical Records; N = sample size.

- Table 11. Mock table shell: Treatment characteristics recorded among the resistant hypertension population in prespecified time window during the study period (*objective 2*).

Variable name	Data source name						
	UZA	DK-DHR	FinOMOP-HUS	IMASIS	SIDIAP	IPCI	IQVIA US - AmbEMR
Prespecified antihypertensive drug 1, n (%)							
Prespecified antihypertensive drug 2, n (%)							
Prespecified antihypertensive drug							
Prespecified antihypertensive drug 4/5/etc., n (%)							

UZA = Universitair Ziekenhuis Antwerpen Structured Data, DK-DHR = Danish Data Health Registries, FinOMOP-HUS = Hospital District of Helsinki and Uusimaa, IMASIS = Institut Municipal d'Assistència Sanitària Information System, SIDIAP = Information System for Research in Primary Care, IPCI = Integrated Primary Care Information, IQVIA US - AmbEMR = IQVIA United States Ambulatory Electronic Medical Records; N = sample size.

Please note that all tables will also be made for the results of the sensitivity analyses.

An interactive dashboard will be generated by incorporating all the results (tables) included in the PDF report mentioned above.

8.9. Evidence synthesis

Results from analyses described in **Section 8.8** will be presented separately for each data source. No meta-analysis of results will be conducted.

9. STRENGTHS AND LIMITATIONS

Strengths

This study leverages several real-world data sources from 5 European countries and the United States, covering primary care, hospital care, and registry data sources. These data sources enable a comprehensive characterisation of patients with resistant hypertension across varied healthcare settings. Use of the OMOP CDM ensures standardised data structuring, harmonised variables, and consistent cohort definitions across sources. This facilitates reproducible analyses and enables robust characterisation of demographics, clinical profiles, and treatment characteristics of individuals with resistant hypertension during the study period.

Limitations

The study will be informed by routinely collected healthcare data, which introduces several important considerations that may influence the interpretation of the findings.

The results from this study only reflect the populations captured within the 7 included data sources and may not be generalisable to other countries or healthcare systems. Differences in healthcare systems, coding practices, and data capture may also contribute to heterogeneity across data sources. In addition, the availability of key variables, including blood pressure and laboratory measurements, may differ across data sources, which may influence cohort identification and characterisation. Please note that in DK-DHR, blood pressure measurement values are only available from hospital records and not from primary care data.

Electronic health records are primarily designed for clinical purposes rather than research, and as such, may contain incomplete, inconsistent, or variably recorded information. Therefore, the documentation of clinical characteristics, concomitant conditions, and treatment characteristics necessary for patient-level characterisation may vary across data sources. For several variables, analyses rely on the presence of recorded information rather than detailed clinical values or classifications. The actual clinical indication for prescribing drugs is not directly recorded in the data sources. Instead, proxy measures based on diagnosis codes around the time of beta-blocker treatment during the study period are used to assess whether beta-blockers are prescribed for resistant hypertension or another indication. Consequently, the estimation of potential indications may be incomplete or imprecise, as this might not be recorded for all patients. To minimise incompleteness of concomitant conditions, a large-scale characterisation will also be performed in different time windows around the index date.

As repeated confirmation of elevated blood pressure ≥140/90 mmHg cannot be consistently ascertained, the study population may include individuals with pseudo resistant hypertension. In addition, the timing and frequency of blood pressure measurements may be driven by routine clinical practice and are not standardised across data sources, resulting in irregular measurement intervals. Consequently, identification of resistant hypertension may be delayed, and the index date should be interpreted as the recorded observation of elevated blood pressure following treatment intensification.

The sequencing and overlap of antihypertensive treatments are inferred from prescription records and may not fully reflect actual patterns. Treatment initiation, discontinuation, and switching may be misclassified due to gaps in recording or differences in prescribing and dispensing practices across data sources. In addition, treatment history prior to entry into the data source may not be fully captured, and recorded prescriptions do not guarantee that the medications were dispensed or consumed; therefore, assumptions of actual use are made. As not all data sources reliably capture drug duration, assumptions of treatment

duration are made. While the aim is to characterise all individuals with resistant hypertension, some individuals with prescriptions that last longer than the allowed gap of three months will be missed.

No minimum prior observation time is required, as this criterion may reduce the study size, especially in data sources which do not have a long duration of longitudinal registration. Consequently, only a subset of included patients may contribute to time windows prior to the index date. The proportion of included individuals with at least 365 days of continuous observation prior to the index date will be assessed to gain more insight into how many individuals with resistant hypertension are longitudinally observed.

10. REFERENCES

1. McEvoy, J.W., et al., *2024 ESC Guidelines for the management of elevated blood pressure and hypertension: Developed by the task force on the management of elevated blood pressure and hypertension of the European Society of Cardiology (ESC) and endorsed by the European Society of Endocrinology (ESE) and the European Stroke Organisation (ESO).* European Heart Journal, 2024. **45**(38): p. 3912-4018.
2. Flack, J.M., et al., *Efficacy and Safety of Baxdrostat in Uncontrolled and Resistant Hypertension.* New England Journal of Medicine, 2025. **393**(14): p. 1363-1374.

11. ANNEXES

ANNEX I. Description of data sources

Belgium: Universitair Ziekenhuis Antwerpen Structured Data (UZA)

#	Section	Description
1	Data source identification and country	UZA (Universitair Ziekenhuis Antwerpen Structured Data) Flanders, Belgium
2	Data partner information section	Universitair Ziekenhuis Antwerpen (UZA) Data Science
3	Coverage and timespan	Data collection since: 2021 Extent: Regional. There are several hospitals located in the catchment area of broad Antwerp. Many patients visit several hospitals (for different diseases, or changing during the treatment of a chronic disease). Some patients will visit UZA from beyond its geographic catchment area.
4	Healthcare setting / type of data	Secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. Diagnoses, medications, some procedures, lab values, visits, cancer register, vital parameters. All structured data from in- and outpatient visits. • The laboratory information system has been operational since 1996. • Chemotherapies from chemopro are collected since 2012 • The electronic health record (EHR) system has been live since June 2021.
5	Data collection process	Insurance/administrative claims. Data is captured from multiple sources, including CERNER EPD, GE and Philips monitors, and the MKG datasets. This data is ingested either continuously using tools like Mirth Connect and RabbitMQ or in batches via Data Factory and Python ETL scripts. Once ingested, the data is processed within an integrated platform into source, clean, and master datasets. Finally, the processed data is used for analytics through tools like Power BI and mapped to the OMOP CDM for collaborative research.
6	General representativeness	The data is representative of Antwerp population that have visited the hospital.
7	Data content /source coding	CNK (unique, seven-digit national code assigned to each pack size or form of a product), multum, atc, icdo3, icd9, icd10, Nomenclatuur van de geneeskundige verstrekkingen (official list of medical procedures and services covered by the Belgian health insurance system), SNOMED.
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5 and the OMOP standard vocabularies. The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. Patients are considered to be unique and not have different id numbers.
9	Quality control (data source specific)	Raw data sources are processed and unified in an ETL procedure to a data lake, where data is vetted and compared with the source. Quality checks are developed to detect gaps in processing (missing days), to see how certain sampled data points from source are processed and if they keep resulting in the same outcome. Daily processing counts are used to detect abnormal ETL behaviour. For each source, unit tests are developed to chart the behaviour of ETL.
10	Linkage	The UZA data is linked to the national death registry.
11	Vital status	The hospital receives death data quarterly from the government.
12	Limitations	Our OMOP observation period ends at the latest hospital stay date. If death occurs outside of the hospital, the death will fall outside of our OMOP observation period. End dates are not available for drug source data, this includes administration and prescription drugs. Prescriptions given in outpatient can also be prolonged by the GP. Unstructured / semi-structured data, primary care, outpatient drug dispensing, outpatient drug treatment schedule are not covered. • Medical administrations from the previous EHR system (before June 2021) • ICU data

#	Section	Description
		(metavision) • Procedures partly mapped (nomenclatuur der geneeskundige verstrekkingen) • Radiotherapy (unstructured reports)
13	Main references	No main reference provided.
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1000000627 Website: https://www.uza.be/en

Denmark: Danish Data Health Registries (DK-DHR)

#	Section	Description
1	Data source identification and country	DK-DHR (Danish Data Health Registries) Denmark
2	Data partner information section	Danish Medicines Agency (DKMA) Data & Data Analysis (DAD)
3	Coverage and timespan	Data collection since: 1995 Extent: Nation-wide. The data is representative of the entire Danish population.
4	Healthcare setting / type of data	Community pharmacists, and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. The following data elements are collected: diagnosis (including rare diseases and pregnancy data), hospital admissions, discharge and ICU data, Cause of death, Drug prescriptions, dispensing, vaccination and contraception, Procedures (surgical and non-surgical hospital), and Sociodemographic information (sex and age only).
5	Data collection process	Outpatient electronic health records, and Inpatient hospital electronic health records, and Registries, and Other. All causes of deaths, all retrieved drug prescriptions, all records of vaccinations, all hospital inpatient and outpatients contacts including disease diagnoses and hospital surgical and non-surgical procedures, histologically confirmed incident cancers, laboratory test results for the entire Danish population from 1/1/1995 onwards.
6	General representativeness	The data is representative of the entire Danish population. Healthcare is free in Denmark, so we do not expect any bias in data collection based on socio-economic status.
7	Data content /source coding	Diagnoses and causes of death are collected using the ICD-10 vocabulary. ATC and RxNorm are used for Drugs. SNOMED codes are used for Procedures.
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5 and the OMOP standard vocabularies. The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. Patients have unique identifiers used to link datasets.
9	Quality control (data source specific)	The data we have received relating to nationwide Danish Health Data registries offer an opportunity for large-scale, population-based studies with several advantages 1) Their large size improves the precision of estimates and enables the study of rare exposures and outcomes with long-term latency, 2) Inclusion of nearly all individuals in the target population ensures that the data reflect routine clinical care and all clinical segments of the source population, 3) Data are collected independently of each research study, thus minimising certain types of bias, e.g., non-response, and the influence from attention to the research question on the diagnostic process. Before the source data is sent to us, the Danish Health Data Authority runs and does comprehensive checks of the registry table data validity of the variables, breaks in data, changes in variable coding, missingness, etc. We perform checks of missingness/completeness in relation to requested variables. In essence, we are receiving a dump of a mirror of the data that is controlled by the SDS. The documentation performed by SDS is available online, in Danish primarily https://www.esundhed.dk/Dokumentation (all variables), but also in

#	Section	Description
		English https://sundhedsdatastyrelsen.dk/da/english/health_data_and_registers/national_health_registers
10	Linkage	There is no linkage in this data source.
11	Vital status	The Cause of Death registry (DAR) is used, the cause of death is collected using ICD-10 codes.
12	Limitations	- Lifestyle factors (such as smoking status and weight) are only complete for pregnant females. - Stillborn children will not have any records in our CDM. This means that e.g. birth weight and birth length of stillborns is not recorded. - There is no information on key socio-economic status (SES) factors, such as occupation, education, and income. - Patient contacts in primary care (visits to the GP) are not included. Consequently, the incidence of chronic diseases like Type 2 Diabetes (T2D) and asthma must be determined using drug prescriptions as a proxy.
13	Main references	Schmidt M, Schmidt SAJ,Adelborg K,Sundbøll J,Laugesen K,Ehrenstein V,Sørensen HT "The Danish health care system and epidemiological research: from health care contacts to database records." Clinical epidemiology (2019): 31372058
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111217 Website: https://sundhedsdatastyrelsen.dk/da/english/health_data_and_registers/healthdatadenmark

Finland: Hospital District of Helsinki and Uusimaa (FinOMOP-HUS)

#	Section	Description
1	Data source identification and country	FinOMOP-HUS (Hospital District of Helsinki and Uusimaa) Uusimaa, Finland
2	Data partner information section	FinOMOP (FinOMOP) HUS Tietohallinto (Information Management)
3	Coverage and timespan	Data collection since: 2004 Extent: Regional. HUS is responsible for specialized healthcare in Finland's Uusimaa region and treatment of many rare and severe diseases that are nationally centralized to HUS.
4	Healthcare setting / type of data	Secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. All visits, examinations, laboratory test, procedures, and treatments are recoded in the HUS IT systems and integrated into the data lake. The data lake stores decades of clinical information in digital format, and data from both past and current source systems are available. Systems providing data into the data lake include: CGI Uranus and Epic Apotti (EHR: visits, diagnoses, medication, etc.), Opera (procedure records), Kemokur and Beacon (cancer-specific medications), Marela (hospital pharmacy), Multilab/Mylab+ (laboratory system), Qpati/Mylab+ (pathology records system). MerlotMedi (prehospital emergency care), Disease specific quality registers (disease specific). Obstetrix (birth register).
5	Data collection process	Outpatient electronic health records, and Inpatient hospital electronic health records, and Other. All visits, all procedures, and all given treatments have been recorded systematically in the electronic format. We use more than hundred different operational IT systems.
6	General representativeness	The data reflects patients treated at the hospital and is thus not a generic population.
7	Data content /source coding	Drugs: ATC, generic name, brand name + dosage and strength information in the data + local route mapping Nordic product number (VNR) Procedures: Nomesco NCSP Genetic: special mappings to OMOP Genomics Diagnosis: ICD10 + Finnish adaptation “ICD10fi”, ICD9fi, ICD8fi Laboratory tests: National laboratory test numbers (LABfi kuntaliitto) + local HUS variants National microbe listings, local “additional information” -

#	Section	Description
		coding Various observations during care: LOINC and local coding Pathology: SNOMED2
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5 and the OMOP standard vocabularies. The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. All patients are identified with a unique, national social security number, which is a permanent person identifier. Based on this, the derived patient identifiers (pseudonyms) in the HUS data lake also are unique.
9	Quality control (data source specific)	The HUS data lake and IT management is ISO 13485:2016 and ISO 9001:2015 certified and has capabilities in developing, validating, and CE-marking medical devices and software. Data passes through three layers of validations and transformations (bronze, silver, cold i.e. medallion lakehouse architecture) before being stored in a layout optimized for efficient analytics. The data is monitored for data freshness and data anomaly. OMOP data is converted from silver level. DARWIN EU® CC data quality dashboard SQL queries are implemented HUS ETL pipeline in addition of the standard quality control that is being done together with the DARWIN EU® CC upon onboarding.
10	Linkage	HUS is linked with Digital and Population Data Services Agency from where national death registry, socicaal security number and address are imported. Data from many nation-wide registries can be combined at HUS, which is subject to obtaining special data permits.
11	Vital status	The source of vital status (date of death) is the Digital and Population Data Services Agency.
12	Limitations	No data source specific limitations documented. General limitations for the data type applicable.
13	Main references	No main reference provided.
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111134 Website: https://www.hus.fi/en

Spain: Institut Municipal Assistència Sanitària Information System (IMASIS)

#	Section	Description
1	Data source identification and country	IMASIS (Institut Municipal Assistència Sanitària Information System) Catalunya, Spain
2	Data partner information section	Parc de Salut Mar Barcelona (PSMAR) Direcció de Control de Gestió (Management and Control Department)
3	Coverage and timespan	Data collection since: 1990 Extent: Regional. The primary catchment area covers the Barcelona city area, in Catalonia (Spain). However, all recorded visits are included, which may also include patients residing outside this area.
4	Healthcare setting / type of data	Secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. The available information includes diagnoses, procedures, drug administrations, dispensations and prescriptions, laboratory tests, visit types, cancer data, and vaccinations. These data are linked to primary care centers, which currently contribute information on diagnoses, and treatments, as well as to mortality data.
5	Data collection process	Outpatient electronic health records, and Inpatient hospital electronic health records, and Other. The data collection process begins at the point of care, when a patient is attended in any facility within our hospital network or associated primary care centers. Clinical information is recorded directly into the electronic health records (EHRs) and subsequently complemented with data from external sources, including the regional vaccine registry, the cancer registry, the national mortality registry, and outpatient treatment records. All information is consolidated into a relational database using standardized formats and subjected to multiple quality control

#	Section	Description
		procedures to ensure completeness and consistency. The resulting datasets are then extracted, transformed, and loaded (ETL) into an OMOP Common Data Model (CDM) database on a quarterly to semi-annual basis.
6	General representativeness	The data source includes patients across all socioeconomic levels, ages, and genders, suggesting that the dataset is representative of the population of Barcelona requiring secondary (hospital-based) healthcare at any given time.
7	Data content /source coding	Health outcomes, diagnoses, and procedures are recorded using ICD-9-CM and ICD-10-CM classification. Laboratory tests are recorded in the source data using LOINC terminology and drugs using the national codes classification. Cancer registry is based on ICD-O-3 classification. Additionally, a set of demographics data are available such as age and sex.
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5 and the OMOP standard vocabularies. The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. A patient cannot be registered under different ID numbers. They are registered under the same ID, even when changing practice of re-registration or admission.
9	Quality control (data source specific)	Several quality control procedures have been implemented during the creation and update of IMASIS. The process includes a series of filters that exclude any data whose format or content do not match the specifications of the target variables. Some examples of these actions include the following. - Records containing codes that do not belong to the standardized vocabularies used for encoding are reviewed within their respective domains and, if necessary, discarded. - Records with implausible values, such as dates preceding a patient's birthdate, are identified and corrected or removed. - Before each IMASIS update, the number of records in the source files is verified. Record counts from the new IMASIS instance are compared with both the source files and previous IMASIS versions to detect potential inconsistencies or anomalies in the source data. - In addition, if the Data Quality Dashboard (DQD) package identifies quality issues in the OMOP instance, for example diagnoses recorded for patients with an implausible gender, these discrepancies are reviewed and corrected in the source data.
10	Linkage	Information of diagnosis and treatments received in outpatient settings and in primary care as well as national mortality registry, cancer registry and regional vaccination registry.
11	Vital status	The hospital's death registry records the age at death for each patient when the event occurs within the hospital. In addition, deaths occurring outside the hospital are captured through linkage with the national mortality registry, however, in these cases, the specific cause of death is not currently available.
12	Limitations	No data source specific limitations documented. General limitations for the data type applicable.
13	Main references	Mayer MA, Furlong LI,Torre P,Planas I,Cots F,Izquierdo E,Portabella J,Rovira J,Gutierrez-Sacristan A,Sanz F "Reuse of EHRs to Support Clinical Research in a Hospital of Reference." Studies in health technology and informatics (2015): 25991136
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/104316 Website: https://www.parcdesalutmar.cat/en/

Spain: The Information System for Research in Primary Care (SIDIAP)

#	Section	Description
1	Data source identification and country	SIDIAP (The Information System for Research in Primary Care) Catalunya, Spain
2	Data partner information section	Institut d'Investigació en Atenció Primària de Salut Jordi Gol (IDIAPJGol)

#	Section	Description
3	Coverage and timespan	Data collection since: 2006 Extent: Regional. SIDIAP is a database of primary care electronic health records of the population of Catalonia, North-East Spain. It contains pseudo-anonymised records of more than 8 million people , of which 6.1 million are active as of 2022, representing around 76% of the Catalan population.
4	Healthcare setting / type of data	Primary care – General Practitioner, and hospital inpatient care. SIDIAP captured data includes routine visits, socio-demographics, diagnoses, laboratory tests, drugs (prescribed and dispensed), referrals, sick leaves and lifestyle information.
5	Data collection process	Outpatient electronic health records, and Inpatient hospital electronic health records. Data is entered by primary care physicians upon healthcare contact, supplemented with hospital discharge records. The Institut Catala de la Salut (Catalan Health Institute) is the data controller.
6	General representativeness	It was previously shown that the captured SIDIAP population is highly representative of the entire Catalan region in terms of geographic, age, and sex distributions.
7	Data content /source coding	SIDIAP data covers all services that occur at the Primary Care Centres, as well as support services, such as sexual and reproductive health or home end-of-life care. Drugs are coded in ATC-WHO terminology in the source data. Health outcomes are captured in ICD-10CM codes. The SIDIAP contains all laboratory tests and results performed in primary health centres. Demographics, geographical, as well as socio-economic factors are recorded for each patient.
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5 and the OMOP standard vocabularies. The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. A patient has a unique id, also upon changing practice or re-registration.
9	Quality control (data source specific)	Internal and external validation processes are carried out to determine the data quality of the SIDIAP information at each data update. These include stratifying the data by geographical regions and year in order to identify differences in data collection that need to be harmonized (e.g. recording of specific information under different codes). The measurement units of variables measuring one characteristic are also homogenized (e.g. transformation of the data from every laboratory that measures haemoglobin to grams per decilitre). Visual inspection of all data included in the database by week is also conducted, allowing one to see temporal patterns in the registry of a certain variable. With this information, the SIDIAP team can issue recommendations to researchers about the most common variable(s) where certain information is recorded (e.g., there are several variables with information concerning the female's menopausal status and with these visual inspection tools the SIDIAP team can inform the researchers about which related variables have the largest number of records and could be more helpful to capture menopause). Data availability (longitudinally and reliability), plausibility (range checks and unusual values), and consistency are inspected through visualisation tools. In addition, before accessing the data for a requested project, research teams have access to a quality-control report. This document contains counts, years, percentiles, maximums and minimums, incidences, and prevalence of the data requested for the project, allowing detection of inconsistencies in the data extraction prior to data delivery. External validation processes of the SIDIAP database mainly include assessing the data recorded in SIDIAP through linkage to external gold standard data sources, by analysing free text, or by sending questionnaires to health professionals.
10	Linkage	SIDIAP is linked to a hospital discharge database, pharmacy dispensation, and primary care laboratories. It can also be linked to other registries in Catalonia on a project by project basis.
11	Vital status	Mortality is fully captured in SIDIAP. The cause of death is not available, but can be linked to the Spanish death registry on a project by project basis.
12	Limitations	The SIDIAP data is not representative of individuals not using public primary care, and conditions that are usually followed by specialist care might not be properly captured. In addition, there is limited information on lifestyle variables. Patients are followed until Death or when transferring to another primary health care centre that does not contribute to SIDIAP.

#	Section	Description
13	Main references	Recalde M, Rodríguez C,Burn E,Far M,García D,Carrere-Molina J,Benítez M,Moleras A,Pistillo A,Bolíbar B,Aragón M,Duarte-Salles T "Data Resource Profile: The Information System for Research in Primary Care (SIDIAP)." International journal of epidemiology (2022): 35415748
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/50190 Website: https://www.sidiap.org/index.php/en

The Netherlands: Integrated Primary Care Information (IPCI)

#	Section	Description
1	Data source identification and country	IPCI (Integrated Primary Care Information) Netherlands
2	Data partner information section	Erasmus University Medical Center (EMC) Department of Medical Informatics
3	Coverage and timespan	Data collection since: 2006 Extent: Nation-wide. IPCI is a Dutch database that contains patient records from 2006 onwards. However, it mainly covers the central part of the country, including the most densely populated area (the 'Randstad') and non-urban areas. IPCI contains information on all patients registered with GPs responsible for non-emergency care and referrals. A patient is registered at birth or at first encounter with the GP.
4	Healthcare setting / type of data	Primary care – General Practitioner. Data is collected from primary care EHR. This includes demographic information, complaints and symptoms, diagnoses, laboratory test results, lifestyle factors (in limited amount), and correspondence with secondary care, such as referral and discharge letters.
5	Data collection process	Outpatient electronic health records. Data is entered into the EHR system by the GPs, during or after the visit. The patientdossiers are collected by Erasmus MC data managers and combined in one harmonized database. Several checks are done on this database to ensure correct data processing. Persons can have dossiers at multiple GPs.
6	General representativeness	More than 99% of the Dutch population has health insurance, and almost all citizens are registered with a general practitioner. Over 12 months, around 78% of the population has at least one contact with their GP. IPCI included around 350 GP practices out of around 5000 in the country (~ 7%). The demographic composition of the IPCI population mirrors that of the general Dutch population in terms of age and sex.
7	Data content /source coding	Dutch GPs use mainly Dutch standard codes, like ICPC-1 and Diagnostische Bepalingen maintained by NHG. And for therapy the G-Standard is used, maintained by ZIndex.
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5 and the OMOP standard vocabularies. The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. Patients can be registered under different IDs, but since a patient can only be registered at one GP at a time, the observations periods will not overlap.
9	Quality control (data source specific)	Prior to each data release, extensive quality control steps are performed, e.g., comparison of patient characteristics between practices, and checks to identify abnormal temporal data patterns in practices. For each practice, around 200 quality indicators are obtained. Of these indicators, a quarter refer to population characteristics, e.g. number of birth and mortalities relative to practice size, temporal consistency. The other indicators are based on medical data, e.g. distribution of measurement values, frequencies of diagnoses and procedures relative to age, completeness of data. The indicators are combined in a couple of quality scores for each practice. For these scores, cut-off values for acceptable quality have been defined. Practices with a score below a cut-off are excluded for research. This approach has shown to be very important, for example to check if data from practices that just joined the database are at an

#	Section	Description
		acceptable level of quality. The details of the approach, like the cut-off values for acceptance, are based on years of experience. In addition, trends are compared with the previous database release. Extensive quality control steps are performed before each data release. These include comparing patient characteristics between practices and checks to identify abnormal temporal data patterns in practices. Additional checks include over 200 indicators related to population characteristics (e.g., reliability of birth and mortality rates) and medical data (e.g., availability of durations of prescriptions and completeness of laboratory results). Records of low quality are excluded from the database.
10	Linkage	Linkage requires additional approval steps and needs to be assessed on a case-by-case basis. IPCI is not routinely linked with other databases.
11	Vital status	Vital status (death date and cause) is collected based on GP records.
12	Limitations	The main limitation comes with the fact that IPCI is limited to GP records, and although it contains information on referrals and discharge letters, it may not fully capture specific hospital information. IPCI does not include coded/detailed data about medications/procedures/test results from the hospital or other care-providers.
13	Main references	de Ridder MAJ, de Wilde M,de Ben C,Leyba AR,Mosseveld BMT,Verhamme KMC,van der Lei J,Rijnbeek PR "Data Resource Profile: The Integrated Primary Care Information (IPCI) database, The Netherlands." International journal of epidemiology (2022): 35182144
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/42618 Website: http://www.ipci.nl

The United States: IQVIA US - Ambulatory EMR (IQVIA US - AmbEMR)

#	Section	Description
1	Data source identification and country	IQVIA US - AmbEMR (IQVIA US - Ambulatory EMR) United States
2	Data partner information section	IQVIA (IQVIA) -
3	Coverage and timespan	Data collection since: 2006 Extent: Nation-wide. Over 100 million patients, with data sourced from more than 800 ambulatory practices and 100,000+ physicians across the United States.
4	Healthcare setting / type of data	Primary care – General Practitioner, and primary care specialists (e.g. paediatricians). The IQVIA Ambulatory EMR-US dataset primarily covers ambulatory care settings, which include: Outpatient physician practices, Primary care and specialty clinics, Independent and group practices, and non-hospital-based care environments.
5	Data collection process	Outpatient electronic health records. Anonymized patient records collected from patient management software used by GPs and selected specialists to document patients' clinical encounters. These records are sent to a central location then processed and aggregated.
6	General representativeness	The data is representative of the US population, with a variety of socioeconomic status patients that is not limited to specific insurances.
7	Data content /source coding	The following source vocabularies are used: CPT, HCPCX, Dx, NDC.
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5 and the OMOP standard vocabularies. The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. IQVIA's patient tokenization process generates a consistent, anonymous patient identifier by

#	Section	Description
		applying multi-layered, irreversible hashing to encrypted patient attributes, enabling privacy-protected linkage of de-identified data across care settings without exposing personally identifiable information. This enables longitudinal tracking of patients.
9	Quality control (data source specific)	Integrity constraints including • 230+ IQVIA Internal QC checks • DQD & Achilles review • IQVIA internal business validation check
10	Linkage	Linkage to other IQVIA datasets is possible, e.g. PharmetricsPlus, but has to be done on a project basis.
11	Vital status	There is no linkage to a death registry.
12	Limitations	No data source specific limitations documented. General limitations for the data type applicable.
13	Main references	No main reference provided.
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111118 Website: https://www.iqvia.com/locations/united-states/library/fact-sheets/iqvia-ambulatory-emr-us

ANNEX II. Fitness for use assessment

Data source justification for inclusion and key characteristics

Belgium: Universitair Ziekenhuis Antwerpen Structured Data (UZA)

UZA will be included in this study because it is a hospital data source that provides relevant information on demographic, clinical, and treatment characteristics in the adult population (>18) with resistant hypertension.

Based on a preliminary feasibility assessment, the expected number of record counts for blood pressure measurements (All objectives) in UZA is 2,406,600.

Moreover, data availability and follow-up in UZA is sufficient, as data availability starts in 2021, and the date of the most recent data extraction is 9/2025, which aligns with the study period. The median follow-up of the first observation period in UZA is 345 (1–345) days.

There are no study specific limitations present in UZA.

Lastly, Institutional Review Board (IRB) approval for UZA is estimated to take 1 month, which makes the execution of this study feasible within the current study timelines.

Denmark: Danish Data Health Registries (DK-DHR)

DK-DHR will be included in this study because it is a registry data source that provides relevant information on demographic, clinical, and treatment characteristics in the adult population (>18) with resistant hypertension.

Based on a preliminary feasibility assessment, the expected number of record counts for blood pressure measurements (All objectives) in DK-DHR is 29,838,700.

Moreover, data availability and follow-up in DK-DHR is sufficient, as data availability starts in 1995, and the date of the most recent data extraction is 12/2024, which aligns with the study period. The median follow-up of the first observation period in DK-DHR is 7,920 (2,610 –10,900) days.

There are some study specific limitations present in DK-DHR, namely that blood pressure measurement values are only available for hospital records and not for primary care records (All objectives).

Lastly, DK-DHR has blanket approval, which makes the execution of this study feasible within the current study timelines.

Finland: Hospital District of Helsinki and Uusimaa (FinOMOP-HUS)

FinOMOP-HUS will be included in this study because it is a hospital data source that provides relevant information on demographic, clinical, and treatment characteristics in the adult population (>18) with resistant hypertension.

Based on a preliminary feasibility assessment, the expected number of record counts for blood pressure measurements (All objectives) in FinOMOP-HUS is 9,880,400.

Moreover, data availability and follow-up in FinOMOP-HUS is sufficient, as data availability starts in 2004, and the date of the most recent data extraction is 12/2025, which aligns with the study period. The median follow-up of the first observation period in FinOMOP-HUS is 3,040 (420 –6,120) days.

There are no study specific limitations present in FinOMOP-HUS.

Lastly, IRB approval for FinOMOP-HUS is estimated to take 1 to 3 months, which makes the execution of this study feasible within the current study timelines.

Spain: Institut Municipal Assistència Sanitària Information System (IMASIS)

IMASIS will be included in this study because it is a hospital data source that provides relevant information on demographic, clinical, and treatment characteristics in the adult population (>18) with resistant hypertension.

Based on a preliminary feasibility assessment, the expected number of record counts for blood pressure measurements (All objectives) in IMASIS is 15,346,800.

Moreover, data availability and follow-up in IMASIS is sufficient, as data availability starts in 1990, and the date of the most recent data extraction is 06/2025, which aligns with the study period. The median follow-up of the first observation period in IMASIS is 1,310 (0–6,340) days.

There are no study specific limitations present in IMASIS.

Lastly, IRB approval for IMASIS is estimated to take 1 to 3 months, which makes the execution of this study feasible within the current study timelines.

Spain: The Information System for Research in Primary Care (SIDIAP)

SIDIAP will be included in this study because it is a primary care data source that provides relevant information on demographic, clinical, and treatment characteristics in the adult population (>18) with resistant hypertension.

Based on a preliminary feasibility assessment, the expected number of record counts for blood pressure measurements (All objectives) in SIDIAP is 74,180,400.

Moreover, data availability and follow-up in SIDIAP is sufficient, as data availability starts in 2006, and the date of the most recent data extraction is 12/2024, which aligns with the study period. The median follow-up of the first observation period in SIDIAP is 5,760 (2,170–6,940) days.

There are no study specific limitations present in SIDIAP.

Lastly, IRB approval for SIDIAP is estimated to take 2 months, which makes the execution of this study feasible within the current study timelines.

The Netherlands: Integrated Primary Care Information (IPCI)

IPCI will be included in this study because it is a primary care data source that provides relevant information on demographic, clinical, and treatment characteristics in the adult population (>18) with resistant hypertension.

Based on a preliminary feasibility assessment, the expected number of record counts for blood pressure measurements (All objectives) in IPCI is 10,933,300.

Moreover, data availability and follow-up in IPCI is sufficient, as data availability starts in 2006, and the date of the most recent data extraction is 06/2025, which aligns with the study period. The median follow-up of the first observation period in IPCI is 1,790 (791–3,160) days.

There are no study specific limitations present in IPCI.

Lastly, IPCI has blanket approval, which makes the execution of this study feasible within the current study timelines.

The United States: IQVIA US - Ambulatory EMR (IQVIA US - AmbEMR)

IQVIA US - AmbEMR will be included in this study because it is a primary care data source that provides relevant information on demographic, clinical, and treatment characteristics in the adult population (>18) with resistant hypertension.

Based on a preliminary feasibility assessment, the expected number of record counts for blood pressure measurements (All objectives) in IQVIA US - AmbEMR is 3,900.

Moreover, data availability and follow-up in IQVIA US - AmbEMR is sufficient, as data availability starts in 2006, and the date of the most recent data extraction is 05/2025, which aligns with the study period. The median follow-up of the first observation period in IQVIA US - AmbEMR is 830 (107–2,240) days.

There are no study specific limitations present in IQVIA US - AmbEMR.

Lastly, IQVIA US - AmbEMR has blanket approval, which makes the execution of this study feasible within the current study timelines.

Table S1. Fitness-for-use assessment of data sources.

Design elements	Operational definition	Data elements for valid capture	Criticality of the quality of the element, including justification where relevant	Suggested extensiveness assessment	Suggested assessment of other quality dimensions	Suggested substantiation by documentation
Study population	Population with resistant hypertension (*obj. 1–2*)	Concomitant use of prespecified antihypertensive drug classes	High	100% will have concomitant prescriptions of prespecified antihypertensive drugs	100% of prescriptions have been mapped to RxNorm	N/A
		Blood pressure measurement	High	100% will have a blood pressure measurement of ≥140/90 mmHg measured recorded at least four weeks after initiation of third antihypertensive drug class.	N/A	N/A
Treatment/ exposure	Concomitant prescriptions of prespecified antihypertensive drugs	Prescription of prespecified antihypertensive treatment (RxNorm code)	High	100% will have concomitant prescriptions of prespecified antihypertensive drugs	100% of prescriptions have been mapped to RxNorm	N/A
Comparator group (not relevant)	N/A	N/A	N/A	N/A	N/A	N/A
Outcomes	Characteristics of individuals with resistant hypertension	Concept IDs for sex, age, and various clinical characteristics, conditions, and medication	Medium	N/A (to be assessed on a research question basis)	N/A	N/A
Covariates (including confounders if relevant)	N/A	N/A	N/A	N/A	N/A	N/A
Follow-up time (if relevant)	N/A	N/A	N/A	N/A	N/A	N/A

N/A = Not applicable, RxNorm = Medical prescription normalised.

EMA Data Quality Framework for EU medicines regulation: application to Real-World Data for more information (https://www.ema.europa.eu/system/files/documents/other/data-quality-framework-eu-medicines-regulation-application-real-world-data_en.pdf).

ANNEX III. Operational and reporting considerations

DATA MANAGEMENT

Data management

All data sources have previously mapped their data to the OMOP common data model. This enables the use of standardised analytics and using DARWIN EU® tools across the network, since the structure of the data and the terminology system is harmonised. The OMOP CDM was 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 and will use standardised analytics wherever possible. Each data partner will execute the study code against their data source containing patient-level data and then return the results (csv files), 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.

Data storage and protection

For this study, participants from various EU member states will process personal data from individuals that is collected in national/regional electronic health record data sources. 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 data sources 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 EU® Digital Research Environment (DRE). These output files do not contain any data that allow identification of subjects included in the study. The DRE implements further security measures 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.

QUALITY CONTROL

Data source quality control

When defining cohorts for indications, a systematic search of possible codes for inclusion will be identified using the *CodelistGenerator* R package (https://github.com/darwin-eu/CodelistGenerator). This package allows the user to define a search strategy and will use this to query the vocabulary tables of the OMOP common data model so as to find potentially relevant codes. In addition, the *CohortDiagnostics* (https://github.com/OHDSI/CohortDiagnostics), *DrugExposureDiagnostics* (https://cran.r-project.org/web/packages/DrugExposureDiagnostics/index.html) and the *PhenotypeR* (https://github.com/OHDSI/phenotypeR) R packages will be run, to assess the use of different codes across the data sources contributing to the study and identify any codes potentially omitted in error. The *DrugExposureDiagnostics* package evaluates ingredient-specific attributes and patterns in drug exposure records.

The study code will be based on DARWIN EU® R package *CohortCharacteristics* to characterise the cohort by indication. This package 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.

PLANS FOR DISSEMINATING AND COMMUNICATING STUDY RESULTS

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.

ANNEX IV. Lists with concept definitions and overview of variables

List of conditions and procedures concepts definition can be provided in the table below:

Table S2. Preliminary list of conditions and procedures definitions.

Phenotype	Concept name	Concept ID (including descendants)	Exclude concept ID	Vocabulary
Premature menopause	Premature menopause	198715	-	SNOMED
Premature menopause	Postsurgical menopause	4154697	-	SNOMED
Premature menopause	Primary ovarian failure	4279913	-	SNOMED
Premature menopause	Excision of bilateral ovaries	4297990	-	SNOMED
Premature menopause	Bilateral acquired absence of ovary	36716936	-	SNOMED
Premature menopause	46,XX ovarian dysgenesis, short stature syndrome	37165047	Yes	SNOMED
Premature menopause	Laparoscopic excision of bilateral ovarian cysts	43531188	Yes	SNOMED
Premature menopause	Lung fibrosis, immunodeficiency, 46,XX gonadal dysgenesis syndrome	36716112	Yes	SNOMED
Premature menopause	Ovarian dysgenesis	4004647	Yes	SNOMED
Premature menopause	Ovarioleukodystrophy	608037	Yes	SNOMED
Premature menopause	Pure gonadal dysgenesis 46,XX	4317840	Yes	SNOMED
Premature menopause	Bilateral salpingo-oophorectomy with omentectomy, total abdominal hysterectomy and radical dissection for debulking	2110268	Yes	CPT4
Premature menopause	Bilateral salpingo-oophorectomy with omentectomy, total abdominal hysterectomy and radical dissection for debulking; with pelvic lymphadenectomy and limited para-aortic lymphadenectomy	2110269	Yes	CPT4
Premature menopause	Bilateral salpingo-oophorectomy with total omentectomy, total abdominal hysterectomy for malignancy	2110270	Yes	CPT4
Premature menopause	Laparoscopic removal of remaining ovary	2004251	Yes	ICD9Proc
Premature menopause	Laparoscopic removal of remaining ovary and tube	2004266	Yes	ICD9Proc
Premature menopause	Laparoscopy, surgical, supracervical hysterectomy, for uterus 250 g or less	2110210	Yes	CPT4
Premature menopause	Laparoscopy, surgical, supracervical hysterectomy, for uterus greater than 250 g	2110212	Yes	CPT4

Phenotype	Concept name	Concept ID (including descendants)	Exclude concept ID	Vocabulary
Premature menopause	Laparoscopy, surgical, with total hysterectomy, for uterus 250 g or less	2110230	Yes	CPT4
Premature menopause	Laparoscopy, surgical, with total hysterectomy, for uterus greater than 250 g	2110232	Yes	CPT4
Premature menopause	Laparoscopy, surgical, with vaginal hysterectomy, for uterus 250 g or less	2110218	Yes	CPT4
Premature menopause	Laparoscopy, surgical, with vaginal hysterectomy, for uterus greater than 250 g	2110220	Yes	CPT4
Premature menopause	Oophorectomy, partial or total, unilateral or bilateral	2110264	Yes	CPT4
Premature menopause	Other removal of both ovaries and tubes at same operative episode	2004263	Yes	ICD9Proc
Premature menopause	Other removal of both ovaries at same operative episode	2004248	Yes	ICD9Proc
Premature menopause	Other removal of remaining ovary and tube	2004264	Yes	ICD9Proc
Premature menopause	Resection (initial) of ovarian, tubal or primary peritoneal malignancy with bilateral salpingo-oophorectomy and omentectomy	2110265	Yes	CPT4
Premature menopause	Resection (initial) of ovarian, tubal or primary peritoneal malignancy with bilateral salpingo-oophorectomy and omentectomy; with radical dissection for debulking (ie, radical excision or destruction, intra-abdominal or retroperitoneal tumors)	2110267	Yes	CPT4
Premature menopause	Resection (initial) of ovarian, tubal or primary peritoneal malignancy with bilateral salpingo-oophorectomy and omentectomy; with total abdominal hysterectomy, pelvic and limited para-aortic lymphadenectomy	2110266	Yes	CPT4
Premature menopause	Other removal of remaining ovary	2004249	Yes	ICD9Proc
Premature menopause	Wedge resection of bilateral ovaries	4195462	Yes	SNOMED
Renal artery stenosis	Renal artery stenosis	4118795	no	SNOMED
Hyperthyroidism	Hyperthyroidism	4142479	no	SNOMED
Hypothyroidism	Hypothyroidism	140673	No	SNOMED
pheochromocytoma	pheochromocytoma	4118993	No	SNOMED
pheochromocytoma	pheochromocytoma	4311087	no	SNOMED
Hypercortisolism (cushing's syndrome)	Hypercortisolism	195212	no	SNOMED
Aortic coarctation	Coarctation of aorta	321119	no	SNOMED
Diabetes mellitus	Diabetes mellitus	201820	No	SNOMED
Diabetes mellitus	Complication due to diabetes mellitus	442793	No	SNOMED

Phenotype	Concept name	Concept ID (including descendants)	Exclude concept ID	Vocabulary
Left ventricular outflow obstruction	Aortic valve stenosis	4189343	No	SNOMED
Left ventricular outflow obstruction	Aortic valve regurgitation	315564	No	SNOMED
Left ventricular outflow obstruction	Left ventricular outflow tract obstruction	4103137	No	SNOMED
Left ventricular outflow obstruction	Hypertrophic obstructive cardiomyopathy	316428	No	SNOMED
Atrial fibrillation	Atrial fibrillation	313217	No	SNOMED
Hepatic impairment	Hepatic failure	4245975	No	SNOMED
Hepatic impairment	Cirrhosis of liver	4064161	No	SNOMED
Hepatic impairment	Chronic necrosis of liver	3655440	No	SNOMED
left bundle branch block and cardiac arrhythmia	Left bundle branch block	316998	No	SNOMED
left bundle branch block and cardiac arrhythmia	Cardiac arrhythmia	44784217	No	SNOMED
long QT syndrome	Acquired long QT syndrome	40479264	Yes	SNOMED
long QT syndrome	long QT syndrome	314664	No	SNOMED
Hypertrophic cardiomegaly	Hypertrophic cardiomegaly	40483223	No	SNOMED
Hypertrophic cardiomegaly	Cardiac hypertrophy secondary to systemic hypertension	42599748	Yes	SNOMED Veterinary
Pregnancy or breastfeeding	Pregnancy, childbirth and puerperium finding	4088927	No	SNOMED
Pregnancy or breastfeeding	Pregnancy	4299535	No	SNOMED
Pregnancy or breastfeeding	Mother currently breastfeeding	3655578	No	SNOMED
Pregnancy or breastfeeding	Breastfeeding (mother)	4188824	No	SNOMED
Pregnancy or breastfeeding	2 miscarriages	1340198	Yes	OMOP Extension
Pregnancy or breastfeeding	3 miscarriages	1340199	Yes	OMOP Extension
Pregnancy or breastfeeding	4 miscarriages	1340200	Yes	OMOP Extension
Pregnancy or breastfeeding	5 miscarriages	1340201	Yes	OMOP Extension
Pregnancy or breastfeeding	6 miscarriages	1340202	Yes	OMOP Extension

Phenotype	Concept name	Concept ID (including descendants)	Exclude concept ID	Vocabulary
Pregnancy or breastfeeding	7 or more miscarriages	1340203	Yes	OMOP Extension
Pregnancy or breastfeeding	Not pregnant	4243056	Yes	SNOMED
Pregnancy or breastfeeding	Pregnancy not yet confirmed	45757490	Yes	SNOMED
Adrenal insufficiency	Adrenal cortical hypofunction	435508	No	SNOMED
COVID-19	Acute bronchitis caused by SARS-CoV-2	3661405	no	SNOMED
COVID-19	Acute bronchitis caused by SARS-CoV-2	3661405	no	SNOMED
COVID-19	Acute hypoxemic respiratory failure due to COVID-19	3655976	no	SNOMED
COVID-19	Acute kidney injury due to COVID-19	3661748	no	SNOMED
COVID-19	Acute respiratory distress syndrome due to COVID-19	3661406	no	SNOMED
COVID-19	Asymptomatic SARS-CoV-2	3662381	no	SNOMED
COVID-19	Bronchitis caused by COVID-19	756031	no	OMOP Extension
COVID-19	Cardiomyopathy due to COVID-19	3656667	no	SNOMED
COVID-19	Conjunctivitis due to COVID-19	3656668	no	SNOMED
COVID-19	Coronavirus infection	439676	no	SNOMED
COVID-19	COVID-19	37311061	no	SNOMED
COVID-19	Disease caused by Coronaviridae	4100065	no	SNOMED
COVID-19	Dyspnea caused by SARS-CoV-2	3656669	no	SNOMED
COVID-19	Encephalopathy due to COVID-19	37310284	no	SNOMED
COVID-19	Fever caused by SARS-CoV-2	3661885	no	SNOMED
COVID-19	Gastroenteritis caused by SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2)	37310283	no	SNOMED
COVID-19	Infection of upper respiratory tract caused by SARS-CoV-2	37310286	no	SNOMED
COVID-19	Infection of lower respiratory tract caused by SARS-CoV-2	3663281	no	SNOMED
COVID-19	Lymphocytopenia due to COVID-19	3661631	no	SNOMED
COVID-19	Myocarditis due to COVID-19	37310287	no	SNOMED
COVID-19	Otitis media due to COVID-19	37310254	no	SNOMED
COVID-19	Patient meets COVID-19 clinical diagnostic criteria	704995	no	OMOP Extension
COVID-19	Patient meets COVID-19 laboratory confirmation criterion (detection of specific RNA in a clinical specimen using a molecular amplification detection test)	700297	no	OMOP Extension
COVID-19	Patient meets COVID-19 laboratory diagnostic criteria	704996	no	OMOP Extension

Phenotype	Concept name	Concept ID (including descendants)	Exclude concept ID	Vocabulary
COVID-19	Patient meets COVID-19 presumptive laboratory evidence criteria (detection of specific antigen in a clinical specimen, OR detection of specific antibody in serum, plasma, or whole blood indicative of a new or recent infection)	700296	no	OMOP Extension
COVID-19	Pneumonia caused by Human coronavirus	37016927	no	SNOMED
COVID-19	Pneumonia caused by SARS-CoV-2	3661408	no	SNOMED
COVID-19	Pneumonia caused by Severe acute respiratory syndrome coronavirus	40479642	no	SNOMED
COVID-19	Respiratory infection caused by COVID-19	756039	no	OMOP Extension
COVID-19	Rhabdomyolysis due to COVID-19	3655977	no	SNOMED
COVID-19	Sepsis due to disease caused by COVID-19	3655975	no	SNOMED
COVID-19	Severe acute respiratory syndrome	320651	no	SNOMED
COVID-19	Severe acute respiratory syndrome of upper respiratory tract	37396171	no	SNOMED
COVID-19	Suspected coronavirus infection	45763724	no	SNOMED
COVID-19	Suspected COVID-19	37311060	no	SNOMED
COVID-19	Thrombocytopenia due to COVID-19	3661632	no	SNOMED
Stroke	Intracranial hemorrhage	439847	No	SNOMED
Stroke	Cerebrovascular accident	381316	no	SNOMED
hypertensive encephalopathy	Hypertensive encephalopathy	312938	No	SNOMED
acute heart failure	Acute heart failure	442310	No	SNOMED
acute coronary syndrome	Acute coronary syndrome	4215140	No	SNOMED
Percutaneous coronary intervention	Percutaneous coronary intervention	4216130	No	SNOMED
coronary artery bypass grafting	Coronary artery bypass graft	4336464	No	SNOMED
Migraine	Migraine	318736	No	SNOMED
Heart failure	Heart failure	316139	No	SNOMED
Coronary artery disease	Obliterative coronary artery disease	4168972	No	SNOMED
Coronary artery disease	Multi vessel coronary artery disease	4155007	No	SNOMED
Coronary artery disease	Coronary artery disease due to type 2 diabetes mellitus	37309630	No	SNOMED
Obesity	Laparoscopic bypass of stomach	46270974	no	SNOMED
Obesity	Maintenance of gastric band	40482679	no	SNOMED
Obesity	Printen and Mason operation, high gastric bypass	4148762	no	SNOMED

Phenotype	Concept name	Concept ID (including descendants)	Exclude concept ID	Vocabulary
Obesity	Removal of gastric band	4142551	no	SNOMED
Obesity	Revision of gastrojejunostomy	4307444	no	SNOMED
Obesity	Roux-en-Y gastrojejunostomy	4069249	no	SNOMED
Obesity	Single anastomosis duodeno-ileal bypass with sleeve gastrectomy	507338	no	SNOMED
Obesity	Sleeve resection of stomach	4227605	no	SNOMED
Obesity	Obese	4215968	no	SNOMED
Obesity	Obesity monitoring status	4087250	no	SNOMED
Obesity	Bypass of stomach	40483096	no	SNOMED
Obesity	Duodenal switch	4145504	no	SNOMED
Obesity	Gastric band procedure complication	45757691	no	SNOMED
Obesity	Gastric banding for obesity	3170866	no	Nebraska Lexicon
Obesity	Gastric stapling for obesity	4307794	no	SNOMED
Obesity	Infection due to gastric band procedure	46270494	no	SNOMED
Obesity	Intestinal bypass for morbid obesity	4245767	no	SNOMED
Obesity	Jejunocecostomy for obesity	4135346	no	SNOMED
Obesity	Jejunocolostomy for obesity	4232489	no	SNOMED
Obesity	Jejunoileostomy bypass shunt for obesity	4234135	no	SNOMED
Obesity	Laparoscopic adjustable gastric banding	4190525	no	SNOMED
Obesity	Attention to connection of stomach to jejunum	44789822	yes	SNOMED
Obesity	Attention to connection of stomach to transposed jejunum	44789820	yes	SNOMED
Chronic kidney disease	Chronic kidney disease	46271022	no	SNOMED
Chronic kidney disease stage 1	Chronic kidney disease stage 1	443614	no	SNOMED
Chronic kidney disease stage 2	Chronic kidney disease stage 2	443601	no	SNOMED
Chronic kidney disease stage 3	Chronic kidney disease stage 3	443597	no	SNOMED
Chronic kidney disease stage 4	Chronic kidney disease stage 4	443612	no	SNOMED
Chronic kidney disease stage 5	Chronic kidney disease stage 5	443611	no	SNOMED
Cardiovascular disease	Acute coronary syndrome	4215140	no	SNOMED
Cardiovascular disease	Atrial fibrillation	313217	no	SNOMED
Cardiovascular disease	Chronic ischemic heart disease	315286	no	SNOMED

Phenotype	Concept name	Concept ID (including descendants)	Exclude concept ID	Vocabulary
Cardiovascular disease	Heart failure	316139	no	SNOMED
Cardiovascular disease	Cerebrovascular accident	381316	no	SNOMED
Cardiovascular disease	Peripheral arterial disease	3654996	no	SNOMED
Cardiovascular disease	Disorder of coronary artery	4187067	no	SNOMED
Cardiovascular disease	Myocardial infarction	4329847	no	SNOMED
Cardiovascular disease	Intracranial hemorrhage	439847	no	SNOMED
Aldosteronism	Primary aldosteronism	434000	no	SNOMED
Coronary artery disease and myocardial infarction	Old myocardial infarction	314666	yes	SNOMED
Coronary artery disease and myocardial infarction	Myocardial infarction	4329847	no	SNOMED
Coronary artery disease and myocardial infarction	Coronary arteriosclerosis	317576	no	SNOMED
Coronary artery disease and myocardial infarction	Chronic ischemic heart disease	315286	no	SNOMED
Arrhythmias	Ventricular tachycardia	4103295	no	SNOMED
Arrhythmias	Cardiac pacemaker procedure	4051938	no	SNOMED
Arrhythmias	Cardiac arrhythmia	44784217	no	SNOMED
Arrhythmias	Ablation operation for arrhythmia	4050568	no	SNOMED

List of medicines concepts definition can be provided in the table below:

Table S3. Preliminary list of medicines definitions.

Substance Name	Concept name	Class	ATC code	Ingredient Concept ID	Include descendants
althiazide	althiazide	diuretics	C03	19018435	yes
amiloride	amiloride	diuretics	C03	991382	yes
bendroflumethiazide	bendroflumethiazide	diuretics	C03	1316354	yes
bumetanide	bumetanide	diuretics	C03	932745	yes
buthiazide	buthiazide	diuretics	C03	19039770	yes
chlorthalidone	chlorthalidone	diuretics	C03	1395058	yes
cicletanine	cicletanine	diuretics	C03	19051444	yes

Substance Name	Concept name	Class	ATC code	Ingredient Concept ID	Include descendants
clopamide	clopamide	diuretics	C03	19099181	yes
Epithiazide	Epithiazide	diuretics	C03	37493796	yes
eplerenone	eplerenone	diuretics	C03	1309799	yes
finerenone	finerenone	diuretics	C03	1537451	yes
FENQUIZONE	FENQUIZONE	diuretics	C03	36859209	yes
furosemide	furosemide	diuretics	C03	956874	yes
hydrochlorothiazide	hydrochlorothiazide	diuretics	C03	974166	yes
hydroflumethiazide	hydroflumethiazide	diuretics	C03	1376289	yes
indapamide	indapamide	diuretics	C03	978555	yes
metolazone	metolazone	diuretics	C03	907013	yes
piretanide	piretanide	diuretics	C03	19025498	yes
spironolactone	spironolactone	diuretics	C03	970250	yes
tolvaptan	tolvaptan	diuretics	C03	19036797	yes
torsemide	torsemide	diuretics	C03	942350	yes
triamterene	triamterene	diuretics	C03	904542	yes
trichlormethiazide	trichlormethiazide	diuretics	C03	904639	yes
cilazapril	cilazapril	ACE inhibitors	C09A	19050216	yes
spirapril	spirapril	ACE inhibitors	C09A	19040051	yes
temocapril hydrochloride	temocapril hydrochloride	ACE inhibitors	C09A	43009010	yes
enalaprilat	enalaprilat	ACE inhibitors	C09A	1342001	yes
lisinopril	lisinopril	ACE inhibitors	C09A	1308216	yes
ramipril	ramipril	ACE inhibitors	C09A	1334456	yes
moexipril	moexipril	ACE inhibitors	C09A	1310756	yes
fosinopril	fosinopril	ACE inhibitors	C09A	1363749	yes
enalapril	enalapril	ACE inhibitors	C09A	1341927	yes
trandolapril	trandolapril	ACE inhibitors	C09A	1342439	yes
captopril	captopril	ACE inhibitors	C09A	1340128	yes
imidapril	imidapril	ACE inhibitors	C09A	19122327	yes
quinapril	quinapril	ACE inhibitors	C09A	1331235	yes
delapril hydrochloride	delapril hydrochloride	ACE inhibitors	C09A	36878917	yes
zofenopril	zofenopril	ACE inhibitors	C09A	19102107	yes
benazepril	benazepril	ACE inhibitors	C09A	1335471	yes
perindopril	perindopril	ACE inhibitors	C09A	1373225	yes
azilsartan	azilsartan	ARB	C09C	40235485	yes
olmesartan	olmesartan	ARB	C09C	40226742	yes
eprosartan	eprosartan	ARB	C09C	1346686	yes

Substance Name	Concept name	Class	ATC code	Ingredient Concept ID	Include descendants
irbesartan	irbesartan	ARB	C09C	1347384	yes
candesartan	candesartan	ARB	C09C	1351557	yes
losartan	losartan	ARB	C09C	1367500	yes
valsartan	valsartan	ARB	C09C	1308842	yes
fimasartan potassium	fimasartan potassium	ARB	C09C	43009001	yes
telmisartan	telmisartan	ARB	C09C	1317640	yes
amlodipine	amlodipine	calcium channel blockers	C08	1332418	yes
benidipine hydrochloride	benidipine hydrochloride	calcium channel blockers	C08	43009040	yes
bepridil	bepridil	calcium channel blockers	C08	1319751	yes
cilnidipine	cilnidipine	calcium channel blockers	C08	43009017	yes
clevidipine	clevidipine	calcium channel blockers	C08	19089969	yes
diltiazem	diltiazem	calcium channel blockers	C08	1328165	yes
etripamil	etripamil	calcium channel blockers	C08	1253597	yes
felodipine	felodipine	calcium channel blockers	C08	1353776	yes
fendiline	fendiline	calcium channel blockers	C08	19053866	yes
gallopamil	gallopamil	calcium channel blockers	C08	19057715	yes
isradipine	isradipine	calcium channel blockers	C08	1326012	yes
lacidipine	lacidipine	calcium channel blockers	C08	19004539	yes
lercanidipine	lercanidipine	calcium channel blockers	C08	19015802	yes
levamlodipine	levamlodipine	calcium channel blockers	C08	1146278	yes
lidoflazine	lidoflazine	calcium channel blockers	C08	19124331	yes
manidipine	manidipine	calcium channel blockers	C08	19071995	yes
mepirodipine	mepirodipine	calcium channel blockers	C08	19102106	yes
mibefradil	mibefradil	calcium channel blockers	C08	1345141	yes

Substance Name	Concept name	Class	ATC code	Ingredient Concept ID	Include descendants
nicardipine	nicardipine	calcium channel blockers	C08	1318137	yes
nifedipine	nifedipine	calcium channel blockers	C08	1318853	yes
nilvadipine	nilvadipine	calcium channel blockers	C08	19113063	yes
nimodipine	nimodipine	calcium channel blockers	C08	1319133	yes
nisoldipine	nisoldipine	calcium channel blockers	C08	1319880	yes
nitrendipine	nitrendipine	calcium channel blockers	C08	19020061	yes
perhexiline	perhexiline	calcium channel blockers	C08	19032359	yes
verapamil	verapamil	calcium channel blockers	C08	1307863	yes
acebutolol	acebutolol	beta blocking agents	C07	1319998	yes
alprenolol	alprenolol	beta blocking agents	C07	19081284	yes
atenolol	atenolol	beta blocking agents	C07	1314002	yes
betaxolol	betaxolol	beta blocking agents	C07	1322081	yes
bevantolol hydrochloride	bevantolol hydrochloride	beta blocking agents	C07	43009042	yes
bisoprolol	bisoprolol	beta blocking agents	C07	1338005	yes
bopindolol	bopindolol	beta blocking agents	C07	19018640	yes
bupranolol	bupranolol	beta blocking agents	C07	19018489	yes
fcarteolol	carteolol	beta blocking agents	C07	950370	yes
carvedilol	carvedilol	beta blocking agents	C07	1346823	yes
celiprolol	celiprolol	beta blocking agents	C07	19049145	yes
CLORANOLOL	CLORANOLOL	beta blocking agents	C07	36860637	yes
epanolol	epanolol	beta blocking agents	C07	36855238	yes
esmolol	esmolol	beta blocking agents	C07	19063575	yes

Substance Name	Concept name	Class	ATC code	Ingredient Concept ID	Include descendants
labetalol	labetalol	beta blocking agents	C07	1386957	yes
landiolol hydrochloride	landiolol hydrochloride	beta blocking agents	C07	35197852	yes
mepindolol	mepindolol	beta blocking agents	C07	19072028	yes
metoprolol	metoprolol	beta blocking agents	C07	1307046	yes
nadolol	nadolol	beta blocking agents	C07	1313200	yes
nebivolol	nebivolol	beta blocking agents	C07	1314577	yes
oxprenolol	oxprenolol	beta blocking agents	C07	19024904	yes
penbutolol	penbutolol	beta blocking agents	C07	1327978	yes
pindolol	pindolol	beta blocking agents	C07	1345858	yes
practolol	practolol	beta blocking agents	C07	19135791	yes
propranolol	propranolol	beta blocking agents	C07	1353766	yes
sotalol	sotalol	beta blocking agents	C07	1370109	yes
talinolol	talinolol	beta blocking agents	C07	19100435	yes
tertatolol	tertatolol	beta blocking agents	C07	19100451	yes
timolol	timolol	beta blocking agents	C07	902427	yes
Progestogens and estrogens, sequential preparations	Progestogens and estrogens, sequential preparations	Hormonal contraception	G03AB	21602488	yes
Progestogens and estrogens, fixed combinations	Progestogens and estrogens, fixed combinations	Hormonal contraception	G03AA	21602473	yes
Progestogens	Progestogens	Hormonal contraception	G03AC	21602496	yes
plastic IUD with progestogen	plastic IUD with progestogen	Hormonal contraception	G02BA03	21602449	yes
Intravaginal contraceptives	Intravaginal contraceptives	Hormonal contraception	G02BB	21602450	yes
cyproterone and estrogen; oral	cyproterone and estrogen; oral	Hormonal contraception	G03HB01	21602617	yes

Substance Name	Concept name	Class	ATC code	Ingredient Concept ID	Include descendants
quinidine	quinidine	QT prolongation medication	C01B	1360421	yes
procainamide	procainamide	QT prolongation medication	C01B	1351461	yes
amiodarone	amiodarone	QT prolongation medication	C01B	1309944	yes
sotalol	sotalol	QT prolongation medication	C07A	1370109	yes
dofetilide	dofetilide	QT prolongation medication	C01B	1362979	yes
haloperidol	haloperidol	QT prolongation medication	N05A	766529	yes
ziprasidone	ziprasidone	QT prolongation medication	N05A	712615	yes
quetiapine	quetiapine	QT prolongation medication	N05A	766814	yes
thioridazine	thioridazine	QT prolongation medication	N05A	700299	yes
olanzapine	olanzapine	QT prolongation medication	N05A	785788	yes
risperidone	risperidone	QT prolongation medication	N05A	735979	yes
flecainide	flecainide	QT prolongation medication	C01B	1354860	yes
Macrolides	Macrolides	QT prolongation medication	J01F	21602969	yes
Fluoroquinolones	Fluoroquinolones	QT prolongation medication	J01M	21603007	yes
imipramine	imipramine	QT prolongation medication	N06A	778268	yes
citalopram	citalopram	QT prolongation medication	N06A	797617	yes
sumatriptan	sumatriptan	QT prolongation medication	N02C	1140643	yes
ondansetron	ondansetron	QT prolongation medication	A04A	1000560	yes
cisapride	cisapride	QT prolongation medication	A03F	932196	yes
amitriptyline	amitriptyline	QT prolongation medication	N06A	710062	yes
ANTIARRHYTHMICS, CLASS I AND III	ANTIARRHYTHMICS, CLASS I AND III	antiarrhythmic medications or potassium-sparing diuretics	C01B	21600248	Yes

Substance Name	Concept name	Class	ATC code	Ingredient Concept ID	Include descendants
ALDOSTERONE ANTAGONISTS AND OTHER POTASSIUM-SPARING AGENTS	ALDOSTERONE ANTAGONISTS AND OTHER POTASSIUM-SPARING AGENTS	antiarrhythmic medications or potassium-sparing diuretics	C03D	21601532	yes
sodium zirconium cyclosilicate	sodium zirconium cyclosilicate	Potassium binders	V03A	1510687	yes
sodium polystyrene sulfonate	sodium polystyrene sulfonate	Potassium binders	V03A	19078126	yes
patiromer	patiromer	Potassium binders	V03A	35602991	yes
carbamazepine	carbamazepine	CYP3A strong inducers, NSAIDs, mineralocorticoid receptor antagonists, systemic steroids	N03A	740275	yes
phenytoin	phenytoin	CYP3A strong inducers, NSAIDs, mineralocorticoid receptor antagonists, systemic steroids	N03A	740910	yes
rifampin	rifampin	CYP3A strong inducers, NSAIDs, mineralocorticoid receptor antagonists, systemic steroids	J04A	1763204	yes
apalutamide	apalutamide	CYP3A strong inducers, NSAIDs, mineralocorticoid receptor antagonists, systemic steroids	L02B	963987	yes
enzalutamide	enzalutamide	CYP3A strong inducers, NSAIDs, mineralocorticoid receptor antagonists, systemic steroids	L02B	42900250	yes
ivosidenib	ivosidenib	CYP3A strong inducers, NSAIDs, mineralocorticoid receptor antagonists, systemic steroids	L01X	1560123	yes
mitotane	mitotane	CYP3A strong inducers, NSAIDs, mineralocorticoid receptor	L01X	1309161	yes

Substance Name	Concept name	Class	ATC code	Ingredient Concept ID	Include descendants
		antagonists, systemic steroids			
St. John's wort extract	St. John's wort extract	CYP3A strong inducers, NSAIDs, mineralocorticoid receptor antagonists, systemic steroids	N06A	1398039	yes
ivacaftor / lumacaftor Oral Product	ivacaftor / lumacaftor Oral Product	CYP3A strong inducers, NSAIDs, mineralocorticoid receptor antagonists, systemic steroids	R07A	36248596	yes
ivacaftor / lumacaftor Granule Product	ivacaftor / lumacaftor Granule Product	CYP3A strong inducers, NSAIDs, mineralocorticoid receptor antagonists, systemic steroids	R07A	35200175	yes
Aldosterone antagonists	Aldosterone antagonists	CYP3A strong inducers, NSAIDs, mineralocorticoid receptor antagonists, systemic steroids	C03DA	21601533	yes
CORTICOSTEROIDS FOR SYSTEMIC USE	CORTICOSTEROIDS FOR SYSTEMIC USE	CYP3A strong inducers, NSAIDs, mineralocorticoid receptor antagonists, systemic steroids	H02	21602722	yes
ANTIINFLAMMATORY AND ANTIRHEUMATIC PRODUCTS, NON-STEROIDS	ANTIINFLAMMATORY AND ANTIRHEUMATIC PRODUCTS, NON-STEROIDS	CYP3A strong inducers, NSAIDs, mineralocorticoid receptor antagonists, systemic steroids	M01A	21603933	yes
ALKYLATING AGENTS	ALKYLATING AGENTS	Cytotoxic therapy	L01A	21601388	yes
ANTIMETABOLITES	ANTIMETABOLITES	Cytotoxic therapy	L01B	21601421	yes
PLANT ALKALOIDS AND OTHER NATURAL PRODUCTS	PLANT ALKALOIDS AND OTHER NATURAL PRODUCTS	Cytotoxic therapy	L01C	21601445	yes
CYTOTOXIC ANTIBIOTICS AND RELATED SUBSTANCES	CYTOTOXIC ANTIBIOTICS AND RELATED SUBSTANCES	Cytotoxic therapy	L01D	21603728	yes
Other antineoplastic agents	Other antineoplastic agents	Cytotoxic therapy	L01X	21603746	yes

List of measurement concepts definition can be provided in the table below:

Table S4. Preliminary list of measurement definitions.

Phenotype	Concept name	Concept ID (including descendants)	Exclude concept ID	Vocabulary
Systolic blood pressure	Systolic blood pressure	3004249	No	LOINC
Systolic blood pressure	Systolic blood pressure	4152194	No	SNOMED
Systolic blood pressure	Systolic blood pressure of neonate at birth	45769778	Yes	SNOMED
Systolic blood pressure	Systolic blood pressure centile	44806887	Yes	SNOMED
Systolic blood pressure	Self reported systolic blood pressure	37160570	Yes	SNOMED
Diastolic blood pressure	Diastolic blood pressure	3012888	No	LOINC
Diastolic blood pressure	Diastolic blood pressure	4154790	No	SNOMED
Diastolic blood pressure	Diastolic blood pressure of neonate at birth	45769783	Yes	SNOMED
Diastolic blood pressure	Diastolic blood pressure centile	44806886	Yes	SNOMED
Diastolic blood pressure	Self reported diastolic blood pressure	37160568	Yes	SNOMED
Glomerular filtration rate	Glomerular filtration rate	4213477	no	SNOMED
Serum potassium	Serum potassium measurement	4154489	no	SNOMED
Serum potassium	Serum potassium level below reference range	4043088	no	SNOMED
Serum potassium	Serum potassium level below reference range	4041897	no	SNOMED
Serum sodium	Serum sodium level	37392172	no	SNOMED
Cortisol	Cortisol measurement	4265057	no	SNOMED
Electrocardiogram	Electrocardiographic procedure	4163951	no	SNOMED
Electrocardiogram	Electrocardiogram finding	4117134	no	SNOMED
Q-T interval corrected	Q-T interval corrected	3026258	no	LOINC
Heart rate	Heart rate	3027018	No	LOINC
Heart rate	Resting heart rate	40481601	no	SNOMED
Hemoglobin A1c	Hemoglobin A1c measurement	4184637	no	SNOMED

Phenotype	Concept name	Concept ID (including descendants)	Exclude concept ID	Vocabulary
Body mass index	Measurement of body mass index	44783982	no	SNOMED
Body mass index	Body mass index (BMI) [Ratio]	3038553	no	LOINC
Body mass index	Body mass index	4245997	no	SNOMED
Body weight	General body weight	1003383	no	LOINC
Body weight	Body weight	4099154	no	SNOMED
Body weight	Body mass index (BMI) | Patient | General body weight	1002813	yes	LOINC
Body weight	Body weight | Fetus | General body weight	1003642	yes	LOINC
Body weight	Body weight | Fetus | OB ultrasound	1003206	yes	LOINC
Body weight	Body weight | Mother | Specific body weight	1003088	yes	LOINC
Body height	General body height (length)	1003116	no	LOINC
Body height	Body height measure	4177340	no	SNOMED
Body height	Body height | Father | General body height (length)	1003434	yes	LOINC
Body height	Body height | Mother | General body height (length)	1003546	yes	LOINC
Body height	Body height Father	42527660	yes	LOINC
Body height	Body height Measured --at birth	36304231	yes	LOINC
Body height	Body height Mother	42527661	yes	LOINC
Body height	Body height --lying	3014149	yes	LOINC
Body height	Body height --post partum	3013842	yes	LOINC
Body height	Body height --preoperative	3008989	yes	LOINC
Body height	Body height --sitting	37020651	yes	LOINC
Body height	Body height of biological father	1077435	yes	SNOMED
Body height	Body height of biological mother	1077436	yes	SNOMED
Body height	Height from demispan	4087493	yes	SNOMED
Body height	Mid-parental height	45770285	yes	SNOMED
Body height	Predicted adult height	45770286	yes	SNOMED
Body height	Pubis to ground height	4093977	yes	SNOMED
Body height	Sitting height	4093976	yes	SNOMED
Body height	Subischial leg length	4173009	yes	SNOMED
COVID 19 measurement	2019-ncov coronavirus, sars-cov-2/2019-ncov (covid-19), any technique, multiple types or subtypes (includes all targets), non-cdc	40218804	no	HCPCS
COVID 19 measurement	Cdc 2019 novel coronavirus (2019-ncov) real-time rt-pcr diagnostic panel	40218805	no	HCPCS
COVID 19 measurement	Coronavirus nucleic acid detection	44789510	no	SNOMED
COVID 19 measurement	Coronavirus nucleic acid detection assay	44811805	no	SNOMED

Phenotype	Concept name	Concept ID (including descendants)	Exclude concept ID	Vocabulary
COVID 19 measurement	Coronavirus RNA (ribonucleic acid) detection assay	45770687	no	SNOMED
COVID 19 measurement	Coronavirus RNA (ribonucleic acid) measurement by NAAT (nucleic acid amplification test)	44807536	no	SNOMED
COVID 19 measurement	Detection of ribonucleic acid of SARS-CoV-2 using polymerase chain reaction	3667069	no	SNOMED
COVID 19 measurement	Human coronavirus 229E RNA [Presence] in Lower respiratory specimen by NAA with non-probe detection	36660491	no	LOINC
COVID 19 measurement	Human coronavirus HKU1 RNA [Presence] in Lower respiratory specimen by NAA with non-probe detection	36659667	no	LOINC
COVID 19 measurement	Human coronavirus NL63 RNA [Presence] in Lower respiratory specimen by NAA with non-probe detection	36660329	no	LOINC
COVID 19 measurement	Human coronavirus OC43 RNA [Presence] in Lower respiratory specimen by NAA with non-probe detection	36660364	no	LOINC
COVID 19 measurement	Infectious agent antigen detection by immunoassay technique, qualitative or semiquantitative, multiple-step method	742224	no	CPT4
COVID 19 measurement	Infectious agent detection by nucleic acid (DNA or RNA)	700360	no	CPT4
COVID 19 measurement	Infectious disease (bacterial or viral respiratory tract infection), pathogen-specific nucleic acid (DNA or RNA), 22 targets including severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), qualitative RT-PCR, nasopharyngeal swab	742218	no	CPT4
COVID 19 measurement	Infectious disease (bacterial or viral respiratory tract infection), pathogen-specific nucleic acid (DNA or RNA), 22 targets including severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), qualitative RT-PCR, nasopharyngeal swab	742219	no	CPT4
COVID 19 measurement	Influenza virus A and B and SARS-CoV-2 (COVID-19) and SARS-related CoV RNA panel - Respiratory system specimen by NAA with probe detection	36661384	no	LOINC
COVID 19 measurement	Influenza virus A and B and SARS-CoV-2 (COVID-19) identified in Respiratory system specimen by NAA with probe detection	36661375	no	LOINC
COVID 19 measurement	Influenza virus A and B and SARS-CoV-2 (COVID-19) RNA panel - Respiratory system specimen by NAA with probe detection	36661376	no	LOINC
COVID 19 measurement	Measurement of Coronavirus (Coronavirinae subfamily species)	705104	no	OMOP Extension
COVID 19 measurement	Measurement of Coronavirus (Coronavirinae subfamily species) antigen	705105	no	OMOP Extension
COVID 19 measurement	Measurement of Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)	756055	no	OMOP Extension
COVID 19 measurement	Measurement of Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) Genetic material using Molecular method	586310	no	OMOP Extension

Phenotype	Concept name	Concept ID (including descendants)	Exclude concept ID	Vocabulary
COVID 19 measurement	Measurement of Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in Blood	704991	no	OMOP Extension
COVID 19 measurement	Measurement of Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in Respiratory specimen	756029	no	OMOP Extension
COVID 19 measurement	Measurement of Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in Saliva	586307	no	OMOP Extension
COVID 19 measurement	Measurement of Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in Sample from nose	705107	no	OMOP Extension
COVID 19 measurement	Measurement of Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in Sample from oropharynx	704976	no	OMOP Extension
COVID 19 measurement	Measurement of Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in Specified specimen	586309	no	OMOP Extension
COVID 19 measurement	Measurement of Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in Unspecified specimen	756065	no	OMOP Extension
COVID 19 measurement	Measurement of Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) specific cell-mediated immune response in Blood	702834	no	OMOP Extension
COVID 19 measurement	Measurement of Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) using Culture method	704992	no	OMOP Extension
COVID 19 measurement	Measurement of Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) using Nucleic acid amplification technique	705001	no	OMOP Extension
COVID 19 measurement	Measurement of Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) using Nucleic acid amplification technique in Blood	705000	no	OMOP Extension
COVID 19 measurement	Measurement of Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) using Nucleic acid amplification technique in Respiratory specimen	756085	no	OMOP Extension
COVID 19 measurement	Measurement of Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) using Nucleic acid amplification technique in Saliva	586308	no	OMOP Extension
COVID 19 measurement	Measurement of Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) using Nucleic acid amplification technique in Sample from nose	705106	no	OMOP Extension
COVID 19 measurement	Measurement of Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) using Nucleic acid amplification technique in Sample from oropharynx	704975	no	OMOP Extension
COVID 19 measurement	Measurement of Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) using Nucleic acid amplification technique in Unspecified specimen	756084	no	OMOP Extension
COVID 19 measurement	Measurement of Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) using Sequencing	704993	no	OMOP Extension
COVID 19 measurement	Measurement of SARS-CoV-2 antigen	37310257	no	SNOMED

Phenotype	Concept name	Concept ID (including descendants)	Exclude concept ID	Vocabulary
COVID 19 measurement	SARS-CoV+SARS-CoV-2 (COVID-19) Ag [Presence] in Respiratory system specimen by Rapid immunoassay	757685	no	LOINC
COVID 19 measurement	SARS-CoV-2 (COVID-19) [Presence] in Specimen by Organism specific culture	586516	no	LOINC
COVID 19 measurement	SARS-CoV-2 (COVID-19) Ag [Presence] in Respiratory system specimen by Rapid immunoassay	723477	no	LOINC
COVID 19 measurement	SARS-CoV-2 (COVID-19) N gene [Cycle Threshold #] in Specimen by NAA with probe detection	706167	no	LOINC
COVID 19 measurement	SARS-CoV-2 (COVID-19) N gene [Cycle Threshold #] in Specimen by Nucleic acid amplification using CDC primer-probe set N1	706157	no	LOINC
COVID 19 measurement	SARS-CoV-2 (COVID-19) N gene [Cycle Threshold #] in Specimen by Nucleic acid amplification using CDC primer-probe set N2	706155	no	LOINC
COVID 19 measurement	SARS-CoV-2 (COVID-19) N gene [Presence] in Nasopharynx by NAA with probe detection	715272	no	LOINC
COVID 19 measurement	SARS-CoV-2 (COVID-19) N gene [Presence] in Nose by NAA with probe detection	757678	no	LOINC
COVID 19 measurement	SARS-CoV-2 (COVID-19) N gene [Presence] in Respiratory system specimen by NAA with probe detection	706161	no	LOINC
COVID 19 measurement	SARS-CoV-2 (COVID-19) N gene [Presence] in Respiratory system specimen by Nucleic acid amplification using CDC primer-probe set N1	586524	no	LOINC
COVID 19 measurement	SARS-CoV-2 (COVID-19) N gene [Presence] in Respiratory system specimen by Nucleic acid amplification using CDC primer-probe set N2	586525	no	LOINC
COVID 19 measurement	SARS-CoV-2 (COVID-19) N gene [Presence] in Saliva (oral fluid) by NAA with probe detection	36661378	no	LOINC
COVID 19 measurement	SARS-CoV-2 (COVID-19) N gene [Presence] in Serum or Plasma by NAA with probe detection	586520	no	LOINC
COVID 19 measurement	SARS-CoV-2 (COVID-19) N gene [Presence] in Specimen by NAA with probe detection	706175	no	LOINC
COVID 19 measurement	SARS-CoV-2 (COVID-19) N gene [Presence] in Specimen by Nucleic acid amplification using CDC primer-probe set N1	706156	no	LOINC
COVID 19 measurement	SARS-CoV-2 (COVID-19) N gene [Presence] in Specimen by Nucleic acid amplification using CDC primer-probe set N2	706154	no	LOINC
COVID 19 measurement	SARS-CoV-2 (COVID-19) ORF1ab region [Cycle Threshold #] in Respiratory system specimen by NAA with probe detection	723469	no	LOINC
COVID 19 measurement	SARS-CoV-2 (COVID-19) ORF1ab region [Cycle Threshold #] in Specimen by NAA with probe detection	706168	no	LOINC
COVID 19 measurement	SARS-CoV-2 (COVID-19) ORF1ab region [Presence] in Respiratory system specimen by NAA with probe detection	723478	no	LOINC
COVID 19 measurement	SARS-CoV-2 (COVID-19) ORF1ab region [Presence] in Specimen by NAA with probe detection	723464	no	LOINC

Phenotype	Concept name	Concept ID (including descendants)	Exclude concept ID	Vocabulary
COVID 19 measurement	SARS-CoV-2 (COVID-19) RdRp gene [Cycle Threshold #] in Respiratory system specimen by NAA with probe detection	723471	no	LOINC
COVID 19 measurement	SARS-CoV-2 (COVID-19) RdRp gene [Cycle Threshold #] in Specimen by NAA with probe detection	723470	no	LOINC
COVID 19 measurement	SARS-CoV-2 (COVID-19) RdRp gene [Presence] in Respiratory system specimen by NAA with probe detection	706160	no	LOINC
COVID 19 measurement	SARS-CoV-2 (COVID-19) RdRp gene [Presence] in Specimen by NAA with probe detection	706173	no	LOINC
COVID 19 measurement	SARS-CoV-2 (COVID-19) RNA [Cycle Threshold #] in Respiratory system specimen by NAA with probe detection	586528	no	LOINC
COVID 19 measurement	SARS-CoV-2 (COVID-19) RNA [Cycle Threshold #] in Specimen by NAA with probe detection	586529	no	LOINC
COVID 19 measurement	SARS-CoV-2 (COVID-19) RNA [Log #/volume] (viral load) in Specimen by NAA with probe detection	715262	no	LOINC
COVID 19 measurement	SARS-CoV-2 (COVID-19) RNA [Presence] in Nasopharynx by NAA with non-probe detection	723476	no	LOINC
COVID 19 measurement	SARS-CoV-2 (COVID-19) RNA [Presence] in Nasopharynx by NAA with probe detection	586526	no	LOINC
COVID 19 measurement	SARS-CoV-2 (COVID-19) RNA [Presence] in Nose by NAA with probe detection	757677	no	LOINC
COVID 19 measurement	SARS-CoV-2 (COVID-19) RNA [Presence] in Respiratory system specimen by NAA with probe detection	706163	no	LOINC
COVID 19 measurement	SARS-CoV-2 (COVID-19) RNA [Presence] in Respiratory system specimen by Sequencing	36661377	no	LOINC
COVID 19 measurement	SARS-CoV-2 (COVID-19) RNA [Presence] in Saliva (oral fluid) by NAA with probe detection	715260	no	LOINC
COVID 19 measurement	SARS-CoV-2 (COVID-19) RNA [Presence] in Saliva (oral fluid) by Sequencing	715261	no	LOINC
COVID 19 measurement	SARS-CoV-2 (COVID-19) RNA [Presence] in Serum or Plasma by NAA with probe detection	723463	no	LOINC
COVID 19 measurement	SARS-CoV-2 (COVID-19) RNA [Presence] in Specimen by NAA with probe detection	706170	no	LOINC
COVID 19 measurement	SARS-CoV-2 (COVID-19) RNA panel - Respiratory system specimen by NAA with probe detection	706158	no	LOINC
COVID 19 measurement	SARS-CoV-2 (COVID-19) RNA panel - Specimen by NAA with probe detection	706169	no	LOINC
COVID 19 measurement	SARS-CoV-2 (COVID-19) S gene [Cycle Threshold #] in Respiratory system specimen by NAA with probe detection	723467	no	LOINC
COVID 19 measurement	SARS-CoV-2 (COVID-19) S gene [Cycle Threshold #] in Specimen by NAA with probe detection	723468	no	LOINC
COVID 19 measurement	SARS-CoV-2 (COVID-19) S gene [Presence] in Respiratory system specimen by NAA with probe detection	723465	no	LOINC

Phenotype	Concept name	Concept ID (including descendants)	Exclude concept ID	Vocabulary
COVID 19 measurement	SARS-CoV-2 (COVID-19) S gene [Presence] in Serum or Plasma by NAA with probe detection	586519	no	LOINC
COVID 19 measurement	SARS-CoV-2 (COVID-19) S gene [Presence] in Specimen by NAA with probe detection	723466	no	LOINC
COVID 19 measurement	SARS-CoV-2 (COVID-19) whole genome [Nucleotide sequence] in Isolate or Specimen by Sequencing	586517	no	LOINC
COVID 19 measurement	SARS-related coronavirus E gene [Cycle Threshold #] in Specimen by NAA with probe detection	706166	no	LOINC
COVID 19 measurement	SARS-related coronavirus E gene [Presence] in Respiratory system specimen by NAA with probe detection	586523	no	LOINC
COVID 19 measurement	SARS-related coronavirus E gene [Presence] in Serum or Plasma by NAA with probe detection	586518	no	LOINC
COVID 19 measurement	SARS-related coronavirus E gene [Presence] in Specimen by NAA with probe detection	706174	no	LOINC
COVID 19 measurement	SARS-related coronavirus N gene [Cycle Threshold #] in Specimen by Nucleic acid amplification using CDC primer-probe set N3	706172	no	LOINC
COVID 19 measurement	SARS-related coronavirus N gene [Presence] in Specimen by Nucleic acid amplification using CDC primer-probe set N3	706171	no	LOINC
COVID 19 measurement	SARS-related coronavirus RNA [Presence] in Respiratory system specimen by NAA with probe detection	706165	no	LOINC
COVID 19 measurement	SARS-related coronavirus RNA [Presence] in Specimen by NAA with probe detection	723472	no	LOINC
COVID 19 measurement	SARS-related coronavirus+MERS coronavirus RNA [Presence] in Respiratory system specimen by NAA with probe detection	706159	no	LOINC

For measurement variables other than blood pressure, analyses will be restricted to the availability of measurement records rather than the assessment of measurement values. This reflects expected variability in data completeness across data sources and ensures a consistent and feasible approach for real-world data within the scope of the study.

Table S5. Overview of in- and exclusion criteria used in Baxdrostat trial and of variables in the current study.

In- and exclusion criteria Baxdrostat trial	Variables DARWIN EU® P5-C1-002/P5-C2-003 study
Age ≥18 years	Age ≥18 years
Seated systolic blood pressure ≥140 mmHg and <170 mmHg	Systolic blood pressure ≥140 mmHg and/or diastolic blood pressure ≥90 mmHg
A stable regimen of ≥3 antihypertensive medications from different therapeutic classes (at least one was a diuretic) at maximum tolerated doses in the judgment of the Investigator. Beta-blockers used to treat other conditions (i.e. migraine, heart failure, coronary artery disease) were not counted as antihypertensive medications for the purpose of qualifying for the study	Concomitant use of: 1) a diuretic, and an ACE inhibitor or ARB, and a calcium channel antagonist (*primary definition*), or 2) a diuretic, and an ACE inhibitor or ARB, and a beta-blocker (*alternative definition*)
Estimated glomerular filtration rate ≥45 ml/min/1.73m2	Glomerular filtration rate measurement
Serum potassium level ≥3.5 and <5.0 mmol/l	Serum potassium measurement

In- and exclusion criteria Baxdrostat trial	Variables DARWIN EU® P5-C1-002/P5-C2-003 study
Morning cortisol levels >3 µg/dl	Cortisol measurements
Females should be permanently sterilised, postmenopausal, or use highly effective form of birth control	Female reproductive status: 1) Postmenopausal females, defined as i) females aged ≥50 to 65 years or ii) females aged 18–49 years with a prior record of conditions or procedures indicating premature or early menopause (primary ovarian insufficiency/failure, premature menopause, early menopause, bilateral oophorectomy, or hysterectomy with ovarian removal); 2) Females of child-bearing potential, defined as females aged 18–49 years at the index date who do not meet criteria for postmenopausal status. Additionally, use of hormonal contraception will be assessed among these females of child-bearing potential.
Females of childbearing potential have a negative serum pregnancy test	–
Capable of giving signed informed consent	–
Seated systolic blood pressure ≥170 mmHg	Systolic blood pressure ≥140 mmHg
Seated diastolic blood pressure ≥110 mmHg	Diastolic blood pressure ≥90 mmHg
Treatment with an angiotensin receptor blocker (ARB) and an angiotensin-converting enzyme inhibitor (ACE), both taken simultaneously	Treatment with an ARB and ACEi, both taken simultaneously
Serum sodium level <135 mmol/l	Serum sodium measurement
Following known secondary causes of hypertension: renal artery stenosis, uncontrolled or untreated hyperthyroidism, uncontrolled or untreated hypothyroidism, pheochromocytoma, Cushing's syndrome, aortic coarctation	Following known secondary causes of hypertension: renal artery stenosis, hyperthyroidism, hypothyroidism, pheochromocytoma, Cushing's syndrome, aortic coarctation
New York Heart Association functional heart failure class IV	–
Medical history of stroke, acute coronary syndrome, hypertensive encephalopathy, or hospitalisation for heart failure	Medical history of stroke, acute coronary syndrome, hypertensive encephalopathy, or heart failure
Planned percutaneous coronary intervention/coronary artery bypass grafting, or percutaneous coronary intervention/coronary artery bypass grafting done	Percutaneous coronary intervention/coronary artery bypass grafting done
Known current severe left ventricular outflow obstruction, such as obstructive hypertrophic cardiomyopathy and/or severe aortic valvular disease	Left ventricular outflow obstruction, such as obstructive hypertrophic cardiomyopathy and/or aortic valvular disease
Left bundle branch block and any cardiac arrhythmia requiring treatment	Left bundle branch block and any cardiac arrhythmia
Persistent atrial fibrillation	Atrial fibrillation
Known severe hepatic impairment, defined as Child-Pugh Class C, based on records that confirm documented medical history	Hepatic impairment
Uncontrolled diabetes with HbA1c >10.0% (86 mmol/mol)	Diabetes mellitus, and HbA1C measurement
Screening QTcF value >470 msec	A record of 1) electrocardiogram or 2) QT corrected interval (if available)
Family history of Long QT syndrome	Long QT syndrome
Heart rate 110 beats/min in a resting position	Heart rate measurement
Suspected severe cardiac hypertrophy	Cardiac hypertrophy

In- and exclusion criteria Baxdrostat trial	Variables DARWIN EU® P5-C1-002/P5-C2-003 study
Pregnant or breastfeeding	Pregnant or breastfeeding
Adrenal insufficiency	Adrenal insufficiency
Suspected (as judged by the principal investigator) or confirmed COVID-19 infection. Or hospitalisation for COVID-19	COVID-19 infection
Medical treatment with any mineralocorticoid receptor antagonist, antiarrhythmic medications, or potassium-sparing diuretics	Medical treatment with any mineralocorticoid receptor antagonist, antiarrhythmic medications, or potassium-sparing diuretics
Treatment with potassium binders	Treatment with potassium binders
Receive or was receiving any of the exclusionary drugs, such as strong inducers of cytochrome P450 (CYP) 3A, chronic (taken more than three times a week for more than 3 months) use of nonsteroidal anti-inflammatory drugs (NSAIDs) (use of low-dose aspirin was permitted, as per medical judgment), mineralocorticoid receptor antagonists, and/or chronic use of systemic steroids	Treatment with strong inducers of cytochrome P450 (CYP) 3A, nonsteroidal anti-inflammatory drugs (NSAIDs), mineralocorticoid receptor antagonists, and/or use of systemic steroids
Drugs that prolong QT were avoided, if possible, and if there were other alternatives without the QT liability, these were preferred. If such QT-prolonging drugs were still needed, the investigator ensured appropriate monitoring of electrocardiograms and electrolytes were performed, as per clinical judgment	Use of medication associated with QT prolongation
Treatment with cytotoxic therapy	Treatment with cytotoxic therapy
Treatment with potassium supplements were not prohibited but were continuously assessed and monitored throughout the trial	–
Known hypersensitivity to Baxdrostat or drugs of the same class or any of its excipients	–
Participation in another clinical study with an investigational product administered in the 3 months before randomisation into this study	–
Participants working shifts (i.e. shifts that comprise working hours at different times on different days)	–

– = not feasible to assess in current study.

ANNEX V. ENCePP checklist for study protocols

ENCePP Checklist for Study Protocols (Revision 4)

Study title: DARWIN EU® - Representativeness of the clinical trial population compared with the real-world population with resistant hypertension

EU PAS Register® number: EUPAS1000001065 **Study reference number (if applicable):** P5-C1-002/P5-C2-003

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]	☒	☐	☐	
1.1.2 End of data collection[2]	☒	☐	☐	
1.1.3 Progress report(s)	☐	☐	☒	5
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:	☒	☐	☐	
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)	☒	☐	☐	7
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?	☒	☐	☐	8.4
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:

Section 4: Source and study populations		**Yes**	**No**	**N/A**	**Section Number**
4.1	Is the source population described?	☒	☐	☐	8
4.2	Is the planned study population defined in terms of:				
	4.2.1 Study time period	☒	☐	☐	8.5
	4.2.2 Age and sex	☒	☐	☐	8.6
	4.2.3 Country of origin	☒	☐	☐	8.4
	4.2.4 Disease/indication	☒	☐	☐	8.6
	4.2.5 Duration of follow-up	☒	☐	☐	8.2
4.3	Does the protocol define how the study population will be sampled from the source population? (e.g. event or inclusion/exclusion criteria)	☒	☐	☐	8.3

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)	☒	☐	☐	8.6.1

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)	☒	☐	☐	8.8
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:

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?	☒	☐	☐	8.6.2
6.2	Does the protocol describe how the outcomes are defined and measured?	☒	☐	☐	8.6.2
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)	☐	☐	☒	
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)	☐	☐	☒	
7.3	Does the protocol address information bias? (e.g. misclassification of exposure and outcomes, time-related bias)	☐	☐	☒	

Comments:

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:

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)	☒	☐	☐	Annex I
	9.1.2 Outcomes? (e.g. clinical records, laboratory markers or values, claims data, self-report, patient interview including scales and questionnaires, vital statistics)	☒	☐	☐	Annex I
	9.1.3 Covariates and other characteristics?	☒	☐	☐	Annex I
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)	☒	☐	☐	Annex I
	9.2.2 Outcomes? (e.g. date of occurrence, multiple event, severity measures related to event)	☒	☐	☐	Annex I
	9.2.3 Covariates and other characteristics? (e.g. age, sex, clinical and drug use history, co-morbidity, co-medications, lifestyle)	☒	☐	☐	Annex I
9.3	Is a coding system described for:				
	9.3.1 Exposure? (e.g. WHO Drug Dictionary, Anatomical Therapeutic Chemical (ATC) Classification System)	☒	☐	☐	Annex I
	9.3.2 Outcomes? (e.g. International Classification of Diseases (ICD), Medical Dictionary for Regulatory Activities (MedDRA))	☒	☐	☐	Annex I
	9.3.3 Covariates and other characteristics?	☒	☐	☐	Annex I
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?	☒	☐	☐	8.8
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)	☒	☐	☐	Annex III
11.2 Are methods of quality assurance described?	☒	☐	☐	Annex III
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:				
12.1.1 Selection bias?	☒	☐	☐	
12.1.2 Information bias?	☒	☐	☐	9
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)	☒	☐	☐	Annex II

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?	☒	☐	☐	Annex II
13.2 Has any outcome of an ethical review procedure been addressed?	☒	☐	☐	Annex II
13.3 Have data protection requirements been described?	☒	☐	☐	Annex II

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)?	☒	☐	☐	Annex III
15.2 Are plans described for disseminating study results externally, including publication?	☒	☐	☐	Annex III

Comments:

Name of the main author of the protocol: Ellen Gerritsen

Date: 20/05/2026

Signature: 

ANNEX VI. Glossary

Additional definitions are available in the EMA Glossary of terms https://www.ema.europa.eu/en/about-us/glossaries.

Aggregated Data

Data collected and combined from multiple sources to generate summary information, typically anonymised.

Benefit-Risk Assessment

Evaluation of the positive therapeutic effects of a medicine compared to its risks (e.g. side effects).

Common Data Model (CDM)

A standardised data structure that enables data from multiple sources to be harmonised, making analysis consistent and reproducible. DARWIN EU® utilises the OMOP CDM maintained by the OHDSI community.

Complex Studies (C3)

Studies requiring the development or customisation of specific study designs, protocols, and Statistical Analysis Plans (SAPs), with extensive collection or extraction of data. Examples include etiological studies measuring the strength and determinants of an association between an exposure and the occurrence of a health outcome in a defined population considering sources of bias, potential confounding factors, and effect modifiers.

Coordination Centre (CC)

The central hub responsible for managing and overseeing the activities within DARWIN EU®. It is based at Erasmus University Medical Center in Rotterdam, the Netherlands.

Data Access

The process of obtaining permission to use specific datasets for regulatory or scientific studies.

Data Quality Framework

A set of standards and procedures to ensure accuracy, completeness, timeliness, and consistency of data used in DARWIN EU®.

Data Source

A data source or repository of structured health-related data, such as electronic health records (EHRs), insurance claims, or registries.

DARWIN EU®

The European Medicines Agency's (EMA) federated network of real-world data sources designed to generate evidence to support regulatory decision-making.

EMA (European Medicines Agency)

The regulatory body responsible for the evaluation and supervision of medicinal products in the EU, overseeing DARWIN EU®.

Evidence Generation

The process of analysing real-world data to produce scientific information that can inform healthcare or regulatory decisions.

Federated Network

A data infrastructure where data remain at their original location but can be analysed in a harmonised way across multiple partners using a common model and tools.

GDPR (General Data Protection Regulation)

The EU regulation governing the protection of personal data and privacy, crucial to how DARWIN EU® handles health data.

Health Technology Assessment (HTA)

A systematic evaluation of properties and impacts of health technology, often using DARWIN EU® data to support assessments.

Metadata

Descriptive information about a data source (e.g. its content, quality, and structure), essential for identifying relevant data sources in DARWIN EU® studies.

Off-the-Shelf Studies (OTS)

Studies for which a standard protocol per study/analysis type and standardised analytics may be developed and applied or adapted, typically relating to a descriptive research question. This includes studies on disease epidemiology, for example, the estimation of the prevalence or incidence of health outcomes in defined time periods and population groups, or drug utilisation studies at the population or patient level.

OHDSI (Observational Health Data Sciences and Informatics)

An open-science collaborative community that develops tools and standards (including the OMOP CDM) to enable large-scale analytics of observational health data. OHDSI provides the technical and scientific foundation for DARWIN EU®'s analytical ecosystem.

Patient-Level Data

Data related to individuals, de-identified, used for longitudinal or detailed analyses.

OMOP (Observational Medical Outcomes Partnership)

A common data model (CDM) that standardises the structure and content of observational healthcare data, enabling systematic analysis across disparate datasets. DARWIN EU® uses the OMOP CDM to ensure interoperability and consistency in real-world evidence generation.

Real-World Data (RWD)

Data relating to individual health status or healthcare delivery that is collected from routine clinical practice rather than from randomised controlled trials.

Real-World Evidence (RWE)

Clinical evidence derived from the analysis of RWD, used to inform decisions by regulators, payers, or clinicians.

Regulatory Decision-Making

The process by which authorities like EMA assess data to authorise, monitor, or modify the use of medicines in the EU.

Routine Repeated Studies (RR)

Studies that are either Off-the-Shelf or Complex studies repeated on a regular basis, following the same protocol and study code, but with updated data and/or different data partners.

Study Protocol

A detailed plan describing how a specific real-world study will be conducted, including objectives, design, data sources, and analyses.

Very Complex Studies (C4)

Studies which cannot rely only on electronic health care data sources, or which would require complex methodological work, for example, due to the occurrence of events that cannot be defined by existing diagnosis codes, including events that do not yet have a diagnosis code, where it may be necessary to combine a diagnosis code with other data such as results of laboratory investigations. These studies might require the collection of data prospectively, or the inclusion of new (not previously onboarded) data sources.