Study Protocol P2 C1-008

20/10/2023

Version 3.1

TABLE OF CONTENTS

DOCUMENT HISTORY

Version	Date	Description
V1.0	14/08/2023	**Submission to EMA**
V2.0	14/09/2023	**Second version following comments from EMA**
V3.0	26/09/2023	**Third version following comments from EMA**
V3.1	20/10/2023	**EUPAS register number added**

Study Title	DARWIN EU® - Rates of occurrence of treatment-related *intercurrent events* in patients with major depressive disorder
Protocol version identifier	V3.0
Date of last version of protocol	26th September 2023
EU PAS register number	EUPAS106685
Active substance	See table below
Medicinal product	N/A
Research question and objectives	See below

Drug Class	ATC code
I. Main drugs of interest (antidepressants)	
Non-selective monoamine reuptake inhibitors (NSRIs)	N06AA
Selective serotonin reuptake inhibitors (SSRIs)	N06AB
Other Antidepressants (excluding N06AX25 and N06AX27)	N06AX
II. Concomitant drugs of interest (psycholeptics)	
Antipsychotics	N05A
Anxiolytics	N05B
Hypnotics and sedatives	N05C

Research question

What is the incidence in clinical practice of treatment-related intercurrent events common in clinical trials in patients with major depressive disorder?

Study objectives

1. To examine the proportion of patients with newly diagnosed major depressive disorder who start treatment with antidepressants (NSRIs, SSRIs, or other anti-depressants), and of those the proportion who switch or discontinue treatment by specific timepoints (4, 6, 8, 12, and 24 weeks) after treatment initiation, stratified by age group, sex, and country/database during the study period (2013 - 2022).
2. To estimate the duration of antidepressant use in patients with newly diagnosed major depressive disorder who initiate treatment with antidepressants (NSRIs, SSRIs, or other antidepressants), stratified by age group, sex, and country/database during the study period (2013 - 2022).
3. To assess the proportions of patients with newly diagnosed major depressive disorder who initiate, switch, or discontinue treatment with psycholeptics (antipsychotics, anxiolytics, hypnotics, and sedatives) by specific timepoints (4, 6, 8, 12, and 24 weeks) after starting antidepressant therapy, stratified by age group, sex, and country/database during the study period (2013 - 2022).

Countries of study	Germany, Netherlands, Spain, and the United Kingdom
Author	Johnmary T. Arinze (j.arinze@darwin-eu.org) Katia Verhamme (k.verhamme@darwin-eu.org)

LIST OF ABBREVIATIONS

Acronyms/term	Description
ATC	Anatomical Therapeutic Chemical Classification
CDM	Common Data Model
CPRD GOLD	Clinical Practice Research Datalink GOLD
DA	Disease Analyzer
DARWIN EU®	Data Analysis and Real World Interrogation Network
DUS	Drug Utilization Study
EHR	Electronic Health Records
EMA	European Medicines Agency
EU	European Union
GDPR	General Data Protection Regulation
GP	General Practitioner
ID	Index date
IMASIS	Institut Municipal Assistència Sanitària Information System
IEs	Intercurrent Events
IP	Inpatient
IPCI	Integrated Primary Care Information Project
MDD	Major Depressive Disorder
NSRIs	Non-Selective monoamine Reuptake Inhibitors
OHDSI	Observational Health Data Sciences and Informatics
OP	Outpatient
OMOP	Observational Medical Outcomes Partnership
SIDIAP	Sistema d'Informació per al Desenvolupament de la Investigació en Atenció Primària
SSRIs	Selective Serotonin Reuptake Inhibitors
SNOMED	Systematized Nomenclature of Medicine
WHO	World Health Organisation

1. TITLE

DARWIN EU® - Rates of occurrence of treatment-related *intercurrent events* in patients with major depressive disorder

2. RESPONSIBLE PARTIES – STUDY TEAM

Table 1: Description of Study Team

Study team Role	Names	Organisation
Principal Investigator(s)/ Clinical Epidemiologists	Johnmary Arinze Katia Verhamme	Erasmus MC
Data Scientist(s)	Cesar Barboza Gutierrez Maarten van Kessel	Erasmus MC
Data Partner*	**Names**	**Organization**
Local Study Coordinator/Data Analyst	Antonella Delmestri James Brash Núria Mercadé Mees Mosseveld Miguel-Angel Mayer	University of Oxford – CPRD IQVIA - DA Germany IDIAPJGol - SIDIAP Erasmus MC – IPCI PSMAR - IMASIS

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

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

Title

DARWIN EU® - Rates of occurrence of treatment-related *intercurrent events* in patients with major depressive disorder

Rationale and Background

In clinical trials involving patients with major depressive disorder, participants who start treatment may experience *intercurrent events* (IEs) during follow-up, such as treatment discontinuation, switch to alternative therapies, or changes in background/concomitant therapies (e.g., sleep aids). The ICH E9(R1) guideline defines IEs as events that occur after treatment initiation and influence the interpretation of the outcome of interest or after which the outcome no longer exists (e.g., death).

While target estimands in these trials may adopt a treatment policy or composite strategy to handle these IEs, it is crucial to recognize that the rate at which these intercurrent events occur significantly impacts the interpretation of estimated treatment effects.

To gain a more comprehensive understanding of the external validity of clinical trials in this indication, it is essential to assess whether the rate of occurrence of these IEs is similar in real-life settings compared to what is observed in the clinical trials. By obtaining such insights, the results of this study aim to provide valuable information regarding the generalisability of clinical trial findings to real-world scenarios.

Research question and Objectives

Research question

What is the incidence in clinical practice of treatment-related intercurrent events (IEs) common in clinical trials in patients with major depressive disorder?

Study objectives

1. To examine the proportion of patients with newly diagnosed major depressive disorder who start treatment with antidepressants (NSRIs, SSRIs, or other anti-depressants), and of those the proportion who switch or discontinue treatment by specific timepoints (4, 6, 8, 12, and 24 weeks after treatment initiation), stratified by age group, sex, and country/database during the study period (2013 - 2022).
2. To estimate the duration of antidepressant use in patients with newly diagnosed major depressive disorder who initiate treatment with antidepressants (NSRIs, SSRIs, or other antidepressants), stratified by age group, sex, and country/database during the study period (2013 - 2022).
3. To assess the proportions of patients with newly diagnosed major depressive disorder who initiate, switch, or discontinue treatment with psycholeptics (antipsychotics, anxiolytics, hypnotics, and sedatives) by specific timepoints (4, 6, 8, 12, and 24 weeks) after starting antidepressant therapy, stratified by age group, sex, and country/database during the study period (2013 - 2022).

Research Methods

Study design

- Patient-level characterisation (Objective 1 and 3, Patient-level characterization of use patterns and sequences, including initiation, discontinuation, and switching, of antidepressants and psycholeptics in patients with newly diagnosed major depressive disorder).

- Patient-level drug utilization (Objective 2, Patient-level drug utilization analyses to assess the duration of antidepressant use in patients with newly diagnosed major depressive disorder).

Population

Patient-level characterisation: Patient-level characterisation analyses will include all patients with newly diagnosed with major depressive disorder who are aged 12 years and above in the respective databases from 2013 to 2022 (or the latest available date if earlier), with a minimum of 1 year of data visibility before their diagnosis, and no previous record of major depressive disorder in the year preceding their diagnosis.

Patient-level utilization: Patient-level drug utilization analyses will include all patients aged 12 years and above with newly diagnosed major depressive disorder who are new users of any of the antidepressant class of interest in the respective databases from 2013 to 2022 (or the latest available date if earlier), with a minimum of 1 year of data visibility before the index date..

Variables

Drug class of interest:

- Non-selective monoamine reuptake inhibitors
- Selective serotonin reuptake inhibitors
- Other antidepressants (excluding esketamine and Hyperici herba)
- Concomitant medications – Psycholeptics
  - ✓ Antipsychotics
  - ✓ Anxiolytics
  - ✓ Hypnotics and sedatives

Condition of interest:

- Major depressive disorder (MDD)

Data sources

1. Clinical Practice Research Datalink GOLD (CPRD GOLD), United Kingdom
2. IQVIA Disease Analyzer Germany (IQVIA DA Germany), Germany
3. Institut Municipal Assistencia Sanitaria Information System (IMASIS), Spain (for this study, we will only use the outpatients)
4. Integrated Primary Care Information Project (IPCI), The Netherlands
5. The Information System for Research in Primary Care (SIDIAP), Spain

Sample size

No sample size was calculated for this descriptive study, as our primary focus is to summarise the use pattern of antidepressants and psycholeptics in adolescents/adults with newly diagnosed MDD. Based on a preliminary feasibility assessment, the expected number of patients with MDD in the included databases for this study will be approximately 380,000.

Data analyses

Objective 1: The number and percentage of patients with newly diagnosed MDD initiating treatment with antidepressants will be estimated at 4, 6, 8, 12, and 24 weeks after date of new MDD diagnosis; and the number and percentage of those discontinuing and switching treatment at 4, 6, 8, 12, and 24 weeks following start of antidepressant therapy. They will be reported in tabular form as absolute numbers as well as

proportions, and depicted using Sunburst plots and Sankey diagrams. The statistical analyses will be performed based on OMOP-CDM mapped data using the "TreatmentPatterns" R package.

Objective 2: The duration (mean, median, quantiles 25% and 75%, minimum and maximum) of antidepressant use in patients with newly diagnosed MDD during the first treatment era will be estimated. The index date will be determined as the date of the first prescription of the specific antidepressant class for each individual. Statistical analyses will be conducted using the "DrugUtilization" R package based on OMOP-CDM mapped data.

Objective 3: The number and percentage of patients with newly diagnosed MDD initiating, switching, or discontinuing treatment with psycholeptics) will be estimated at 4, 6, 8, 12, and 24 weeks after the date of treatment initiation with antidepressants; and at 4, 6, 8, 12, and 24 weeks after starting psycholeptics (for switching or discontinuing treatment). This will be reported in tabular form as absolute numbers as well as proportions and depicted using Sunburst plots and Sankey diagrams. The statistical analyses will be performed based on OMOP-CDM mapped data using the "TreatmentPatterns" R package.

.

For all analyses, results will be reported with a minimum cell count of 5, and any counts smaller than 5 will be reported as <5 to ensure privacy and confidentiality.

4. AMENDMENTS AND UPDATES

Number	Date	Section of study protocol	Amendment or update	Reason
1	14th September 2023	Throughout the whole protocol	Update	The study protocol was updated in response to EMA feedback.
2.	26th September 2023	Abstract and Research Methods	Update	The study protocol was updated in response to EMA feedback.
V3.1	20/10/2023	Document history	Update	EUPAS register number added

5. MILESTONES

STUDY SPECIFIC DELIVERABLE	TIMELINE
Draft Study Protocol	14th August 2023

Final Study Protocol	26th September 2023
Creation of Analytical code	October 2023
Execution of Analytical Code on the data	October 2023
Interim Study Report (if applicable)	Not applicable
Draft Study Report	November 2023
Final Study Report	To be confirmed

6. RATIONALE AND BACKGROUND

Major depressive disorder (MDD) is a significant and escalating global health burden, ranked as the third leading cause of disease burden worldwide in 2008 and projected to be the first by 2030. (1) The prevalence of MDD in Europe is 2.1%,(2) with higher rates in women and a global lifetime prevalence ranging from 5% to 17%.(3) Recent trends show an alarming increase in MDD cases among younger populations due to substance abuse. Comorbidity is common in MDD, often involving concurrent substance use disorders, anxiety disorders, or other psychiatric conditions, increasing the risk of suicide.(4)

MDD is characterized by persistent low or depressed mood, anhedonia, guilt or worthlessness feelings, lack of energy, impaired concentration, appetite changes, psychomotor disturbances, sleep disturbances, and, in severe cases, suicidal thoughts.(4) Its etiology is multifactorial, involving complex interactions between biological, genetic, environmental, and psychosocial factors. Early theories focused on neurotransmitter abnormalities, particularly involving serotonin, norepinephrine, and dopamine, leading to the development of antidepressants targeting these systems. Thyroid and growth hormone abnormalities, as well as childhood adversity and trauma, are also linked to increased susceptibility to major depression later in life.(4-6)

The management of MDD requires a comprehensive and multimodal approach, including pharmacological, psychotherapeutic, interventional, and lifestyle modifications.(4) Selective serotonin reuptake inhibitors (SSRIs) and serotonin-norepinephrine reuptake inhibitors (SNRIs) are common first-line pharmacological agents, with cognitive-behavioural therapy and interpersonal therapy proving highly effective as psychotherapeutic interventions. For treatment-resistant cases, interventions such as electroconvulsive therapy (ECT), transcranial magnetic stimulation (TMS), and vagus nerve stimulation (VNS) offer alternative options.(7-9)

Despite the available treatments, a considerable proportion of MDD patients, estimated at 10% to 30%, exhibit limited or no response to medications, highlighting the need for novel therapeutic approaches. (10) Clinical trials (CTs) play a crucial role in evaluating treatment efficacy and safety but understanding intercurrent events (IEs) occurring after starting therapy is essential for translating CT findings to real-life clinical practice.(11) IEs, events that arise after treatment initiation and impact the interpretation of outcomes, can include treatment discontinuation, switches to alternative therapies, and modifications in concomitant treatments, introducing complexity in the definition and the estimation of a treatment effect.(11-13) Addressing IEs in CTs requires the definition of a target estimand. Where such estimands include strategies defined as treatment policy (considering the outcomes regardless of the occurrence of IEs) or composite (as a signal of treatment failure), their incidence in the trial is a determinant of the treatment effect estimated, raising a question around the external validity of such estimate.

This study aims to provide valuable insights into the incidence in clinical practice of such events, and in particular of treatment discontinuation, initiation of (other) treatments, and switching among newly diagnosed MDD patients prescribed antidepressants and psycholeptics in various European clinical settings. Understanding the impact of IEs in real-life clinical practice will facilitate extrapolation of CT findings to improve everyday patient care and outcomes in MDD management.

7. RESEARCH QUESTION AND OBJECTIVES

Research question

What is the incidence in clinical practice of treatment-related intercurrent events common in clinical trials in patients with major depressive disorder?

Study objectives

1. To examine the proportion of patients with newly diagnosed major depressive disorder who start treatment with antidepressants (NSRIs, SSRIs, or other anti-depressants), as well as those who switch or discontinue treatment by specific timepoints (4, 6, 8, 12, and 24 weeks after treatment initiation), stratified by age, sex, and country/database during the study period (2013 - 2022).

2. To estimate the duration of antidepressant use in patients with newly diagnosed major depressive disorder who initiate treatment with antidepressants (NSRIs, SSRIs, or other antidepressants), stratified by age, sex, and country/database during the study period (2013 - 2022).

3. To assess the proportions of patients with newly diagnosed major depressive disorder who initiate, switch, or discontinue treatment with psycholeptics (antipsychotics, anxiolytics, hypnotics, and sedatives) by specific timepoints (4, 6, 8, 12, and 24 weeks) after starting antidepressant therapy, stratified by age, sex, and country/database during the study period (2013 - 2022).

Table 2: Primary and secondary research questions and objective

A. Primary research question and objectives

Objective:	To examine the proportion of patients with newly diagnosed major depressive disorder who start treatment with antidepressants (NSRIs, SSRIs, or other anti-depressants) and *psycholeptics (antipsychotics, anxiolytics, hypnotics and sedatives),* and of those the proportion who switch or discontinue treatment by specific timepoints (4, 6, 8, 12, and 24 weeks) after treatment initiation, stratified by age group, sex, and country/database during the study period (2013 - 2022).
Hypothesis:	Not applicable
Population *(mention key inclusion-exclusion criteria):*	All patients with newly diagnosed major depressive disorder who are aged 12 years and above in the respective databases from 2013 to 2022 (or the latest available date if earlier), with a minimum of 1 year of data visibility before their diagnosis, and no previous record of major depressive disorder in the year preceding their diagnosis.
Exposure:	• Antidepressants: 1. Non-selective monoamine reuptake inhibitors 2. Selective serotonin reuptake inhibitors 3. Other Antidepressants (excluding esketamine and 14ypericin herba) • Psycholeptics: 1. Antipsychotics 2. Anxiolytics

	3. Hypnotics and sedatives
Comparator:	None
Outcome:	Treatment initiation, treatment discontinuation and treatment switch.
Time *(when follow up begins and ends):*	Follow-up will start on the date of MDD diagnosis (index date) until the earliest of loss to follow-up, end of data availability or death, or end of study period (31st December 2022).
Setting:	Inpatient and outpatient setting using data from the following 5 data sources: CPRD GOLD (UK), IQVIA DA Germany (Germany), IMASIS (Spain), IPCI (the Netherlands) and SIDIAP (Spain).
Main measure of effect:	Proportions (%), patient level characterisation to assess the use patterns of the respective drugs of interest.

B. Secondary research question and objective

Objective:	To estimate the duration of antidepressant use in patients with newly diagnosed major depressive disorder who initiate treatment with antidepressants (NSRIs, SSRIs, or other antidepressants), stratified by age, sex, and country/database during the study period (2013 - 2022)
Hypothesis:	Not applicable
Population *(mention key inclusion-exclusion criteria):*	Individuals aged 12 and above who are new users of antidepressants, with new diagnosis of MDD, in the period between 1st of January 2032 and 31st of December 2022 (or the latest available date if earlier), with at least 1 year of data visibility prior they become eligible for study inclusion.
Exposure:	Antidepressants: • Non-selective monoamine reuptake inhibitors • Selective serotonin reuptake inhibitors • Other Antidepressants (excluding esketamine and hyperici herba)
Comparator:	None
Outcome:	None
Time *(when follow up begins and ends):*	Follow-up will start on the date of the first prescription and/or dispensation of the antidepressant class of interest (index date) until the earliest of loss to follow-up, end of data availability or death, or end of study period (31st December 2022).

Setting:	Inpatient and outpatient setting using data from the following 5 data sources: CPRD GOLD (UK), IQVIA DA Germany (Germany), IMASIS (Spain), IPCI (the Netherlands) and SIDIAP (Spain).
Main measure of effect:	Duration summary descriptive statistics (mean, median, quantiles 25% and 75%, minimum and maximum) of antidepressant use, patient level drug utilisation to assess duration of use of the first exposure episode of the respective antidepressant class of interest.

8. RESEARCH METHODS

8.1 Study Design

A disease epidemiology/ drug utilisation study will be conducted using routinely collected health data from 5 databases. The study will comprise two consecutive parts:

- A patient-level characterisation study will be conducted to address objective 1 and 3, assessing the proportions of treatment initiation, switching, and discontinuation of antidepressants and psycholeptics as classes of interest in patients newly diagnosed with MDD who are aged 12 years and above.
- A patient-level drug utilisation study will be used to address objective 2; estimating the duration (in days) of antidepressant use among new users of the specified drug class, in patients with newly diagnosed MDD who are aged 12 years and above.

Table 3. Description of Potential Study Types and Related Study Designs

STUDY TYPE	STUDY DESIGN	STUDY CLASSIFICATION
Patient Level Characterisation/ Disease Epidemiology	Cohort analysis	Off the shelf (C1)
Patient Level DUS	New drug/s user cohort	Off the shelf (C1)

8.2 Study Setting and Data Sources

This study will be conducted using routinely collected data from 5 databases in 4 European countries (3 EU countries and United Kingdom). All databases were previously mapped to the OMOP Common Data Model (CDM).

1. Clinical Practice Research Datalink GOLD (CPRD GOLD), United Kingdom
2. IQVIA Disease Analyzer Germany (IQVIA DA Germany), Germany
3. Institut Municipal Assistencia Sanitaria Information System (IMASIS), Spain (for this study, we will only use the outpatients)
4. Integrated Primary Care Information Project (IPCI), The Netherlands
5. The Information System for Research in Primary Care (SIDIAP), Spain

For this study, we selected 5 out of the 10 databases available in the DARWIN EU® Database Catalogue. The selection process was based on the size of the databases, the number of individuals with MDD and the counts of antidepressant class of interest across databases, as well as the geographical spread.

These selected databases fulfil the requirements for conducting both a patient-level characterisation study and a patient-level drug utilization study. This enables us to estimate the proportions of utilization related to treatment initiation, switching, and discontinuation for antidepressants in the context of incident MDD. Moreover, it facilitates the assessment of utilization proportions for treatment initiation, switching, and discontinuation of concomitant drug classes (specifically psycholeptics) within this patient cohort. Importantly, this selection encompasses databases from diverse clinical settings, thus allowing us to capture both inpatient and outpatient prescriptions or dispensing of drugs.

Additionally, these selected databases possess comprehensive data on MDD, along with representation of at least five ingredients from each of the three main antidepressant classes of interest. The rationale and justification for selecting these specific data sources, underpinned by their capacity to capture pertinent information, are detailed in **Table 4**.

In general, drug utilization studies have been extensively conducted across the selected databases(14-16). Consequently, these databases serve as highly suitable resources for examining intercurrent events linked to the pharmacological management of MDD. Furthermore, prior research on MDD has been undertaken in a minimum of three of the participating databases, namely CPRD GOLD, IQVIA DA Germany, and SIDIAP (17-19). This observation holds significant relevance in validating the accuracy of MDD cases within the scope of this study.

When it comes to assessing the reliability of data sources, the data partners are asked to describe their internal data quality process on the source data as part of the onboarding procedure. In addition, they are asked to share the results from three data quality assurance package: CdmOnboarding, Data Quality Dashboard (DQD) and DashboardExport. The latter exports a subset of analyses from the Achilles tool (https://github.com/OHDSI/Achilles), which systematically characterizes the data and presents it in a dashboard format to ease the detection of potential quality issues. The generated data characteristics such as age distribution, condition prevalence per year, data density, measurement value distribution can be compared against the national healthcare data. CdmOnboarding creates a report with select characterisation and mapping coverage statistics that are closely inspected upon onboarding. DQD provides more objective checks on conformance and plausibility, applied consistently across the data sources.

Additionally, the data quality dashboard (DQD) provides more objective checks on plausibility consistently across the data sources. In terms of relevance, a more study-specific diagnostic tool, CohortDiagnostics, was developed. This package evaluates phenotype algorithms for OMOP CDM datasets, offering a standard set of analytics for understanding patient capture including data generation. It provides additional insights into cohort characteristics, record counts and index event misclassification. Furthermore, timeliness is guarded by extracting the release dates for each dataset in the network and monitoring when data are out-of-date with the expected refresh cycle (typically quarterly or half-yearly). In addition, it is important to have clear understanding of the time period covered by each released database, as this can vary across different domains. To facilitate this, the CdmOnboarding (and Achilles) packages contain a 'data density' plot. This plot displays the number of records per OMOP domain on a monthly basis. This allows to get insights when data collection started, when new sources of data were added and when until when data was `.

Table 4. Description of the selected Data Sources.

Country	Name of Database	Justification for Inclusion	Health Care setting	Type of Data	Number of active subjects	Data lock for the last update
UK	CPRD GOLD	Database covers primary care where antidepressants may be prescribed/dispensed.	Primary care	EHR	3 million	20/03/2023
Germany	IQVIA DA Germany	Databases covers primary care / outpatient specialist care setting where antidepressants may be prescribed/dispensed.	Primary care and outpatient specialist care	EHR	8.5 million	13/03/2023
Spain	IMASIS	Database covers hospital care setting where antidepressants may be prescribed/dispensed. For this study, we will only use outpatients from IMASIS	Secondary care (in and outpatient)	EHR	0.6 million	31/12/2022
The Netherlands	IPCI	Database covers primary care where antidepressants may be prescribed/dispensed.	Primary care	EHR	1.4 million	01/12/2022
Spain	SIDIAP	Database covers primary care where medication antidepressants may be prescribed/dispensed.	Primary care with hospital linkage	EHR	5.8 million	31/12/2022

UK = United Kingdom; CPRD GOLD = Clinical Practice Research Datalink GOLD; IMASIS = Institut Municipal Assistencia Sanitaria Information System, IPCI = Integrated Primary Care Information Project; SIDIAP = Sistema d'Informació per al Desenvolupament de la Investigació en Atenció Primària; DA = Disease Analyzer; EHR = Electronic Heath record;

Clinical Practice Research Datalink GOLD (CPRD GOLD), United Kingdom (University of Oxford)

The Clinical Practice Research Datalink (CPRD) is a governmental, not-for-profit research service, jointly funded by the National Institute for Health and Care Research and the Medicines and Healthcare products Regulatory Agency, a part of the Department of Health, United Kingdom (UK) (https://cprd.com). CPRD GOLD(20) comprises computerized records of all clinical and referral events in primary care in addition to comprehensive demographic information and medication prescription data in a sample of UK patients (predominantly from Scotland (52% of practices) and Wales (28% of practices). The prescription records include information on the type of product, date of prescription, strength, dosage, quantity, and route of administration. Data from contributing practices are collected and processed into research databases. Quality checks on patient and practice level are applied during the initial processing. Data are available for 20 million patients, including 3.2 million currently registered patients.

Access to CPRD GOLD data requires approval via the Research Data Governance Process. CPRD GOLD will be onboarded as Data Partner for DARWIN EU® in 2023.

IQVIA Disease Analyser (DA) Germany, Germany

DA Germany is collected from extracts of patient management software used by GPs and specialists practicing in ambulatory care settings.(21) Data coverage includes more than 34M distinct person records out of at total population of 80M (42.5%) in the country and collected from 2,734 providers. Patient visiting more than one provider are not cross identified for data protection reasons and therefore recorded as separate in the system. Dates of service include from 1992 through present. Observation time is defined by the first and last consultation dates. Germany has no mandatory GP system and patient have free choice of specialist. As a result, data are collected from visits to 28.8% General, 13.4% Orthopaedic Surgery, 11.8% Otolaryngology, 11.2% Dermatology, 7.7% Obstetrics/Gynaecology, 6.2% various Neurology and Psychiatry 7.0% Paediatric, 4.6% Urology, 3.7% Cardiology, 3.5% Gastroenterology, 1.5% Pulmonary and 0.7% Rheumatology practices. Drugs are recorded as prescriptions of marketed products. No registration or approval is required for drug utilisation studies.

Institut Municipal Assistència Sanitària Information System (IMASIS), Spain

The Institut Municipal Assistència Sanitària Information System (IMASIS) is the Electronic Health Record (EHR) system of Parc de Salut Mar Barcelona (PSMar) which is a complete healthcare services organisation.(22) Currently, this information system includes and shares the clinical information of two general hospitals (Hospital del Mar and Hospital de l'Esperança), one mental health care centre (Centre Dr. Emili Mira) and one social-healthcare centre (Centre Fòrum) including emergency room settings, which are offering specific and different services in the Barcelona city area (Spain). At present, IMASIS includes clinical information more than 1 million patients with at least one diagnosis and who have used the services of this healthcare system since 1990 and from different settings such as admissions, outpatients, emergency room and major ambulatory surgery. The diagnoses are coded using The International Classification of Diseases ICD-9-CM and ICD-10-CM. The average follow-up period per patient in years is 6.37 (SD±6.82). IMASIS-2 is the anonymized relational database of IMASIS which is used for mapping to OMOP including additional sources of information such as the Tumours Registry.

Integrated Primary Care Information Project (IPCI), The Netherlands

IPCI is collected from electronic health records (EHR) of patients registered with their general practitioners (GPs) throughout the Netherlands.(23) The selection of 374 GP practices is representative of the entire country. The database contains records from 2.8 million patients out of a Dutch population of 17M starting in 1996.(23) The median follow-up is 4.7 years. The observation period for a patient is determined by the date of registration at the GP and the date of leave/death. The observation period start date is refined by many quality indicators, e.g., exclusion of peaks of conditions when registering at the GP. All data before the observation period is kept as history data. Drugs are captured as prescription records with product, quantity, dosing directions, strength and indication. Drugs not prescribed in the GP setting might be underreported. Indications are available as diagnoses by the GPs and, indirectly, from secondary care providers but the latter might not be complete. Approval needs to be obtained for each study from the Governance Board. (23)

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

SIDIAP is collected from EHR records of patients receiving primary care delivered through Primary Care Teams, consisting of GPs, nurses and non-clinical staff.(24) The Catalan Health Institute manages 328 out of 370 such Primary Care Teams with a coverage of 5.8M patients, out of 7.8M people in the Catalan population (74%). The database started to collect data in 2006. The mean follow-up is 15 years. The observation period for a patient can be the start of the database (2006), or when a person is assigned to a Catalan Health Institute primary care centre. Date of exit can be when a person is transferred-out to a primary care centre that does not pertain to the Catalan Health Institute, or date of death, or date of end of follow-up in the database. Drug information is available from prescriptions and from dispensing records in pharmacies. Drugs not prescribed in the GP setting might be underreported; and disease diagnoses made at specialist care settings are not included. Studies using SIDIAP data require previous approval by both a Scientific and an Ethics Committee.

8.3 Study Period

The study period will be from 1st of January 2013 to 31st December 2022 or the end of available data in each of the data sources if earlier (see Table 4 for more details).

8.4 Follow-up

Patient-level characterisation: study participants will be followed from the date of MDD diagnosis (index date) until the earliest of loss to follow-up, end of data availability, death, or end of study period (31st December 2022). Additionally:

- To assess the proportions of patients with newly diagnosed MDD initiating antidepressant therapy, incident MDD cases will be followed from the date of MDD diagnosis (index date) until the first prescription of the antidepressant class of interest, up to 24 weeks following MDD diagnosis.
- To assess the proportions of patients with newly diagnosed MDD switching or discontinuing specific antidepressant therapy, incident MDD cases on antidepressant therapy will be followed from the date of the first prescription of the specific antidepressant class of interest up to the date of specific intervals (4, 6, 8, 12, and 24 weeks) after starting antidepressant therapy.
- To assess the proportions of patients with newly diagnosed MDD starting,switching and discontinuing concomitant therapy with psycholeptics, incident MDD cases on antidepressant therapy will be followed from the date of the first prescription of the specific antidepressant class of interest up to the date of specific intervals (4, 6, 8, 12, and 24 weeks) after starting antidepressant therapy.

Patient-level utilization of antidepressants: study participants will be followed up from the date of incident prescription and/or dispensation of antidepressant class of interest (index date) until the earliest of loss to follow-up, end of data availability, death, or end of study period (31st December 2022).

Table 5: Operational Definition of Time 0 (index date) and other primary time anchors

Study population name(s)	Time Anchor Description (e.g., time 0)	Number of entries	Type of entry	Washout window	Care Setting[1]	Code Type[2]	Diagnosis position	Incident with respect to…	Measurement characteristics / validation	Source of algorithm
All patients from the database eligible for the study and newly diagnosed with MDD	Patient present in the database during the study period (2013-2022) and with at least 1 year of valid database history, and no record MDD diagnosis in the previous year.	Single entry	Incident	[-365, ID]	IP and OP	n/a	n/a	MDD	n/a	n/a
All patients from the database eligible for the study – Analysis of incident antidepressant use in patients with newly diagnosed MDD	Patient present in the database during the study period (2013-2022) and with at least 1 year of valid database history.	Single entry	Incident	[-365, ID]	IP and OP	n/a	n/a	Specific drug class	n/a	n/a

[1]ID = index date, IP = inpatient, OP = outpatient, n/a = not applicable, MDD = Major Depressive Disorder

8.5 Study Population with inclusion and exclusion criteria

8.5.1 Patient-level characterization of MDD treatment

The study cohort will comprise all patients, aged ≥ 12 years, with newly diagnosed MDD present in the respective databases during the study period (2013-2022) and with at least 365 days of data availability before the day they become eligible for study inclusion.

8.5.2 Patient-level utilization of antidepressants

All individuals aged 12 and above who are new users of antidepressants with incident MDD diagnosis, after 1 year of no use of the specific medication, in the period between 1st of January 2013 and 31st of December 2022 (or latest date available), with at least 1 year of data visibility prior to the date of their first prescription of the antidepressant class of interest.

Table 6. Operational Definitions of Inclusion Criteria

Criterion	Details	Order of application	Assessment window	Care Settings[1]	Code Type	Diagnosis position[2]	Applied to study populations:	Measurement characteristics/ validation	Source for algorithm
Incident MDD	Patients with newly diagnosed MDD during the study period, that is individuals without a diagnosis of MDD 1 year prior.	After	1 year	IP, OP, OT	SNOMED	First	All study participants aged 12 years and above	N/A	N/A
Prior database history	Study participants will be required to have a year of prior history observed before contributing observation time	After	1 year	IP and OP	n/a	n/a	All patients, aged ≥ 12 years, with newly diagnosed MDD in the selected databases	n/a	n/a

[1] IP = inpatient, OP = outpatient, ED = emergency department, OT = other, n/a = not applicable
[2] Specify whether a diagnosis code is required to be in the primary position (main reason for encounter)

8.6 Variables

8.6.1. Exposure

For this study, exposure of interest is use (during study period) of antidepressants and concomitant medications (psycholeptics). Exposure will be assessed at drug class level. The calculation of duration of the exposures is described in section 8.8.3 Statistical model specification and assumptions of the analytical approach considered.

The list of drug classes of interest is described in Table 7.

Table 7: Exposure of interest

Drug Class of interest	ATC code
I. Main drugs of interest (antidepressants)	
Non-selective monoamine reuptake inhibitors (NSRIs)	N06AA
Selective serotonin reuptake inhibitors (SSRIs)	N06AB
Other Antidepressants (excluding N06AX25 and N06AX27)	N06AX
II. Concomitant drugs of interest (psycholeptics)	
Antipsychotics	N05A
Anxiolytics	N05B
Hypnotics and sedatives	N05C

Details of exposure are described in Table 8.

Table 8. Operational Definitions of Exposure

Exposure group name(s)	Details	Washout window	Assessment Window	Care Setting[1]	Code Type	Diagnosis position[2]	Applied to study populations:	Incident with respect to…	Measurement characteristics/ validation	Source of algorithm
Antidepressants (classes)	Preliminary code list provided in Table 7	[-365, ID]	Calendar year	Primary and secondary care	RxNorm	n/a	All patients with newly diagnosed MDD present in the database during the study period.	Previous use of antidepressants of interest	n/a	n/a
Psycholeptics (classes)	Preliminary code list provided in Table 7	[-365, ID]	Calendar year	Primary and secondary care	RxNorm	n/a	All patients with newly diagnosed MDD present in the database during the study period and being treated with an antidepressant.	Previous use of psycholeptics of interest	n/a	n/a

[1] IP = inpatient, OP = outpatient, ED = emergency department, OT = other, n/a = not applicable; [2] Specify whether a diagnosis code is required to be in the primary position (main reason for encounter)

8.6.2. Outcomes

This study will examine the following four primary outcomes of interest.

- Treatment Initiation

Initiation of treatments will be assessed within a window of 4, 6, 8, 12, and 24 weeks following diagnosis. In relation to the outcome, a predefined list of MDD treatments will be compiled to address Objectives 1 and 3. MDD treatments will include NSRIs, SSRIs, other antidepressants, and concomitant drugs of interest (Psycholeptics - N05A, N05B, and N05C) as listed in Table 7.

- Treatment Switching

Switching of treatments will be assessed within a window of 4, 6, 8, 12, and 24 weeks following treatment initiation. In relation to the outcome, a predefined list of MDD treatments will be compiled to address Objectives 1 and 3. MDD treatments will include NSRIs, SSRIs, other antidepressants, and concomitant drugs of interest (Psycholeptics - N05A, N05B, and N05C) as listed in Table 7.

- Treatment Discontinuation

Discontinuation of treatments will be assessed within a window of 4, 6, 8, 12, and 24 weeks following treatment initiation. In relation to the outcome, a predefined list of MDD treatments will be compiled to address Objectives 1 and 3. MDD treatments will include NSRIs, SSRIs, other antidepressants, and concomitant drugs of interest (Psycholeptics - N05A, N05B, and N05C) as listed in Table 7.

Table 9. Operational Definitions of Outcome

Outcome name	Details	Primary outcome?	Type of outcome?	Wahout window	Care Setting	Code Type	Diagnosis Position	Applied to study populations	Measurement characteristics/validation	Source of algorithm
Treatment initiation	Preliminary code lists provided in Table 7	Yes	Binary: Counts and percentages	365 days as incident use only	IP and OP care	RxNorm	N/A	All patients with incident MDD	N/A	N/A
Treatment Switching	Preliminary code lists provided in Table 7	Yes	Binary: Counts and percentages	365 days as incident use only	IP and OP care	RxNorm	N/A	All patients with incident MDD undergoing treatment	N/A	N/A
Treatment Discontinuation	Preliminary code lists provided in Table 7	Yes	Binary: Counts and percentages	365 days as incident use only	IP and OP care	RxNorm	N/A	All patients with incident MDD undergoing treatment	N/A	N/A

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

8.6.3. Other covariates, including confounders, effect modifiers and other variables (where relevant)

Age at MDD diagnosis will be described. The following age grouping will be used: 12-17, 18 – 44, 45 – 64, and 65 years and above. The sex (male/ female) of study participants will also be identified.

All co-morbidities and concomitant-medications recorded prior (any time prior to the ID, 365 to 31 days prior to the ID and -30 to 1 day before ID) will be used for large-scale patient characterisation, identified as concept/code and descendants.

In addition, the distribution of the initial quantity of the antidepressant of interest at time of first prescribing will be provided. This will be presented by minimum. P25, median, P75 and maximum.

Table 10. Operational Definitions of Covariates

Characteristic	Details	Type of variable	Assessment window	Care Settings[1]	Code Type	Diagnosis Position[2]	Applied to study populations:	Measurement characteristics/ validation	Source for algorithm
Large-scale summary characteristics of new users	Large-scale patient-level characterization with regard to baseline covariates	Binary/ categorical: Counts and percentages	Any time prior to the ID, 365 to 31 days prior to the ID and -30 to 1 day before ID	Primary and secondary care	RxNorm	n/a	Persons with new use during the study period	n/a	n/a
Initial quantity of antidepressant of interest	Characterisation	Min, P25, median, P75, max	At Index date (date of first prescribing of the antidepressant of interest)	Primary and secondary care	RxNorm	n/a	Persons with new use during the study period	n/a	n/a

[1] ID = index date, IP = inpatient, OP = outpatient, ED = emergency department, OT = other, n/a = not applicable

[2] Specify whether a diagnosis code is required to be in the primary position (main reason for encounter)

8.7 Study size

A formal sample size calculation was not undertaken for this descriptive study, given that our main objective is to summarise the characteristics and utilization patterns of antidepressants and psycholeptics in newly diagnosed MDD patients. Based on a preliminary feasibility assessment, the expected number of patients with MDD records in the included databases for this study will be approximately 380,000.

8.8 Data analysis

Table 11. Description of Study Types and Type of analysis

STUDY TYPE	STUDY CLASSIFICATION	TYPE OF ANALYSIS
Disease Epidemiology - Patient Level Characterisation	Off-the-shelf (C1)	- Percentage of patients initiating treatment with specific antidepressants or psycholeptics - Percentage of patients switching treatment to specific antidepressants or psycholeptics - Percentage of patients discontinuing treatment with specific antidepressants or psycholeptics
Patient Level DUS	Off-the-shelf (C1)	- Estimation of the mean, median, 25th percentile, 75th percentile, minimum, and maximum durations of antidepressant use. - Estimation of the median, 25th percentile, 75th percentile, minimum, and maximum of the drug quantity at first prescription of the antidepressant of interest

8.8.1 Federated Network Analyses

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

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

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

8.8.2 Patient privacy protection

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

8.8.3 Statistical model specification and assumptions of the analytical approach considered

R-packages

We will utilize the R packages "TreatmentPatterns" for patient-level treatment characterizations and "DrugUtilization" for patient-level drug utilization analyses.

Patient-level characterisation: Treatment initiation, switching, and discontinuation

To characterise treatment of MDD, the number and percentages of patients receiving each of a pre-specified list of MDD treatments (as listed in section 8.6.1) will be described at index date, and 4, 6, 8, 12, and 24 weeks post index date. The criteria for treatment switching and discontinuation will be based on the options currently available in the standard analytical tools that will be used in this project, specifically, the "TreatmentPattern" R package. This analysis will produce sunburst plots and Sankey diagrams to describe treatment initiation, switching and discontinuation over time (objectives 1 and 3).

Patient-level characterisation: Duration of antidepressant use

The calculation of individual prescription/dispensing durations (in days) will be derived from the DRUG_EXPOSURE table within the CDM. This table contains the drug_exposure_start_date and the drug_exposure_end_date, which are populated during the Extraction Transform and Load (ETL) process based on available source data. The advantage of this approach is that the drug exposure duration is directly obtained, eliminating the need to infer it from other information during analysis. This ensures a consistent analytical pipeline across all databases. New users will be selected based on their first prescription of the respective antidepressant class of interest after the start of the study. For each patient, at least 1 year of data visibility will be required prior to that prescription. New users will be required to not have been exposed to the drug of interest for at least 1 year prior the current prescription. If the start date of a prescription does not fulfil the exposure washout criteria of 1 year of no use, the whole exposure is eliminated.

Drug exposure calculations

Drug eras will be defined as follows: Exposure starts at date of the first prescription after the first MDD diagnosis. For each prescription, the estimated duration of use is retrieved from the drug exposure table in the CDM. Subsequent prescriptions for the same drug will be combined into continuous exposed episodes (drug eras) using the following specifications:

Two drug eras will be merged into one continuous drug era if the distance in days between end of the first era and start of the second era is ≤ 7 days. The time between the two joined eras will be considered as exposed to the first era as show in in Figure 1.

Gap era joint mode	Schematics	Dose in between	Cumulative dose	Cumulative time
"first"		d_1	$d_1 \cdot (x_1 + x_{12}) + d_2 \cdot x_2$	$x_1 + x_{12} + x_2$
"second"		d_2	$d_1 \cdot x_1 + d_2 \cdot (x_2 + x_{12})$	$x_1 + x_{12} + x_2$
"zero"		0	$d_1 \cdot x_1 + d_2 \cdot x_2$	$x_1 + x_{12} + x_2$
"join"		NA	$d_1 \cdot x_1 + d_2 \cdot x_2$	$x_1 + x_2$

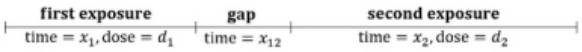

Figure 1. Gap era joint mode

If two eras overlap, the overlap time will be considered exposed to the first era (Figure 2). No time will be added at the end of the combined drug era to account for the overlap.

If two eras start at the same date, the overlapping period will be considered exposed to both. We will not consider repetitive exposure.

Overlap mode	Schematics	Dose overlap
"first"		d_1
"second"		d_2
"both"		$d_1 + d_2$
"maximum"		$\max(d_1, d_2)$

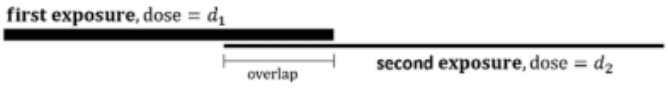

Figure 2. Gap era overlap mode

To construct treatment pathways, various parameters can be defined in the TreatmentPatterns package (Figure 3). In this figure, an example medical file is presented, highlighting the key decisions required to formulate a treatment pathway for an individual undergoing treatments A, B, and C. For a comprehensive understanding of the parameters integrated into the treatment pathways, please refer to the table provided below for a detailed description:

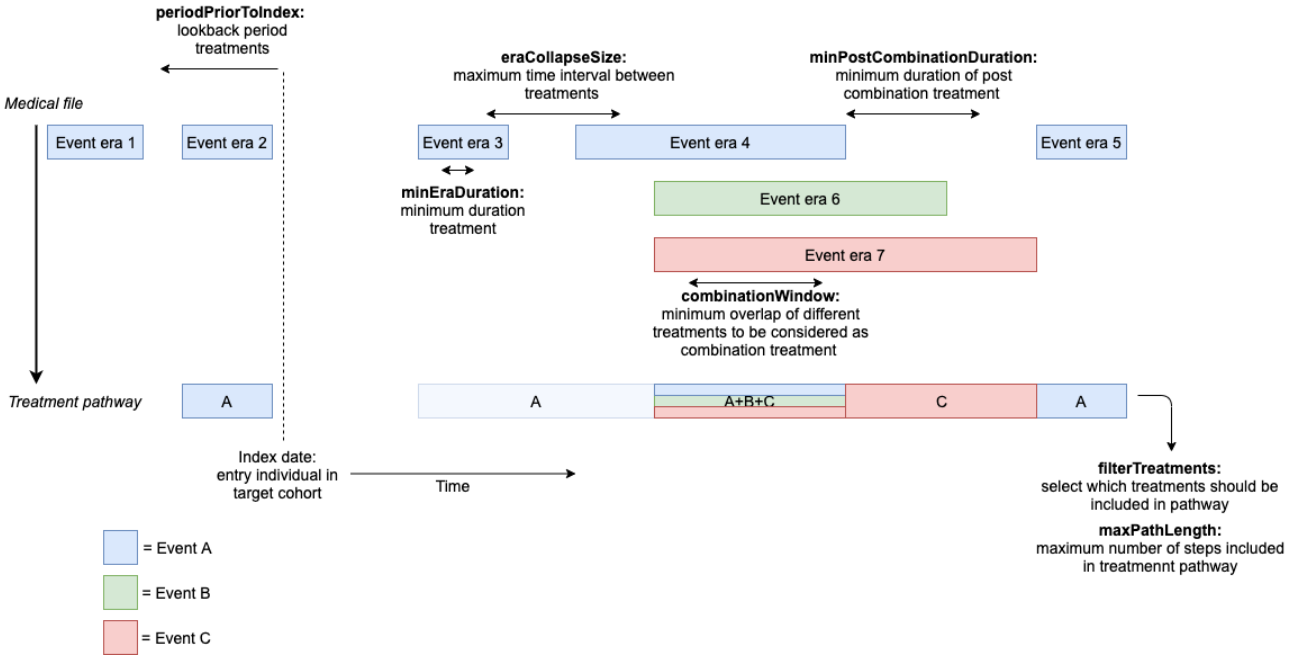

Figure 3. Parameters in TreatmentPatterns package (25)

The following parameters will be defined in this study.

Parameters	Values (days)	Description
periodPriorToIndex	0	Number of days prior to the index date of the target cohort
minEraDuration	0	Minimum time an event era should last to be included in analysis
eraCollapseSize	7	Window of time between which two eras of the same event cohort are collapsed into one era
combinationWindow	7	Window of time two event cohorts need to overlap to be considered a combination treatment
minPostCombinationDuration	7	Minimum time an event era before or after a generated combination treatment should last to be included in analysis
filterTreatments	All	Select all treatments including first time occurrences and sequential repeated treatments
maxPathLength	5	Maximum number of steps included in treatment pathway

For all continuous variables, summary descriptive statistics will be reported: median and interquartile interval will be reported. For all categorical analyses, counts and percentages will be reported. A minimum cell count of 5 will be used when reporting results, with any smaller counts reported as "<5". All analyses will be

reported by country/database, overall and stratified by age and sex when possible (minimum cell count reached).

8.8.4 Sensitivity analysis

If the percentage of people who dropped from the analysis at week 24 is high (where high is defined as more than 50% of individuals participating at week 4 do not longer participate at week 24), we will do a sensitivity analysis where i) we assume that the people who were lost for follow-up continued treatment and ii) and assume that people who were lost to follow-up discontinued treatment. As this analysis can not be generated by the treatment pattern package, we will only do this for the overall analysis (thus not stratified by age and sex). (see also section on limitation)

8.8.5 Evidence synthesis

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

9. DATA MANAGMENT

9.1 Data management

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

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

9.2 Data storage and protection

For this study, participants from various EU member states will process personal data from patients which is collected in national/regional electronic health record databases. Due to the sensitive nature of this personal medical data, it is important to be fully aware of ethical and regulatory aspects and to strive to take all reasonable measures to ensure compliance with ethical and regulatory issues on privacy.
All databases used in this study are already used for pharmaco-epidemiological research and have a well-developed mechanism to ensure that European and local regulations dealing with ethical use of the data and adequate privacy control are adhered to. In agreement with these regulations, rather than combining person level data and performing only a central analysis, local analyses will be run, which generate non-identifiable aggregate summary results.
The output files are stored in the DARWIN Digital Research Environment. These output files do not contain any data that allow identification of subjects included in the study. The DRE implements further security measures in order to ensure a high level of stored data protection to comply with the local implementation of the General Data Protection Regulation (GDPR) (EU) 679/20161 in the various member states.

10. QUALITY CONTROL

General database quality control
A number of open-source quality control mechanisms for the OMOP CDM have been developed (see Chapter 15 of The Book of OHDSI http://book.ohdsi.org/DataQuality.html). In particular, it is expected that data partners will have run the OHDSI Data Quality Dashboard tool (https://github.com/OHDSI/DataQualityDashboard). This tool provides numerous checks relating to the conformance, completeness and plausibility of the mapped data. Conformance focuses on checks that describe the compliance of the representation of data against internal or external formatting, relational, or

computational definitions, completeness in the sense of data quality is solely focused on quantifying missingness, or the absence of data, while plausibility seeks to determine the believability or truthfulness of data values. Each of these categories has one or more subcategories and are evaluated in two contexts: validation and verification. Validation relates to how well data align with external benchmarks with expectations derived from known true standards, while verification relates to how well data conform to local knowledge, metadata descriptions, and system assumptions.

Study specific quality control

When defining cohorts for drugs, a systematic search of possible codes for inclusion will be identified using CodelistGenerator R package (https://github.com/darwin-eu/CodelistGenerator). A pharmacist will review the codes of the opioids of interest. This software allows the user to define a search strategy and using this will then query the vocabulary tables of the OMOP common data model so as to find potentially relevant codes. In addition, DrugExposureDiagnostics (26) will be run if needed to assess the use of different codes across the databases contributing to the study.

The study code will be based on two R packages currently being developed to (1) estimate treatment patterns (initiation, switching or discontinuation) and (2) characterize drug utilization (duration of use) using the OMOP common data model. These packages will include numerous automated unit tests to ensure the validity of the codes, alongside software peer review and user testing. The R package will be made publicly available via GitHub.

11. LIMITATION OF THE RESERACH METHODS

The study will be informed by routinely collected health care data and so data quality issues must be considered, despite the data quality checks. One crucial aspect pertains to the identification of patients with MDD. It is worth noting that the accuracy of these records may vary across different databases. While a relatively low rate of false positives is expected (i.e., cases recorded with a condition they do not truly have), the likelihood of false negatives (i.e., cases with the condition not being recorded) may be higher, particularly in databases lacking patient-level linkage between primary care and secondary care data. In this study, the MDD phenotype will be defined solely based on physician-diagnosed cases identified using relevant SNOMED codes, rather than relying on standardized depression rating scales or individual clinical interviews, often regarded as the "gold standard" for MDD diagnosis. It is important to note that previous studies in CPRD GOLD, IQVIA DA Germany, and SIDIAP have more likely included well-phenotyped MDD cases.

Furthermore, the documentation of medication use and co-morbidities, necessary for patient -level characterization, may vary across databases. In databases where information on MDD treatment is available, the recording of medication use may be incomplete, especially in primary care databases like CPRD GOLD that lack linkage to hospital data. Additionally, it is essential to highlight that the mere recording of a prescription or dispensing does not necessarily imply that the patient used the prescribed medication. Furthermore, assumptions around the duration of drug use will be unavoidable such as using prescribed duration which may not accurately reflect the actual duration of drug exposure. For databases, where duration cannot be calculated due to e.g. missing information on quantity or dosing, treatment duration will not be provided.

For the treatment pattern analysis, we will look at specific time intervals following treatment initiation namely at week 4, 6, 8, 12 and 24 following treatment initiation. Only patients with sufficient follow-up at these intervals will be included for this analysis, implying that patients for instance which were initially included for the analysis at week 4 might not longer be included at week 6 because of lost to follow-up. As part of the results, we will provide the number of individuals included in the analysis at the different time intervals which will provide information on the extent of people who dropped out from the analysis. If the percentage of people who dropped from the analysis at week 24 is high (where high is defined as more than 50% of individuals participating at week 4 do not longer participate at week 24), we will do a sensitivity analysis where i) we assume that the people who were lost for follow-up continued treatment and ii) we assume that the people who were lost to follow-up discontinued treatment. As this analysis can not be generated by the treatment pattern package, we will only do this for the overall analysis (thus not stratified by age and sex).
The number of pills (per package) of the antidepressant of interest might vary by countries which might bias the proportion of individuals continuing treatment. Indeed, if the antidepressant is dosed as one pill daily, and the initial prescription is a package of 30 pills, by definition the patient will be considered as having continued treatment at week 4 even if the patient discontinued treatment in the meantime. To contextualize this potential misclassification, we will provide information on the median quantity (with min, max, P25 and P75) by type of antidepressant and by database.

12. MANAGEMENT AND REPORTING OF ADVERSE EVENTS/ADVERSE REACTIONS

Adverse events/adverse reactions will not be collected or analyzed as part of this evaluation. The nature of this non-interventional evaluation, through the use of secondary data, does not fulfill the criteria for reporting adverse events, according to module VI, VI.C.1.2.1.2 of the Good Pharmacovigilance Practices

13. GOVERNANCE BOARD ASPECTS

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

14. PLANS FOR DISSEMINATING AND COMMUNICATING STUDY RESULTS

14.1 Study Report

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

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

15. OTHER RESULTS

n/a

16. REFERENCES

1. Malhi GS, Mann JJ. Depression. Lancet. 2018;392(10161):2299-312.
2. Fischer F, Zocholl D, Rauch G, Levis B, Benedetti A, Thombs B, et al. Prevalence estimates of major depressive disorder in 27 European countries from the European Health Interview Survey: accounting for imperfect diagnostic accuracy of the PHQ-8. BMJ Ment Health. 2023;26(1).
3. Pedersen CB, Mors O, Bertelsen A, Waltoft BL, Agerbo E, McGrath JJ, et al. A comprehensive nationwide study of the incidence rate and lifetime risk for treated mental disorders. JAMA Psychiatry. 2014;71(5):573-81.
4. Bains N, Abdijadid S. Major Depressive Disorder. 2023.
5. Bradley RG, Binder EB, Epstein MP, Tang Y, Nair HP, Liu W, et al. Influence of child abuse on adult depression: moderation by the corticotropin-releasing hormone receptor gene. Arch Gen Psychiatry. 2008;65(2):190-200.
6. Sullivan PF, Neale MC, Kendler KS. Genetic epidemiology of major depression: review and meta-analysis. Am J Psychiatry. 2000;157(10):1552-62.
7. Cuijpers P, Dekker J, Hollon SD, Andersson G. Adding psychotherapy to pharmacotherapy in the treatment of depressive disorders in adults: a meta-analysis. J Clin Psychiatry. 2009;70(9):1219-29.
8. Pagnin D, de Queiroz V, Pini S, Cassano GB. Efficacy of ECT in depression: a meta-analytic review. J Ect. 2004;20(1):13-20.
9. Gartlehner G, Wagner G, Matyas N, Titscher V, Greimel J, Lux L, et al. Pharmacological and non-pharmacological treatments for major depressive disorder: review of systematic reviews. BMJ Open. 2017;7(6):e014912.
10. Al-Harbi KS. Treatment-resistant depression: therapeutic trends, challenges, and future directions. Patient Prefer Adherence. 2012;6:369-88.
11. Mitroiu M, Teerenstra S, Oude Rengerink K, Pétavy F, Roes KCB. Estimation of treatment effects in short-term depression studies. An evaluation based on the ICH E9(R1) estimands framework. Pharm Stat. 2022;21(5):1037-57.
12. Stensrud MJ, Dukes O. Translating questions to estimands in randomized clinical trials with intercurrent events. Stat Med. 2022;41(16):3211-28.
13. Polverejan E, O'Kelly M, Hefting N, Norton JD, Lim P, Walton MK. Defining Clinical Trial Estimands: A Practical Guide for Study Teams with Examples Based on a Psychiatric Disorder. Ther Innov Regul Sci. 2023;57(5):911-39.
14. Tan EH, Robinson DE, Jödicke AM, Mosseveld M, Bødkergaard K, Reyes C, et al. Drug utilization analysis of osteoporosis medications in seven European electronic health databases. Osteoporos Int. 2023;34(10):1771-81.
15. Hedenmalm K, Quinten C, Kurz X, Bradley M, Lee H, Eworuke E. A collaborative study of the impact of N-nitrosamines presence and ARB recall on ARB utilization - results from IQVIA™ Disease Analyzer Germany. Eur J Clin Pharmacol. 2023;79(6):849-58.
16. Voss EA, Shoaibi A, Yin Hui Lai L, Blacketer C, Alshammari T, Makadia R, et al. Contextualising adverse events of special interest to characterise the baseline incidence rates in 24 million patients with COVID-19 across 26 databases: a multinational retrospective cohort study. EClinicalMedicine. 2023;58:101932.
17. Lane JCE, Weaver J, Kostka K, Duarte-Salles T, Abrahao MTF, Alghoul H, et al. Risk of depression, suicide and psychosis with hydroxychloroquine treatment for rheumatoid arthritis: a multinational network cohort study. Rheumatology. 2020;60(7):3222-34.

18. Denee T, Kerr C, Ming T, Wood R, Tritton T, Middleton-Dalby C, et al. Current treatments used in clinical practice for major depressive disorder and treatment resistant depression in England: A retrospective database study. J Psychiatr Res. 2021;139:172-8.
19. Roca M, Bonelli A, Cattaneo A, Comandini A, Di Dato G, Heiman F, et al. Factors associated with sick leave duration in patients suffering from major depressive disorder initiating antidepressant treatment: a real-world evidence study in Germany and Spain. Int J Psychiatry Clin Pract. 2023;27(1):59-68.
20. Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, van Staa T, et al. Data Resource Profile: Clinical Practice Research Datalink (CPRD). Int J Epidemiol. 2015;44(3):827-36.
21. Planas Domingo J, Gallen Castillo M, Malats Riera N, Porta Serra M, Guasch Jordan I, Cardona Hernandez T, et al. [Tumors registered in Hospital del Mar (Barcelona). Descriptive analysis from 1978 to 1986]. Rev Clin Esp. 1988;183(4):175-9.
22. Martín-Baranera M, Planas I, Palau J, Sanz F. IMASIS computer-based medical record project: dealing with the human factor. Medinfo. 1995;8 Pt 1:333.
23. Vlug AE, van der Lei J, Mosseveld BM, van Wijk MA, van der Linden PD, Sturkenboom MC, et al. Postmarketing surveillance based on electronic patient records: the IPCI project. Methods Inf Med. 1999;38(4-5):339-44.
24. Recalde M, Manzano-Salgado CB, Diaz Y, Puente D, Garcia-Gil MDM, Marcos-Gragera R, et al. Validation Of Cancer Diagnoses In Electronic Health Records: Results From The Information System For Research In Primary Care (SIDIAP) In Northeast Spain. Clin Epidemiol. 2019;11:1015-24.
25. Markus AF, Verhamme KMC, Kors JA, Rijnbeek PR. TreatmentPatterns: An R package to facilitate the standardized development and analysis of treatment patterns across disease domains. Comput Methods Programs Biomed. 2022;225:107081.
26. al GIe. Diagnostics for OMOP Common Data Model Drug Records: Package 'DrugExposureDiagnostics'. Version 0.4.1. 2023 [Available from: https://cran.r-project.org/web//packages//DrugExposureDiagnostics/DrugExposureDiagnostics.pdf.

17. ANNEXES

Appendix I: List with preliminary concept definitions (exposure)

Appendix II: List with preliminary concept definitions for major depressive disorder

Appendix III: ENCePP checklist for study protocols

APPENDIX I – LIST WITH PRELIMINARY CONCEPT DEFINITIONS FOR EXPOSURE

Prescriptions will be identified based on the relevant ingredient. Non-systemic products will be excluded from the code list.

CONCEPT ID	Name	ATC code
21604687	Non-selective monoamine reuptake inhibitors (NSRIs)	N06AA
21604709	Selective serotonin reuptake inhibitors (SSRIs)	N06AB
21604729	Other antidepressants (excluding N06AX25 and N06AX27)	N06AX
21604489	Psycholeptics	N05
21604490	Antipsychotics	N05A
21604564	Anxiolytics	N05B
21604606	Hypnotics and sedatives	N05C

APPENDIX II – LIST WITH PRELIMINARY CONCEPT DEFINITIONS FOR MAJOR DEPRESSIVE DISORDER

Preliminary list – list to be reviewed prior to parametrisation of study code (using phenotype deck)

Major Depressive Disorder

concept_id	concept_name
4152280	Major depressive disorder
4094358	Chronic recurrent major depressive disorder
45757195	Major depressive disorder in mother complicating childbirth
45757196	Major depressive disorder in mother complicating pregnancy
4304140	Recurrent major depressive disorder with atypical features
4220023	Recurrent major depressive disorder with catatonic features
4205471	Recurrent major depressive disorder with melancholic features
4324959	Recurrent major depressive disorder with postpartum onset
4031328	Chronic major depressive disorder, single episode
35615151	Recurrent mild major depressive disorder co-occurrent with anxiety
35615153	Recurrent moderate major depressive disorder co-occurrent with anxiety
35615152	Recurrent severe major depressive disorder co-occurrent with anxiety
37109052	Mild major depressive disorder co-occurrent with anxiety single episode
37109053	Moderate major depressive disorder co-occurrent with anxiety single episode
37109054	Severe major depressive disorder co-occurrent with anxiety single episode

APPENDIX III – ENCEPP CHECKLIST FOR STUDY PROTOCOLS

Study title:
DARWIN EU® - DUS in patients with major depressive disorder

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

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

Comments:

Section 2: Research question	**Yes**	**No**	**N/A**	**Section Number**
2.1 Does the formulation of the research question and objectives clearly explain:	☒	☐	☐	
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 generalized)	☒	☐	☐	
2.1.4 Which hypothesis(-es) is (are) to be tested?	☐	☐	☒	
2.1.5 If applicable, that there is no *a priori* hypothesis?	☐	☐	☒	

Comments:

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

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

Comments:

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

Comments:

Section 5: Exposure definition and measurement		**Yes**	**No**	**N/A**	**Section Number**
5.1	Does the protocol describe how the study exposure is defined and measured? (e.g. operational details for defining and categorizing exposure, measurement of dose and duration of drug exposure)	☒	☐	☐	8.6
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 categorized according to time windows?	☒	☐	☐	
5.4	Is intensity of exposure addressed? (e.g., dose, duration)	☒	☐	☐	
5.5	Is exposure categorized based on biological mechanism of action and taking into account the pharmacokinetics and pharmacodynamics of the drug?	☐	☐	☒	
5.6	Is (are) (an) appropriate comparator(s) identified?	☐	☐	☒	

Comments:

Section 6: Outcome definition and measurement		**Yes**	**No**	**N/A**	**Section Number**
6.1	Does the protocol specify the primary and secondary (if applicable) outcome(s) to be investigated?	☒	☐	☐	8.6
6.2	Does the protocol describe how the outcomes are defined and measured?	☒	☐	☐	
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 utilization, 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)	☐	☐	☒	

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

Comments:

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

Comments:

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

Section 9: Data sources	**Yes**	**No**	**N/A**	**Section Number**
9.3.3 Covariates and other characteristics?	☒	☐	☐	8.6
9.4 Is a linkage method between data sources described? (e.g. based on a unique identifier or other)	☐	☐	☒	

Comments:

Section 10: Analysis plan	**Yes**	**No**	**N/A**	**Section Number**
10.1 Are the statistical methods and the reason for their choice described?	☒	☐	☐	8.8
10.2 Is study size and/or statistical precision estimated?	☐	☐	☒	
10.3 Are descriptive analyses included?	☒	☐	☐	8.8
10.4 Are stratified analyses included?	☒	☐	☐	
10.5 Does the plan describe methods for analytic control of confounding?	☐	☐	☒	
10.6 Does the plan describe methods for analytic control of outcome misclassification?	☐	☐	☒	
10.7 Does the plan describe methods for handling missing data?	☒	☐	☐	8.8.4
10.8 Are relevant sensitivity analyses described?	☒	☐	☐	8.8.4

Comments:

Section 11: Data management and quality control	**Yes**	**No**	**N/A**	**Section Number**
11.1 Does the protocol provide information on data storage? (e.g., software and IT environment, database maintenance and anti-fraud protection, archiving)	☒	☐	☐	9.2
11.2 Are methods of quality assurance described?	☒	☐	☐	10.0
11.3 Is there a system in place for independent review of study results?	☐	☐	☒	

Comments:

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

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

Comments:

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

Comments:

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

Comments:

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

Comments:

Name of the main author of the protocol: Johnmary T. Arinze

Date: 26/09/2023

Signature: Johnmary T. Arinze