Study Protocol

P4-C3-005

DARWIN EU® - Assessing the potential association between venlafaxine and heart failure among adults

23/04/2026

Version 5.0

Authors: Martí Català Sabaté, Annika Jödicke

Public

CONTENTS

Study title	DARWIN EU® - Assessing the potential association between venlafaxine and heart failure among adults.
Protocol version	V5.0
Date	23/04/2026
EUPAS number	EUPAS1000000974
Active Substance	Venlafaxine ATC-code: N06AX16
Medicinal Product	N/A
Research question and objectives	The study aims to assess the potential association between venlafaxine and heart failure among adults. Specific objectives are: 1) To assess the risk of incident heart failure and incident cardiomyopathy between new users of venlafaxine vs. new users of mirtazapine. 2) To assess the risk of heart failure exacerbation between new users of venlafaxine vs. new users of mirtazapine who have been diagnosed with heart failure before treatment initiation. 3) To characterise both new venlafaxine and new mirtazapine users at the time of treatment start (in terms of demographics, pre-defined comorbidities/comedication/conditions of interest, and drug utilisation) for contextualisation of Objectives 1+2. Analyses will be conducted overall and stratified for condition of interest, age group, sex , history of CV disease and previous SSRI use.
Countries of study	Denmark, Finland, Spain, Sweden, the United Kingdom
Authors	Martí Català Sabaté m.catalasabate@darwin-eu.org Annika Jödicke a.jodicke@darwin-eu.org

LIST OF ABBREVIATIONS

Acronyms/terms	Description
ATC	Anatomical Therapeutic Chemical
ATT	Average treatment effects in the treated
BIFAP	Pharmacoepidemiological Research Database for Public Health Systems
CC	Coordination centre
CDM	Common Data Model
CPRD GOLD	Clinical Practice Research Datalink GOLD
CV	Cardiovascular
DARWIN EU®	Data Analysis and Real-World Interrogation Network
DK-DHR	Danish Data Health Registries
DRE	Digital Research Environment
DTZ	Data Transfer Zone
ED	Emergency Department
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
FAERS	FDA Adverse Event Reporting System
FinOMOP-THL	Finnish Care Register for Health Care
GDPR	General Data Protection Regulation
GP	General Practitioner
HF	Heart Failure
HI-SPEED	Health Impact - Swedish Population Evidence Enabling Data-linkage
HR	Hazard ratio
ICD	International Classification of Diseases
IHD	ischemic heart disease
IP	Inpatient
IR	Incidence Rates
IRB	Institutional Review Board
IRR	Incidence Rate Ratios
IQR	Interquartile Range
ITT	Intention to treat
LASSO	Least Absolute Shrinkage and Selection Operator [LASSO]
MDRR	Minimum detectable relative risk
NCO	Negative control outcomes

Acronyms/terms	Description
N/A	Not applicable
OHDSI	Observational Health Data Sciences and Informatics
OMOP	Observational Medical Outcomes Partnership
OP	Outpatient
PP	Per protocol
PS	Propensity score
RxNorm	Medical prescription normalised
SIDIAP	The Information System for the Development of Research in Primary Care
SmPC	Summaries of product characteristics
SNOMED	Systemised Nomenclature of Medicine
SNRI	Serotonin and norepinephrine reuptake inhibitor
SSRI	Selective Serotonin Reuptake Inhibitor
WHO	World Health Organisation

1. TITLE

DARWIN EU® - Assessing the potential association between venlafaxine and incident heart failure among adults.

2. DESCRIPTION OF THE STUDY TEAM

Study team role	Names	Organisation
Principal Investigators	Annika Jodicke	University of Oxford
Data Scientists	Marti Catala Sabate Cecilia Campanile Xintong Li	University of Oxford
Epidemiologists	Wanning Wang Marti Catala Sabate Xintong Li	University of Oxford
Clinical Domain Experts	Daniel Prieto-Alhambra Annika Jodicke Anna Camps Vilaro Anna Saura Lazaro Alejandro Ballve	University of Oxford
Study Manager	Natasha Yefimenko	Erasmus Medical Centre
Data source	**Names**	**Data Partner Organisation***
DK-DHR	Elvira Bräuner Susanne Bruun	Danish Data Health Registries
FinOMOP-THL	Toni Lehtonen	Finnish Care Register for Health Care
BIFAP	Belén Castillo Cano Cristina Justo Astorgano Alicia Peñaranda Navazo Ana Llorente Garcia Miguel Angel Macia Martinez	Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público
SIDIAP	Elena Roel Herranz Agustina Giuliodori Picco Laura Granés Irene López Sánchez Anna Palomar Cros	The Information System for the Development of Research in Primary Care
HI-SPEED	Marcel Ballin Huiqi Li Fredrik Nyberg	Health Impact - Swedish Population Evidence Enabling Data-linkage

	Mats Talbäck Rickard Ljung	
CPRD GOLD	Antonella Delmestri	Clinical Practice Research Datalink GOLD

*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® - Assessing the potential association between venlafaxine and heart failure in adults.

Rationale and background

Venlafaxine is a dual-acting serotonin and norepinephrine reuptake inhibitor (SNRI) indicated for the treatment of depression, prevention of relapse and prevention of recurrence of depression, anxiety or generalized anxiety disorder, social anxiety disorder, and panic disorder.

Venlafaxine has a dose-dependent pharmacodynamic profile. At lower doses it primarily inhibits serotonin reuptake, while at higher doses (≥150 mg/day), it additionally inhibits norepinephrine reuptake and exerts weak inhibitory effects on dopamine reuptake.

Several cardiovascular effects are already labelled in section 4.8 of the Summaries of product characteristics (SmPC) in Europe, e.g., tachycardia and hypertension, Torsade de pointes, ventricular tachycardia, ventricular fibrillation, prolonged QT, and stress cardiomyopathy (Takotsubo cardiomyopathy).

A published case series and spontaneously reported cases (e.g., EudraVigilance, Vigibase, company global pharmacovigilance safety database) suggest an association between venlafaxine and cardiotoxicity (heart failure, cardiomyopathy other than Takotsubo). An observational study using high-quality data from the DARWIN EU® network could add important further evidence to evaluate the potential association.

The European Medicines Agency (EMA) has therefore requested a study to estimate the association between venlafaxine and heart failure/cardiomyopathy, using mirtazapine, which is indicated for the treatment of depression, as active comparator.

Research question and objectives

Research questions

Is new use of venlafaxine associated with an increased risk of incident heart failure compared to new use of mirtazapine?

Is new use of venlafaxine associated with exacerbation of pre-existing heart failure compared to new use of mirtazapine?

Objectives

The study aims to assess the potential association between venlafaxine and heart failure among adults.

The specific objectives of this study are:

1) To assess the risk of incident heart failure and incident cardiomyopathy between new users of venlafaxine vs. new users of mirtazapine.
2) To assess the risk of heart failure exacerbation between new users of venlafaxine vs. new users of mirtazapine who have been diagnosed with heart failure before treatment initiation.
3) To characterise both new venlafaxine and new mirtazapine users at the time of treatment start (in terms of demographics, pre-defined comorbidities/comedication/conditions of interest, and drug utilisation) for contextualisation of Objectives 1+2.

Analyses will be conducted overall and stratified for condition of interest, age group, sex, history of cardiovascular (CV) disease and previous Selective Serotonin Reuptake Inhibitor (SSRI) use.

Methods

Study design

- New user active comparator cohort studies (Objectives 1+2)
- Patient-level characterisation (Objective 3)

Index date will be the date of first prescription of venlafaxine or mirtazapine, and individuals are followed up until date of last data availability, drug discontinuation, outcome occurrence, death, or end of study period.

Population

The study population will include all individuals with a recorded first prescription for venlafaxine or mirtazapine within the study period, from 01/01/2010 up to the end of data availability, who meet the eligibility criteria at study entry.

Other eligibility criteria are:

- At least 730 days of data source history prior to index date.
- Aged ≥18 years at index date.
- No prescription / dispensation of other antidepressants recorded in the data source in the year before index date, except SSRIs, which are permitted.
- No diagnosis or record of heart failure or cardiomyopathy ever recorded in the data source before index date (Objective 1).
- At least one record/diagnosis of heart failure recorded in the data source before index date (Objective 2).
- No prescription/dispensation record of both venlafaxine and mirtazapine at index date.

For Objectives 1+2, a sensitivity analysis will be conducted, requiring 5 years of data source history prior to index date.

Variables

Exposure:

- Venlafaxine (exposure) [ATC-code: N06AX16]
- Mirtazapine (active comparator) [ATC-code: N06AX11]

Outcome:

- *Objective 1:*
  - Primary outcome: Incident heart failure (broad)
  - Secondary outcomes: broad/narrow definitions for incident heart failure; incident cardiomyopathy (broad, primary), heart failure (HF) hospitalisation
- *Objective 2:*
  - Primary outcome: HF hospitalisation
  - Secondary outcomes: Acute pulmonary oedema, recurrent heart failure (broad)
- *Objective 3:* Summarised patient characteristics at index date

Data sources

1. Denmark: Danish Data Health Registries (DK-DHR)
2. Finland: Finnish Care Register for Health Care (FinOMOP-THL)
3. Spain: Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público (BIFAP)
4. Spain: The Information System for Research on Primary Care (SIDIAP)
5. Sweden: Health Impact - Swedish Population Evidence Enabling Data-linkage (HI-SPEED)
6. The United Kingdom: Clinical Practice Research Datalink GOLD (CPRD GOLD)

The outcome of heart failure hospitalisation will only be conducted in DK-DHR, FinOMOP-THL, SIDIAP, and HI-SPEED, where hospital data is available.

Study size

Based on a preliminary feasibility assessment, the expected number of person counts for venlafaxine in the data sources included in this study range from 184,600 (HI-SPEED) to 303,200 (CPRD GOLD). The expected number of person counts for mirtazapine in the data sources included in this study range from 266,900 (SIDIAP) to 1,088,000 (BIFAP). The expected number of person counts for heart failure in the data sources included in this study range from 252,200 (CPRD GOLD) to 936,600 (BIFAP).

In the DARWIN EU® P3-C3-001 (Adverse events of special interest) study, background rates of incident heart failure were estimated to be 138 per 100,000 person-years in CPRD GOLD and 386 per 100,000 person-years in SIDIAP.

The required sample size to estimate a hazard ratio (HR) of 1.25 with relative precision of 10% would be 150,363 (assuming an average event rate of 1.125%) and 75,182 (assuming an average event rate of 2.25%) for incident heart failure outcomes. The corresponding sample size for recurrent heart failure events would be 15,036 (assuming an average event rate of 11.25%) and 7,518 (assuming an average event rate of 22.5%)..

Statistical analysis

A minimum cell count of 5 will be used when reporting results, with smaller counts reported as “<5” and zero counts as “0”. All results will be presented separately by data source. For objectives 1+2, random-effects meta analyses will be conducted to pool results.

Objectives 1–2:

We will use large-scale propensity score matching (PS) to address confounding from observed covariates, and match venlafaxine with mirtazapine patients based on a combination of pre-defined key variables for exact matching [age bands, sex, calendar time, history of CV disease, previous SSRI use, condition of interest (depression/anxiety)], and data-driven (Least Absolute Shrinkage and Selection Operator [LASSO]) analyses to select confounders.

Among the PS matched population, patient-level characterisation analyses will be conducted.

Poisson regression will be used to estimate incidence rates for the outcomes of interest among the two treatment groups of new users of venlafaxine and mirtazapine.

Cox proportional hazards modelling will then be used to estimate hazard ratios (HRs) for all outcomes of interest, respectively. Results will be stratified for age, sex, history of CV disease, previous SSRI use, and condition of interest (depression/anxiety).

The proportional hazards assumption will be tested using visual inspection of log-log plots. In the scenario that the assumption is violated, Incidence Rate Ratios (IRR) will be calculated using Poisson regression and presented alongside the HR.

To detect residual confounding, Negative Control Outcome (NCO) analyses will be conducted for each outcome among the overall population and subgroups of previous SSRI use and condition of interest.

A sensitivity analysis with a population restricted to having at least 5 years of data source history prior to index date will be conducted.

Objective 3:

Patient-level characterisation analyses of all new venlafaxine and mirtazapine will be conducted. Covariates of interest will also be reported as counts and proportions, including demographics, pre-defined comorbidities (including cardiovascular diseases) and comedication, and drug utilisation, stratified separately by indication (depression/anxiety), age group, sex, history of CV disease (yes/no), and previous SSRI use (yes/no). *CohortCharacteristics* and *DrugUtilisation* R packages will be used.

4. AMENDMENTS AND UPDATES

None.

5. MILESTONES

Study milestones and deliverables	Planned dates*
Final Study Outline	09 January 2026
Final Study Protocol approved	28 February 2026
Creation of Analytical code	February – March 2026
Execution of Analytical Code on the data	End of March – beginning of April 2026
Draft Study Report	Early May 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.

6. RATIONALE AND BACKGROUND

Venlafaxine is a dual-acting serotonin and norepinephrine reuptake inhibitor (SNRI) indicated for the treatment of depression, prevention of relapse and prevention of recurrence of depression, anxiety or generalized anxiety disorder, social anxiety disorder, and panic disorder. Mirtazapine is indicated for the treatment of depression.

Venlafaxine has a dose-dependent pharmacodynamic profile. At lower doses it primarily inhibits serotonin reuptake, while at higher doses (≥150 mg/day), it additionally inhibits norepinephrine reuptake and exerts weak inhibitory effects on dopamine reuptake.

Several cardiovascular effects are already labelled in section 4.8 of the Summaries of product characteristics (SmPC) in Europe, e.g., tachycardia and hypertension, Torsade de pointes, ventricular tachycardia, ventricular fibrillation, prolonged QT, and stress cardiomyopathy (Takotsubo cardiomyopathy).

A published case series and spontaneously reported cases (e.g., EudraVigilance, Vigibase, US FDA Adverse Event Reporting System (FAERS), company global pharmacovigilance safety database) suggest an association between venlafaxine and cardiotoxicity (heart failure, cardiomyopathy, ischemic heart disease). [1, 2] In addition, serious cardiac events were reported with venlafaxine overdose. However, a previous observational study did not find an increased risk of adverse cardiac events with low-to-moderate dose venlafaxine compared to sertraline.[3]

An observational study using high-quality data from the DARWIN EU® network could add important further evidence to evaluate the potential association.

The European Medicines Agency (EMA) has therefore requested a study to estimate the association between venlafaxine and heart failure/cardiomyopathy, using mirtazapine, which is indicated for the treatment of depression, as active comparator.

7. RESEARCH QUESTION AND OBJECTIVES

Research question

Is new use of venlafaxine associated with an increased risk of incident heart failure compared to new use of mirtazapine?

Is new use of venlafaxine associated with worsening of pre-existing heart failure compared to new use of mirtazapine?

Objectives

The study aims to assess the potential association between venlafaxine and heart failure among adults.

The specific objectives of this study are:

1. To assess the risk of incident heart failure and incident cardiomyopathy between new users of venlafaxine vs. new users of mirtazapine.
2. To assess the risk of heart failure exacerbation between new users of venlafaxine vs. new users of mirtazapine who have been diagnosed with heart failure before treatment initiation.
3. To characterise both new venlafaxine and new mirtazapine users at the time of treatment start (in terms of demographics, pre-defined comorbidities/comedication/conditions of interest, and drug utilisation) for contextualisation of Objectives 1+2

Analyses will be conducted overall and stratified for condition of interest, age group, sex, history of cardiovascular (CV) disease and previous Selective Serotonin Reuptake Inhibitor (SSRI) use.

8. RESEARCH METHODS

8.1. Study design

The study will comprise of two new user active comparator cohort studies (Objectives 1+2) and a patient-level characterisation (Objective 3).

All studies will be conducted using routinely collected health data from 6 data sources from 5 countries across Europe and in 4 European Union (EU) member states.

8.1.1. Objectives 1+2

A new user active comparator cohort study design will be used to assess

1) The risk of incident heart failure and incident cardiomyopathy between new users of venlafaxine vs. new users of mirtazapine.
2) The risk of heart failure exacerbation between new users of venlafaxine vs. new users of mirtazapine who have been diagnosed with heart failure before treatment initiation

This study design is recognised as a gold standard method in the ENCEPP Methodological guidelines and proposed as a standard analysis in the DARWIN EU® Catalogue of Standard Analyses (https://darwin-eu.org/index.php/methods/standardised-analytics).

The studies were designed using the Target trial emulation framework, with the target trial and plan for emulation in DARWIN EU® data sources provided in **Table 1** below. An illustration of the study design is provided in **Figures 1 and 2** below.

Table 1. Study Design Specification for Objectives 1 and 2.

	Target trial	Emulation of target trial in DARWIN EU® data sources
	Population	
Inclusion	• Patient consents to participate • Aged ≥18 years at study inclusion/randomisation • Prior history of heart failure [Objective 2 only]	• Aged ≥18 years at index date • At least 730 days of data source history prior to the date of treatment initiation with venlafaxine or mirtazapine (index date).

	Target trial	Emulation of target trial in DARWIN EU® data sources
		• Index date occurs between 01/01/2010 and date of last data availability in the respective data source • Diagnosis or record of heart failure or cardiomyopathy ever recorded in the data source before index date [Objective 2 only]
Exclusion	• Use of other antidepressants in the year before study start, except SSRIs, which are permitted. • Prior history of heart failure or cardiomyopathy [Objective 1 only]	• Prescription/dispensation of other antidepressants recorded in the data source in the year before index date, except SSRIs, which are permitted. • Prescription/dispensation record of both venlafaxine and mirtazapine at index date • Diagnosis or record of heart failure or cardiomyopathy ever recorded in the data source before index date [Objective 1 only]
	Intervention	
Target	Venlafaxine	Prescription/dispensation record of venlafaxine
Comparator	Mirtazapine	Prescription/dispensation record of mirtazapine
Treatment Assignment	Randomisation to either venlafaxine or mirtazapine.	Participants will be classified according to the treatment strategy that their data is compatible at baseline (first prescription of venlafaxine or mirtazapine). Randomisation will be emulated using PS matching.
	Outcome	
Time of outcome	First heart failure (or cardiomyopathy) event after drug therapy has started [Objective 1] Hospitalisation for exacerbation of pre-existing heart failure, acute pulmonary oedema, or recurrent heart failure after drug therapy has started. [Objective 2]	First heart failure (or cardiomyopathy) diagnosis recorded in the data source after index date [Objective 1] Hospitalisation for heart failure, acute pulmonary oedema, or recurrent heart failure recorded in the data source after index date [Objective 2]
	Causal estimands	
Censoring	Treat continuously until the earliest occurrence of either an outcome event, death, end of study (24 months), or consent withdrawal.	Follow-up will end on the earliest of loss to follow-up, death, end of observation period (the latest available data) or occurrence of outcome of interest (per-protocol analysis). Follow-up will end on the earliest of loss to follow-up, death, end of observation period (the latest available data), occurrence of outcome of interest, or 24 months after index date (Intention-to-treat and additional per-protocol analysis).
Intercurrent events	• Treatment discontinuation (due to adverse events or resolution of depression). • Treatment switch (due to adverse events or lack of efficacy).	• Treatment discontinuation (end of first continuous drug era) • Treatment switch, defined as a first record of prescription/dispensation of another antidepressant without continuation of

	Target trial	Emulation of target trial in DARWIN EU® data sources
	• Treatment escalation/add on treatments are allowed	venlafaxine/mirtazapine, as well as a first record of prescription/dispensation of mirtazapine among venlafaxine new users or a first record of prescription/dispensation of venlafaxine among mirtazapine new users • Treatment escalation is allowed (add on of any additional antidepressants or dose escalation)
Artificial Censoring	N/A	Treatment discontinuation and switching will be handled through artificial censoring at their occurrence.
Causal contrast	1) Per-protocol (PP) analysis, based on the treatment that was received. 2) Intention-to-treat (ITT) analysis, based on the treatment that was assigned.*	1) Observational analogue of the per-protocol analysis. Artificial censoring will occur for treatment discontinuation and switching. 2) Observational analogue of the ITT analysis, with maximum follow-up of 24months from index date. No artificial censoring for treatment discontinuation and switching.*
Statistical analyses		
Statistical analyses	Survival analysis in randomised groups	Among the PS matched population, we will estimate average treatment effects in the treated (ATT). Poisson regression will be used to estimate incidence rates of the outcome events in each treatment group. Cox proportional hazard models will estimate HRs for the outcome events, overall and stratified by age, sex, CV history, previous SSRI use, and condition of interest (depression/anxiety). Incidence rate ratios estimated through Poisson regression analyses will be used if proportional hazard assumptions are violated. NCO and calibrated HRs will be used to detect residual confounding, for the overall population, previous SSRI use, and condition of interest subgroups will be reported in addition. Random effect meta-analyses to pool results from individual data sources. I^2 will be reported to assess heterogeneity.
Sensitivity analyses		Sensitivity analyses will be conducted for each outcome: • Restricted to patients with at least 5 years of prior observation prior to index date instead of 730 days

SSRIs = Selective serotonin reuptake inhibitors; FinOMOP-THL = Finnish Care Register for Health Care; N/A = Not applicable; ITT = Intention to treat; ATT = average treatment effects in the treated; HRs = Hazard ratios; CV = Cardiovascular; NCO = Negative control outcomes.
*Details on the causal contrast are provided in Section **8.6.3 Intercurrent events/censoring**

Choice of active comparator:

We expect venlafaxine to be predominantly prescribed as a second-line treatment, potentially following treatment with SSRIs as first line treatment for depression. Mirtazapine is considered the most likely alternative treatment option for venlafaxine for treatment of depression and chosen as active comparator for the comparative safety study. Mirtazapine does not have warnings on heart failure/cardiomyopathy in the SmPC. However, peripheral oedema is labelled as side effect.

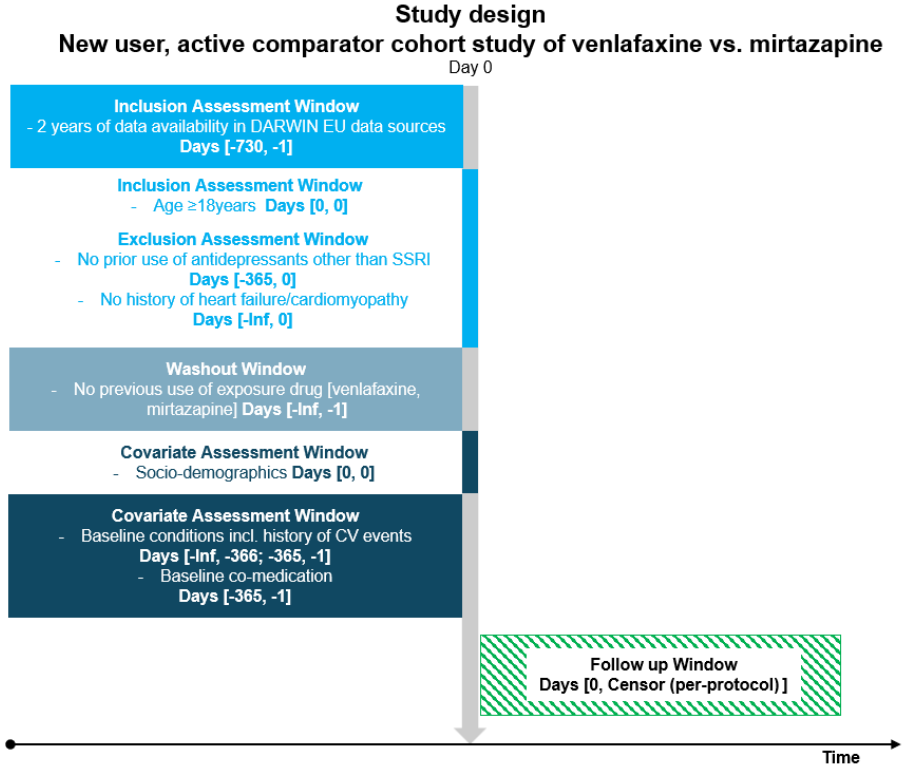

Figure 1. Graphical depiction of the study design for Objective 1.

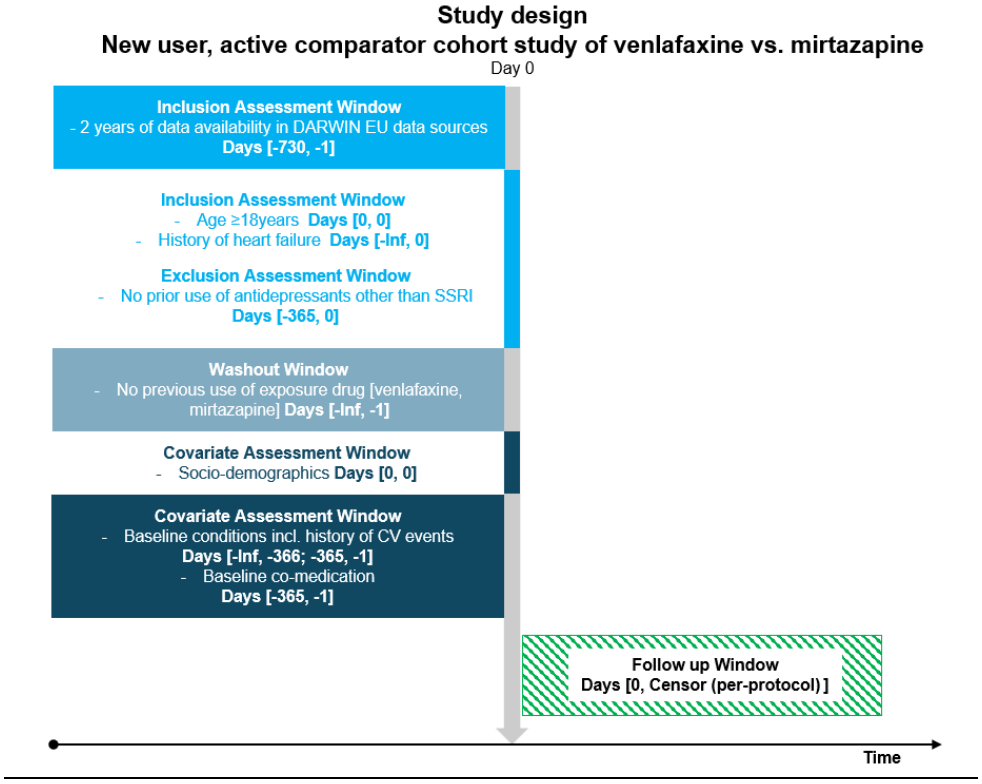

Figure 2. Graphical depiction of the study design for Objective 2.

8.1.2. Objective 3

We will summarise patient characteristics for new venlafaxine and new mirtazapine users, respectively. Patient-level characterisation will include

- Demographics
- Pre-defined comorbidities (e.g., cardiovascular diseases/history of outcomes)
- Comedication (e.g., previous use of antidepressants)
- Utilisation of venlafaxine and mirtazapine (e.g., time to treatment from depression/anxiety diagnosis, treatment dose, and duration)

Stratifications will be conducted for condition of interest/indication (depression/anxiety), age groups, sex, history of CV disease (yes/no), and previous SSRI use (yes/no).

8.2. Follow-up

For characterisation, baseline characteristics will be assessed for venlafaxine or mirtazapine users in pre-defined time windows before/at index date.

Follow-up for the new user active comparator cohort studies will start on the date of first prescription of venlafaxine or mirtazapine. Follow-up will end on the earliest of

- Loss to follow-up
- Death
- Outcome occurrence
- End of observation period (the latest available data).

In addition, for per-protocol analyses follow-up will be artificially censored at the time of treatment discontinuation (either stopping of treatment or switching to a different treatment).

Intention-to-treat analyses will have a maximum follow-up of 24 month after index date. Similarly, in addition to the overall analysis, follow-up will also be restricted to 24 month for a per-protocol analysis.

8.3. Study population with inclusion and exclusion criteria

The study population will include all individuals with a recorded first prescription for venlafaxine or mirtazapine within the study period from 01/01/2010 up to the end of data availability who meet the eligibility criteria at study entry.

Other eligibility criteria are:

Inclusion criteria

- At least 730 days of data source history prior to index date.
- Aged ≥18 years at index date
- At least one record/diagnosis of heart failure recorded in the data source before index date (Objective 2 only).

Exclusion criteria

- Prescription / dispensation of other antidepressants recorded in the data source in the year before index date, except SSRIs, which are permitted.
- Prescription/dispensation record of both venlafaxine and mirtazapine at index date.
- Diagnosis or record of heart failure or cardiomyopathy ever recorded in the data source before index date (Objective 1 only).

For Objectives 1+2, a sensitivity analysis will be conducted requiring 5 years of data source history prior to index date.

8.4. Study setting and data sources

This study will be conducted using routinely collected data from 6 primary/secondary care data sources in the DARWIN EU® network of data partners from 5 European countries, of which 4 are EU member states. All data were *a priori* mapped to the Observational Medical Outcomes Partnership Common Data Model (OMOP CDM).

1. Denmark: Danish Data Health Registries (DK-DHR)
2. Finland: Finnish Care Register for Health Care (FinOMOP-THL)
3. Spain: Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público (BIFAP)
4. Spain: The Information System for Research on Primary Care (SIDIAP)
5. Sweden: Health Impact - Swedish Population Evidence Enabling Data-linkage (HI-SPEED)
6. The United Kingdom: Clinical Practice Research Datalink GOLD (CPRD GOLD)

An overview of the 6 data sources is provided in **Table 2** below.

Table 2. Data sources.

Country	Name of Data source	Health Care setting	Type of Data	Number of active individuals	Calendar period covered by each data source	Contributing to Objectives
Denmark	DK-DHR	Community pharmacists, secondary Care and Hospital in-patient care	National Registry	6M	1995–2025	1, 2, 3
Finland	FinOMOP-THL	Primary care, secondary care, hospital inpatient care	EHR, registries	5.7M	2016–2023	1, 2, 3
Spain	BIFAP	Primary care, no use of hospital linkage	Insurance claims, EHR, registries	23M	2005–2024	1, 2, 3
Spain	SIDIAP	Primary care, hospital inpatient care	EHR	8M	2006–2025	1, 2, 3
Sweden	HI-SPEED	Primary care, secondary care, hospital inpatient care	National Registry	4.5M	2015–2025	1, 2, 3
The United Kingdom	CPRD GOLD	Primary Care	EHR	17M	1987–2025	1, 2, 3

DK-DHR = Danish Data Health Registries; FinOMOP-THL = Finnish Care Register for Health Care; BIFAP = Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público; SIDIAP = The Information System for the Development of Research in Primary Care; HI-SPEED = Health Impact - Swedish Population Evidence Enabling Data-linkage; CPRD GOLD = Clinical Practice Research Datalink GOLD; EHR = Electronic Health Records

Selection of data sources

Out of 40 data partners onboarded into the DARWIN EU® network as of November 2025, 10 covered the required setting for the study, namely population-level (regional or national) coverage of primary care, and where available, linked to secondary care or registries. Six data sources had an acceptable estimated size of study population, fulfilled data source recency/lookback period requirements, and covered key variables.

The 6 selected data sources therefore fulfil the criteria required in terms of data quality, completeness, timeliness, and representativeness for a new user active comparator cohort study while covering different regions of Europe. Additional details on the data sources are provided in **Annex II**.

As CPRD GOLD and BIFAP do not have inpatient data, those data sources will not assess the outcome of heart failure hospitalisation in Objectives 1 and 2.

8.5. Study period

The study period will range from 01/01/2010 to the most recent data available for each contributing data source (see **Table 2**).

It should be noted that in the FinOMOP-THL data source, the availability of the accurate data ends in 12/2023, as prescription records are only available until the end of 2023. The study period will be restricted to between 2016–2023. For HI-SPEED, data availability starts in 2015.

8.6. Variables

8.6.1. Exposure

Exposure:

- Venlafaxine (exposure) [ATC-code: N06AX16]
- Mirtazapine (active comparator) [ATC-code: N06AX11]

In this study, venlafaxine or mirtazapine will be identified as the first prescription/dispensing recorded (defined as no previous exposure to venlafaxine or mirtazapine in an individual's history) using RxNorm codes for the respective ingredient. The standard ingredient concepts used for the identification of venlafaxine and mirtazapine will be 743670 and 725131, respectively. Desvenlafaxine will not be included in the venlafaxine cohort.

Drug eras will be defined as follows: Exposure starts at date of the first prescription, e.g., the index date the individual entered the cohort. For each prescription, the estimated duration of use is retrieved from the drug exposure table in the CDM, using start and end date of the respective prescription. Subsequent prescriptions will be combined into continuous exposed episodes (drug eras) using the following specifications: Treatment episodes of sequential prescriptions (i.e., drug era) will have a maximum of 90 days between the end date of one prescription and the start date of the next prescription (i.e., definition of gap). The drug era will therefore account for all periods when the individual is likely to be continuously using venlafaxine or mirtazapine starting from the day of their first prescription ("index date").

8.6.2. Outcome

Objective 1:

The outcomes for objective 1 will be separated into primary and secondary outcomes:

Heart failure and cardiomyopathy will be defined as described in the DARWIN EU® study P3-C3-001 (Adverse events of special interest), with the first ever event/diagnosis in the patient history being used as the event date for incident outcomes.

The heart failure (broad) definition included all terms directly related to heart failure, with high sensitivity and moderate specificity, which would reasonably be included in notes for a patient with a clinical diagnosis of heart failure.

Concepts included in the code list for cardiomyopathy (broad) included both codes related to primary and secondary causes, whereas cardiomyopathy (primary) excludes diagnoses related to secondary causes.

Preliminary code lists are available in **Annex IV.** These definitions were developed following the DARWIN EU® phenotyping standard processes, which involved the development of a clinical description, the review of code lists by clinical experts and the review of phenotypes after their execution in the participating data sources (of which CPRD GOLD and SIDIAP are included in the present study). We will conduct large-scale diagnostics to test the definition in all data sources. Cohort code use for each data source will be reported as part of the study results.

Primary outcome: incident heart failure

The following definitions will be used:

- Incident heart failure (broad)

  Cohort inclusion event will be the first ever record of heart failure.

Secondary outcomes: broad/narrow definitions for incident heart failure; incident cardiomyopathy

The following definitions will be used:

- Incident heart failure (broad), *without* a preceding alternative cause of heart failure

  We will use the "Incident heart failure (broad)" definition. Cohort inclusion event will be the first ever record of heart failure. In the case that any of the redefined cardiac conditions (ischemic heart disease, valvular disease, pulmonary heart disease) are recorded in the month before the respective first ever record of heart failure, we will not consider the event as outcome.

- Incident heart failure (narrow)

  Building on the definition of heart failure (broad), we will exclude heart failure cases with a clearly identified primary cause unrelated to cardiomyopathy or acquired drug-related effect ("narrow definition"). Cohort inclusion event will be the first ever record of heart failure.

- Incident heart failure (narrow, *without* a preceding alternative cause of heart failure)

  We will use the narrow definition of heart failure. Cohort inclusion event will be the first ever record of heart failure (narrow). In the case that any of the redefined cardiac conditions (ischemic heart disease, valvular disease, pulmonary heart disease) are recorded in the month before the respective first ever record of heart failure, we will not consider the event as outcome.

- Incident cardiomyopathy (broad)

  A broad definition of cardiomyopathy, including both primary and secondary causes, will be used. Cohort inclusion event will be the first ever record of cardiomyopathy.

- Incident cardiomyopathy (primary)

  We will use a definition restricted to records of primary cardiomyopathies, which are typically those where the condition is predominantly confined to the heart muscle. Cohort inclusion event will be the first ever record of cardiomyopathy.

- HF hospitalisation

  Heart failure hospitalisation will be defined as a record of heart failure (broad definition) with an inpatient visit (HI-SPEED, FinOMOP-THL, DK-DHR) or a record of heart failure identified from linked hospital records (SIDIAP).

Objective 2:

Proxies for exacerbation of previously diagnosed heart failure will be assessed. The following definition will be used:

Primary outcome: HF hospitalisation

- HF hospitalisation

  Heart failure hospitalisation will be defined as a record of heart failure (broad definition) with an inpatient visit (HI-SPEED, FinOMOP-THL, DK-DHR) or a record of heart failure identified from linked hospital records (SIDIAP).

Secondary outcomes: Acute pulmonary oedema, repeated event of heart failure

- Acute pulmonary oedema

  Acute pulmonary oedema will be defined as a (first or recurrent) record of acute pulmonary oedema

- Recurrent heart failure (broad)

  We will define recurrent heart failure as a subsequent record of heart failure (broad) diagnosis following the incident diagnosis.

The outcome of recurrent HF diagnosis will be assessed carefully at the diagnostic stage due to a high risk of re-recordings of diagnosis which might not accurately reflect a worsening of the condition. The inclusion of secondary outcome will be conditional on the results from diagnostics review, e.g., plausibility based on prescriptions/procedures in the immediate time around re-recordings of heart failure.

Objective 3:

The outcome for the patient-level characterisation will be summarised patient characteristics, including

- Demographics, including age, sex, previous observation time in the data source
- Conditions of interest, defined as potential indications including depression, anxiety, and other conditions, such as insomnia
- Pre-defined comorbidities, including cardiovascular diseases, CV outcomes, and hypertension
- Comedication, including previous use of antidepressants
- Utilisation of venlafaxine/mirtazapine,
    - time between new treatment from last depression/anxiety diagnosis, respectively
    - treatment dose (initial, cumulative)
    - treatment duration

Analyses for Objective 3 will be stratified by condition of interest (for depression/anxiety), sex, age group, history of CV events, and previous SSRI use, respectively.

8.6.3. Intercurrent events/censoring

Intercurrent events are events that occur after the start of follow-up (e.g., treatment initiation or cohort entry) and that may affect the interpretation or the existence of the outcome of interest.

For Objectives 1 and 2, the primary estimand is the per-protocol causal effect (i.e. the effect under full adherence), and the secondary estimand is the intention-to-treat effect (i.e. the effect of treatment assignment regardless of adherence).. For both, we will censor follow-up at the time of the following events:

Censoring:

- Loss to follow-up
- End of observation period (latest available data)
- Death
- Outcome occurrence
- 24 months (for intention-to-treat and a per-protocol analysis with restricted follow-up)

If an individual in the matched pair is censored for any of the reasons above, follow-up for the affected individual will end but not for the matched counterpart.

Artificial censoring:

As per the PP approach and focusing on the time individuals are exposed to the exposure/active comparator treatment, we will censor follow-up at the time an individual discontinues treatments (i.e., end of the continuous treatment era) with venlafaxine and mirtazapine, respectively.

Artificial censoring will be conducted at

- Treatment discontinuation (end of first continuous drug era)

- Treatment switch (defined as a first record of prescription/dispensation of another antidepressant without continuation of venlafaxine/mirtazapine, as well as a first record of prescription/dispensation of mirtazapine among venlafaxine new users or a first record of prescription/dispensation of venlafaxine among mirtazapine new users)

Treatment escalation will be allowed, and no censoring will be done in case of the addition of any additional antidepressant or dose escalation. If an individual in the matched pair gets censored for any of the reasons above, follow-up for the matched pair will be censored. Sensitivity analyses without censoring the pair will be conducted.

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

Objectives 1+ 2

Covariates to be included in the large-scale PS will be selected from comedications and comorbidities recorded prior to index date using a data-driven approach (LASSO regression). Pre-defined key variables will be used for exact matching, namely age, sex, calendar year, history of CV disease, previous SSRI use, and conditions of interest.

History of CV disease will be defined as history of CV events, namely ischaemic heart disease, atrial fibrillation, valvular heart disease, myocardial infarction, and stroke.

Analyses will be conducted overall and stratified for age groups (18–39, 40–59, 60–79, 80+), sex, history of CV disease (yes/no), previous SSRI use (yes/no), and clinical indication of interest (depression, anxiety).

Objective 3

The covariates for the new venlafaxine/mirtazapine user characterisation are as follows:

- Demographics
    - Age [median, Interquartile Range (IQR), range, mean, SD]
    - Sex [binary: female/male]
    - Calendar year [median, IQR, range, mean, SD]
    - Previous observation time in data source [median, IQR, range, mean, SD]
- Conditions of interest
    - Depression (with/without concomitant recording of anxiety)
    - Anxiety (without concomitant recording of depression)
    - Other (i.e., insomnia)
    - Assessment window before/at index date: [-Inf, -366; -365, 0]
- Medication prior to index date (comedication)
    - *Previous antidepressant use, including SSRIs, MAOs, and TCAs*
    - *Previous use of CV medication: anti-arrhythmic medications, diuretics (aldosterone antagonists), RAAS-inhibitors, beta-blockers, calcium channel blockers.*
    - *Non-steroidal anti-inflammatory drugs (NSAIDs), corticosteroids*
    - Assessment window before index date: [-365, -1]
- Comorbidities
    - *History of outcome events: heart failure, cardiomyopathy*

  - *History of CV events: ischaemic heart disease, atrial fibrillation and valvular heart disease, MI, stroke*
  - *Hypertension, diabetes mellitus, obesity, chronic kidney disease*
  - Assessment window before index date: [-Inf, -366; -365, -1]
- Healthcare visits
  - Outpatient visits [primary care/outpatient specialises care] (where available in the respective data source)
  - Inpatient visits (where available in the respective data source)
  - Assessment window before index date: [-365, -1]
- Drug utilisation for venlafaxine/mirtazapine:
  - Time between new treatment (index date) and last record of depression diagnosis [median, IQR, range, mean, SD]
  - Time between new treatment (index date) and last record of anxiety diagnosis [median, IQR, range, mean, SD]
  - The number of records/episodes for depression/anxiety, with a washout of ≥365 days between repeated recordings (where possible) [median, IQR, range, mean, SD]
  - Initial treatment dose [median, IQR, range, mean, SD]
  - Number of prescriptions in the first treatment era [median, IQR, range, mean, SD]
  - Cumulative dose for first treatment era [median, IQR, range, mean, SD]
  - Duration of first continuous treatment era [median, IQR, range, mean, SD]

Analyses will be conducted overall and stratified for age groups (18–39, 40–59, 60–79, 80+), sex, history of CV disease (yes/no), previous SSRI use (yes/no), and clinical indication of interest (depression, anxiety).

8.6.5. Negative Control Outcomes

Negative control outcomes (NCO) are outcomes not causally associated with the exposure of interest,(4) here venlafaxine. However, their association with venlafaxine is ideally impacted by the same type of unmeasured confounding, e.g., healthcare-seeking behaviour, as the venlafaxine-outcome association.(5) NCO will be used as an additional measure to measure any potential residual confounding.

A list of NCOs (**Annex IV**) used in previous safety and effectiveness studies will be used as a starting point. In addition, NCO will be identified through large-scale characterisation at the study diagnostics stage: Among all venlafaxine/mirtazapine users in each data source, the 50 most commonly recorded conditions in the 180 days after treatment start will be extracted. The list will then be used by clinical experts to identify potential NCOs. This process will be done to ensure sufficient counts of NCOs in the respective data source and study cohort.

8.7. Study size

Based on a preliminary feasibility assessment, the expected number of person counts for venlafaxine in the data sources included in this study range from 184,600 (HI-SPEED) to 303,200 (CPRD GOLD). The expected number of person counts for mirtazapine in the data sources included in this study range from 266,900 (SIDIAP) to 1,088,000 (BIFAP). The expected number of person counts for heart failure in the data sources included in this study range from 252,200 (CPRD GOLD) to 936,600 (BIFAP).

In the DARWIN EU® P3-C3-001 (Adverse events of special interest) study, background rates of heart failure were estimated to be 138 per 100,000 person-years in CPRD GOLD and 386 per 100,000 person-years in SIDIAP.

Estimation of precision

Table 3 below shows the required number of events and sample size to estimate a HR of 1.25 with a relative precision of 5, 10 and 20%. Depending on the expected (difference in) event rates of incident heart failure outcomes in the venlafaxine and mirtazapine groups, the required sample size ranged between 13697 and 573793.

Table 3. Relative precision for HR incident heart failure.

Expected HR	Event rate venlafaxine	Event rate mirtazepine	Relative precision (%)	Lower limit 95%CI	Upper limit 95%CI	Events	Sample size
1.25	0.0125	0.01	5	1.19	1.31	6455	573793
1.25	0.0125	0.01	10	1.14	1.38	1692	150363
1.25	0.0125	0.01	20	1.04	1.50	462	41091
1.25	0.025	0.02	5	1.19	1.31	6455	286896
1.25	0.025	0.02	10	1.14	1.38	1692	75182
1.25	0.025	0.02	20	1.04	1.50	462	20545
1.25	0.0375	0.03	5	1.19	1.31	6455	191264
1.25	0.0375	0.03	10	1.14	1.38	1692	50121
1.25	0.0375	0.03	20	1.04	1.50	462	13697

CI = 95% Confidence interval. Estimated based on Rothman KJ, Greenland S. Planning Study Size Based on Precision Rather Than Power. Epidemiology (Cambridge, Mass). 2018;29(5):599-603. and Schoenfeld D. Sample-size formula for the proportional-hazards regression model. Biometrics. 1983.

Table 4 below shows the required number of events and sample size to estimate a HR of 1.25 with a relative precision of 5, 10 and 20%. Depending on the expected (difference in) event rates of heart failure (repeated events) outcomes in the venlafaxine and mirtazapine groups, the required sample size ranged between 1370 and 57379.

Table 4. Relative precision for HR heart failure.

Expected HR	Event rate venlafaxine	Event rate mirtazepine	Relative precision (%)	Lower limit 95%CI	Upper limit 95%CI	Events	Sample size
1.25	0.125	0.1	5	1.19	1.31	6455	57379
1.25	0.125	0.1	10	1.14	1.38	1692	15036
1.25	0.125	0.1	20	1.04	1.50	462	4109
1.25	0.25	0.2	5	1.19	1.31	6455	28690
1.25	0.25	0.2	10	1.14	1.38	1692	7518
1.25	0.25	0.2	20	1.04	1.50	462	2055
1.25	0.375	0.3	5	1.19	1.31	6455	19126
1.25	0.375	0.3	10	1.14	1.38	1692	5012
1.25	0.375	0.3	20	1.04	1.50	462	1370

CI = 95% Confidence interval. Estimated based on Rothman KJ, Greenland S. Planning Study Size Based on Precision Rather Than Power. Epidemiology (Cambridge, Mass). 2018;29(5):599-603. and Schoenfeld D. Sample-size formula for the proportional-hazards regression model. Biometrics. 1983.

Minimum detectable relative risk (MDRR)

MDRRs estimated based on preliminary feasibility counts and event rate of 100/100'000 person years and/or 200/100'000 person years over a 5-year follow-up with an alpha of 0.05, power of 80% are presented in **Table 5** below.

Table 5. MDRR.

Event rate		DK-DHR	FinOMOP-THL	BIFAP	SIDIAP	HI-SPEED	CPRD GOLD
100/100,000 person years	MDRR	<0.91; >1.10	<0.92; >1.10	<0.91; >1.10	<0.90; >1.13	<0.90; >1.11	<0.92; >1.09
200/100,000 person years	MDRR	<0.94; >1.07	<0.94; >1.08	<0.94; >1.08	<0.92; >1.10	<0.93; >1.09	<0.94; >1.06

DK-DHR = Danish Data Health Registries; FinOMOP-THL = Finnish Care Register for Health Care; BIFAP = Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público; SIDIAP = The Information System for the Development of Research in Primary Care; HI-SPEED = Health Impact - Swedish Population Evidence Enabling Data-linkage; CPRD GOLD = Clinical Practice Research Datalink GOLD; MDRR = Minimum detectable relative risk

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 DTZ (Data Transfer Zone). 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 new user active comparator cohort studies will be conducted using OMOP CDM mapped data.

Matching

All people meeting the eligibility criteria will be included in the venlafaxine or mirtazapine group based on the first prescription they received, respectively. This prescription will define the "index date", which will be used for Propensity Score (PS) calculation. PS will estimate the probability of receiving target treatment.

We will use PS matching (ratio 1:1) to minimize observed confounding: venlafaxine vs. mirtazapine users will be matched based on a combination of 1) pre-defined key variables for exact matching, namely:

- Age (e.g., 5-year age bands)
- Sex (exact)
- Calendar time (+/- 365days from index date)
- History of CV disease [yes/no] (exact) [-365 , -1]
- Condition of interest [depression (yes/no), anxiety (yes/no)] (exact) [-365 , 0]
- previous SSRI use [yes/no] (exact) [-365 , 0]

In addition, 2) calliper matching (0.2 Standard Deviation) on large-scale PS will be conducted: Covariates to be included in the large-scale PS will be selected from comedications [assessment windows: -365, -31; -30, -1] and comorbidities recorded prior to index date [assessment windows: -Inf, -366; -365, -31; -30, -1]. Among those, covariates with a prevalence below 0.5% in the study population will be omitted. Logistic regression with LASSO regularisation will then be used for variable selection. The list of selected covariates will be manually screened by 2 epidemiologists/clinical domain experts to exclude potential instrumental variables and variables already used for exact matching.

Propensity Scores: Diagnostics

Large-scale covariate balance will be assessed before/after PS matching through visual inspection: For each data source, a plot will be produced depicting the standardised mean differences between venlafaxine and mirtazapine cohorts for all available covariates before (x axis) and after propensity score matching (y axis)

Similarly, equipoise will be assessed for each data source though plots of the distribution of the propensity score stratified by venlafaxine vs. mirtazapine cohort.

Characterisation

Matched cohorts will be characterised using the same methods outlined in Objective 3: Covariates of interest will include demographics, conditions of interest, pre-defined comorbidities and previous medication, healthcare visits, as well as analyses on the utilisation of venlafaxine and mirtazapine.

Incidence rates of outcome events

Following PS matching as described above, we will estimate incidence rates as events per 10,000 person-years and 95% confidence intervals for each of the outcomes of interest in the respective treatment group (venlafaxine/mirtazapine). Poisson regression will be used.

In addition to the overall analyses, analysis with a pre-defined follow-up of up to 24 months after index date will be conducted. This analysis will not be stratified.

Comparative Safety: Survival analyses

Cox proportional hazard models will be used to estimate hazard ratios (HRs) for the outcomes of interest, respectively. Kaplan Meier plots and/or cumulative incidence plots will be used to illustrate survival analyses. Corresponding 95% confidence intervals will be provided.

Per-protocol analyses will be conducted with artificial censoring at treatment discontinuation or switch to restrict the analyses to treatment-exposed times and avoid treatment exposure misclassification. In addition to the overall analyses, analysis with a pre-defined follow-up of up to 24 months after index date will be conducted.

Intention-to-treat analyses will be conducted for a pre-defined follow-up of up to 24 months after index date without artificial censoring, allowing for capture of potential clinical complications/late cardiac effects following discontinuation of venlafaxine/mirtazapine.

HRs will be calculated for each outcome overall and stratified by age group, sex, history of CV disease, previous SSRI use, and condition of interest (depression/anxiety).

The proportional hazard assumption will be tested through visual investigation of log(-log) plots. In the scenario of substantial violation of the proportional hazard assumptions, incidence rate ratios will be calculated from Poisson regression and presented alongside the HR.

A list of negative control outcomes (NCO) will be used to identify unmeasured confounding. Cox proportional hazard models will be used to estimate hazard ratios (HRs) for NCOs, which will be presented in forest plots. Additionally, calibrated HRs will be provided for each outcome in the overall study population and in the stratified analyses of previous SSRI use and conditions of interest (depression/anxiety) to help understand whether unmeasured confounding is present.

Sensitivity analyses

1) Prior observation time

A sensitivity analysis restricting the study population to individuals with at least 5 years of data availability prior to index date. This analysis will not be stratified.

Table 4. Objective 1+2 Specification of statistical analyses.

Research question:	Is new use of venlafaxine associated with an increased risk of heart failure compared to new use of mirtazapine?
Treatment assignment:	First ever prescription of venlafaxine or mirtazapine
Outcome:	Objective 1 Primary: incident heart failure (broad) Secondary: • Incident heart failure (broad, without preceding alternative cause) • Incident heart failure (narrow) • Incident heart failure (narrow, without preceding alternative cause)

	• HF hospitalisation • Incident Cardiomyopathy (broad) • Incident Cardiomyopathy (primary) Objective 2 • HF hospitalisation • Acute pulmonary oedema • Recurrent HF (broad)
Follow-up time:	Per-protocol analyses with artificial censoring: overall and up to 24 months ITT (without artificial censoring): up to 24 months
Model:	• Poisson regression for incidence rates of outcome events • Cox proportional hazards outcome model: follow up time*status (0) = exposure
Confounding adjustment method	
	PS matching (exact matching on key variables + large-scale PS) to account for baseline confounding. Negative control outcome analyses to detect residual confounding.
Subgroup Analyses	
	Analyses for each outcome will be stratified for: • Age groups (18–39, 40–59, 60–79, 80+) • Sex (female; male) • History of CV disease (yes/no) • Previous SSRI use (yes/no) • Condition of interest (depression/anxiety)
Sensitivity Analyses	
	Sensitivity analyses will be conducted for each outcome: Restricted to patients with at least 5 years of observation prior to index date instead of 730 days

Objective 3

The patient-level characterisation analyses will be calculated based on OMOP CDM mapped data.

Covariates of interest will include demographics, conditions of interest, pre-defined comorbidities and previous medication, healthcare visits, as well as analyses on the utilisation of venlafaxine and mirtazapine.

Analyses will be stratified separately by conditions of interest (depression/anxiety), age group, sex, history of CV disease (yes/no), and previous SSRI use (yes/no). For categorical variables, counts and proportions will be reported. For numeric variables, median/IQR, mean/SD, and range will be reported.

The R packages *CohortCharacteristics(6)* and *DrugUtilisation(7),* developed by DARWIN EU®, will be used for these analyses.

8.8.4. Output

Output will include the following:

A PDF report including an executive summary and the following tables and figures:

Objective 1+2:

- Tables
  - Attrition tables for venlafaxine and mirtazapine cohorts
  - Baseline characteristics of individuals in the venlafaxine and mirtazapine cohorts after matching, overall
  - Number of events and incidence rates in the venlafaxine and mirtazapine cohorts after matching, overall and stratified
  - HRs for outcomes, overall and stratified
  - Meta-analytic HRs for outcomes, overall and stratified
- Figures
  - Forrest Plot for meta-analytic HR for outcomes
  - Forrest Plot for calibrated, meta-analytic HR for outcomes
  - Forrest Plot for HR of all NCOs

Objective 3:

- Table: Attrition tables for venlafaxine and mirtazapine cohorts
- Table: Patient characteristics of new venlafaxine and mirtazapine users, overall
- Table: Patient characteristics of new venlafaxine and mirtazapine users, stratified by clinical indication (depression/anxiety) [per data source]

An interactive dashboard will be generated by incorporating all the results included in the PDF report mentioned above, as well as stratifications for Objective 3, database snapshots, cohort code use, diagnostic plots of PS covariate balance, diagnostic plot of PS equipoise, HR for NCOs, as well as log-log plots to assess proportionate hazard assumption.

Mock Table 1. Number of individuals with venlafaxine or mirtazapine included for matching.

Mock table shown for 3 data sources only. Separate Tables for Objectives 1 and 2.

	BIFAP		CPRD GOLD		DK-DHR	
	venlafaxine	**mirtazapine**	**venlafaxine**	**mirtazapine**	**venlafaxine**	**mirtazapine**
Qualifying initial subjects						
First ever venlafaxine/mirtazapine (index date)						
Prior observation of 730 days before index date						
Index date after 01/01/2010						
Age at index date ≥18						

	BIFAP		CPRD GOLD		DK-DHR	
	venlafaxine	mirtazapine	venlafaxine	mirtazapine	venlafaxine	mirtazapine
No use of antidepressants with known cardiac effects in the year before index date						
No history of heart failure/cardiomyopathy before index date						
No record of both venlafaxine and mirtazapine on index date						
Number of individuals after PS Matching						

BIFAP = Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público , CPRD GOLD = Clinical Practice Research Datalink GOLD, DK-DHR = Danish Data Health Registries.

Mock Table 2. Patient baseline characteristics of matched venlafaxine and mirtazapine users.

Mock table shown for three data sources only. Separate Tables for Objectives 1 and 2.

		BIFAP		CPRD GOLD		DK-DHR	
		venlafaxine	mirtazapine	venlafaxine	mirtazapine	venlafaxine	mirtazapine
Demographics							
Age	median, IQR, range, mean, SD						
Sex: female	N, %						
Calendar year	Median, IQR, range, mean, SD						
Previous observation time	median, IQR, range, mean, SD						
Follow-up time	median, IQR, range, mean, SD						
Conditions of interest [-365 , 0]							
Depression	N, %						
Anxiety	N, %						
Other	N, %						
Line of therapy							
SSRI: previous use (yes)	N, %						
History of CV disease							
History of CV disease (yes)	N, %						
Hypertension	N, %						

		BIFAP		CPRD GOLD		DK-DHR	
		venlafaxine	mirtazapine	venlafaxine	mirtazapine	venlafaxine	mirtazapine
Ischaemic heart disease	N, %						
…							
Reasons for follow-up censoring							
Censoring: death, end of observation	N, %						
Censoring: outcome occurrence	N, %						
Artificial censoring: treatment discontinuation/switching	N, %						

BIFAP = Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público, CPRD GOLD = Clinical Practice Research Datalink GOLD, DK-DHR = Danish Data Health Registries, N = Number of persons, IQR = interquartile range, SSRI = Selective Serotonin Reuptake Inhibitors, CV = cardiovascular, SD = standard deviation.

Mock Table 3. Number of events and incidence rates in the venlafaxine and mirtazapine cohorts after matching, overall and stratified.

Mock table shown for venlafaxine only, and for one data source only. Separate Tables for Objectives 1 and 2.

		Venlafaxine																	
		Primary Outcome			Secondary Outcome			Secondary Outcome			Secondary Outcome			Secondary Outcome			Secondary Outcome		
		N	PY	IR, 95% CI	N	PY	IR, 95% CI	N	PY	IR, 95% CI	N	PY	IR, 95% CI	N	PY	IR, 95% CI	N	PY	IR, 95% CI
		BIFAP																	
Overall	overall																		
	24 months																		
Overall: 5 yrs previous data availability	Overall																		
Overall: stratified																			

		Venlafaxine																	
		Primary Outcome			Secondary Outcome			Secondary Outcome			Secondary Outcome			Secondary Outcome			Secondary Outcome		
		N	PY	IR, 95% CI	N	PY	IR, 95% CI	N	PY	IR, 95% CI	N	PY	IR, 95% CI	N	PY	IR, 95% CI	N	PY	IR, 95% CI
		BIFAP																	
Age groups	18–39																		
	40–59																		
	60–79																		
	80+																		
Sex	female																		
	male																		
History of CV disease	yes																		
	no																		
Previous SSRI use	yes																		
	no																		
Condition of interest	depression																		
	anxiety																		

N = Number of persons, PY = person years, IR = Incidence Rate, 95% CI = 95% Confidence Intervals, BIFAP = Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público, CV = cardiovascular, SSRI = Selective serotonin receptor inhibitor.

Mock Table 4. HRs for outcomes of interest, overall and stratified.

Mock table shown for one data source only. Separate Tables for Objectives 1 and 2.

		Primary Outcome		Secondary Outcome		Secondary Outcome		Secondary Outcome		Secondary Outcome		Secondary Outcome	
		HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI
BIFAP													
Overall	overall												
	24 months												
Overall: 5 yrs previous data availability	overall												
Overall: stratified													
Age groups	18–39												
	40–59												
	60–79												
	80+												
Sex	female												
	male												
History of CV disease	yes												
	no												
Previous SSRI use	yes												
	no												
Condition of interest	depression												
	anxiety												

HR = Hazard Ratio, 95% CI = 95% Confidence Intervals, BIFAP = Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público, CV = cardiovascular, SSRI = Selective serotonin receptor inhibitor.

Mock Table 5. Meta-analytic HRs for outcomes of interest, overall and stratified.

Mock table shown for one data source only. Separate Tables for Objectives 1 and 2.

		Primary Outcome		Secondary Outcome		Secondary Outcome		Secondary Outcome		Secondary Outcome		Secondary Outcome	
		HR_{meta}	95% CI_{meta}	HR_{meta}	95% CI_{meta}	HR_{meta}	95% CI_{meta}	HR_{meta}	95% CI_{meta}	HR_{meta}	95% CI_{meta}	HR_{meta}	95% CI_{meta}
BIFAP													
Overall	overall												
	24 months												
Overall: 5 yrs previous data availability	overall												
Overall: stratified													
Age groups	18–39												
	40–59												
	60–79												
	80+												
Sex	female												
	male												
History of CV disease	yes												
	no												
Previous SSRI use	yes												
	no												
Condition of interest	depression												
	anxiety												

HR = Hazard Ratio, 95% CI = 95% Confidence Intervals, BIFAP = Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público, CV = cardiovascular, SSRI = Selective Serotonin Reuptake Inhibitors.

Mock Table 6. Number of individuals and records with venlafaxine or mirtazapine included in the study, by data source (Objective 3).

Mock tables showing records counts. Results tables will show both subject and record counts.

	BIFAP	CPRD GOLD	DK-DHR	FinOMOP-THL	HI-SPEED	SIDIAP
Venlafaxine						
Qualifying initial records						
Records of first ever venlafaxine (index date)						
Prior observation of 365 days before index date						
Index date after 01/01/2010						
Age at index date ≥18						
No record of both venlafaxine and mirtazapine on index date						
Mirtazapine						
Qualifying initial records						
Records of first ever mirtazapine (index date)						
Prior observation of 365 days before index date						
Index date after 01/01/2010						
Age at index date ≥18						
No record of both venlafaxine and mirtazapine on index date						

BIFAP = Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público, CPRD GOLD = Clinical Practice Research Datalink GOLD, DK-DHR = Danish Data Health Registries, FinOMOP-THL = Finnish Care Register for Health Care, HI-SPEED = Health Impact - Swedish Population Evidence Enabling Data-linkage, SIDIAP = The Information System for the Development of Research in Primary Care.

Mock Table 7. Patient characteristics of new venlafaxine users, overall.

Mock table shown for venlafaxine only.

		BIFAP	CPRD GOLD	DK-DHR	FinOMOP-THL	HI-SPEED	SIDIAP
Demographics							
Age	median, IQR, range, mean, SD						
Sex: female	N, %						
Calendar year	Median, IQR, range, mean, SD						
Previous observation time	median, IQR, range, mean, SD						
Conditions of interest [-365 , 0]							
Depression	N, %						
Anxiety	N, %						
Insomnia	N, %						
Concomitant medications [-365 , -1]							
SSRI	N, %						
Other antidepressants	N, %						
Other comedication	N, %						
Comorbidities [-Inf, -365] [-365 , -1]							
Heart failure (history of)	N, %						
Cardiomyopathy (history of)	N, %						
Hypertension	N, %						
Ischaemic heart disease	N, %						
…							
Healthcare visits [-365 , -1]							

		BIFAP	CPRD GOLD	DK-DHR	FinOMOP-THL	HI-SPEED	SIDIAP
Outpatient GP visits	N, %						
Outpatient specialist visits	N, %						
Inpatient visits	N, %						
Drug utilisation							
Time between new treatment and last depression record	median, IQR, range, mean, SD						
Time between new treatment and last anxiety record	median, IQR, range, mean, SD						
Initial treatment dose	median, IQR, range, mean, SD						
Cumulative treatment dose for first treatment era	median, IQR, range, mean, SD						
Duration of first continuous treatment era	median, IQR, range, mean, SD						

BIFAP = Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público, CPRD GOLD = Clinical Practice Research Datalink GOLD, DK-DHR = Danish Data Health Registries, FinOMOP-THL = Finnish Care Register for Health Care, HI-SPEED = Health Impact - Swedish Population Evidence Enabling Data-linkage, SIDIAP = The Information System for the Development of Research in Primary Care, N = Number of persons, IQR = interquartile range, SSRI = Selective Serotonin Reuptake Inhibitors, CV = cardiovascular, GP = General Practitioner, SD = standard deviation.

Mock Table 8. Patient characteristics of new venlafaxine and mirtazapine users, stratified by condition of interest.

Mock table shown for one data source only.

		venlafaxine		mirtazapine	
		depression	anxiety	depression	anxiety
Demographics					
Age	median, IQR, range, mean, SD				
Sex: female	N, %				
Calendar year	Median, IQR, range, mean, SD				
Previous observation time	median, IQR, range, mean, SD				
Conditions of interest [-365 , 0]					
Depression	N, %				
Anxiety	N, %				
Insomnia	N, %				
Concomitant medications [-365 , -1]					
SSRI	N, %				
Other antidepressants	N, %				
Other comedication	N, %				
Comorbidities [-Inf, -365] [-365 , -1]					
Heart failure (history of)	N, %				
Cardiomyopathy (history of)	N, %				
Hypertension	N, %				
Ischaemic heart disease	N, %				
…					
Healthcare visits [-365 , -1]					
Outpatient GP visits	N, %				
Outpatient specialist visits	N, %				
Inpatient visits	N, %				
Drug utilisation					
Time between new treatment and last depression record	median, IQR, range, mean, SD				
Time between new treatment and last anxiety record	median, IQR, range, mean, SD				
Initial treatment dose	median, IQR, range, mean, SD				
Cumulative treatment dose for first treatment era	median, IQR, range, mean, SD				

		venlafaxine		mirtazapine	
		depression	anxiety	depression	anxiety
Duration of first continuous treatment era	median, IQR, range, mean, SD				

N = Number of persons, IQR = interquartile range, SSRI = Selective Serotonin Reuptake Inhibitors, CV = cardiovascular, GP = General Practitioner, SD = standard deviation.

8.9. Evidence synthesis

For objectives 1+2, we will report all analyses separately for each data source and outcome. Additionally, pooled effect estimates across data sources will be estimated using random effect meta-analyses for analyses for which at least 3 data sources have event counts ≥5. I^2 for heterogeneity will be reported. Forest plots will be used to show results from meta-analyses.

Results from analyses described in Objective 3 will be presented separately for each data source. No meta-analysis of results will be conducted.

9. STRENGTHS AND LIMITATIONS

General limitations of the study

The study will be informed by routinely collected health care data and so data quality issues must be considered. In particular, a recording of a prescription or dispensation does not mean that the patient actually took the drug. In addition, assumptions around the duration of drug use will be unavoidable.

Moreover, the documentation of comorbidities may vary across data sources.

Additionally, the results estimated from this study will only reflect the populations from the included data sources. Electronic health records have certain inherent limitations because they were collected for clinical purposes rather than primarily for research use. Consequently, generalisability to countries outside the scope of the study with different healthcare utilisation and prescribing patterns may be limited.

Limitations related to individual data sources

FinOMOP-THL: As prescription records are only available until the end of 2023, and data on cardiovascular outcomes will only be available until October 2024, the study period will be restricted to between 2016–2023 for FinOMOP. In addition, the primary source dataset for prescriptions (Kanta Prescription Centre) includes a prescription start date but no actual end date. For each prescription, the end date is therefore imputed the standard value given in the THEMIS conventions (30 days) for the end date. No calculations based on the package size and e.g., defined daily doses have been performed.

BIFAP: Availability of linkage to hospital data in BIFAP varies by region and time. As bias might be introduced with varying availability of HF inpatient diagnoses over time, we will not use linked hospital data for this study for BIFAP.

CPRD and BIFAP will not have linkage to hospital data for this study. While both the UK and Spain have GPs as gatekeepers in their national health systems, and hence HF admissions in hospitals are usually re-recorded in primary care, some degree of index date misclassification cannot be ruled out. However, this is not expected to be differential between venlafaxine and mirtazapine users.

Limitations related to the study design/outcome definitions

- Mirtazapine was selected as active comparator in order to address potential confounding by indication and assumes the lack of significant cardiovascular effects. This is expected due to its differing expected mechanism of action and lack of significant known cardiovascular adverse drug reactions.
- Previous studies have shown potential for index date misclassification for heart failure diagnoses (DARWIN EU® P3-C3-001 (Adverse events of special interest)). As a subgroup analysis, we will exclude individuals with history of CV disease, in case index date misclassification is non-differential between the intervention and comparator cohorts.
- For repeated HF diagnoses there is a high risk that re-recordings of diagnosis might not reflect a worsening of the condition. At the diagnostics stage, we will carefully assess whether repeated

diagnoses have recordings of prescriptions/procedures in the immediate time around them that would point towards an exacerbation of heart failure. Based on the expected high proportion of missingness measurements of ejection fractions and/or N-terminal pro brain natriuretic peptide in the data sources, such lab measurements were not considered for the study This might limit the information available on the severity of HF available in the respective data sources (e.g., for Objective 2, when selecting a study population with previous HF)

- The terminology of heart failure and cardiomyopathy might overlap: Heart failure as a functional diagnosis might be the presenting symptom for cardiomyopathy (structural diagnosis) and could hence be the diagnosis recorded. To address this uncertainty, we will use 7 different variations of heart failure/cardiomyopathy phenotypes in this study.
- Heart failure can be caused by several other conditions, including ischemic heart disease (IHD), valvular dysfunction, pulmonary disease. We will include an outcome definition for heart failure without preceding alternative causes of heart failure.
- Worsening of pre-existing HF will be assessed in Objective 2, and we have selected hospitalisation of heart failure as primary outcome, which would potentially serve as a proxy for sudden worsening of cardiac function. Changes in HF medication (add-on therapies, dose escalations) or measurements/procedures related to heart failure severity (e.g., ejection fraction) are not included in this study.
- We will use large-scale propensity scores to minimize measured confounding and NCOs to assess potential residual confounding. However, given the observational nature of our data, we cannot rule out residual confounding, which could partially account for findings in this study.
- This study will be estimating a per-protocol causal contrast, with censoring of follow-up time at treatment discontinuation or switching.
    - Therefore, outcomes recorded after treatment discontinuation (e.g., potential late effects) will not be assessed.
    - Add-on treatments might affect the risk for outcome occurrence: The study will build on the assumption of conditional exchangeability after PS matching, with no substantial difference in the occurrence of add-on therapies expected. We therefore do not account for add-on treatments as intercurrent events in this study. Similarly, we do not expect a high degree of venlafaxine to mirtazapine (or vice versa) switching.
- Intention-to-treat analysis will be conducted to assess the impact of artificial censoring and include potential late effects in case of cumulative effects or delayed development/complication of the disease resulting from venlafaxine/mirtazapine treatment. Follow-up will be limited to a maximum of 24 months to reduce bias from potential exposure misclassification.
- The comparative safety studies will not take into account venlafaxine or mirtazapine dose or changes in dose over time. Therefore, potential dose-dependent effects cannot be assessed in this study.
- The feasibility assessment for this study was based upon Objective 1. It is therefore uncertain what number of individuals with venlafaxine and mirtazapine with pre-existing HF are available for analysis of Objective 2.

10. REFERENCES

1. Wang Z, Lu H, Li Y, Dewei S. Effects ofage, sex, daily dose, comorbidity and co-medication onvenlafaxine-associated cardiovascular adverse events: Apharmacovigilance analysis of the FDA Adverse EventReporting System database. . Br J Clin Pharmacol. 2024.

2. Sopi B, Yonekawa K, Russmann S, Attenhofer Jost C. Cardiotoxicity Associated with Venlafaxine—Defining Features in a Series of Five Cases and a Call for Proactive Monitoring. Journal of Clinical Medicine. 2025;14(8):2792.

3. Ho J, Gomes T, Straus S, Juurlink DN. Adverse Cardiac Events in Older People Receiving Venlafaxine: A Population-Based Study J Clin Psychiatry. 2014;75:6.

4. Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. Epidemiology. 2010;21(3):383-8.

5. Shi X, Miao W, Tchetgen ET. A Selective Review of Negative Control Methods in Epidemiology. Curr Epidemiol Rep. 2020;7(4):190-202.

6. Catala M, Guo Y, Du M. CohortCharacteristics: Summarise and Visualise Characteristics of Patients in the OMOP CDM 2024 [

7. Catala M, Du M, Guo Y, Lopez-Guell K, Burn E, Li X. DrugUtilisation: Summarise Patient-Level Drug Utilisation in Data Mapped to the OMOP Common Data Model [Available from: https://cran.r-project.org/web/packages/DrugUtilisation/index.html.

11. ANNEXES

ANNEX I. Description of data sources

DATA SOURCES DESCRIPTIONS

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 Analytics Centre (DAC)
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 v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). 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 English https://sundhedsdatastyrelsen.dk/da/english/health_data_and_registers/national_health_registers
10	Linkage	There is no linkage in this data source.

#	Section	Description
11	Vital status	The Cause of Death registry (DAR) is used, the cause of death is collected using ICD-10 codes.
12	Limitations	DK-DHR has the following limitations, which may be relevant confounders for certain complex Darwin EU studies: - We lack information on key socio-economic status (SES) factors, such as occupation, education, and income. These variables may be important for analysis in some studies. - We only have complete data on lifestyle factors (such as smoking status and weight) for pregnant women. - We have no information on patient contacts in primary care (visits to the GP). Consequently, the incidence of chronic diseases like Type 2 Diabetes (T2D) and asthma must be determined using drug prescriptions as a proxy. Stillborn children will not have any records in our CDM. This means that e.g. birth length of stillborns is not recorded.
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

Finnish Care Register for Health Care (FinOMOP-THL)

#	Section	Description
1	Data source identification and country	FinOMOP-THL (Finnish Care Register for Health Care) Finland
2	Data partner information section	Finnish Institute for Health and Welfare (THL) The Department of Data and Analytics
3	Coverage and timespan	Data collection since: 1998 Extent: Nation-wide. The current CDM population comprises all persons having been alive and residing in Finland since the beginning of 2011.
4	Healthcare setting / type of data	Primary care – General Practitioner, and primary care specialists (e.g. paediatricians), and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care, and other (specify). THL maintains health registers that cover both public and private, primary, and specialised inpatient, urgent and outpatient health care encounters in Finland, starting from 2011. The entire public sector and private inpatient encounters have been included since 2011, while private outpatient encounters, including occupational care, are included since 2020. Since 1998, the Care Register for Health Care (TerveysHilmo) has covered both public outpatient and inpatient specialized care and private inpatient care. Since 2009, the Finnish National Vaccination Register has covered all vaccinations from the public sector and from a large part of private vaccination providers, with the data coverage from both sections being very good to complete from 2020 onwards. Since 2011, the Register of Primary Health Care Visits (AvoHilmo) has covered public primary care. Since 2020, the register has also covered private outpatient care and occupational care. In addition, the CDM also contains positive COVID-19 test results from the Finnish National Infectious Diseases Register. The register itself covers all laboratory confirmations for around 70 specific microbes from 1995 onwards, but only COVID-19 has currently been mapped to the CDM. The CDM also includes prescription records from multiple different sources. Both Care Register for Health Care and Register of Primary Health Care Visits contain very basic prescription data recorded during health care encounters that include just the ATC-code and trade name of medication. More comprehensive prescription data from Kanta Prescription Centre, maintained by Social Insurance Institution of Finland (Kela), has been integrated into the CDM since its 2024

#	Section	Description
		release. The Kanta Prescription records are based electronic prescriptions, which were adopted by most public health care providers in 2010 and by most private providers by 2017.
5	Data collection process	Outpatient electronic health records, and Inpatient hospital electronic health records, and Registries. Data is entered by clinicians upon healthcare contact and processed by THL (Kela in the case of Kanta Prescription Centre).
6	General representativeness	The THL data has national coverage and is therefore well representative of the Finnish population. Using the complete population as a basis for the person table also serves to facilitate calculations on a population level, e.g. incidence rates.
7	Data content /source coding	The following coding systems have been OMOP-mapped, typically to a good level of completeness: ICD10fi Finnish Extension, ICPC-2, ATC, Toimenpideluokitus (procedure classification adapted from the Nordic Classification of Surgical Procedures (NCSP)), Terveydenhuollon erikoisalat (Hilmo specific provider speciality), Rokotustapa (AR/YDIN National classification for vaccine administration), Tupakointistatus (AR/YDIN National classification for smoking status). Vaccinations are identified on product level based on batch number, trade name, vaccine title, and ATC-code. This is mapped on brand and type in the OMOP CDM.
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. Each patient in THL has a unique identifier.
9	Quality control (data source specific)	The source data collection undergoes a structural and semantic validation before entry into the source database. Additionally, some coded variables undergo quality assessment against the respective code systems post entry into the database. The source registers are also assessed for completeness and coverage, with the aim of improving future collection in the areas where data is lacking.
10	Linkage	THL is already a linkage of multiple Finnish registries (see above).
11	Vital status	The National Population registry data forms the basis for forming the patient population. This ensures an up-to-date location (municipality of residence) of patients, as well as complete death occurrences (although not the cause of death).
12	Limitations	All drug records in the CDM are currently based on prescriptions. Kanta Prescription Centre also includes information on drug dispensings, but these have not currently been converted into OMOP CDM. Depending on the type of the medication, a single prescription can be for up to two years dosage of the drug. This means a patient can have up to two-year breaks between observations while actively using the drug. Observation of a prescription also does not mean that the patient necessarily bought or used the medication. The CDM does not currently have meaningful end dates or days of supply for drug exposure. This information is not available for Care Register for Health Care or Register of Primary Health Care Visits, and for Kanta Prescription Centre it is only available in unstructured, free-form format that has not been converted into OMOP CDM. The source system for prescription drug data is only available to the DP until Dec-2023, therefore drug_exposure records from this source are only available in that time range. Not all private health care records are covered for the CDM's entire follow-up time from 2011 onwards. For Register of Primary Health Care Visits and Finnish National Vaccination Register, records from private health care have been available from 2020 onwards. For Kanta Prescription Centre, the coverage of private health care records has been good from 2017 onwards. The inclusion of private health care mainly presents itself as an increase in the number of observations, meaning that it has to be accounted for when interpreting any time series data from the CDM.
13	Main references	Häkkinen, Pirjo; Mölläri, Kaisa; Saukkonen, Sanna-Mari; Väyrynen, Riikka; Mielikäinen, Lasse; Järvelin, Jutta "Hilmo - Sosiaali- ja terveydenhuollon hoitoilmoitus 2020 : Määrittelyt ja ohjeistus : Voimassa 1.1.2020 alkaen" Terveyden ja hyvinvoinnin laitos (2019):

#	Section	Description
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111187;https://catalogues.ema.europa.eu/data-source/1111191 Website: https://thl.fi/en/statistics-and-data/data-and-services/register-descriptions;https://www.kanta.fi/en/research-and-knowledge-management

Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público (BIFAP)

#	Section	Description
1	Data source identification and country	BIFAP (Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público (Pharmacoepidemiological Research Database for Public Health Systems)) Spain
2	Data partner information section	AEMPS Pharmacoepidemiology and Pharmacovigilance Division - Medicines for human use Department
3	Coverage and timespan	Data collection since: 2001 Extent: Regional. Spanish National Health Service (SNS) from 9 of the 17 regions in Spain. The population currently included represents 36% of the total Spanish population.
4	Healthcare setting / type of data	Primary care – General Practitioner, and community pharmacists, and primary care specialists (e.g. paediatricians), and hospital inpatient care. BIFAP includes a collection of databases linked at individual patient level. The main one is the Primary care Database, given the central role of PCPs in the SNS. Linked, there are additional important structural databases, like the medicines dispensed at community pharmacies and the patients' hospital diagnosis at discharge. 7 out of the 9 regions have linkage to hospital data. However, hospital data is available for different time periods for each region. From 2014 onwards, linkage to hospital data is available for >68% of patients.
5	Data collection process	Insurance/administrative claims, and Outpatient electronic health records, and Inpatient hospital electronic health records, and Registries. Data in BIFAP is collected from Primary Care and Hospital EHR.
6	General representativeness	Spain has a SNS that provides universal access to health services through the Regional Healthcare Services. Primary care physicians (PCPs), both general practitioners and paediatricians, have a central role. They act as gatekeepers of the system and exchange information with other levels of care to ensure the continuity of care. Most of the population (98.9%) is registered with a PCP and, in addition, most drug prescriptions are written at the primary care level.
7	Data content /source coding	The BIFAP source data is coded in SNOMED, ICD, ICPC-2 (diagnoses), AEMPS (drugs), and local lab codes.
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. Pseudonymized ID numbers are generated at regional level. The Personal Identification Code for the Autonomous Community (CIPA) is used to perform the pseudonymisation procedure. Therefore, upon changing practice or de- and re-registration within the same region, (Autonomous Community) the patient in BIFAP is correctly identified as the same person with the same ID number. However, the same patient would obtain different ID numbers if the patient moves to a different region and is registered in a primary care practice in the new region. The percentage of people who are de-registered due to moving to other region in relation to the BIFAP population is, for example, 5% in Madrid and 4% Castilla y Leon. This situation would have a very limited impact on the data analysis due to the following: - The proportion is low (less than 5%) in relation to the overall population in BIFAP. - In BIFAP, only stable residents are included. This means that patients living in another region for a foreseen short time period and are provisionally assigned to a primary care practice are not

#	Section	Description
		included in BIFAP. - Medical events of those patients who have more than one ID do not overlap in time, since dates of events correspond to different periods. This means that counts of these events are never duplicated. - A number of study designs allows the same patient to be part of different cohorts or to be selected both as case and as control, provided that their person-time experience correspond to a different period of time. In all these cases, the impact in study analysis of duplicated IDs would be negligible.
9	Quality control (data source specific)	Patients who meet any of the following disability criteria are discarded: - Non-owners of the individual health card - Date of birth before 01/01/1801 - Active patients over 115 years of age - Patients without clinical records (only contains administrative information) - Patients marked as “fictitious” in the clinical history - Badly coded sex - Inactive without termination date - Start date = End date - Clinical records prior to date of birth
10	Linkage	The following data are also linked at individual patient level and available. For a subset of the BIFAP population (regions and/or periods of time): • Information on dispensation of medicines at hospital pharmacies from outpatients and inpatients. • Registration of Causes of Death by the National Institute for Statistics. From the start of the COVID-19 pandemic: • Vaccines COVID-19 Administration Registry linked to patients included in BIFAP. • Diagnosis Tests of COVID-19 linked to patients included in BIFAP, for some regions.
11	Vital status	Source for vital status unknown.
12	Limitations	Primary care is available from 2001 but is considered complete since 2005. Hospital discharge has different coverage periods per region Spain, with most starting between 2014–2016. This means that for different regions and different time periods there is a different coverage of healthcare events. In the release of July 2025, the laboratory results are not covered. These will be added again at the next release, expected at the end of 2025.
13	Main references	Maciá-Martínez MA, Gil M,Huerta C,Martín-Merino E,Álvarez A,Bryant V,Montero D "Base de Datos para la Investigación Farmacoepidemiológica en Atención Primaria (BIFAP): A data resource for pharmacoepidemiology in Spain." Pharmacoepidemiology and drug safety (2020): 32337840
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/21501 Website: http://www.bifap.org/index_EN.html

The Information System for the Development of Research in Primary Care (SIDIAP)

#	Section	Description
1	Data source identification and country	SIDIAP (The Information System for the Development of Research in Primary Care) Catalunya, Spain
2	Data partner information section	IDIAPJGol
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,

#	Section	Description
		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 v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). 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 women'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

Health Impact - Swedish Population Evidence Enabling Data-linkage (HI-SPEED)

#	Section	Description
1	Data source identification and country	HI-SPEED (Health Impact - Swedish Population Evidence Enabling Data-linkage) Sweden
2	Data partner information section	SMPA-GU, Läkemedelsverket, Box 26, 751 03 Uppsala, Sweden - Box 469, 405 30 Gothenburg, Sweden Pharmacoepidemiology and Analysis Department (FeA) - School of Public Health and Community Medicine, Institute of Medicine
3	Coverage and timespan	Data collection since: 2015 Extent: Nation-wide. The catchment area includes the whole of Sweden, covering the full population of approximately 11.7 million.
4	Healthcare setting / type of data	Primary care – General Practitioner, and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. Primary care (GPs) is available only for the 2 largest regions (~40% of national population) The following data elements are collected: Socio-demographics, dispensed drug prescriptions, cause of death, diagnoses and procedures from secondary (specialist) care and inpatient visits or clinical events, as well as from primary care visits (40%pop only).
5	Data collection process	Registries. The data is acquired from the relevant Swedish national and regional registries, only once all legislative, GDPR and ethical approvals have been granted. Therefore, only relevant data is passed on, which will then be entered and processed by the study team. The data are updated several times annually.
6	General representativeness	The coverage includes all patients of all sociodemographic characteristics. Therefore, it should mirror the source population to a very good extent.
7	Data content /source coding	Medicines are coded with ATC and NPLID (National Product ID), ICD10-SE is used for diagnoses, and the Swedish procedure coding system (KVA) is used for clinical procedures.
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. Patients have a uniquely id across datasets.
9	Quality control (data source specific)	The source data are obtained from the relevant Swedish National and Regional Registers. The registers perform some regular quality controls on their data. After receiving the data, we perform additional checks and cleaning. We also run regular quality checks on the data we manage.
10	Linkage	Data on specialist care is acquired from the National Patient Register; mortality information is provided by the Cause-Of-Death Registry. Drug data is provided by the Patient Drug Register. And similarly for other registers. Data are linked very accurately using the national personal ID number and pseudonymized before delivery to HI-SPEED. All data are updated 2-4 times per year.

#	Section	Description
11	Vital status	Data on death and underlying + contributing causes-of-death are extracted from the Cause-of-Death registry (i.e. based on death certificates).
12	Limitations	General limitations for the data type applicable. This is a research project where all studies require ethics approval. Data collection since: 2015 for most data, except prescribed drug register (from 2018), and some COVID-related data (tests, vaccination) from 2020 Primary care is only available for a subset.
13	Main references	Nyberg F, Franzén S, Lindh M, Vanfleteren L, Hammar N, Wettermark B, Sundström J, Santosa A, Björck S, Gisslén M "Swedish Covid-19 Investigation for Future Insights - A Population Epidemiology Approach Using Register Linkage (SCIFI-PEARL)." Clinical epidemiology (2021): 34354377
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/node/4463/ Website: https://www.gu.se/en/research/scifi-pearl

Clinical Practice Research Datalink GOLD (CPRD GOLD)

#	Section	Description
1	Data source identification and country	CPRD GOLD (Clinical Practice Research Datalink GOLD) The United Kingdom
2	Data partner information section	University of Oxford NDORMS
3	Coverage and timespan	Data collection since: 1987 Extent: Nation-wide. CPRD GOLD consists of patients in contributing practices using Vision software. Historically this covered the whole of the UK, but the number of contributing practices in the England is dropping. In January 2025 only 3 practices from England were a part of CPRD GOLD, while historical patient data were from the whole of the UK, and will continue to be so. In the future, no practices from England will be present, only practices from Scotland, Wales, and Northern Ireland.
4	Healthcare setting / type of data	Primary care – General Practitioner, and primary care specialists (e.g. paediatricians), and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. CPRD GOLD data include patient demographics, biological measurements, clinical symptoms and diagnoses, referrals to specialist/hospital and their outcome, laboratory tests/results, and prescribed medications.
5	Data collection process	Outpatient electronic health records. Data are entered by clinicians into the EHR. Data is processed by CPRD that provides data releases for research.
6	General representativeness	In the last 10 years, the CPRD GOLD regional distribution of currently contributing general practitioner (GP) practices has significantly shifted, resulting in many new practices joining from Scotland, Wales, and Northern Ireland, and fewer participating from England. These changes have affected the CPRD GOLD population size, regional coverage, and eligibility for data linkages. CPRD GOLD January 2024 contains >21.3 million historical and current patients (12.9 in England, 3.1 in Wales, 4.7 in Scotland, 0.7 in Northern Ireland). Of these, nearly 3 million are currently registered in a GP practice and represent ~4.3% of the estimated current UK population (0.1% in England, 32.3% in Wales, 28.6% in Scotland, 16.2% in Northern Ireland). Patients currently registered in CPRD GOLD January 2024 are broadly representative of the UK

#	Section	Description
		population with respect to age and sex. Reference: https://doi.org/10.1093/ije/dyaf077
7	Data content /source coding	Gemscript, Read, dm+d
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. In GOLD, a patient can be registered under different ID numbers upon changing practice or re-registration. Researchers are not able to identify these patients, as the data are anonymised. However, GOLD covers less than 5% of the current UK GP practices and it is unlikely that an individual who does change GP practice ends up in another GP practice which uses the Vision software and accepts the CPRD data collection agreement. The very small number of duplicated IDs will have different observation periods and should not have an impact on the data analyses.
9	Quality control (data source specific)	CPRD GOLD only includes practices whose data quality is assessed to be up-to-standard (UTS). Each practice is associated to an UTS date set when the data quality standards become satisfactory, and CPRD recommend using only longitudinal data starting from this UTS date. Every time CPRD collect the EHR from a practice, checks are run for the data quality standards, and if they are not adequate, the EHR is not accepted. When the data quality becomes acceptable again, CPRD updates the practice UTS date. CPRD also checks data quality standards at the patient level, and associates each patient with a flag, reporting if its data are acceptable for clinical research. Only patients with acceptable data quality are included in the population to be mapped to CDM.
10	Linkage	CPRD GOLD can be linked to several sources, however our Oxford OMOP CDM is only linked to the CPRD GOLD Ethnicity Record and to the CPRD Townsend Deprivation Index at the Practice Level.
11	Vital status	The date of death in CPRD GOLD has been validated against the Population registry (ONS) mortality data. Reference: https://doi.org/10.1002/pds.4747
12	Limitations	The main limitation is due to the fact that CPRD GOLD is limited to GP records, and although it contains information on referrals and discharge letters, it may not fully capture specific hospital information. Events from hospital and specialist care are not covered.
13	Main references	Sanchez-Santos MT, Axson EL,Dedman D,Delmestri A "Data Resource Profile Update: CPRD GOLD." International journal of epidemiology (2025): 40499193
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111113 Website: https://www.cprd.com/data/primary-care-data/cprd-gold

ANNEX II. Fitness for use assessment

Data source justification for inclusion and key characteristics

Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público (BIFAP)

BIFAP will be included in this study because it covers both the primary care and secondary care settings. For this study we will not use hospital linkage. The data source provides relevant information on venlafaxine and mirtazapine prescriptions in the general adult population and covers the settings in which heart failure and cardiomyopathies are diagnosed and treated.

Based on a preliminary feasibility assessment, the expected number of person counts for venlafaxine and mirtazapine prescriptions are 229,200 and 1,088,000, respectively. 936,600 and 144300 people have a record of heart failure and/or cardiomyopathy.

Moreover, data availability and follow-up in BIFAP is currently available from the start of the study period (2010) until the date of the most recent data extraction, which is 06/2025.

The study will need IRB approval, which is estimated to take around 1–3months. This makes the execution of this study feasible, assuming we receive approvals within the current study timelines.

Danish Data Health Registries (DK-DHR)

DK-DHR, the Danish National Linked Health registries, will be included in this study because it covers both the primary care (proxy for this using prescription retrievals) and secondary care settings, including hospital data. The data source provides relevant information on venlafaxine and mirtazapine prescriptions in the general adult population and covers the settings in which heart failure and cardiomyopathies are diagnosed and treated.

Based on a preliminary feasibility assessment, the expected number of person counts for venlafaxine and mirtazapine prescriptions are 265,300 and 594,300, respectively. 415,400 and 48,000 people have a record of heart failure and/or cardiomyopathy.

Moreover, data availability and follow-up in DK-DHR are currently available from the start of the study period (2010) until the date of the most recent data extraction, which is 04/2025.

The study will be covered by blanket IRB approvals for DARWIN EU® pharmacoepidemiology studies.

Finnish Care Register for Health Care (FinOMOP-THL)

FinOMOP-THL will be included in this study because it covers both the primary care and secondary care settings. The data source provides relevant information on venlafaxine and mirtazapine prescriptions in the general adult population and covers the settings in which heart failure and cardiomyopathies are diagnosed and treated.

Based on a preliminary feasibility assessment, the expected number of person counts for venlafaxine and mirtazapine prescriptions are 239,900 and 851,600, respectively. 350,500 and 80,400 people have a record of heart failure and/or cardiomyopathy.

Moreover, data availability and follow-up in FinOMOP-THL are currently available from the start of the study period (2010) until the date of the most recent data extraction, which is 10/2024. As prescription data will only be available until 2023, the study period will be adapted accordingly for FinOMOP-THL.

The study will need IRB approval, which is estimated to take around 1–3months. This makes the execution of this study feasible, assuming we receive approvals within the current study timelines.

The Information System for the Development of Research in Primary Care (SIDIAP)

SIDIAP will be included in this study because it covers both the primary care and secondary care setting in Catalonia, Spain. Linkage to hospital data is available. The data source provides relevant information on venlafaxine and mirtazapine prescriptions in the general adult population and covers the settings in which heart failure and cardiomyopathies are diagnosed and treated.

Based on a preliminary feasibility assessment, the expected number of person counts for venlafaxine and mirtazapine prescriptions are 193,100 and 266,900, respectively. 381,100 and 79,300 people have a record of heart failure and/or cardiomyopathy.

Moreover, data availability and follow-up in SIDIAP are currently available from the start of the study period (2010) until the date of the most recent data extraction, which is end of 2024.

The study will need IRB approval, which is estimated to take around 2 months. This makes the execution of this study feasible within the current study timelines.

Health Impact - Swedish Population Evidence Enabling Data-linkage (HI-SPEED)

HI-SPEED will be included in this study because it is based on the Swedish National and Regional Health Registries, covering both the primary care and secondary care setting. The data source provides relevant information on venlafaxine and mirtazapine prescriptions in the general adult population and covers the settings in which heart failure and cardiomyopathies are diagnosed and treated.

Based on a preliminary feasibility assessment, the expected number of person counts for venlafaxine and mirtazapine prescriptions are 184,600 and 706,100, respectively. 492,400 and 48,100 people have a record of heart failure and/or cardiomyopathy.

Moreover, data availability and follow-up in HI-SPEED is sufficient, as data availability starts in 2015, and the date of most recent data extraction is 2025.

Clinical Practice Research Datalink GOLD (CPRD GOLD)

CPRD GOLD will be included in this study because it is a primary care data source that provides relevant information on venlafaxine and mirtazapine prescriptions in the general adult population and covers the settings in which heart failure and cardiomyopathies are diagnosed and treated.

Based on a preliminary feasibility assessment, the expected number of person counts for venlafaxine and mirtazapine prescriptions are 303,200 and 620,400, respectively. 252,200 and 23,100 people have a record of heart failure and/or cardiomyopathy.

Moreover, data availability and follow-up in CPRD GOLD are sufficient, as is availability to cover the study period from 2010 to the date of the most recent data extraction in 06/2025.

The study will need IRB approval, which is estimated to take around 1–2months. This makes the execution of this study feasible within the current study timelines.

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 standardized 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 drug cohorts, non-systemic products will be excluded from the list of included codes summarised on the ingredient level.

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) and *DrugExposureDiagnostics* (https://cran.r-project.org/web/packages/DrugExposureDiagnostics/index.html) R packages will be run, if needed, 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 packages: *IncidencePrevalence* to estimate Incidence and Prevalence, *DrugUtilisation* to characterise the drug use, and *CohortCharacteristics* to characterise the cohort by indication. These packages will include numerous automated unit tests to ensure the validity of

the codes, alongside software peer review and user testing. The R package will be made publicly available via GitHub.

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.

We aim to publish the study results in a peer-reviewed scientific journal, and present findings at international conferences.

ANNEX IV. List of stand-alone documents

Table S1. Preliminary list of conditions definitions (cardiomyopathy, heart failure).

Phenotype	concept_id	Concept Name	domain	Vocabulary	wash out
cardiomyopathy (broad)	4321717	Fatty degeneration of heart	Condition	SNOMED	Inf
cardiomyopathy (broad)	37398931	Tropical endomyocardial fibrosis	Condition	SNOMED	Inf
cardiomyopathy (broad)	36714540	Vici syndrome	Condition	SNOMED	Inf
cardiomyopathy (broad)	43020781	Fetal cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	44804845	Generalised right ventricular dysfunction	Condition	SNOMED	Inf
cardiomyopathy (broad)	36675150	Mitochondrial hypertrophic cardiomyopathy with lactic acidosis due to MTO1 deficiency	Condition	SNOMED	Inf
cardiomyopathy (broad)	4239975	Myocardial disease	Condition	SNOMED	Inf
cardiomyopathy (broad)	36675174	Hypertrophic cardiomyopathy and renal tubular disease due to mitochondrial DNA mutation	Condition	SNOMED	Inf
cardiomyopathy (broad)	36712757	Cardiomyopathy due to hypertension	Condition	SNOMED	Inf
cardiomyopathy (broad)	37110890	Cirrhotic cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	36712927	Congestive heart failure stage C due to ischemic cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	37312019	Diabetic cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	36675178	Infantile hypertrophic cardiomyopathy due to MRPL44 deficiency	Condition	SNOMED	Inf
cardiomyopathy (broad)	37396000	Transthyretin related familial amyloid cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	36712928	Congestive heart failure stage B due to ischemic cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	36712929	Systolic heart failure stage C due to ischemic cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	36713473	Congenital cataract, hypertrophic cardiomyopathy, mitochondrial myopathy syndrome	Condition	SNOMED	Inf
cardiomyopathy (broad)	37312596	Rheumatic cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	36713761	Hypertrophic cardiomyopathy with hypotonia and lactic acidosis syndrome	Condition	SNOMED	Inf
cardiomyopathy (broad)	40483296	Hypertrophic cardiomyopathy associated with another disorder	Condition	SNOMED	Inf

Phenotype	concept_id	Concept Name	domain	Vocabulary	wash out
cardiomyopathy (broad)	36713480	Glycogen storage disease with severe cardiomyopathy due to glycogenin deficiency	Condition	SNOMED	Inf
cardiomyopathy (broad)	36714157	Progressive sensorineural hearing loss and hypertrophic cardiomyopathy syndrome	Condition	SNOMED	Inf
cardiomyopathy (broad)	40479589	Takotsubo cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	36714240	Microcephalus cardiomyopathy syndrome	Condition	SNOMED	Inf
cardiomyopathy (broad)	36714549	Wooly hair and palmoplantar keratoderma with dilated cardiomyopathy syndrome	Condition	SNOMED	Inf
cardiomyopathy (broad)	36715121	Cardiomyopathy with cataract and hip spine disease syndrome	Condition	SNOMED	Inf
cardiomyopathy (broad)	40487104	Right ventricular myocardial noncompaction cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	40487105	Left ventricular myocardial noncompaction cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	36715122	Cardiomyopathy and renal anomaly syndrome	Condition	SNOMED	Inf
cardiomyopathy (broad)	36717359	Systolic heart failure stage B due to ischemic cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	43020652	Inflammatory cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	43020654	Valvular cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	43020658	Hypertrophic cardiomyopathy with genetic marker	Condition	SNOMED	Inf
cardiomyopathy (broad)	36717720	Dilated cardiomyopathy with hypergonadotropic hypogonadism syndrome	Condition	SNOMED	Inf
cardiomyopathy (broad)	43020656	Pacing-induced cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	43020659	Dilated cardiomyopathy with genetic marker	Condition	SNOMED	Inf
cardiomyopathy (broad)	43020660	Ischemic dilated cardiomyopathy due to coronary artery disease	Condition	SNOMED	Inf
cardiomyopathy (broad)	36717761	Tubular renal disease with cardiomyopathy syndrome	Condition	SNOMED	Inf
cardiomyopathy (broad)	43020668	Hypertrophic mitochondrial cardiomyopathy associated with cataracts and lactic acidosis	Condition	SNOMED	Inf
cardiomyopathy (broad)	43020666	Mitochondrial cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	43020667	Histiocytoid mitochondrial cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	37017276	Cardiomyopathy co-occurrent with human immunodeficiency virus infection	Condition	SNOMED	Inf

Phenotype	concept_id	Concept Name	domain	Vocabulary	wash out
cardiomyopathy (broad)	43021903	Ventricular myocardial noncompaction cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	43020669	Fatal infantile mitochondrial cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	37109990	Sensorineural deafness with dilated cardiomyopathy syndrome	Condition	SNOMED	Inf
cardiomyopathy (broad)	43021299	Histiocytoid mitochondrial cardiomyopathy due to cytochrome aa3 deficiency	Condition	SNOMED	Inf
cardiomyopathy (broad)	43021902	Cardiomyopathy due to neuromuscular disorder	Condition	SNOMED	Inf
cardiomyopathy (broad)	43021905	Hypertrophic mitochondrial cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	37119055	Maternally inherited cardiomyopathy and hearing loss syndrome	Condition	SNOMED	Inf
cardiomyopathy (broad)	44784145	Primary hypertrophic cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	43021906	Maternally inherited mitochondrial cardiomyopathy and myopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	45757557	Cardiomyopathy due to viral infection	Condition	SNOMED	Inf
cardiomyopathy (broad)	44782428	Congestive heart failure due to cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	44783568	Secondary nonischemic congestive cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	44807892	Amyloid cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	45765475	Dilated cardiomyopathy 3B	Condition	SNOMED	Inf
cardiomyopathy (broad)	45765411	Early onset myopathy with fatal cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	45766963	Fetal hypertrophic cardiomyopathy due to maternal diabetes mellitus	Condition	SNOMED	Inf
cardiomyopathy (broad)	45766167	Heart failure with reduced ejection fraction due to cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	43020782	Fetal hypertrophic cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	45771323	Mitral valve regurgitation due to cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	46269894	Cardiomyopathy due to diphtheria	Condition	SNOMED	Inf
cardiomyopathy (broad)	43020783	Fetal hypertrophic cardiomyopathy due to twin to twin transfusion syndrome	Condition	SNOMED	Inf
cardiomyopathy (broad)	43020785	Fetal dilated cardiomyopathy	Condition	SNOMED	Inf

Phenotype	concept_id	Concept Name	domain	Vocabulary	wash out
cardiomyopathy (broad)	312383	Postpartum cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	315560	Primary endomyocardial fibrosis cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	316428	Hypertrophic obstructive cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	318773	Dilated cardiomyopathy secondary to alcohol	Condition	SNOMED	Inf
cardiomyopathy (broad)	4037495	Dilated peripartum cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	320746	Cardiomyopathy associated with another disorder	Condition	SNOMED	Inf
cardiomyopathy (broad)	4097992	Danish type familial amyloid cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	321319	Cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	4103513	Dilated cardiomyopathy due to muscular dystrophy	Condition	SNOMED	Inf
cardiomyopathy (broad)	4101863	Dilated cardiomyopathy due to myotonic dystrophy	Condition	SNOMED	Inf
cardiomyopathy (broad)	4108236	Hypertrophic cardiomyopathy without obstruction	Condition	SNOMED	Inf
cardiomyopathy (broad)	4006792	Dilated cardiomyopathy due to phytanic acid storage disease	Condition	SNOMED	Inf
cardiomyopathy (broad)	4108822	Primary dilated cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	4108237	Cardiomyopathy due to mucopolysaccharidosis	Condition	SNOMED	Inf
cardiomyopathy (broad)	4048268	Restrictive cardiomyopathy secondary to infiltrations	Condition	SNOMED	Inf
cardiomyopathy (broad)	4108238	Cardiomyopathy in myotonic dystrophy	Condition	SNOMED	Inf
cardiomyopathy (broad)	4108239	Dystrophic cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	4023964	Dilated cardiomyopathy secondary to toxic reaction	Condition	SNOMED	Inf
cardiomyopathy (broad)	4119592	Restrictive cardiomyopathy without endomyocardial fibrosis	Condition	SNOMED	Inf
cardiomyopathy (broad)	4111569	Cardiomyopathy in Friedreich's ataxia	Condition	SNOMED	Inf
cardiomyopathy (broad)	4006292	Dilated cardiomyopathy due to metabolic disorder	Condition	SNOMED	Inf
cardiomyopathy (broad)	4119959	Familial restrictive cardiomyopathy	Condition	SNOMED	Inf

Phenotype	concept_id	Concept Name	domain	Vocabulary	wash out
cardiomyopathy (broad)	4124692	Congestive obstructive cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	4124693	Hypertrophic cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	4060405	Dilated cardiomyopathy due to rheumatoid arthritis	Condition	SNOMED	Inf
cardiomyopathy (broad)	4129399	Dilated cardiomyopathy secondary to radiation	Condition	SNOMED	Inf
cardiomyopathy (broad)	4139754	Restrictive cardiomyopathy secondary to endocardial fibroelastosis	Condition	SNOMED	Inf
cardiomyopathy (broad)	4068262	Dilated cardiomyopathy secondary to metazoal myocarditis	Condition	SNOMED	Inf
cardiomyopathy (broad)	4141588	Primary eosinophilic endomyocardial restrictive cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	4142185	Dilated cardiomyopathy secondary to deficiency	Condition	SNOMED	Inf
cardiomyopathy (broad)	4071896	Secondary dilated cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	4145696	Ischemic congestive cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	4144463	Tachycardia-induced cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	4076799	Restrictive cardiomyopathy secondary to malignancy	Condition	SNOMED	Inf
cardiomyopathy (broad)	4147604	Dilated cardiomyopathy due to systemic sclerosis	Condition	SNOMED	Inf
cardiomyopathy (broad)	4147982	Dilated cardiomyopathy due to viral myocarditis	Condition	SNOMED	Inf
cardiomyopathy (broad)	4094188	Arrhythmogenic right ventricular cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	4149725	Restrictive cardiomyopathy secondary to familial storage disease	Condition	SNOMED	Inf
cardiomyopathy (broad)	4154093	Cardiomyopathy in Duchenne muscular dystrophy	Condition	SNOMED	Inf
cardiomyopathy (broad)	4172641	Primary endomyocardial fibrosis restrictive cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	4162674	Restrictive cardiomyopathy with endomyocardial fibrosis	Condition	SNOMED	Inf
cardiomyopathy (broad)	4163710	Dilated cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	4177187	Dilated cardiomyopathy due to Friedreich's ataxia	Condition	SNOMED	Inf
cardiomyopathy (broad)	4240306	Dilated cardiomyopathy due to polyarteritis nodosa	Condition	SNOMED	Inf

Phenotype	concept_id	Concept Name	domain	Vocabulary	wash out
cardiomyopathy (broad)	4177188	Cobalt cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	4178584	Primary familial dilated cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	4185450	Dilated cardiomyopathy with connective tissue disorder	Condition	SNOMED	Inf
cardiomyopathy (broad)	4243983	Dilated cardiomyopathy secondary to glycogen storage disease	Condition	SNOMED	Inf
cardiomyopathy (broad)	4189147	Dilated cardiomyopathy due to taurine deficiency	Condition	SNOMED	Inf
cardiomyopathy (broad)	4245506	Dilated cardiomyopathy due to familial storage disease	Condition	SNOMED	Inf
cardiomyopathy (broad)	4185379	Dilated cardiomyopathy due to bacterial myocarditis	Condition	SNOMED	Inf
cardiomyopathy (broad)	4246673	Hypertrophic cardiomyopathy secondary to hyperthyroidism	Condition	SNOMED	Inf
cardiomyopathy (broad)	4247261	Dilated cardiomyopathy due to sarcoidosis	Condition	SNOMED	Inf
cardiomyopathy (broad)	4190773	Restrictive cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	4191606	Restrictive cardiomyopathy secondary to sarcoidosis	Condition	SNOMED	Inf
cardiomyopathy (broad)	4195417	Dilated cardiomyopathy due to neuromuscular disorder	Condition	SNOMED	Inf
cardiomyopathy (broad)	4251051	Dilated cardiomyopathy due to malignancy	Condition	SNOMED	Inf
cardiomyopathy (broad)	4197087	Restrictive cardiomyopathy secondary to amyloidosis	Condition	SNOMED	Inf
cardiomyopathy (broad)	4203149	Primary idiopathic dilated cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	4258681	Dilated cardiomyopathy due to fungal myocarditis	Condition	SNOMED	Inf
cardiomyopathy (broad)	4262911	Dilated cardiomyopathy due to dermatomyositis	Condition	SNOMED	Inf
cardiomyopathy (broad)	4204856	Primary eosinophilic endomyocardial cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	4261970	Familial cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	4269448	Dilated cardiomyopathy due to systemic lupus erythematosus	Condition	SNOMED	Inf
cardiomyopathy (broad)	4205593	Hypertrophic cardiomyopathy secondary to Friedreich's ataxia	Condition	SNOMED	Inf
cardiomyopathy (broad)	4270625	Primary idiopathic hypertrophic cardiomyopathy	Condition	SNOMED	Inf

Phenotype	concept_id	Concept Name	domain	Vocabulary	wash out
cardiomyopathy (broad)	4273369	Restrictive cardiomyopathy secondary to glycogen storage disease	Condition	SNOMED	Inf
cardiomyopathy (broad)	4215284	Dilated cardiomyopathy due to granuloma	Condition	SNOMED	Inf
cardiomyopathy (broad)	4273401	Dilated cardiomyopathy due to infiltration	Condition	SNOMED	Inf
cardiomyopathy (broad)	4217551	Secondary restrictive cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	4274635	Primary idiopathic restrictive cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	4218771	Dilated cardiomyopathy secondary to drug	Condition	SNOMED	Inf
cardiomyopathy (broad)	4219376	Dilated cardiomyopathy due to hemochromatosis	Condition	SNOMED	Inf
cardiomyopathy (broad)	4288868	Nonobstructive cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	4313305	Dilated cardiomyopathy due to infectious disease	Condition	SNOMED	Inf
cardiomyopathy (broad)	4222765	Primary familial hypertrophic cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	4314400	Dilated cardiomyopathy due to mucopolysaccharidosis	Condition	SNOMED	Inf
cardiomyopathy (broad)	4223017	Hypertrophic cardiomyopathy secondary to neuromuscular disorder	Condition	SNOMED	Inf
cardiomyopathy (broad)	35621979	Encephalopathy, hypertrophic cardiomyopathy, renal tubular disease syndrome	Condition	SNOMED	Inf
cardiomyopathy (broad)	4345564	Benign scapuloperoneal muscular dystrophy with cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	4224935	Dilated cardiomyopathy due to nutritional deficiency	Condition	SNOMED	Inf
cardiomyopathy (broad)	4225077	Dilated cardiomyopathy due to protozoan myocarditis	Condition	SNOMED	Inf
cardiomyopathy (broad)	4345565	Severe scapuloperoneal muscular dystrophy with cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	35615148	Cardiomyopathy caused by drug	Condition	SNOMED	Inf
cardiomyopathy (broad)	4225117	Restrictive cardiomyopathy due to mucopolysaccharidosis	Condition	SNOMED	Inf
cardiomyopathy (broad)	35624231	Familial dilated cardiomyopathy with conduction defect due to LMNA mutation	Condition	SNOMED	Inf
cardiomyopathy (broad)	4101624	Arrhythmogenic right ventricular dysplasia	Condition	SNOMED	Inf
cardiomyopathy (broad)	4234537	Dilated cardiomyopathy secondary to electrolyte deficiency	Condition	SNOMED	Inf

Phenotype	concept_id	Concept Name	domain	Vocabulary	wash out
cardiomyopathy (broad)	320122	Nutritional and metabolic cardiomyopathies	Condition	SNOMED	Inf
cardiomyopathy (broad)	4232495	Primary cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	4236332	Primary restrictive cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	4240233	Dilated cardiomyopathy secondary to amyloidosis	Condition	SNOMED	Inf
cardiomyopathy (broad)	36715371	Heart-hand syndrome Slovenian type	Condition	SNOMED	Inf
cardiomyopathy (broad)	37397548	TMEM70 related mitochondrial encephalo-cardio-myopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	37399476	Familial isolated arrhythmogenic right ventricular dysplasia	Condition	SNOMED	Inf
cardiomyopathy (broad)	37398927	Naxos disease	Condition	SNOMED	Inf
cardiomyopathy (broad)	46272954	3-methylglutaconic aciduria type 5	Condition	SNOMED	Inf
cardiomyopathy (broad)	321320	Myocardial degeneration	Condition	SNOMED	Inf
cardiomyopathy (broad)	4264027	Keshan disease	Condition	SNOMED	Inf
cardiomyopathy (broad)	4111421	Cardiac glycogenosis	Condition	SNOMED	Inf
cardiomyopathy (broad)	37110708	Muscle and heart glycogen synthase deficiency	Condition	SNOMED	Inf
cardiomyopathy (broad)	4167868	Danon disease	Condition	SNOMED	Inf
cardiomyopathy (broad)	36676853	Polyglucosan body myopathy type 1	Condition	SNOMED	Inf
cardiomyopathy (broad)	43020784	Fetal hypertrophic cardiomyopathy associated with renal disease	Condition	SNOMED	Inf
cardiomyopathy (broad)	4121474	Post-myocarditic cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (broad)	4211517	Restrictive cardiomyopathy secondary to hemochromatosis	Condition	SNOMED	Inf
cardiomyopathy (broad)	4297316	Restrictive cardiomyopathy secondary to granulomas	Condition	SNOMED	Inf
cardiomyopathy (broad)	439399	Endomyocardial fibrosis	Condition	SNOMED	Inf
cardiomyopathy (broad)	36678496	Combined oxidative phosphorylation defect type 17	Condition	SNOMED	Inf
cardiomyopathy (broad)	321502	Chagas' disease with heart involvement	Condition	SNOMED	Inf

Phenotype	concept_id	Concept Name	domain	Vocabulary	wash out
cardiomyopathy (broad)	42872390	Nonischemic congestive cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (primary)	44804845	Generalised right ventricular dysfunction	Condition	SNOMED	Inf
cardiomyopathy (primary)	4239975	Myocardial disease	Condition	SNOMED	Inf
cardiomyopathy (primary)	36715121	Cardiomyopathy with cataract and hip spine disease syndrome	Condition	SNOMED	Inf
cardiomyopathy (primary)	36715122	Cardiomyopathy and renal anomaly syndrome	Condition	SNOMED	Inf
cardiomyopathy (primary)	315560	Primary endomyocardial fibrosis cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (primary)	321319	Cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (primary)	4108822	Primary dilated cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (primary)	4124692	Congestive obstructive cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (primary)	4141588	Primary eosinophilic endomyocardial restrictive cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (primary)	4172641	Primary endomyocardial fibrosis restrictive cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (primary)	4162674	Restrictive cardiomyopathy with endomyocardial fibrosis	Condition	SNOMED	Inf
cardiomyopathy (primary)	4163710	Dilated cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (primary)	4185450	Dilated cardiomyopathy with connective tissue disorder	Condition	SNOMED	Inf
cardiomyopathy (primary)	4190773	Restrictive cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (primary)	4203149	Primary idiopathic dilated cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (primary)	4204856	Primary eosinophilic endomyocardial cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (primary)	4274635	Primary idiopathic restrictive cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (primary)	4218771	Dilated cardiomyopathy secondary to drug	Condition	SNOMED	Inf
cardiomyopathy (primary)	4288868	Nonobstructive cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (primary)	35615148	Cardiomyopathy caused by drug	Condition	SNOMED	Inf
cardiomyopathy (primary)	4232495	Primary cardiomyopathy	Condition	SNOMED	Inf

Phenotype	concept_id	Concept Name	domain	Vocabulary	wash out
cardiomyopathy (primary)	4236332	Primary restrictive cardiomyopathy	Condition	SNOMED	Inf
cardiomyopathy (primary)	321320	Myocardial degeneration	Condition	SNOMED	Inf
cardiomyopathy (primary)	42872390	Nonischemic congestive cardiomyopathy	Condition	SNOMED	Inf
heart failure (broad)	44782718	Acute combined systolic and diastolic heart failure	Condition	SNOMED	Inf
heart failure (broad)	4023479	Acute congestive heart failure	Condition	SNOMED	Inf
heart failure (broad)	312927	Acute cor pulmonale	Condition	SNOMED	Inf
heart failure (broad)	40481042	Acute diastolic heart failure	Condition	SNOMED	Inf
heart failure (broad)	44782655	Acute exacerbation of chronic congestive heart failure	Condition	SNOMED	Inf
heart failure (broad)	442310	Acute heart failure	Condition	SNOMED	Inf
heart failure (broad)	764877	Acute heart failure co-occurrent with normal ejection fraction	Condition	SNOMED	Inf
heart failure (broad)	4108245	Acute left ventricular failure	Condition	SNOMED	Inf
heart failure (broad)	4327205	Acute left-sided congestive heart failure	Condition	SNOMED	Inf
heart failure (broad)	4267800	Acute left-sided heart failure	Condition	SNOMED	Inf
heart failure (broad)	44782733	Acute on chronic combined systolic and diastolic heart failure	Condition	SNOMED	Inf
heart failure (broad)	40481043	Acute on chronic diastolic heart failure	Condition	SNOMED	Inf
heart failure (broad)	764874	Acute on chronic heart failure co-occurrent with normal ejection fraction	Condition	SNOMED	Inf
heart failure (broad)	37309625	Acute on chronic right-sided congestive heart failure	Condition	SNOMED	Inf
heart failure (broad)	40480602	Acute on chronic systolic heart failure	Condition	SNOMED	Inf
heart failure (broad)	4215446	Acute right-sided congestive heart failure	Condition	SNOMED	Inf
heart failure (broad)	4233424	Acute right-sided heart failure	Condition	SNOMED	Inf
heart failure (broad)	40480603	Acute systolic heart failure	Condition	SNOMED	Inf
heart failure (broad)	439698	Benign hypertensive heart disease with congestive cardiac failure	Condition	SNOMED	Inf

Phenotype	concept_id	Concept Name	domain	Vocabulary	wash out
heart failure (broad)	4242669	Biventricular congestive heart failure	Condition	SNOMED	Inf
heart failure (broad)	4205558	Cardiac failure after obstetrical surgery AND/OR other procedure including delivery	Condition	SNOMED	Inf
heart failure (broad)	4146456	Cardiac failure therapy	Procedure	SNOMED	Inf
heart failure (broad)	4233224	Cardiac insufficiency during AND/OR resulting from a procedure	Condition	SNOMED	Inf
heart failure (broad)	4264636	Cardiac insufficiency following cardiac surgery	Condition	SNOMED	Inf
heart failure (broad)	40482857	Cardiorenal syndrome	Condition	SNOMED	Inf
heart failure (broad)	4259490	Cardiorespiratory failure	Condition	SNOMED	Inf
heart failure (broad)	44782719	Chronic combined systolic and diastolic heart failure	Condition	SNOMED	Inf
heart failure (broad)	4229440	Chronic congestive heart failure	Condition	SNOMED	Inf
heart failure (broad)	4195892	Chronic cor pulmonale	Condition	SNOMED	Inf
heart failure (broad)	40479576	Chronic diastolic heart failure	Condition	SNOMED	Inf
heart failure (broad)	444031	Chronic heart failure	Condition	SNOMED	Inf
heart failure (broad)	764876	Chronic heart failure co-occurrent with normal ejection fraction	Condition	SNOMED	Inf
heart failure (broad)	4206009	Chronic left-sided congestive heart failure	Condition	SNOMED	Inf
heart failure (broad)	4009047	Chronic left-sided heart failure	Condition	SNOMED	Inf
heart failure (broad)	4284562	Chronic right-sided congestive heart failure	Condition	SNOMED	Inf
heart failure (broad)	4014159	Chronic right-sided heart failure	Condition	SNOMED	Inf
heart failure (broad)	40479192	Chronic systolic heart failure	Condition	SNOMED	Inf
heart failure (broad)	4108244	Compensated cardiac failure	Condition	SNOMED	Inf
heart failure (broad)	4071869	Congenital cardiac failure	Condition	SNOMED	Inf
heart failure (broad)	319835	Congestive heart failure	Condition	SNOMED	Inf
heart failure (broad)	44784345	Congestive heart failure as early postoperative complication	Condition	SNOMED	Inf

Phenotype	concept_id	Concept Name	domain	Vocabulary	wash out
heart failure (broad)	762002	Congestive heart failure as post-operative complication of cardiac surgery	Condition	SNOMED	Inf
heart failure (broad)	762003	Congestive heart failure as post-operative complication of non-cardiac surgery	Condition	SNOMED	Inf
heart failure (broad)	44782428	Congestive heart failure due to cardiomyopathy	Condition	SNOMED	Inf
heart failure (broad)	4139864	Congestive heart failure due to left ventricular systolic dysfunction	Condition	SNOMED	Inf
heart failure (broad)	4142561	Congestive heart failure due to valvular disease	Condition	SNOMED	Inf
heart failure (broad)	36712928	Congestive heart failure stage B due to ischemic cardiomyopathy	Condition	SNOMED	Inf
heart failure (broad)	43021826	Congestive heart failure stage C	Condition	SNOMED	Inf
heart failure (broad)	36712927	Congestive heart failure stage C due to ischemic cardiomyopathy	Condition	SNOMED	Inf
heart failure (broad)	43021825	Congestive heart failure stage D	Condition	SNOMED	Inf
heart failure (broad)	44782713	Congestive heart failure with right heart failure	Condition	SNOMED	Inf
heart failure (broad)	315295	Congestive rheumatic heart failure	Condition	SNOMED	Inf
heart failure (broad)	4307356	Cor pulmonale	Condition	SNOMED	Inf
heart failure (broad)	4111554	Decompensated cardiac failure	Condition	SNOMED	Inf
heart failure (broad)	4311437	Decompensated chronic heart failure	Condition	SNOMED	Inf
heart failure (broad)	443587	Diastolic heart failure	Condition	SNOMED	Inf
heart failure (broad)	43021842	Diastolic heart failure stage C	Condition	SNOMED	Inf
heart failure (broad)	43021841	Diastolic heart failure stage D	Condition	SNOMED	Inf
heart failure (broad)	44805683	Discharge from heart failure nurse service	Observation	SNOMED	Inf
heart failure (broad)	4215511	Emergency hospital admission for heart failure	Observation	SNOMED	Inf
heart failure (broad)	43022068	Exacerbation of congestive heart failure	Condition	SNOMED	Inf
heart failure (broad)	36204172	Felt discouraged or down in the dumps because of heart failure over the past 2 weeks [KCCQ]	Observation	LOINC	Inf
heart failure (broad)	44813162	Has heart failure management plan	Observation	SNOMED	Inf

Phenotype	concept_id	Concept Name	domain	Vocabulary	wash out
heart failure (broad)	316139	Heart failure	Condition	SNOMED	Inf
heart failure (broad)	44790972	Heart failure 6 month review	Observation	SNOMED	Inf
heart failure (broad)	4193010	Heart failure annual review	Observation	SNOMED	Inf
heart failure (broad)	4124705	Heart failure as a complication of care	Condition	SNOMED	Inf
heart failure (broad)	44808957	Heart failure clinical pathway	Observation	SNOMED	Inf
heart failure (broad)	4215689	Heart failure confirmed	Condition	SNOMED	Inf
heart failure (broad)	43020657	Heart failure due to end stage congenital heart disease	Condition	SNOMED	Inf
heart failure (broad)	44808862	Heart failure initial assessment	Observation	SNOMED	Inf
heart failure (broad)	44789102	Heart failure review completed	Observation	SNOMED	Inf
heart failure (broad)	44806371	Heart failure self-management plan agreed	Observation	SNOMED	Inf
heart failure (broad)	44806125	Heart failure self-management plan review	Procedure	SNOMED	Inf
heart failure (broad)	37311948	Heart failure with mid range ejection fraction	Condition	SNOMED	Inf
heart failure (broad)	40486933	Heart failure with normal ejection fraction	Condition	SNOMED	Inf
heart failure (broad)	45766164	Heart failure with reduced ejection fraction	Condition	SNOMED	Inf
heart failure (broad)	45766167	Heart failure with reduced ejection fraction due to cardiomyopathy	Condition	SNOMED	Inf
heart failure (broad)	45766165	Heart failure with reduced ejection fraction due to coronary artery disease	Condition	SNOMED	Inf
heart failure (broad)	45773075	Heart failure with reduced ejection fraction due to heart valve disease	Condition	SNOMED	Inf
heart failure (broad)	45766166	Heart failure with reduced ejection fraction due to myocarditis	Condition	SNOMED	Inf
heart failure (broad)	4004279	High output heart failure	Condition	SNOMED	Inf
heart failure (broad)	36305423	History of Acute heart failure within 24 hours prior to procedure	Observation	LOINC	Inf
heart failure (broad)	36204171	How would you feel if you had to spend the rest of your life with your heart failure the way it is right now [KCCQ]	Observation	LOINC	Inf
heart failure (broad)	36204159	Hurrying or jogging (as if to catch a bus) was limited by heart failure over the past 2 weeks [KCCQ]	Observation	LOINC	Inf

Phenotype	concept_id	Concept Name	domain	Vocabulary	wash out
heart failure (broad)	44784621	Hypertensive heart and chronic kidney disease	Condition	SNOMED	Inf
heart failure (broad)	44782728	Hypertensive heart AND chronic kidney disease with congestive heart failure	Condition	SNOMED	Inf
heart failure (broad)	439696	Hypertensive heart and renal disease with (congestive) heart failure	Condition	SNOMED	Inf
heart failure (broad)	439694	Hypertensive heart and renal disease with both (congestive) heart failure and renal failure	Condition	SNOMED	Inf
heart failure (broad)	314378	Hypertensive heart disease with congestive heart failure	Condition	SNOMED	Inf
heart failure (broad)	444101	Hypertensive heart failure	Condition	SNOMED	Inf
heart failure (broad)	43530961	Induced termination of pregnancy complicated by cardiac failure	Condition	SNOMED	Inf
heart failure (broad)	439846	Left heart failure	Condition	SNOMED	Inf
heart failure (broad)	4103448	Low output heart failure	Condition	SNOMED	Inf
heart failure (broad)	316994	Malignant hypertensive heart disease with congestive heart failure	Condition	SNOMED	Inf
heart failure (broad)	44782439	Nutrition therapy for congestive heart failure	Procedure	SNOMED	Inf
heart failure (broad)	46269995	Nutritional therapy for congestive heart failure done	Observation	SNOMED	Inf
heart failure (broad)	44811840	On optimal heart failure therapy	Condition	SNOMED	Inf
heart failure (broad)	36204173	Participation in hobbies, recreational activities limited by heart failure over the past 2 weeks [KCCQ]	Observation	LOINC	Inf
heart failure (broad)	36204176	Participation in intimate relationships with loved ones limited by heart failure over the past 2 weeks [KCCQ]	Observation	LOINC	Inf
heart failure (broad)	36204175	Participation in visiting family or friends out of your home limited by heart failure over the past 2 weeks [KCCQ]	Observation	LOINC	Inf
heart failure (broad)	36204174	Participation in working or doing household chores limited by heart failure over the past 2 weeks [KCCQ]	Observation	LOINC	Inf
heart failure (broad)	4236658	Pleural effusion due to congestive heart failure	Condition	SNOMED	Inf
heart failure (broad)	4293582	Red half-moon nail in congestive heart failure	Condition	SNOMED	Inf
heart failure (broad)	764873	Reduced ejection fraction co-occurrent and due to acute heart failure	Condition	SNOMED	Inf
heart failure (broad)	764871	Reduced ejection fraction co-occurrent and due to acute on chronic heart failure	Condition	SNOMED	Inf

Phenotype	concept_id	Concept Name	domain	Vocabulary	wash out
heart failure (broad)	764872	Reduced ejection fraction co-occurrent and due to chronic heart failure	Condition	SNOMED	Inf
heart failure (broad)	4199500	Refractory heart failure	Condition	SNOMED	Inf
heart failure (broad)	44809279	Rehabilitation for heart failure	Observation	SNOMED	Inf
heart failure (broad)	4184497	Rheumatic left ventricular failure	Condition	SNOMED	Inf
heart failure (broad)	4138307	Right heart failure due to pulmonary hypertension	Condition	SNOMED	Inf
heart failure (broad)	4195785	Right heart failure secondary to left heart failure	Condition	SNOMED	Inf
heart failure (broad)	4273632	Right ventricular failure	Condition	SNOMED	Inf
heart failure (broad)	35615055	Saddle embolus of pulmonary artery with acute cor pulmonale	Condition	SNOMED	Inf
heart failure (broad)	4079695	Sepsis-associated left ventricular failure	Condition	SNOMED	Inf
heart failure (broad)	4079296	Sepsis-associated right ventricular failure	Condition	SNOMED	Inf
heart failure (broad)	44784442	Symptomatic congestive heart failure	Condition	SNOMED	Inf
heart failure (broad)	40760358	Symptoms in heart failure patients [CMS Assessment]	Observation	LOINC	Inf
heart failure (broad)	443580	Systolic heart failure	Condition	SNOMED	Inf
heart failure (broad)	43020421	Systolic heart failure stage C	Condition	SNOMED	Inf
heart failure (broad)	36712929	Systolic heart failure stage C due to ischemic cardiomyopathy	Condition	SNOMED	Inf
heart failure (broad)	43021840	Systolic heart failure stage D	Condition	SNOMED	Inf
heart failure (broad)	36676642	X-linked intellectual disability, cardiomegaly, congestive heart failure syndrome	Condition	SNOMED	Inf
heart failure (broad)	4110961	Generalized ischemic myocardial dysfunction	Condition	SNOMED	Inf
heart failure (broad)	43020910	Pulmonary hypertension due to left heart disease	Condition	SNOMED	Inf

Table S2. Preliminary list of medicines definitions.

Substance Name	Concept name	Class	ATC code	Ingredient Concept ID	Include descendants
Venlafaxine	venlafaxine	SNRI	N06AX16	743670	Yes
Mirtazapine	mirtazapine	Tetracyclic antidepressants	N06AX11	725131	Yes

Table S3. Preliminary list of Negative Control Outcomes (NCO).

Phenotype	Concept name	Concept ID (including descendants)
NCO	Constipation	75860
NCO	Ulcer of lower extremity	197304
NCO	Cellulitis of lower limb	42709838
NCO	Iron deficiency anemia	436659
NCO	Wax in ear canal	4155902
NCO	Glaucoma	437541
NCO	Otitis externa	380731
NCO	Osteopenia	4195039
NCO	Dry eyes	4036620
NCO	Benign prostatic hyperplasia	198803
NCO	Open wound of lower leg	4053604
NCO	Vulval irritation	4060207
NCO	Acute conjunctivitis	376707
NCO	Vaginal irritation	4058568
NCO	Prostatism	4016155
NCO	Vitamin D deficiency	436070
NCO	Basal cell carcinoma of skin	4112752
NCO	Hemorrhoids	195562
NCO	Senile hyperkeratosis	141932
NCO	Intraocular pressure left eye	4217260
NCO	Traumatic wound	46287159
NCO	Gallstone	196456
NCO	Pressure ulcer	135333
NCO	Polyp of colon	4285898
NCO	Impacted cerumen	374375
NCO	Hearing loss	377889
NCO	Hypothyroidism	140673
NCO	Rectal hemorrhage	4026112
NCO	Foot pain	4169905

Phenotype	Concept name	Concept ID (including descendants)
NCO	Urinary incontinence	197672
NCO	Squamous cell carcinoma of skin	4111921
NCO	Acquired hypothyroidism	138384
NCO	Age related macular degeneration	374028
NCO	Acid reflux	44783954
NCO	Laceration of lower leg	4155040
NCO	Ulcer of lower extremity	197304
NCO	Cellulitis of lower limb	42709838
NCO	Iron deficiency anemia	436659
NCO	Wax in ear canal	4155902
NCO	Actinic keratosis	138825
NCO	Otitis externa	380731
NCO	Osteopenia	4195039
NCO	Dry eyes	4036620
NCO	Benign prostatic hyperplasia	198803
NCO	Blepharitis	378425
NCO	Vulval irritation	4060207
NCO	Acute conjunctivitis	376707
NCO	Vaginal irritation	4058568
NCO	Prostatism	4016155

ANNEX V. ENCePP checklist for study protocols

ENCePP Checklist for Study Protocols (Revision 4)

Study title: DARWIN EU® - Assessing the potential association between venlafaxine and incident heart failure among adults

EU PAS Register® number: ***"Study not registered yet".***
Study reference number (if applicable):

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

Comments:

Section 2: Research question	**Yes**	**No**	**N/A**	**Section Number**
2.1 Does the formulation of the research question and objectives clearly explain:	☒	☐	☐	
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)	☒	☐	☐	6 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))	☒	☐	☐	8.8
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)	☐	☐	☒	Annex III

Comments:

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

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
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?	☒	☐	☐	8.8

Section 5: Exposure definition and measurement	**Yes**	**No**	**N/A**	**Section Number**
5.4 Is intensity of exposure addressed? (e.g. dose, duration)	☐	☐	☒	
5.5 Is exposure categorised based on biological mechanism of action and taking into account the pharmacokinetics and pharmacodynamics of the drug?	☒	☐	☐	8.6.1
5.6 Is (are) (an) appropriate comparator(s) identified?	☒	☐	☐	8.6.1

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)	☒	☐	☐	8.8
7.2 Does the protocol address selection bias? (e.g. healthy user/adherer bias)	☒	☐	☐	8.8
7.3 Does the protocol address information bias? (e.g. misclassification of exposure and outcomes, time-related bias)	☒	☐	☐	8.8

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)	☒	☐	☐	8.8

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)	☒	☐	☐	8.4
9.1.2 Outcomes? (e.g. clinical records, laboratory markers or values, claims data, self-report, patient interview including scales and questionnaires, vital statistics)	☒	☐	☐	8.4
9.1.3 Covariates and other characteristics?	☒	☐	☐	8.4
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)	☒	☐	☐	8.6.1 Annex IV
9.3.2 Outcomes? (e.g. International Classification of Diseases (ICD), Medical Dictionary for Regulatory Activities (MedDRA))	☒	☐	☐	8.6.2 Annex IV
9.3.3 Covariates and other characteristics?	☐	☐	☒	
9.4 Is a linkage method between data sources described? (e.g. based on a unique identifier or other)	☒	☐	☐	Annex IV

Comments:

Section 10: Analysis plan	Yes	No	N/A	Section Number
10.1 Are the statistical methods and the reason for their choice described?	☒	☐	☐	8.8.3
10.2 Is study size and/or statistical precision estimated?	☒	☐	☐	8.7

Section 10: Analysis plan	**Yes**	**No**	**N/A**	**Section Number**
10.3 Are descriptive analyses included?	☒	☐	☐	8.8.3
10.4 Are stratified analyses included?	☒	☐	☐	8.8.3
10.5 Does the plan describe methods for analytic control of confounding?	☒	☐	☐	8.8.3
10.6 Does the plan describe methods for analytic control of outcome misclassification?	☒	☐	☐	8.6.2
10.7 Does the plan describe methods for handling missing data?	☐	☐	☒	
10.8 Are relevant sensitivity analyses described?	☒	☐	☐	8.8.3

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?	☐	☐	☒	
13.3 Have data protection requirements been described?	☒	☐	☐	8.8.1, 8.8.2

Comments:

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

Comments:

Section 15: Plans for communication of study results	**Yes**	**No**	**N/A**	**Section Number**
15.1 Are plans described for communicating study results (e.g. to regulatory authorities)?	☒	☐	☐	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: Annika Jodicke

Date: 12/02/2026

Signature: Annika Jodicke

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 standardized data structure that enables data from multiple sources to be harmonized, 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 Centre 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.